

Stochastik

Vorlesungsskript

Thorsten Dickhaus
Universität Bremen
Sommersemester 2023
Version: 9. März 2023

Vorbemerkungen

Das Material zu diesem Skript habe ich im Wesentlichen dem Buch von Georgii (2007) sowie Vorlesungsskripten von Gerhard Osius entnommen. Sollten sich in den übernommenen Teilen Fehler finden, so bin dafür natürlich ich verantwortlich. Lob und positive Kritik gebührt indes den Original-Autoren.

Für die Manuskripterstellung und hilfreiche Korrekturen danke ich Natalia Sirotko-Sibirskaya, Jonathan von Schroeder, Konstantin Schildknecht und Kevin Henke.

Übungsaufgaben und R-Programme zu diesem Kurs stelle ich auf Anfrage gerne zur Verfügung. Einige Referenzen dazu finden sich im Text an den zugehörigen Stellen.

Inhaltsverzeichnis

0	Zufall und Mathematik, motivierende Beispiele	1
1	Wahrscheinlichkeitsräume	4
1.1	Mengensysteme	5
1.2	Wahrscheinlichkeitsmaße	7
1.3	Klassen von Wahrscheinlichkeitsräumen	10
1.3.1	Endliche Wahrscheinlichkeitsräume	10
1.3.2	Abzählbare Wahrscheinlichkeitsräume	11
1.3.3	Geometrische Wahrscheinlichkeiten	11
1.3.4	Reelle Wahrscheinlichkeitsräume mit Lebesguedichten	12
2	Kombinatorik	17
3	Zufallsvariablen und ihre Verteilungen	23
4	Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit	27
4.1	Bedingte Wahrscheinlichkeit	27
4.2	Stochastische Unabhängigkeit von Ereignissen	32
4.3	Produkte diskreter Wahrscheinlichkeitsräume	33
4.4	Produkte stetiger Wahrscheinlichkeitsräume	35
5	Verteilungsfunktionen und Dichten, Transformationsformel	38
5.1	Verteilungsfunktionen und Dichten	38
5.2	Transformationsformel	40
5.3	Zufallsvektoren	43
6	Stochastische Unabhängigkeit von Zufallsvariablen	46
7	Faltungen von Verteilungen	50
7.1	Faltungen diskreter Verteilungen	50
7.2	Faltungen stetiger Verteilungen mit Lebesguedichten	53

7.3	Ergebnisse für nicht notwendigerweise stochastisch unabhängige Zufallsvariablen	56
8	Momente von Verteilungen, Integralungleichungen	58
8.1	Der Erwartungswert	58
8.1.1	Erwartungswert diskreter Verteilungen	58
8.1.2	Erwartungswert stetiger Verteilungen	61
8.1.3	Allgemeine Eigenschaften des Erwartungswertes	62
8.2	Momente und Varianz	66
8.3	Momente von Zufallsvektoren	70
8.4	Integralungleichungen	72
9	Erzeugende Funktion, Laplacetransformierte, Charakteristische Funktion	75
9.1	Erzeugende Funktion	75
9.2	Laplace-Transformierte	77
9.3	Charakteristische Funktion	79
10	Folgen von Zufallsvariablen, Grenzwertsätze	88
	Tabellenverzeichnis	102
	Abbildungsverzeichnis	103
	Literaturverzeichnis	104

Kapitel 0

Zufall und Mathematik, motivierende Beispiele

Das Wort „Stochastik“ kommt aus dem Griechischen und bedeutet „Lehre vom Zufall“ bzw., als mathematische Teildisziplin, „Lehre von den Gesetzmäßigkeiten des Zufalls“.

Beispiel (Gesetz der großen Zahlen):

Wirft man eine „faire“ Münze sehr oft, so wird in etwa der Hälfte der Fälle „Kopf“ fallen. Dies ist ein mathematischer Satz, der in der Stochastik bewiesen wird. Dazu braucht es einen eigenen Wahrscheinlichkeitskalkül.

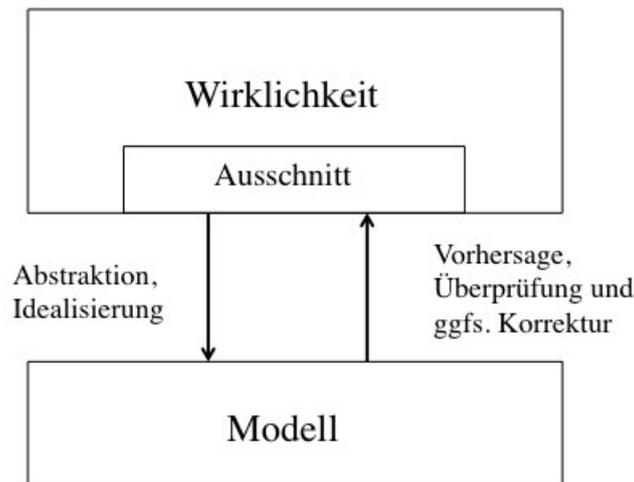
Ursachen von „Zufall“:

- (i) Naturinhärente Indeterminiertheit (z. B. in der Quantenphysik)
- (ii) unsere Unkenntnis über die genauen Rahmenbedingungen der Situation

Aufgaben der Mathematik:

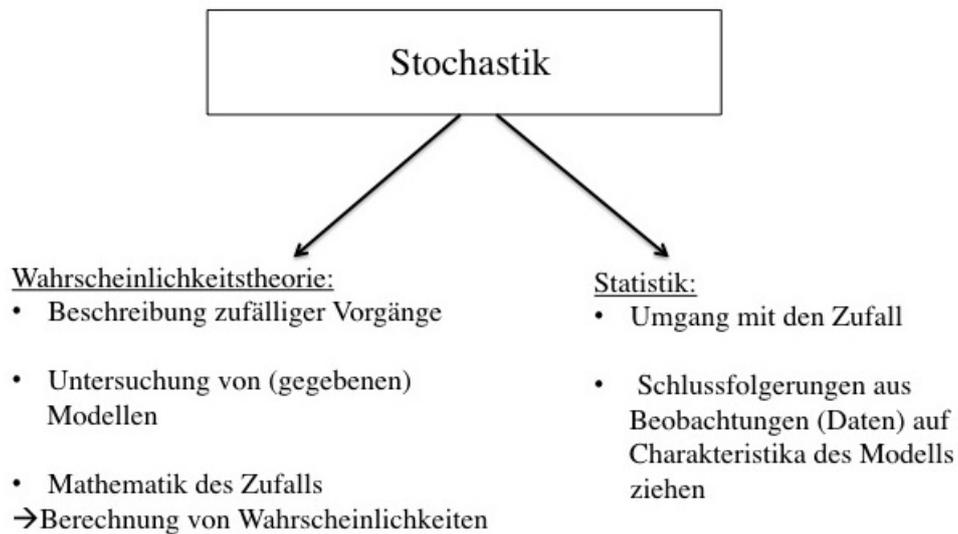
- (i) Abstraktion der Wirklichkeit, Modellbildung
- (ii) stochastischer Kalkül im aufgestellten Modell
- (iii) Rückschluss auf die Wirklichkeit

Schema 0.1



Die Stochastik gliedert sich in zwei Teilgebiete, die Wahrscheinlichkeitstheorie und die (mathematische) Statistik.

Schema 0.2



Beispiele für zufällige Vorgänge, die in der Stochastik untersucht werden können:

Beispiel 0.3 (Problem des abgebrochenen Spiels, Pacioli 1494, Fermat/Pascal 17. Jhdt)

Zwei Spieler spielen um einen hälftigen Einsatz ein faires Spiel. Den Einsatz bekommt der Spieler, der zuerst sechs Runden gewonnen hat. Beim Stand von 5:3 für Spieler 1 muss das Spiel abgebrochen werden. Wie sollten sich die Spieler den Einsatz aufteilen?

Lösung (später): Die „gerechte“ Aufteilung ist 7:1 zu Gunsten von Spieler 1.

Beispiel 0.4 (geometrische Wahrscheinlichkeit)

Wähle zwei Punkte x, y zufällig, jeweils im Einheitsintervall $[0, 1]$. Betrachte das (möglicherweise degenerierte) Rechteck in $[0, 1]^2$ mit den Eckpunkten $(0, 0)$, $(x, 0)$, $(0, y)$ und (x, y) . Wie groß ist die Wahrscheinlichkeit, dass dieses Rechteck eine Fläche von mehr als $1/2$ besitzt?

Lösung (später): Die gesuchte Wahrscheinlichkeit beträgt $1/2(1 - \log 2) \cdot 100\%$.

Beispiel 0.5 (Sekretärsproblem)

Gegeben seien N Bewerber/innen für eine freie Stelle, die ihrer Eignung nach unterscheidbar sind. Wir können ihnen also aufgrund ihrer Eignung einen Rang zuordnen, wobei der Rang 1 der besten Eignung entsprechen soll. Die endgültige Zuordnung der Ränge ist natürlich erst möglich, nachdem sich alle N Bewerber/innen einem Eignungstest unterzogen haben.

Um dieses Verfahren abzukürzen, hat die Firmenleitung sich entschlossen, über eine Einstellung sofort nach dem Einstellungsgespräch zu entscheiden. Natürlich ist es unmöglich, eine einmal getroffene Entscheidung wieder rückgängig zu machen.

Problem/Fragestellung: Gibt es eine optimale Strategie? Falls ja, wie lautet sie?

Lösung (später): Die optimale¹ Strategie lautet:

- (1) Prüfe $(n^* - 1)$ Bewerber/innen, ohne eine(n) von ihnen einzustellen.
- (2) Anschließend wähle den- bzw. diejenige, der bzw. die besser als alle Vorgänger/innen ist.

Dabei ist $n^* = \min\{1 \leq n \leq N : \sum_{k=n}^{N-1} 1/k \leq 1\}$.

¹„optimal“ bedeutet: die Wahrscheinlichkeit, den/die beste(n) Kandidat/in zu wählen, ist maximal.

Kapitel 1

Wahrscheinlichkeitsräume

Die mathematische Modellierung von Zufallssituationen geschieht in drei Schritten:

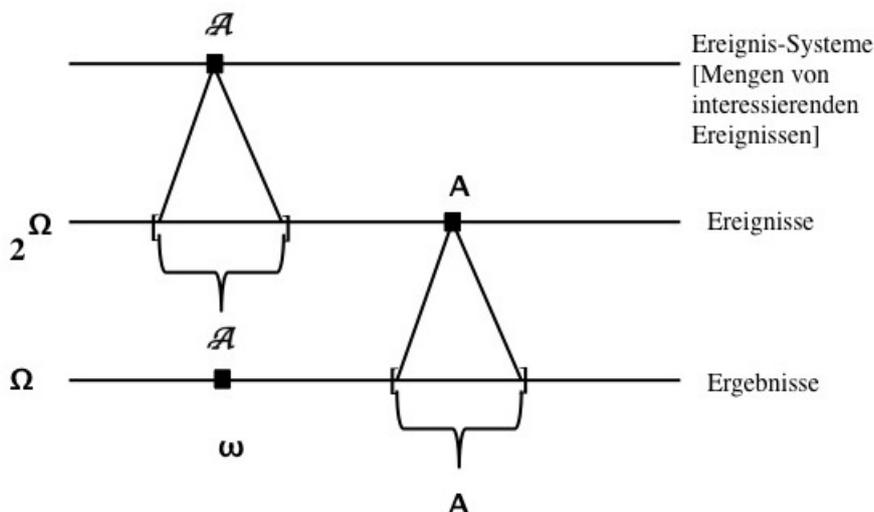
- 1) Festlegung eines Ergebnisraums (Grundraums): Ω
- 2) Festlegung der Menge der interessierenden Ereignisse: $\mathcal{A} \subseteq 2^\Omega$, Ereignis: $A \subseteq \Omega$
- 3) Wahrscheinlichkeitsbewertung der Ereignisse: $\mathbb{P}(A)$, $A \in \mathcal{A}$ (\mathbb{P} : „probability“)

Beispiel 1.1

<u>Stochastischer Vorgang</u>	<u>Grundraum Ω</u>	<u>Ereignis $A \subseteq \Omega$</u>
Einfacher Würfelwurf	$\Omega = \{1, 2, \dots, 6\}$	„gerade Zahl“: $A = \{2, 4, 6\}$
Roulette-Spiel	$\Omega = \{0, 1, \dots, 36\}$	„1. Dutzend“: $A = \{1, 2, \dots, 12\}$
Messung eines Körpergewichts [Kilogramm]	$\Omega = \mathbb{R}_{>0}$	„Übergewicht laut Fahrstuhl“: $A = \{x \in \mathbb{R} : x > 75\}$
Unendlich oft wiederholter Münzwurf	$\Omega = \{0, 1\}^{\mathbb{N}} = \{w = (w_i)_{i \in \mathbb{N}} : w_i \in \{0, 1\}\}$ (Menge der Binärfolgen)	$A = \{w \in \Omega : w_i = 0 \forall 1 \leq i \leq 5\}$

Im Allgemeinen ist der Grundraum Ω irgend eine nicht-leere Menge. Die maximal mögliche Ereignismenge ist die Potenzmenge von Ω , in Zeichen $2^\Omega := \{A : A \subseteq \Omega\}$, wobei dazu per Definition auch die leere Menge \emptyset gehört.

Schema 1.2



Leider ist es nicht immer möglich, als Menge der interessierenden Ereignisse die Potenzmenge 2^Ω selbst zu wählen, siehe Satz von Vitali (später). Dies ist im Allgemeinen nur möglich, falls Ω endlich oder abzählbar („diskret“) ist. Man muss \mathcal{A} also je nach Situation geeignet festlegen.

1.1 Mengensysteme

Sei $\Omega \neq \emptyset$ ein Grundraum. Ein Mengensystem \mathcal{A} über Ω ist (irgend) eine Menge von Teilmengen von Ω , d. h. $\mathcal{A} \subseteq 2^\Omega$.

Definition 1.3 (σ -Algebra)

Ein Mengensystem $\mathcal{A} \subseteq 2^\Omega$ heißt eine σ -Algebra (über Ω), falls gilt:

- (A1) $\Omega \in \mathcal{A}$.
- (A2) \mathcal{A} ist abgeschlossen gegenüber Komplementsbildung, d.h. $\forall A \in \mathcal{A}: A^c := \Omega \setminus A \in \mathcal{A}$.
- (A3) \mathcal{A} ist abgeschlossen gegenüber abzählbarer Vereinigungsbildung: Für jede Folge $(A_n)_{n \in \mathbb{N}}$ mit $A_n \in \mathcal{A}$ für alle $n \in \mathbb{N}$ ist auch $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Das Tupel (Ω, \mathcal{A}) heißt ein messbarer Raum.

Übungsaufgabe: Duale Charakterisierung über abzählbare Durchschnitte.

Es zeigt sich, dass σ -Algebren dem Grundraum eine für die Wahrscheinlichkeitsrechnung geeignete Struktur geben.

Übungsaufgabe: Abgeschlossenheit gegenüber endlicher Durchschnitts- und Vereinigungsbildung.

Bemerkung und Definition 1.4 (Erzeugung von σ -Algebren)

Ist $\Omega \neq \emptyset$ und $\mathcal{F} \subseteq 2^\Omega$ beliebig, so gibt es genau eine kleinste σ -Algebra $\mathcal{A} = \sigma(\mathcal{F})$ über Ω mit $\mathcal{A} \supseteq \mathcal{F}$. Dieses \mathcal{A} heißt die von \mathcal{F} erzeugte σ -Algebra, und \mathcal{F} heißt dann ein Erzeuger von \mathcal{A} .

Beweis:

Sei Σ das System aller σ -Algebren \mathcal{G} über Ω mit $\mathcal{G} \supseteq \mathcal{F}$. Das System Σ ist nicht-leer, denn $2^\Omega \in \Sigma$. Wir setzen $\mathcal{A} := \bigcap_{\mathcal{G} \in \Sigma} \mathcal{G}$. Nach Übungsaufgabe ist \mathcal{A} eine σ -Algebra. Also gehört \mathcal{A} selbst zu Σ und ist offenbar dessen kleinstes Element. ■

Beispiel 1.5

(a) Potenzmenge:

Sei Ω (höchstens) abzählbar und $\mathcal{F} = \{\{\omega\} : \omega \in \Omega\}$ das System der ein-elementigen Teilmengen (Elementarereignisse) von Ω . Dann ist $\sigma(\mathcal{F}) = 2^\Omega$, denn jedes $A \in 2^\Omega$ ist abzählbar und lässt sich darstellen als $A = \bigcup_{\omega \in A} \{\omega\}$.

(b) Borel'sche σ -Algebra über \mathbb{R} , in Zeichen $\mathcal{B}(\mathbb{R})$:

Sei $\mathcal{F} = \{(-\infty, c] : c \in \mathbb{R}\}$. Dann heißt $\sigma(\mathcal{F}) =: \mathcal{B}(\mathbb{R})$ die Borel'sche σ -Algebra über \mathbb{R} (zu Ehren von Émile Borel, 1871 - 1956). Die σ -Algebra $\mathcal{B}(\mathbb{R})$ enthält alle halboffenen Intervalle $(a, b] = (-\infty, b] \setminus (-\infty, a]$, alle kompakten Intervalle $[a, b] = \bigcap_{n \in \mathbb{N}} (a - \frac{1}{n}, b]$ sowie alle offenen Intervalle $(a, b) = (-\infty, b) \cap (-\infty, a]^c$, wobei $(-\infty, b) = \bigcup_{n \in \mathbb{N}} (-\infty, b - \frac{1}{n}]$. Die Elemente von $\mathcal{B}(\mathbb{R})$ heißen Borel-Mengen.

Weitere Borel-Mengen sind:

- alle Elementarereignisse $\{x\}$ von \mathbb{R} ,
- alle endlichen und abzählbaren Teilmengen von \mathbb{R} ,
- alle offenen und alle abgeschlossenen Teilmengen von \mathbb{R} .

Allerdings ist $\mathcal{B}(\mathbb{R}) \neq 2^{\mathbb{R}}$!

(c) Borel'sche σ -Algebra über $\Omega \subset \mathbb{R}$, in Zeichen $\mathcal{B}(\Omega)$:

Für $\emptyset \neq \Omega \subset \mathbb{R}$ ist das System $\mathcal{B}(\Omega) = \{A \cap \Omega : A \in \mathcal{B}(\mathbb{R})\}$ eine σ -Algebra über Ω und heißt Borel'sche σ -Algebra über Ω .

(d) Produkt- σ -Algebra:

Sei Ω ein kartesisches Produkt von Mengen E_i , d. h. $\Omega = \times_{i \in I} E_i$ für eine Indexmenge $I \neq \emptyset$. Sei \mathcal{E}_i eine σ -Algebra auf E_i sowie $\pi_i : \Omega \rightarrow E_i$ die Projektion auf die i -te Koordinate. Betrachte das Mengensystem $\mathcal{F} = \{\pi_i^{-1}(A_i) : i \in I, A_i \in \mathcal{E}_i\}$. (System aller Mengen in Ω , die durch ein Ereignis in einer einzelnen Koordinate bestimmt sind).

Dann heißt $\bigotimes_{i \in I} \mathcal{E}_i := \sigma(\mathcal{F})$ die Produkt- σ -Algebra der \mathcal{E}_i über Ω . Im Fall $I = \{1, \dots, d\}$, $E_i \equiv E$ und $\mathcal{E}_i \equiv \mathcal{E} \forall 1 \leq i \leq d$ schreibt man auch $\mathcal{E}^{\otimes d}$ statt $\bigotimes_{i \in I} \mathcal{E}_i$. Zum Beispiel ist die Borel'sche σ -Algebra über \mathbb{R}^d gegeben als $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{F})$ mit $\mathcal{F} = \{\times_{i=1}^d (-\infty, c_i] : c_i \in \mathbb{Q} \forall 1 \leq i \leq d\}$.

1.2 Wahrscheinlichkeitsmaße

Definition 1.6

Ein Wahrscheinlichkeitsmaß \mathbb{P} auf einer σ -Algebra \mathcal{A} über einem Ergebnisraum $\Omega \neq \emptyset$ ist eine Abbildung $\mathbb{P} : \mathcal{A} \rightarrow \mathbb{R}$ mit den folgenden Eigenschaften.

(P1) Nicht-Negativität: $\mathbb{P}(A) \geq 0 \forall A \in \mathcal{A}$.

(P2) Normiertheit: $\mathbb{P}(\Omega) = 1 = 100\%$ („sicheres Ereignis“).

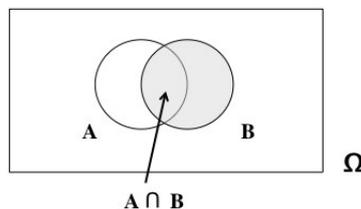
(P3) σ -Additivität: Für jede Folge $(A_n)_{n \in \mathbb{N}}$ mit $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, von paarweise disjunkten Mengen (d.h. $A_i \cap A_j = \emptyset$, $i \neq j$) gilt $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

Das Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ heißt ein Wahrscheinlichkeitsraum. Ist Ω höchstens abzählbar, so heißt $(\Omega, \mathcal{A}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum; ist Ω überabzählbar, so heißt $(\Omega, \mathcal{A}, \mathbb{P})$ ein stetiger Wahrscheinlichkeitsraum.

Korollar 1.7 (Rechenregeln für Wahrscheinlichkeitsmaße)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Dann gilt:

- (a) Das Wahrscheinlichkeitsmaß \mathbb{P} ist endlich additiv: Für paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{A}$ gilt $\mathbb{P}(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mathbb{P}(A_k)$.
- (b) $\forall A \in \mathcal{A} : \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. Insbesondere ist $\mathbb{P}(\emptyset) = 0$ („unmögliches Ereignis“).
- (c) $\forall A \in \mathcal{A} : 0 \leq \mathbb{P}(A) \leq 1$, d.h. $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$.
- (d) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.



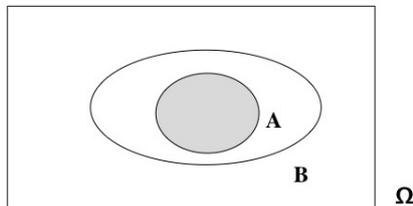
(e) Seien $A_1, \dots, A_n \in \mathcal{A}$, nicht notwendigerweise paarweise disjunkt. Dann gilt

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)$$

(Siebformel von Poincaré und Sylvester, inclusion-exclusion principle, Additionsformel, ...).

(f) Sub-Additivität: Unter den Voraussetzungen von (e) gilt $\mathbb{P}(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^n \mathbb{P}(A_k)$.

(g) Monotonie: $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A)$, $A, B \in \mathcal{A}$.



(h) Stetigkeit von unten: Sei $(A_n)_{n \in \mathbb{N}}$ eine aufsteigende Folge in \mathcal{A} , d. h., $A_n \subseteq A_{n+1}$ für alle $n \in \mathbb{N}$. Dann gilt $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$.

(i) Stetigkeit von oben: Sei $(A_n)_{n \in \mathbb{N}}$ eine abfallende Folge in \mathcal{A} , d. h., $A_{n+1} \subseteq A_n$ für alle $n \in \mathbb{N}$. Dann gilt $\mathbb{P}(\bigcap_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$.

(j) σ -Subadditivität: Für allgemeine Folgen $(A_n)_{n \in \mathbb{N}}$ in \mathcal{A} gilt $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$, wobei die rechte Seite gleich $+\infty$ sein kann.

(k) Bei einer beliebigen Familie $(A_i)_{i \in I}$ paarweise diskjunkter Ereignisse in \mathcal{A} haben höchstens abzählbar viele eine von Null verschiedene Wahrscheinlichkeit. D. h., die Menge $\mathcal{M} = \{i \in I : \mathbb{P}(A_i) > 0\}$ ist höchstens abzählbar.

Nachträglich liefern wir jetzt noch die Begründung, warum nicht stets $\mathcal{A} = 2^\Omega$ gewählt werden kann.

Satz 1.8 (Satz von Vitali)

Sei $\Omega = \{0, 1\}^{\mathbb{N}}$ (Ergebnisraum des unendlich oft wiederholten Münzwurfes, vgl. Beispiel 1.1). Dann gibt es keine Abbildung $\mathbb{P} : 2^\Omega \rightarrow [0, 1]$ mit den folgenden drei Eigenschaften.

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

(ii) σ -Additivität: siehe (P3) in Definition 1.6.

(iii) Invarianz: $\forall A \subseteq \Omega$ und $n \geq 1$ gilt $\mathbb{P}(T_n(A)) = \mathbb{P}(A)$, wobei $T_n : \omega = (\omega_1, \omega_2, \dots) \mapsto (\omega_1, \dots, \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, \dots)$, die Abbildung von Ω auf sich selbst bezeichnet, welche das Ergebnis des n -ten Wurfes umdreht („Flip“), und $T_n(A) = \{T_n(\omega) : \omega \in A\}$ das Bild von A unter T_n bezeichnet.

[Eigenschaft (iii) drückt die Fairness der Münze und die Unabhängigkeit der Würfe aus.]

Beweis:

Definiere eine Äquivalenzrelation „ \sim “ auf Ω wie folgt: $\omega \sim \omega' : \iff \exists n_0 \in \mathbb{N} : \omega_k = \omega'_k \ \forall k \geq n_0$. Damit zerfällt Ω in disjunkte Äquivalenzklassen. Nach dem Auswahlaxiom der Mengenlehre können wir aus jeder Äquivalenzklasse einen Vertreter (Repräsentanten) wählen.

Sei M die Menge dieser Vertreter. Sei $\mathcal{S} = \{S \subset \mathbb{N} : |S| < \infty\} = \bigcup_m \{S \subset \mathbb{N} : \max S = m\}$ die abzählbare Menge der endlichen Teilmengen von \mathbb{N} . Für $S = \{n_1, \dots, n_k\} \in \mathcal{S}$ sei $T_S = T_{n_1} \circ T_{n_2} \circ \dots \circ T_{n_k}$ der Flip zu allen Indizes in S . Dann gilt:

- (a) $\Omega = \bigcup_{S \in \mathcal{S}} T_S(M)$, denn zu jedem $\omega \in \Omega$ existiert ein $\omega' \in M$ mit $\omega \sim \omega'$, und folglich ein $S \in \mathcal{S}$ mit $\omega = T_S(\omega') \in T_S(M)$.
- (b) Die Mengen $(T_S(M))_{S \in \mathcal{S}}$ sind paarweise disjunkt, denn wenn $T_S(M) \cap T_{S'}(M) \neq \emptyset$ für $S, S' \in \mathcal{S}$, so gibt es $\omega, \omega' \in M$ mit $T_S(\omega) = T_{S'}(\omega')$ und daher $\omega \sim T_S(\omega) = T_{S'}(\omega') \sim \omega'$. Nach Konstruktion von M gilt dann aber $\omega = \omega'$ und daher $S = S'$.

Aus diesen Überlegungen und den Voraussetzungen (i) bis (iii) konstruieren wir den Widerspruch $1 = \mathbb{P}(\Omega) = \sum_{S \in \mathcal{S}} \mathbb{P}(T_S(M)) = \sum_{S \in \mathcal{S}} \mathbb{P}(M)$. Diese Gleichungskette kann nicht richtig sein, denn $\sum_{S \in \mathcal{S}} \mathbb{P}(M)$ ist entweder gleich Null oder gleich $+\infty$, je nachdem, ob $\mathbb{P}(M) = 0$ oder $\mathbb{P}(M) > 0$ gesetzt wird. ■

Bemerkung 1.9

Der hier eingeführte Wahrscheinlichkeitsbegriff ist der „axiomatische Wahrscheinlichkeitsbegriff“ nach Kolmogorov (1903 - 1987). Es gibt noch mindestens zwei ältere „Definitionen“, die heutzutage indes nicht mehr als Grundlage der mathematischen Stochastik verwendet werden.

- (a) „Klassischer“ Wahrscheinlichkeitsbegriff (Pascal, Fermat, Bernoulli, Laplace):

Die Wahrscheinlichkeit eines Ereignisses A ist gegeben als das Verhältnis der Zahl der (für A) günstigen Ergebnisse zu der aller möglichen Ergebnisse; vorausgesetzt, alle Ergebnisse (in Ω) sind gleich wahrscheinlich.

in Formeln: $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.

Probleme:

1. Ringschluss: Wahrscheinlichkeit wird „definiert“ darüber, dass alle Ereignisse dieselbe Wahrscheinlichkeit haben.
2. Kann nicht mit Fällen umgehen, in denen die Voraussetzung der gleichen Wahrscheinlichkeit aller Elementarereignisse verletzt ist.

- (b) „Statistischer“ Wahrscheinlichkeitsbegriff (Ellis, Bode, Venn, von Mises):

Ein Ereignis A trete zufällig auf. Dann ist die Wahrscheinlichkeit von A „definiert“ als der „Grenzwert“ der Folge $p_n(A) = \frac{n_A}{n}$ der relativen Häufigkeit des Eintretens von A bei n Versuchen, $n \rightarrow \infty$, wobei $n_A = \#\{1 \leq j \leq n : A \text{ tritt im } j\text{-ten Versuch ein}\}$.

Probleme:

1. $\lim_{n \rightarrow \infty} p_n(A)$ muss weder existieren noch eindeutig bestimmt sein.
2. Viele Vorgänge sind nicht wiederholbar, z. B. $A = \{\text{Herr X war der Täter}\}$.

Satz 1.10 (Konstruktion von Wahrscheinlichkeitsmaßen)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, und es gelte $\mathcal{A} = \sigma(\mathcal{F})$ für ein Erzeugendensystem $\mathcal{F} \subseteq 2^\Omega$. Ist \mathcal{F} \cap -stabil in dem Sinne, dass mit $A, B \in \mathcal{F}$ auch $A \cap B \in \mathcal{F}$ ist, so ist \mathbb{P} bereits durch seine Werte auf \mathcal{F} , d. h., durch seine Einschränkung $\mathbb{P}|_{\mathcal{F}}$ auf \mathcal{F} eindeutig bestimmt.

Beweis: Maßtheorie-Vorlesung

1.3 Klassen von Wahrscheinlichkeitsräumen

1.3.1 Endliche Wahrscheinlichkeitsräume

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit endlichem Ergebnisraum Ω und $\mathcal{A} = 2^\Omega$. Dann nennen wir $(\Omega, \mathcal{A}, \mathbb{P})$ einen endlichen Wahrscheinlichkeitsraum. Aus der Additivität von \mathbb{P} ergibt sich

$$\forall \emptyset \neq A \in \mathcal{A} : \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) =: \sum_{\omega \in A} \mathbb{P}(\omega).$$

Es genügt also, die Elementarwahrscheinlichkeiten $\mathbb{P}(\{\omega\}), \omega \in \Omega$, anzugeben, wobei $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$ gelten muss; vgl. Satz 1.10 in Verbindung mit Beispiel 1.5.(a). Ist umgekehrt eine nicht-negative Abbildung $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$ gegeben mit der Eigenschaft

$$\sum_{\omega \in \Omega} f(\omega) = 1, \tag{1.1}$$

so induziert f ein Wahrscheinlichkeitsmaß \mathbb{P}_f auf 2^Ω mittels

$$\mathbb{P}_f(A) = \sum_{\omega \in A} f(\omega), A \subseteq \Omega. \tag{1.2}$$

Definition 1.11

Die Funktion f aus (1.1) und (1.2) heißt Zähldichte oder Wahrscheinlichkeitsfunktion von \mathbb{P}_f .

Beispiel 1.12

(a) Diskrete Gleichverteilung, Laplace'scher Wahrscheinlichkeitsraum:

Ein endlicher Wahrscheinlichkeitsraum $(\Omega, 2^\Omega, \mathbb{P})$ mit $|\Omega| = n \in \mathbb{N}$ heißt Laplace'scher Wahrscheinlichkeitsraum, falls \mathbb{P} die Zähldichte $f_{\mathbb{P}}$ mit $f_{\mathbb{P}}(\omega) = 1/n \forall \omega \in \Omega$ besitzt. Das Wahrscheinlichkeitsmaß \mathbb{P} heißt die diskrete Gleichverteilung auf Ω . Es gilt: $\forall A \in 2^\Omega : \mathbb{P}(A) = \frac{|A|}{|\Omega|} = |A|/n$, vgl. Bemerkung 1.9.(a). Beispiele für Laplace'sche Wahrscheinlichkeitsräume studieren wir in Kapitel 2 (Kombinatorik) näher.

(b) Bernoulli-Verteilung auf binärem Ergebnisraum: Sei $|\Omega| = 2$ („binärer“ Grundraum), also o. B. d. A. $\Omega = \{0, 1\}$, sowie $\mathcal{A} = 2^\Omega$. Jedes Wahrscheinlichkeitsmaß \mathbb{P} auf (Ω, \mathcal{A}) ist vollständig bestimmt durch eine einzige Zahl, nämlich $p := \mathbb{P}(1) \in [0, 1]$, denn es gilt $f_{\mathbb{P}}(k) = \mathbb{P}(\{k\}) = p^k(1-p)^{1-k}$, $k \in \{0, 1\}$. Ein solches \mathbb{P} heißt Bernoulli-Verteilung mit Parameter p , kurz: Bernoulli(p).

(c) Binomialverteilung auf $\{0, 1, \dots, n\}$: Sei $\Omega = \{0, 1, \dots, n\}$ und $\mathcal{A} = 2^\Omega$. Dann ist die Binomialverteilung mit Parametern n und $p \in [0, 1]$ auf (Ω, \mathcal{A}) , kurz Bin(n, p), gegeben durch die Zähldichte $f_{\text{Bin}(n,p)}(k) \equiv f(k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$, $k \in \Omega$. Für $n = 1$ ergibt sich Bernoulli(p) = Bin($1, p$). Die Binomialverteilung ist die Wahrscheinlichkeitsverteilung der (zufälligen) Anzahl der „Treffer“ bei n unabhängigen Versuchen unter gleichen, standardisierten Bedingungen, falls die „Trefferwahrscheinlichkeit“ bei jedem Einzelversuch jeweils p beträgt.

1.3.2 Abzählbare Wahrscheinlichkeitsräume

Ein abzählbarer Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum $(\Omega, 2^\Omega, \mathbb{P})$ mit abzählbarem Grundraum Ω , etwa $\Omega = \mathbb{N}$ oder $\Omega = \mathbb{N}_0$. Analog zu den Ausführungen in Abschnitt 1.3.1 genügt zur Charakterisierung von \mathbb{P} die Angabe seiner Zähldichte $f_{\mathbb{P}} \equiv f : \Omega \rightarrow \mathbb{R}_{\geq 0}$ mit $\sum_{\omega \in \Omega} f(\omega) = 1$. Wegen dieser Analogie fasst man endliche und abzählbare Wahrscheinlichkeitsräume zur Klasse der diskreten Wahrscheinlichkeitsräume zusammen, siehe Definition 1.6.

Beispiel 1.13 (Poisson-Verteilung auf \mathbb{N}_0)

Sei $\lambda > 0$ eine vorgegebene Konstante, $\Omega = \mathbb{N}_0$, $\mathcal{A} = 2^\Omega$. Dann ist die Poisson-Verteilung mit Intensitätsparameter λ , kurz Poisson(λ), gegeben durch die Zähldichte $f_{\text{Poisson}(\lambda)}$, gegeben durch

$$f_{\text{Poisson}(\lambda)}(k) \equiv f(k|\lambda) = \frac{\lambda^k}{k!} \exp(-\lambda), k \in \Omega = \mathbb{N}_0.$$

Die Zähldichte $f(k|\lambda)$ wächst, solange $k \leq \lambda$ gilt, und fällt danach ab.

Die Poisson-Verteilung ist ein Modell für die Anzahl von Eintritten eines definierten Zielereignisses in einer spezifizierten Grundgesamtheit, vorausgesetzt, dass die Einzelereignisse zufällig und unabhängig voneinander eintreten (z. B.: Anzahl an neu auftretenden Salmonellen-Infektionen in Bremen im Jahre 2016).

Erinnerung:

Jedes Modell abstrahiert/idealisiert die Wirklichkeit. Selbstverständlich sind nicht beliebig viele Infektionen möglich, und es wird in der Praxis Abhängigkeiten geben (z. B. Familien mit selbem Abendessen o. ä.).

1.3.3 Geometrische Wahrscheinlichkeiten

Definition 1.14

Eine Teilmenge A des \mathbb{R}^d , $d \in \mathbb{N}$, heißt geometrisch regulär, falls man ihr ein d -dimensionales

Volumen (Länge, Fläche, ...) $\lambda^d(A)$ zuordnen kann.

Genauer:

A heißt *geometrisch regulär*, falls zu jedem $\varepsilon > 0$ endliche Folgen $(I_j)_{1 \leq j \leq m}$ und $(J_k)_{1 \leq k \leq n}$ von jeweils paarweise disjunkten „Intervallen“ der Form $\times_{i=1}^d [a_i, b_i]$ im \mathbb{R}^d existieren, so dass $\bigcup_{j=1}^m I_j \subseteq A \subseteq \bigcup_{k=1}^n J_k$ sowie

$$0 \leq \sum_{k=1}^n \lambda^d(J_k) - \sum_{j=1}^m \lambda^d(I_j) < \varepsilon$$

gilt.

Anmerkung: Alle wohlbekanntesten geometrischen Objekte wie z. B. Kreis, Dreieck, Kugel, Quader, Pyramide, etc. sind geometrisch reguläre Mengen ($d = 2, 3$).

Definition 1.15

Sei Ω eine nicht-leere, geometrisch reguläre Teilmenge des \mathbb{R}^d mit $\infty > \lambda^d(\Omega) > 0$, und sei \mathcal{A} die Klasse aller geometrisch regulären Teilmengen von Ω . Dann wird durch

$$\mathbb{P}(A) := \frac{\lambda^d(A)}{\lambda^d(\Omega)} = \frac{\text{Volumen von } A}{\text{Volumen von } \Omega}, A \in \mathcal{A},$$

ein Wahrscheinlichkeitsmaß definiert, genannt die Gleichverteilung auf Ω .

Die Zahl $\mathbb{P}(A) \in [0, 1]$ heißt die geometrische Wahrscheinlichkeit von A .

Beispiel 1.16 (siehe Beispiel 0.4)

Sei $\Omega = (0, 1)^2$ und \mathbb{P} die Gleichverteilung auf Ω . Berechne $\mathbb{P}(E)$, wobei $E = \{(x, y) \in \Omega : xy > 1/2\}$.

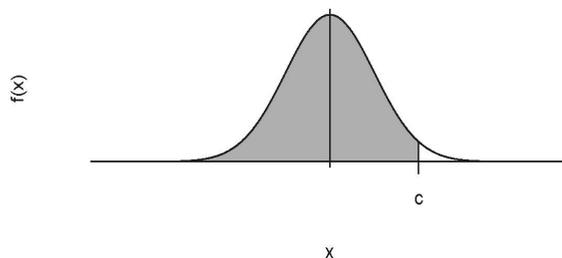
Lösung: Da $\lambda^2(\Omega) = 1 \cdot 1 = 1$ ist, genügt es, $\lambda^2(E)$ zu berechnen. Beachte dazu, dass man äquivalenterweise $E = \{(x, y) \in (0, 1)^2 : x > 1/2, y > 1/(2x)\}$ schreiben kann. Also ist der Flächeninhalt von E gleich $1/2$ minus der Fläche unter der Kurve $x \mapsto \frac{1}{2x}$ für $x \in (1/2, 1)$, und demnach

$$\mathbb{P}(E) = 1/2 - \int_{1/2}^1 \frac{1}{2x} dx = \frac{1}{2}(1 - \log 2).$$

1.3.4 Reelle Wahrscheinlichkeitsräume mit Lebesguedichten

Wir betrachten hier den Grundraum \mathbb{R}^d , versehen mit der Borel'schen σ -Algebren $\mathcal{B}(\mathbb{R}^d)$, vgl. Beispiel 1.5.(b) und 1.5.(d). Wir beachten, dass das Erzeugendensystem \mathcal{F} aus Beispiel 1.5.(d) \cap -stabil ist. Nach Satz 1.10 genügt zur Festlegung eines jeden Wahrscheinlichkeitsmaßes \mathbb{P} auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ die Angabe der Wahrscheinlichkeiten $\mathbb{P}\left(\times_{i=1}^d (-\infty, c_i]\right)$ für reelle Konstanten $c_i, 1 \leq i \leq d$. Diese Wahrscheinlichkeiten sollen über (uneigentliche) Integrale von (stückweise) stetigen Dichtefunktionen f formalisiert werden (daher auch die Bezeichnung „stetiger Wahrscheinlichkeitsraum“, siehe Definition 1.6).

Abbildung 1.17 (Illustration für $d = 1$)



$$\mathbb{P}(\cdot \in]-\infty, c]) = \int_{-\infty}^c f(x) dx$$

Abbildung 1.1: Berechnung von Wahrscheinlichkeiten mit Dichtefunktionen

Dazu ist es erforderlich, den aus der Schule bekannten Begriff des Riemann-Integrals auf den des Lebesgue-Integrals zu verallgemeinern.

Definition und Satz 1.18 (Maß- und Integrationstheorie)

Für jede Funktion $f : \mathbb{R}^d \rightarrow [0, \infty]$, welche die Messbarkeitseigenschaft

$$\{x \in \mathbb{R}^d : f(x) \leq \gamma\} \in \mathcal{B}(\mathbb{R}^d) \text{ für alle } \gamma > 0 \quad (1.3)$$

erfüllt, kann das Lebesgue-Integral $\int f(x) dx = \int_{\mathbb{R}^d} f(x) dx \in [0, \infty]$ so erklärt werden, dass Folgendes gilt:

(a) Für jede Folge f_1, f_2, \dots von nicht-negativen, gemäß (1.3) messbaren Funktionen gilt

$$\int \sum_{n \geq 1} f_n(x) dx = \sum_{n \geq 1} \int f_n(x) dx.$$

(b) Für jede Teilmenge M des \mathbb{R}^d sei $\mathbf{1}_M$, gegeben durch

$$\mathbf{1}_M(x) = \begin{cases} 1, & \text{falls } x \in M \\ 0, & \text{sonst} \end{cases}$$

$x \in \mathbb{R}^d$, die Indikatorfunktion von M . Dann ist das Lebesgue-Integral über $A \in \mathcal{B}(\mathbb{R}^d)$ von f definiert als $\int_A f(x) dx := \int \mathbf{1}_A(x) f(x) dx$.

(c) Ist speziell $d = 1$ und existiert für $f : \mathbb{R} \rightarrow [0, \infty]$ und $a, b \in \mathbb{R}$ das eigentliche Riemann-Integral $\int_a^b f(x) dx$, so existiert das Lebesgue-Integral $\int_{]a,b]} f(x) dx$ und es gilt $\int_{]a,b]} f(x) dx = \int_a^b f(x) dx$, d. h., Lebesgue- und Riemann-Integral stimmen in solchen Fällen überein.

Bemerkung 1.19

- (i) Gemäß Beispiel 1.5.(b) sind halboffene Intervalle der Form $]a, b]$ Borel-Mengen.
- (ii) Wegen Satz 1.18.(a) und 1.18.(c) können Lebesgue-Integrale für stückweise stetige, reellwertige Funktionen über Borel-Mengen, die sich als endliche Vereinigung disjunkter Intervalle darstellen lassen, auf Riemann-Integrale zurückgeführt werden.
- (iii) Für Dimensionen $d \geq 1$ benutzt man typischerweise den Satz von Fubini in Verbindung mit der Konstruktion von Produkt- σ -Algebren; siehe Beispiel 1.5.(d) (später mehr).
- (iv) Elementarereignisse haben in stetigen Wahrscheinlichkeitsräumen die Wahrscheinlichkeit Null.

Satz und Definition 1.20

Ist der Grundraum $\Omega \subseteq \mathbb{R}^d$ eine Borel-Menge, so bestimmt jede Funktion $f : \Omega \rightarrow [0, \infty[$ mit den Eigenschaften

- (i) $\{x \in \Omega : f(x) \leq \gamma\} \in \mathcal{B}(\Omega)$ für alle $\gamma > 0$,
- (ii) $\int_{\Omega} f(x) dx = 1$

ein Wahrscheinlichkeitsmaß \mathbb{P} auf $(\Omega, \mathcal{B}(\Omega))$ vermöge $\mathbb{P}(A) = \int_A f(x) dx$ für $A \in \mathcal{B}(\Omega)$.

Die Funktion f heißt Dichtefunktion von \mathbb{P} , Wahrscheinlichkeitsdichte von \mathbb{P} bzw. Lebesguedichte von \mathbb{P} . Wir nennen $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ dann einen stetigen Wahrscheinlichkeitsraum.

Beweis:

Normiertheit und Nicht-Negativität von \mathbb{P} sind klar. Die σ -Additivität von \mathbb{P} ergibt sich aus der Tatsache, dass für paarweise disjunkte Mengen $(A_i)_{i \geq 1}$ gilt: $\mathbf{1}_{\bigcup_{i \geq 1} A_i} = \sum_{i \geq 1} \mathbf{1}_{A_i}$. Damit liefert Satz 1.18.(a) das Gewünschte.

Bemerkung 1.21

Vergleicht man Definition 1.20 mit Definition 1.11, so stellt man fest, dass im stetigen Fall (gegenüber dem diskreten Fall) Integrale statt Summen zur Berechnung von Wahrscheinlichkeiten dienen. Im Rahmen der Maßtheorie lassen sich Summen ebenfalls als Integrale (bezüglich des Zählmaßes) auffassen. Dies erlaubt dann eine einheitliche Behandlung von Wahrscheinlichkeitsmaßen, die durch eine Dichte f definiert sind, und rechtfertigt den Begriff „Zähldichte“ in Definition 1.11.

Beispiel 1.22

(a) Exponentialverteilungen auf $(0, \infty)$:

Für eine vorgegebene Konstante $\lambda > 0$ ist die Exponentialverteilung mit Intensitätsparameter λ , kurz $\text{Exp}(\lambda)$, auf $\Omega = (0, \infty)$, versehen mit $\mathcal{B}(\Omega)$, gegeben durch die Lebesguedichte $f_{\text{Exp}(\lambda)}$ mit

$$f_{\text{Exp}(\lambda)}(t) \equiv f(t|\lambda) = \lambda \exp(-\lambda t), t \in \Omega.$$

Man verifiziert leicht, dass $f(t|\lambda) > 0$ für alle $t \in \Omega$ sowie $\int_{\Omega} f(t|\lambda) dt = 1$ gelten. Exponentialverteilungen werden häufig zur Modellierung von Lebensdauern bzw. Wartezeiten verwendet.

(b) Stetige Gleichverteilungen auf $\Omega \subset \mathbb{R}^d$:

Ist Ω eine Borelmenge mit Volumen $0 < \mathfrak{M}^d(\Omega) < \infty$, so heißt das Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}(\Omega))$ mit der konstanten Dichtefunktion $f(x) \equiv 1/\mathfrak{M}^d(\Omega)$ die stetige Gleichverteilung auf Ω . Dies verallgemeinert (leicht) das Konzept der geometrischen Wahrscheinlichkeit aus Abschnitt 1.3.3. Ist $d = 1$ und $\Omega = [a, b]$ ein Intervall, so schreiben wir $UNI[a, b]$ für die stetige Gleichverteilung auf Ω .

(c) Normalverteilungen auf \mathbb{R} :

Für vorgegebene Konstanten $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ ist die (Gauß'sche) Normalverteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Parametern μ und σ^2 , kurz $\mathcal{N}(\mu, \sigma^2)$, gegeben durch die Lebesgue-dichte $f_{\mathcal{N}(\mu, \sigma^2)}$ mit

$$f_{\mathcal{N}(\mu, \sigma^2)}(x) \equiv f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R}.$$

Für $\mu = 0$ und $\sigma^2 = 1$ ergibt sich die Standardnormalverteilung $\mathcal{N}(0, 1)$ mit Dichtefunktion $x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$. Die Standardnormalverteilung spielt eine zentrale Rolle in der Stochastik wegen des zentralen Grenzwertsatzes, siehe Kapitel 10.

Um nachzuweisen, dass $f(\cdot|\mu, \sigma^2)$ tatsächlich eine Dichtefunktion ist, genügt es zu zeigen, dass $\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}$ ist (horizontale Verschiebung ändert den Integralwert nicht!). Dazu verwendet man typischerweise den folgenden Trick.

Lemma 1.23

Sei $f : \mathbb{R}^2 \rightarrow (0, \infty)$ gegeben durch $f(x, y) = \exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)$. Dann gilt

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \left[\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right]^2 = 2\pi\sigma^2.$$

Beweis:

Wir wenden die bivariate Substitutionsregel der Integralrechnung an und transformieren dazu auf Polarkoordinaten. Sei also $g : \mathbb{R}^2 \rightarrow [0, \infty) \times [0, 2\pi)$ gegeben durch

$$g(x, y) = (\sqrt{x^2 + y^2}, \arctan(y/x)) =: (r, \varphi) \implies g^{-1}(r, \varphi) = (r \cos(\varphi), r \sin(\varphi)).$$

[Winkelverschiebungen und singuläre Punkte können vernachlässigt werden.]

Es ist $f(g^{-1}(r, \varphi)) = \exp\left(-\frac{r^2}{2\sigma^2}\right)$. Bleibt, die Jacobi-Matrix $J(x, y)$ von g zu berechnen. Dies geschieht durch Betrachten von

$$\begin{aligned}\frac{\partial g_1(x, y)}{\partial x} &= \frac{2x}{2\sqrt{x^2 + y^2}} = \frac{x}{\sqrt{x^2 + y^2}}, \\ \frac{\partial g_1(x, y)}{\partial y} &= \frac{y}{\sqrt{x^2 + y^2}}, \\ \frac{\partial g_2(x, y)}{\partial x} &= \frac{1}{1 + y^2/x^2} \left(-\frac{y}{x^2}\right) = -\frac{y}{x^2 + y^2}, \\ \frac{\partial g_2(x, y)}{\partial y} &= \frac{1}{1 + y^2/x^2} \cdot \frac{1}{x} = \frac{1}{x + y^2/x} = \frac{x}{x^2 + y^2}.\end{aligned}$$

Folglich ergibt sich die Jacobi-Matrix

$$J(x, y) = \begin{pmatrix} \frac{x}{\sqrt{x^2 + y^2}} & \frac{y}{\sqrt{x^2 + y^2}} \\ -\frac{y}{x^2 + y^2} & \frac{x}{x^2 + y^2} \end{pmatrix}, \quad J(g^{-1}(r, \varphi)) = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\frac{\sin(\varphi)}{r} & \frac{\cos(\varphi)}{r} \end{pmatrix}$$

und

$$|J(g^{-1}(r, \varphi))| = \frac{\cos^2(\varphi)}{r} + \frac{\sin^2(\varphi)}{r} = \frac{1}{r}.$$

Damit erhalten wir schließlich

$$\begin{aligned}\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy &= \int_0^{2\pi} \int_0^{\infty} r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr d\varphi \\ &= 2\pi \int_0^{\infty} r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr = 2\pi \left[-\sigma^2 \exp\left(-\frac{r^2}{2\sigma^2}\right) \right]_0^{\infty} \\ &= 2\pi [0 + \sigma^2] = 2\pi\sigma^2. \quad \blacksquare\end{aligned}$$

Kapitel 2

Kombinatorik

Insbesondere für die Behandlung Laplace'scher Wahrscheinlichkeitsräume (siehe Beispiel 1.12.(a)) ist es hilfreich, einige Grundergebnisse der Kombinatorik (Lehre von der Anzahlbestimmung) zu kennen.

Lemma 2.1 (Additionsregel)

Sei A eine endliche Menge und es gelte $A = A_1 \cup A_2$ mit $A_1 \cap A_2 = \emptyset$. Dann ist $|A| = |A_1| + |A_2|$.

Lemma 2.2 (Multiplikationsregel)

Aus k Mengen A_1, \dots, A_k werden geordnete k -Tupel (m_1, \dots, m_k) gebildet, deren j -te Komponente in A_j liegt ($m_j \in A_j, 1 \leq j \leq k$). Außerdem unterliegen die Komponenten der Einschränkung, dass für alle $2 \leq j \leq k$ die j -te Komponente m_j bei gegebenen m_1, \dots, m_{j-1} genau n_j verschiedene Elemente aus A_j annehmen kann, deren Auswahl, nicht aber deren Anzahl, gegebenenfalls von den vorherigen Komponenten m_1, \dots, m_{j-1} abhängt. Sei A die Menge aller möglichen k -Tupel (unter diesen Voraussetzungen).

Dann gilt:

$$|A| = \prod_{j=1}^k n_j = n_1 \cdot n_2 \cdot \dots \cdot n_k.$$

Lemma 2.3 (Anzahl möglicher k -Permutationen von n Objekten mit Wiederholung)

Permutation $\hat{=}$ *geordnetes Tupel!*

$$\begin{aligned} A &= \{(m_1, \dots, m_k) \mid m_j \in M \forall 1 \leq j \leq k, |M| = n\} = M^k \\ \implies |A| &=: Pe^*(n, k) = \underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ Faktoren}} = n^k, \quad k \geq 1. \end{aligned}$$

Beispiel 2.4 (Geburtstagsparadoxon)

Gegeben sei eine Gruppe von k Personen, von denen keine am 29. Februar Geburtstag habe. Es werde angenommen, alle anderen 365 Geburtstage seien gleich wahrscheinlich. Wie groß ist dann

die Wahrscheinlichkeit, dass mindestens zwei der k Personen am gleichen Tag Geburtstag haben?
 Ab welchem Wert von k überschreitet diese Wahrscheinlichkeit den Wert $1/2$?

Lösung:

$$\Omega = \{(m_1, \dots, m_k) | 1 \leq m_j \leq 365 \forall 1 \leq j \leq k\} \implies |\Omega| = 365^k.$$

(Wir nummerieren die 365 Tage des Jahres durch.)

Sei $A := \{\text{Alle } k \text{ Geburtstage sind verschieden}\}$. Dann ist $|A| = 365 \cdot 364 \cdot \dots \cdot (365 - k + 1)$.
 Modellieren wir dieses Experiment mit dem Laplace'schen Wahrscheinlichkeitsraum $(\Omega, 2^\Omega, \mathbb{P})$
 mit der Gleichverteilung \mathbb{P} auf $(\Omega, 2^\Omega)$, so ist

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\prod_{j=0}^{k-1} (365 - j)}{365^k}$$

und

$$\mathbb{P}(A^c) = 1 - \frac{\prod_{j=0}^{k-1} (365 - j)}{365^k} =: q_k.$$

k	q_k
2	$1/365 \approx 0.00274$
5	0.02714
10	0.11695
15	0.2529
20	0.41144
23	0.507297

Tabelle 2.1: Tabelle zum Geburtstagsparadoxon

Lemma 2.5 (Anzahl möglicher k -Permutationen von n Objekten ohne Wiederholung)

$$A = \{(m_1, \dots, m_k) | m_j \in M \forall 1 \leq j \leq k, m_i \neq m_j \text{ für } i \neq j, |M| = n\}$$

$$\implies |A| =: Pe(n, k) = n(n-1) \cdot \dots \cdot (n - (k-1)) = \frac{n!}{(n-k)!}, \quad 1 \leq k \leq n.$$

Für $k = n$ ist $|A| = n!$.

Lemma 2.6 (Anzahl möglicher k -Kombinationen von n Objekten ohne Wiederholung)

Kombination $\hat{=}$ ungeordnetes Tupel! (Reihenfolge spielt keine Rolle)

$$A = \{\{m_1, \dots, m_k\} | m_j \in M \forall 1 \leq j \leq k, m_i \neq m_j \text{ für } i \neq j, |M| = n\}$$

$$\hat{=} \text{Menge der } k\text{-elementigen Teilmengen von } M, 1 \leq k \leq n.$$

Wir schreiben $C(n, k) := |A|$. Jedes Element aus A kann auf $k!$ verschiedene Arten angeordnet werden.

$$\begin{aligned} \implies C(n, k)k! &= Pe(n, k) = \frac{n!}{(n-k)!} \\ \implies C(n, k) &= \frac{n!}{k!(n-k)!} =: \binom{n}{k}. \end{aligned}$$

Definition und Lemma 2.7

Für $n \in \mathbb{N}_0$ und $0 \leq k \leq n$ heißt

$$\binom{n}{k} := \frac{n!}{k!(n-k)!},$$

wobei $0! = 1$, Binomialkoeffizient (sprich: „n über k“).

Es gilt:

(a)
$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = n$$

(b)
$$\binom{n}{k} = \binom{n}{n-k}$$

(c) (Pascal'sches Dreieck)
$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

$k \rightarrow$					
				1	
			1	1	
		1	2	1	
		1	3	3	1
$n \downarrow$	1	4	6	4	1

(d) Binomischer Lehrsatz:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

für $a, b \in \mathbb{R}$ und $n \in \mathbb{N}_0$.

Beweis:

Teile (a) bis (c) zur Übung. Zu Teil (d) führen wir einen Induktionsbeweis.

$$(a + b)^0 = 1 = \sum_{k=0}^0 \binom{0}{k} a^k b^{0-k} = \binom{0}{0} a^0 b^0 = 1 \cdot 1 \cdot 1, \text{ wie gewünscht.}$$

Für den Induktionsschritt beachten wir, dass

$$(a + b)^n = (a + b)^{n-1} (a + b)^1. \quad (2.1)$$

Nach Induktionsvoraussetzung ist die rechte Seite von (2.1) gleich

$$\left[\sum_{k=0}^{n-1} \binom{n-1}{k} a^k b^{n-1-k} \right] (a + b) = \sum_{k=0}^{n-1} \binom{n-1}{k} a^{k+1} b^{n-1-k} + \sum_{j=0}^{n-1} \binom{n-1}{j} a^j b^{n-j} \quad (2.2)$$

Indextransformation:

$$\ell := k + 1 \Leftrightarrow k = \ell - 1$$

$$k = 0 \Rightarrow \ell = 1$$

$$k = n - 1 \Rightarrow \ell = n$$

Damit ist die rechte Seite von (2.2) gleich

$$\begin{aligned} & \sum_{\ell=1}^n \binom{n-1}{\ell-1} a^\ell b^{n-\ell} + \sum_{j=0}^{n-1} \binom{n-1}{j} a^j b^{n-j} \\ &= \binom{n-1}{0} a^0 b^n + \sum_{k=1}^{n-1} a^k b^{n-k} \left[\binom{n-1}{k-1} + \binom{n-1}{k} \right] + \binom{n-1}{n-1} a^n b^0 \\ &= \binom{n}{0} a^0 b^n + \sum_{k=1}^{n-1} \binom{n}{k} a^k b^{n-k} + \binom{n}{n} a^n b^0 = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad \blacksquare \end{aligned}$$

Beispiel 2.8 (Urnenmodell)

Gegeben sei eine Urne mit n nummerierten Kugeln (o. B. d. A: $1, \dots, n$). Wir ziehen gleichzeitig $1 \leq k \leq n$ (unterschiedliche) Kugeln aus dieser Urne. Damit ist

$$\Omega = \{ \{m_1, \dots, m_k\} \mid m_j \in \{1, \dots, n\} \forall 1 \leq j \leq k, m_i \neq m_j \text{ für } i \neq j \} \text{ und } |\Omega| = \binom{n}{k}.$$

Nehmen wir als Modell den Laplace'schen Wahrscheinlichkeitsraum auf $(\Omega, 2^\Omega)$ an, so ist

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{\binom{n}{k}}$$

für alle $\omega \in \Omega$.

Sei nun $A = \{ \text{Kugel } j^* \text{ wird nicht gezogen} \}, 1 \leq j^* \leq n$.

Es gilt:

$$|A| = \binom{n-1}{k}$$

und damit

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \binom{n-1}{k} \binom{n}{k}^{-1} = \frac{(n-1)!k!(n-k)!}{k!(n-1-k)!n!} = \frac{n-k}{n} = 1 - \frac{k}{n}.$$

Folglich ergibt sich, dass

$$\mathbb{P}(\{\text{Kugel } j^* \text{ wird gezogen}\}) = \mathbb{P}(A^c) = \frac{k}{n}.$$

Übungsaufgabe: Die Wahrscheinlichkeit für A bleibt gleich, wenn wir Ziehen mit Berücksichtigung der Reihenfolge betrachten!

Lemma 2.9 (Anzahl möglicher k -Kombinationen von n Objekten mit Wiederholung)

$$A = \{[m_1, \dots, m_k] \mid m_j \in M \forall 1 \leq j \leq k, |M| = n\}$$

Wir kodieren die Elemente von A als (geordnete) $(n+k-1)$ -Tupel um. Zu diesem Zweck sei ω die Menge $M = \{1, \dots, n\}$. Wir starten mit der Auswahlmöglichkeit „1“ und notieren für ein Tupel $\omega = [m_1, \dots, m_k] \in A$ so viele „G“ (gewählt) hintereinander, wie es m_j in ω mit $m_j = 1$ gibt. Danach notieren wir ein „N“ (neues Element). Sodann notieren wir wiederum so oft ein „G“ hintereinander, wie es m_j in ω mit $m_j = 2$ gibt, usw.

Ist etwa $n = 5, k = 3$ und $\omega = [2 \ 1 \ 1]$, so führt das zum Notieren von $G \ G \ N \ G \ N \ N \ N$. Offenbar gibt es sieben (allgemein: $n+k-1$) Positionen, auf die man die drei (allgemein: k) „G“ platzieren kann. Wir haben es also mit einer Auswahl von k Plätzen aus $n+k-1$ Möglichkeiten zu tun.

$$\Rightarrow |A| = C(n+k-1, k) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} := C^*(n, k)$$

Übungsaufgabe: Eisverkäufer/innen-Problem.

Schema 2.10

Berücksichtigung der Reihenfolge	Wiederholung	
	ja	nein
ja	$\text{Pe}^*(n, k) = n^k$	$\text{Pe}(n, k) = \frac{n!}{(n-k)!}$
nein	$C^*(n, k) = \binom{n+k-1}{k}$	$C(n, k) = \binom{n}{k}$

Lemma 2.11 (Anzahl möglicher Permutationen mit vorgegebenen Besetzungszahlen)

Seien $n \in \mathbb{N}$ und $k \in \mathbb{N}$ gegeben. Ferner sei ein Tupel $(n_1, \dots, n_k) \in \mathbb{N}^k$ derart gegeben, dass $0 \leq n_j \leq n \forall 1 \leq j \leq k$ und $\sum_{j=1}^k n_j = n$ gilt. Betrachte

$$A = \{(m_1, \dots, m_n) : m_i \in M \forall 1 \leq i \leq n, |M| = k,$$

jedes Element j von M kommt genau n_j -mal in (m_1, \dots, m_n) vor\}.

Sukzessives Auswählen der jeweils n_j Plätze für die k Elemente ($1 \leq j \leq k$) der Menge M liefert nach Multiplikationsregel, dass

$$\begin{aligned} |A| &= \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdot \dots \cdot \binom{n-\sum_{j=1}^{k-1} n_j}{n_k} \\ &= \frac{n!}{n_1!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \times \dots \times \frac{(n-\sum_{j=1}^{k-1} n_j)!}{n_k!0!} \\ &= \frac{n!}{\prod_{j=1}^k n_j!} =: \binom{n}{n_1, n_2, \dots, n_k} \quad \text{(Multinomialkoeffizient)}. \end{aligned}$$

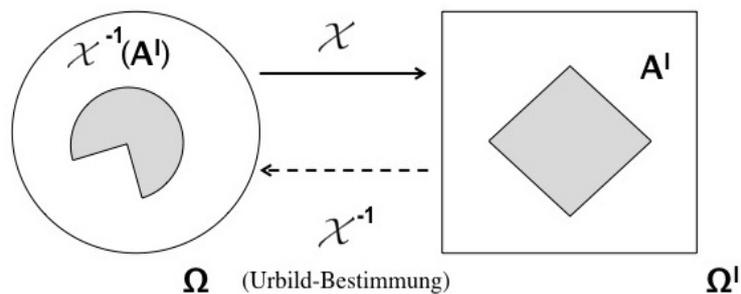
Für $k = 2$ ergibt sich

$$|A| = \binom{n}{n_1, n_2} = \binom{n}{n_1} = C(n, n_1).$$

Kapitel 3

Zufallsvariablen und ihre Verteilungen

Ziel: Studieren von Abbildungen (Transformationen) von einem messbaren Raum (Ω, \mathcal{A}) in einen anderen messbaren Raum (Ω', \mathcal{A}') . In Schema 3.1 wird eine solche Abbildung mit X bezeichnet.



Schema 3.1

Frage: Falls $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum ist, wie kann dann X dazu benutzt werden, ein Wahrscheinlichkeitsmaß \mathbb{P}^X auf (Ω', \mathcal{A}') zu definieren? Da \mathbb{P} auf Mengen in \mathcal{A} operiert, lautet eine nahe liegende Forderung:

$$A' \in \mathcal{A}' \implies X^{-1}(A') \in \mathcal{A}. \quad (3.1)$$

Abbildungen zwischen messbaren Räumen, die (3.1) erfüllen, heißen Zufallsvariablen.

Definition 3.2

Seien (Ω, \mathcal{A}) und (Ω', \mathcal{A}') zwei messbare Räume. Dann heißt jede Abbildung $X : \Omega \rightarrow \Omega'$, die die Messbarkeitseigenschaft

$$A' \in \mathcal{A}' \implies X^{-1}(A') \in \mathcal{A}$$

erfüllt, eine Zufallsvariable von (Ω, \mathcal{A}) nach (Ω', \mathcal{A}') .

Schreibweise:

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

Beispiel 3.3

Betrachte den zweifachen Würfelwurf mit $\Omega = \{1, \dots, 6\}^2$ und $\mathcal{A} = 2^\Omega$, und die Abbildung $X : \Omega \rightarrow \{2, \dots, 12\} = \Omega'$, versehen mit $\mathcal{A}' = 2^{\Omega'}$, die jedem Zweiertupel aus Ω die Augensumme zuordnet. Wir erhalten zum Beispiel

$$X^{-1}(\{2\}) = \{(1, 1)\},$$

$$X^{-1}(\{7\}) = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\},$$

$$X^{-1}(\{2, 7\}) = \{(1, 1), (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

Offenbar ist X eine Zufallsvariable, da $\mathcal{A} = 2^\Omega$ alle Teilmengen von Ω enthält.

Lemma 3.4

Ist in Definition 3.2 $\mathcal{A} = 2^\Omega$, so ist jede Abbildung $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ messbar und somit eine Zufallsvariable.

Lemma 3.5

Wird in Definition 3.2 die σ -Algebra \mathcal{A}' erzeugt von einem Mengensystem \mathcal{F}' , d.h., $\mathcal{A}' = \sigma(\mathcal{F}')$, so ist $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ bereits dann eine Zufallsvariable, wenn die Bedingung $X^{-1}(A') \in \mathcal{A}$ nur für alle $A' \in \mathcal{F}'$ gilt.

Beweis:

Das Mengensystem

$$\mathcal{G}' := \{A' \subseteq \Omega' : X^{-1}(A') \in \mathcal{A}\}$$

ist eine σ -Algebra (siehe Übungsaufgabe). Nach Voraussetzung umfasst \mathcal{G}' das Erzeugendensystem \mathcal{F}' . Da nach Konstruktion jedoch \mathcal{A}' die kleinste solche σ -Algebra ist, gilt $\mathcal{G}' \supseteq \mathcal{A}'$. Demnach ist X messbar. ■

Korollar 3.6 (Stetige Funktionen)

Sei $\Omega \subseteq \mathbb{R}^d$ und $\mathcal{A} = \mathcal{B}(\Omega)$. Dann ist jede stetige Funktion $X : (\Omega, \mathcal{B}(\Omega)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Zufallsvariable.

Beweis:

Die Borel'sche σ -Algebra $\mathcal{B}(\mathbb{R})$ wird erzeugt von dem System der Halbstrahlen

$$\mathcal{F}' = \{] - \infty, c] : c \in \mathbb{R}\},$$

siehe Beispiel 1.5.(b). Nun ist aber für jedes $c \in \mathbb{R}$ die Menge $\{X \leq c\}$ abgeschlossen in Ω , gehört also gemäß Beispiel 1.5.(d) in Analogie zu Beispiel 1.5.(b) zu $\mathcal{B}(\Omega)$. Damit folgt die Aussage aus Lemma 3.5. ■

Satz und Definition 3.7

Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\Omega', \mathcal{A}')$ eine Zufallsvariable von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ in einen messbarem Raum (Ω', \mathcal{A}') . Dann wird durch

$$\mathbb{P}^X(A') := \mathbb{P}(X^{-1}(A')) = \mathbb{P}(\{X \in A'\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A'\})$$

für $A' \in \mathcal{A}'$ ein Wahrscheinlichkeitsmaß \mathbb{P}^X auf (Ω', \mathcal{A}') definiert.

Das Wahrscheinlichkeitsmaß \mathbb{P}^X heißt Bildmaß von \mathbb{P} unter X oder Verteilung von X (auf (Ω', \mathcal{A}')), in Zeichen: $\mathbb{P}^X = \mathbb{P} \circ X^{-1} = \mathcal{L}(X)$ („law of X “).

Bemerkung 3.8

Die Verteilung der Identität $id: (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\Omega, \mathcal{A})$ ist natürlich $\mathbb{P}^{id} = \mathbb{P}$. Folglich existiert zu jedem Wahrscheinlichkeitsmaß \mathbb{P} auf einem messbaren Raum (Ω, \mathcal{A}) eine Zufallsvariable mit Werten in (Ω, \mathcal{A}) , deren Verteilung gerade \mathbb{P} ist. Dies liefert auch die Begründung, warum wir bereits zuvor verschiedentlich Wahrscheinlichkeitsmaße als „Verteilungen“ bezeichnet hatten (Binomialverteilung, Poisson-Verteilung, etc.).

Bemerkung und Definition 3.9

Wegen Satz 1.10 in Verbindung mit Beispiel 1.5.(b) ist jedes Wahrscheinlichkeitsmaß \mathbb{P} auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ bereits eindeutig festgelegt durch die Funktion $F_{\mathbb{P}}$, gegeben durch $F_{\mathbb{P}}(c) = \mathbb{P}(] - \infty, c])$ für $c \in \mathbb{R}$. Ebenso ist die Verteilung $\mathcal{L}(X)$ einer reellwertigen Zufallsvariablen X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ bereits festgelegt durch die Funktion F_X mit

$$F_X(c) := \mathbb{P}(X \leq c) := \mathbb{P}(\{X \leq c\}), \quad c \in \mathbb{R}.$$

Die Funktion $F_{\mathbb{P}}$ heißt Verteilungsfunktion von \mathbb{P} und die Funktion F_X heißt Verteilungsfunktion von X (bzw. von $\mathcal{L}(X)$).

Es gilt:

(i) $F_X = F_{\mathbb{P} \circ X^{-1}}$

(ii) Jede Verteilungsfunktion $F = F_X$ hat die folgenden Eigenschaften:

(a) F ist monoton wachsend.

(b) F ist rechtsseitig stetig.

(c) $\lim_{c \rightarrow -\infty} F(c) = 0$ und $\lim_{c \rightarrow +\infty} F(c) = 1$.

Bemerkung 3.10

Bei konkreten Zufallsvorgängen in der Praxis gibt man oft nur die interessierende (reellwertige) Zufallsvariable X mit ihrer (modellierten) Verteilung(sfunktion) an, z. B.

- Anzahl X an Salmonellen-Neuinfektionen in einer definierten Zielpopulation in einem definierten Zeitraum: $\mathcal{L}(X) = \text{Poisson}(\lambda)$.
- Anzahl X der Stimmen für eine bestimmte Partei bei einer festgelegten Wahl:
 $\mathcal{L}(X) = \text{Bin}(n, p)$.

Ist bei solchen Anwendungen nur noch $\mathcal{L}(X) = \mathbb{P}^X$ von Interesse, so wird oft der zu Grunde liegende Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ gar nicht mehr explizit erwähnt. Lediglich das Symbol \mathbb{P} in Ausdrücken der Form $\mathbb{P}(X \in A')$ erinnert noch an ihn.

Kapitel 4

Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit

4.1 Bedingte Wahrscheinlichkeit

Beispiel 4.1 (Zahlenlotto „6 aus 49“)

Die Ziehung der Lottozahlen „6 aus 49“ lässt sich durch einen Laplace'schen Wahrscheinlichkeitsraum $(\Omega, 2^\Omega, \mathbb{P})$ modellieren mit

$$\Omega = \{\omega \subset \{1, \dots, 49\} : |\omega| = 6\},$$

so dass

$$|\Omega| = \binom{49}{6} \approx 1,4 \cdot 10^7 \quad \text{und} \quad \mathbb{P} : \text{diskrete Gleichverteilung auf } (\Omega, 2^\Omega).$$

Frau N. spielt Lotto. Die Wahrscheinlichkeit dafür, dass sie „6 Richtige“ getippt hat, ist

$$\mathbb{P}(\{\omega^*\}) = \frac{1}{|\Omega|} \approx 7,2 \cdot 10^{-8}, \quad \omega^* = \text{Tipp von Frau N.}$$

Angenommen, Frau N. verfolgt die Ziehung live und hat nach dem Ziehen der ersten fünf Kugeln bereits „5 Richtige“. Gegeben diese Information ist die Wahrscheinlichkeit, dass sie auch „6 Richtige“ getippt hat, gleich $\frac{1}{44}$, weil hierzu nur noch die fehlende Zahl aus den verbleibenden 44 Kugeln gezogen werden muss.

Definition 4.2

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ ein Ereignis. Dann ist die bedingte Wahrscheinlichkeit von $A \in \mathcal{A}$ gegeben (unter der Bedingung) B definiert durch

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

Beispiel 4.3 (Beispiel 4.1 fortgesetzt)

Unter den Gegebenheiten von Beispiel 4.1 sei

$$A = \{\omega^*\} = \text{„6 Richtige“ und } B = \{\text{„5 Richtige“ nach fünf gezogenen Kugeln}\}.$$

Dann gilt:

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{\omega^*\})}{\mathbb{P}(B)} \\ &= \frac{1/\binom{49}{6}}{6/\binom{49}{5}} = \frac{1}{6} \cdot \frac{\binom{49}{5}}{\binom{49}{6}} \\ &= \frac{1}{6} \cdot \frac{49! \cdot 6! \cdot 43!}{5! \cdot 44! \cdot 49!} \\ &= \frac{1}{6} \cdot \frac{6}{44} = \frac{1}{44}. \end{aligned}$$

Satz 4.4

Unter den Voraussetzungen von Definition 4.2 sind die auf B bedingten Wahrscheinlichkeiten für Ereignisse $A \in \mathcal{A}$ bereits festgelegt durch die bedingten Wahrscheinlichkeiten $\mathbb{P}(C|B)$ mit $C \subseteq B$. Das Mengensystem $\mathcal{A}_B := \{C \in \mathcal{A} | C \subseteq B\}$ ist eine σ -Algebra über B . Fasst man B als einen neuen Ergebnisraum auf, so definiert die auf B bedingte Wahrscheinlichkeit ein Wahrscheinlichkeitsmaß, d. h., $\mathbb{P}(\cdot|B) : \mathcal{A}_B \rightarrow [0, 1]$ ist ein Wahrscheinlichkeitsmaß auf (B, \mathcal{A}_B) .

Beweis: Übungsaufgabe.

Korollar 4.5 (Rechenregeln der bedingten Wahrscheinlichkeit)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A, B, A_1, \dots, A_n Ereignisse in \mathcal{A} , so dass alle folgenden bedingten Wahrscheinlichkeiten definiert sind.

Dann gilt:

a) $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$

(b) $\mathbb{P}(\bigcap_{i=1}^n A_i) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \times \dots \times \mathbb{P}(A_n|\bigcap_{j=1}^{n-1} A_j)$
(Kettenfaktorisierung)

(c) Falls $A_1 \supset A_2 \supset \dots \supset A_n$, so folgt

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \mathbb{P}(A_n) = \mathbb{P}(A_1) \prod_{j=2}^n \mathbb{P}(A_j|A_{j-1}).$$

Bemerkung 4.6 (Pfad- bzw. Pfeildiagramm)

Die Eigenschaften (b) und (c) lassen sich grafisch in einem Baum darstellen, dessen Knoten die Ereignisse und dessen (gerichtete) Kanten die Inklusionen repräsentieren.

Beispiel 4.7 (Beispiel 4.1 fortgesetzt)

Bezeichne beim Lotto „6 aus 49“ A_i das Ereignis, dass nach Ziehen der i -ten Kugel bereits „ i Richtige“ für dem Tipp ω^* vorliegen, $1 \leq i \leq 6$. Dann ergibt sich das folgende Schema:

$$A_1 \xrightarrow{\frac{5}{48}} A_2 \xrightarrow{\frac{4}{47}} A_3 \xrightarrow{\frac{3}{46}} A_4 \xrightarrow{\frac{2}{45}} A_5 \xrightarrow{\frac{1}{44}} A_6$$

Es gilt:

$$\begin{aligned} \mathbb{P}(\{\omega^*\}) &= \mathbb{P}(A_6) \\ &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_2) \cdot \dots \cdot \mathbb{P}(A_6|A_5) \\ &= \frac{6}{49} \cdot \frac{5}{48} \cdot \frac{4}{47} \cdot \frac{3}{46} \cdot \frac{2}{45} \cdot \frac{1}{44} \\ &=: \frac{6!}{(49)_6} = \frac{6!}{49!/43!} = \frac{6! \cdot 43!}{49!} = \frac{1}{\binom{49}{6}}. \end{aligned}$$

Satz 4.8 (Satz von der totalen Wahrscheinlichkeit, Zerlegungsformel)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Es sei $(B_i)_{i \in I}$ eine disjunkte Zerlegung von Ω , wobei I eine höchstens abzählbare Indexmenge bezeichnet, und es gelte $\mathbb{P}(B_i) > 0 \forall i \in I$. Es sei also

$$\bigcup_{i \in I} B_i = \Omega \quad \text{und} \quad B_i \cap B_j = \emptyset, \quad i \neq j.$$

Dann gilt für alle $A \in \mathcal{A}$, dass

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i).$$

Beweis:

$$\begin{aligned} \sum_{i \in I} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i) &= \sum_{i \in I} \mathbb{P}(A \cap B_i) = \mathbb{P}\left(\bigcup_{i \in I} \{A \cap B_i\}\right) \\ &= \mathbb{P}\left(A \cap \bigcup_{i \in I} B_i\right) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A), \end{aligned}$$

wobei wir ausgenutzt haben, dass die B_i , $i \in I$, disjunkt sind und das Distributivgesetz der Mengenlehre angewendet worden ist. ■

Beispiel 4.9 (Gambler's Ruin)

Sie betreten ein Spielkasino mit einem Kapital von k Euro. Sie spielen Roulette und setzen in jeder Runde 1 Euro auf das Ereignis $R = \{\text{Rot}\}$. Tritt R ein, so erhalten Sie 2 Euro, anderenfalls wird Ihr Einsatz von 1 Euro von der Spielbank einbehalten.

Laplace'scher Wahrscheinlichkeitsraum liefert:

$$p := \mathbb{P}(R) = \frac{18}{37} < 1/2$$

Sie legen von vorne herein ein Zielkapital $K \geq k$ [Euro] fest und beenden das Spiel, sobald Sie K Euro besitzen oder alles verloren haben.

Gesucht: Ruinwahrscheinlichkeit $p_k := \mathbb{P}(A_k)$, $A_k = \{\text{Ruin bei Anfangskapital } k \text{ [Euro]}\}$

Für $0 < k < K$ liefert der Satz von der totalen Wahrscheinlichkeit, dass

$$\begin{aligned} \mathbb{P}(A_k) &= \mathbb{P}(R) \cdot \mathbb{P}(A_k|R) + \mathbb{P}(R^c) \cdot \mathbb{P}(A_k|R^c) \\ &= \mathbb{P}(R) \cdot \mathbb{P}(A_{k+1}) + \mathbb{P}(R^c) \cdot \mathbb{P}(A_{k-1}) \\ \iff p_k &= p \cdot p_{k+1} + (1-p) \cdot p_{k-1}. \end{aligned} \quad (4.1)$$

Wir lösen (4.1) unter den zwei Randbedingungen $p_0 = 1$ und $p_K = 0$. Definiere dazu für $0 < p < 1$ das Verhältnis $r := \frac{1-p}{p}$ sowie $d_k := p_k - p_{k+1}$. Wir beachten, dass $p_k = p \cdot p_{k+1} + (1-p) \cdot p_{k-1}$. Einsetzen in (4.1) liefert

$$\begin{aligned} p_k - p_{k+1} &= \frac{1-p}{p}(p_{k-1} - p_k) = \left(\frac{1-p}{p}\right)^k (1 - p_1) \\ \implies d_k &= r d_{k-1} = r^k d_0. \end{aligned}$$

Beachte ferner $1 = p_0 - p_K = \sum_{k=0}^{K-1} (p_k - p_{k+1}) = \sum_{k=0}^{K-1} d_k = \sum_{k=0}^{K-1} r^k d_0$.

Geometrische Summenformel liefert daher:

$$1 = \begin{cases} K d_0, & \text{falls } p = 1-p = 1/2 \implies r = 1 \\ \frac{1-r^K}{1-r} d_0, & \text{falls } p \neq 1-p \neq 1/2 \end{cases} \implies d_0 = \begin{cases} 1/K, & p = 1/2 \\ \frac{1-r}{1-r^K}, & p \neq 1/2 \end{cases} \quad (4.2)$$

Analog ergibt sich

$$\begin{aligned}
 p_k &= p_k - p_K = \sum_{i=k}^{K-1} (p_i - p_{i+1}) \\
 &= \sum_{i=k}^{K-1} d_i = \sum_{i=k}^{K-1} r^i d_0 \\
 &= \begin{cases} (K-k)d_0, & \text{falls } p = 1/2 \\ \frac{r^k - r^K}{1-r} d_0, & \text{falls } p \neq 1/2. \end{cases} \quad (4.3)
 \end{aligned}$$

Nehmen wir (4.2) und (4.3) zusammen, ergibt sich schließlich

$$p_k = \begin{cases} \frac{K-k}{K}, & \text{falls } p = 1/2 \\ \frac{r^k - r^K}{1-r^K}, & \text{falls } p \neq 1/2. \end{cases} \quad \blacksquare$$

Satz 4.10 (Satz von Bayes, nach Rev. Thomas Bayes (18. Jhdt.))

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ Ereignisse mit $\mathbb{P}(A) > 0$ und $\mathbb{P}(B) > 0$.

Dann gilt:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}$$

Beweis:

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) \implies \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B|A). \quad \blacksquare$$

Bemerkung 4.11

$\mathbb{P}(B)$ heißt a priori-Wahrscheinlichkeit von B und $\mathbb{P}(B|A)$ heißt a posteriori-Wahrscheinlichkeit von B . Fasst man B als eine Ursache und A als eine Wirkung auf, so liefert $\mathbb{P}(B|A)$ die Wahrscheinlichkeit dafür, dass A aufgrund der Ursache B aufgetreten ist. Der Satz von Bayes (und seine Verallgemeinerungen) bilden die Grundlage der „Bayesianischen Statistik“.

Beispiel 4.12

Drei Maschinen produzieren das gleiche Teil. Die Tagesproduktionen (in Stück) seien gegeben durch

$$\left. \begin{array}{l} 1. \text{ Maschine: } 6000 \\ 2. \text{ Maschine: } 1000 \\ 3. \text{ Maschine: } 3000 \end{array} \right\}.$$

Der durchschnittliche Ausschussanteil (erwartete relative Häufigkeit von produzierten Stücken, die eine gewisse Qualitätsnorm nicht erfüllen) sei

$$\left. \begin{array}{l} 1. \text{ Maschine: } 10\% \\ 2. \text{ Maschine: } 8\% \\ 3. \text{ Maschine: } 15\% \end{array} \right\}.$$

Angenommen, Sie bekommen ein Stück geliefert, das sich als Ausschuss erweist. Berechnen Sie für $1 \leq i \leq 3$ die Wahrscheinlichkeit \mathbb{P} („Dieses Stück wurde von Maschine i produziert“).

Lösung:

Sei $A = \{\text{Ein produziertes Teil ist Ausschuss.}\}$. Der Satz von der totalen Wahrscheinlichkeit liefert

$$\mathbb{P}(A) = \sum_{i=1}^3 \mathbb{P}(A|B_i)\mathbb{P}(B_i),$$

wobei $B_i = \{\text{Stück wurde von Maschine } i \text{ produziert}\}$. Nach Voraussetzungen ist $\mathbb{P}(A) = 0,1 \cdot 0,6 + 0,08 \cdot 0,1 + 0,15 \cdot 0,3 = 0,113$. Nach dem Satz von Bayes ergeben sich somit

$$\begin{aligned} \mathbb{P}(B_1|A) &= \frac{0,6 \cdot 0,1}{0,113} \approx 53\%, \\ \mathbb{P}(B_2|A) &= \frac{0,1 \cdot 0,08}{0,113} \approx 7\%, \\ \mathbb{P}(B_3|A) &= \frac{0,3 \cdot 0,15}{0,113} \approx 40\%. \end{aligned}$$

4.2 Stochastische Unabhängigkeit von Ereignissen

Definition 4.13

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

a) Zwei Ereignisse $A, B \in \mathcal{A}$ heißen stochastisch unabhängig (in Zeichen: $A \perp\!\!\!\perp B$), falls

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

b) Für eine beliebige Indexmenge $I \neq \emptyset$ heißen $(A_i)_{i \in I}$ mit $A_i \in \mathcal{A} \forall i \in I$ stochastisch unabhängig, falls für jede nicht-leere, endliche Teilmenge $K \subseteq I$ gilt, dass

$$\mathbb{P}\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} \mathbb{P}(A_k).$$

Bemerkung 4.14

Gilt in Definition 4.13.a) zusätzlich $\mathbb{P}(B) > 0$, so ist

$$A \perp\!\!\!\perp B \iff \mathbb{P}(A|B) = \mathbb{P}(A).$$

Das bedeutet, dass die Bedingung B die Wahrscheinlichkeitsbewertung von A nicht ändert.

Beispiel 4.15

Sei $(\Omega, 2^\Omega, \mathbb{P})$ mit $\Omega = \{1, \dots, 8\}$ ein Laplace'scher Wahrscheinlichkeitsraum. Seien

$$A_1 = \{1, 2, 3, 4\}, \quad A_2 = \{1, 2, 5, 6\}, \quad A_3 = \{3, 4, 5, 6\}.$$

Dann gilt $\forall 1 \leq i < j \leq 3$, dass $A_i \perp\!\!\!\perp A_j$, denn

$$\mathbb{P}(A_i \cap A_j) = \frac{2}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \frac{4}{8} \cdot \frac{4}{8}, \quad 1 \leq i < j \leq 3.$$

Allerdings sind (A_1, A_2, A_3) nicht stochastisch unabhängig, denn

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}.$$

4.3 Produkte diskreter Wahrscheinlichkeitsräume

Definition 4.16

Seien $n \in \mathbb{N}$ und $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ für $1 \leq i \leq n$ diskrete Wahrscheinlichkeitsräume mit $\mathcal{A}_i = 2^{\Omega_i}$ für $1 \leq i \leq n$. Als Modell für das simultane und voneinander unabhängige Ausführen der n Zufallsexperimente, die zu den n Wahrscheinlichkeitsräumen gehören, definieren wir den Produktraum $(\Omega, \mathcal{A}, \mathbb{P})$ vermittelt

$$\Omega = \prod_{i=1}^n \Omega_i = \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \Omega_i \forall 1 \leq i \leq n\},$$

$$\mathcal{A} = 2^\Omega = \bigotimes_{i=1}^n \mathcal{A}_i \quad (\text{vgl. Beispiel 1.5.(d)})$$

und

$$\mathbb{P}((\omega_1, \omega_2, \dots, \omega_n)) = \prod_{i=1}^n \mathbb{P}_i(\omega_i), \quad \forall (\omega_1, \omega_2, \dots, \omega_n) \in \Omega.$$

Man schreibt

$$\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i \quad \text{und} \quad (\Omega, \mathcal{A}, \mathbb{P}) =: \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i).$$

Korollar und Definition 4.17

Unter den Voraussetzungen von Definition 4.16 heißt die Zufallsvariable

$$\pi_i : \Omega \rightarrow \Omega_i,$$

$$\omega = (\omega_1, \dots, \omega_n) \mapsto \omega_i$$

die Projektion auf die i -te Koordinate.

Das gleichzeitige Eintreten von Ereignissen $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$ lässt sich beschreiben durch

$$\bigcap_{i=1}^n \{\pi_i \in A_i\} = A_1 \times A_2 \times \cdots \times A_n = \bigotimes_{i=1}^n A_i.$$

Für jedes $\emptyset \neq K \subseteq \{1, \dots, n\}$ gilt

$$\mathbb{P}\left(\bigcap_{k \in K} \{\pi_k \in A_k\}\right) = \prod_{k \in K} \mathbb{P}(\pi_k \in A_k),$$

d. h., die Ereignisse $\{\pi_1 \in A_1\}, \dots, \{\pi_n \in A_n\}$ sind stochastisch unabhängig.

Beweis:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k \in K} \{\pi_k \in A_k\}\right) &= \mathbb{P}\left(\bigotimes_{k \in K} A_k\right) = \mathbb{P}\left(\bigotimes_{k \in K} \{\omega_k \in \Omega_k : \omega_k \in A_k\}\right) \\ &= \mathbb{P}\left(\{(\omega_k)_{k \in K} : \omega_k \in A_k \forall k \in K\}\right) \\ &= \sum_{\substack{(\omega_k)_{k \in K} \\ \omega_k \in A_k \\ \forall k \in K}} \prod_{k \in K} \mathbb{P}_k(\omega_k) = \prod_{k \in K} \sum_{\omega_k \in A_k} \mathbb{P}_k(\omega_k) \\ &= \prod_{k \in K} \mathbb{P}_k(A_k) = \prod_{k \in K} \mathbb{P}(\pi_k \in A_k). \quad \blacksquare \end{aligned}$$

Beispiel 4.18 (Binomialverteilung)

Unter den Voraussetzungen von Definition 4.16 sei $\Omega_i = \{0, 1\}$ und $\mathbb{P}_i = \text{Bernoulli}(p)$ für alle $1 \leq i \leq n$, vgl. Beispiel 1.12.(b). Somit ist $\Omega = \{0, 1\}^n$ und

$$\mathbb{P} = (\text{Bernoulli}(p))^{\otimes n} := \bigotimes_{i=1}^n \text{Bernoulli}(p).$$

Folglich ist $\mathbb{P}((\omega_1, \dots, \omega_n)) = p^k(1-p)^{n-k}$ für alle $(\omega_1, \dots, \omega_n) \in \Omega = \{0, 1\}^n$, wobei $k = |\{1 \leq i \leq n : \omega_i = 1\}| = \sum_{i=1}^n \omega_i$ die Anzahl der beobachteten „Treffer“ bezeichnet.

Betrachte nun die (zufällige) Trefferanzahl, d.h., die Zufallsvariable $X : \Omega \rightarrow \{0, \dots, n\}$ mit $X((\omega_1, \dots, \omega_n)) = \sum_{i=1}^n \omega_i$. Dann ist X binomialverteilt mit Parametern n und p , also

$$\forall k \in \{0, 1, \dots, n\} : \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

vgl. Beispiel 1.12.(c).

Beispiel 4.19 (Produktmaß von diskreten Gleichverteilungen)

Unter den Voraussetzungen von Definition 4.16 sei Ω_i endlich und \mathbb{P}_i die diskrete Gleichverteilung auf Ω_i mit

$$\mathbb{P}_i(\{\omega_i\}) = \frac{1}{m_i} \quad \forall \omega_i \in \Omega_i, \quad m_i = |\Omega_i| \implies m := |\Omega| = \left| \bigotimes_{i=1}^n \Omega_i \right| = \prod_{i=1}^n m_i.$$

Es folgt, dass $\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$ die Gleichverteilung auf Ω ist, denn

$$\forall (\omega_1, \dots, \omega_n) \in \Omega : \mathbb{P}((\omega_1, \dots, \omega_n)) = \prod_{i=1}^n \mathbb{P}_i(\omega_i) = \prod_{i=1}^n \frac{1}{m_i} = \frac{1}{m}.$$

Beispielsweise liefert der doppelte Würfelwurf mit zwei homogenen und unabhängig voneinander geworfenen Würfeln die Gleichverteilung auf $\{1, \dots, 6\}^2$ mit

$$\mathbb{P}((i_1, i_2)) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}_1(i_1)\mathbb{P}_2(i_2) \quad \forall (i_1, i_2) \in \{1, \dots, 6\}^2.$$

4.4 Produkte stetiger Wahrscheinlichkeitsräume

Satz 4.20 (Satz von Fubini, siehe Seiten 88-89 in Forster (2012))

Seien k, ℓ natürliche Zahlen mit $k + \ell = n$. Sei ferner $f : \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^\ell \rightarrow \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ eine Funktion, die die Messbarkeitseigenschaft (1.3) (vgl. Lebesgue-Integral) erfüllt. Schreibe $f : (x_1, x_2) \mapsto f(x_1, x_2)$ mit $x_1 \in \mathbb{R}^k$ und $x_2 \in \mathbb{R}^\ell$.

Dann gilt für das Lebesgue-Integral:

$$\int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}^\ell} \left[\int_{\mathbb{R}^k} f(x_1, x_2) dx_1 \right] dx_2 = \int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^\ell} f(x_1, x_2) dx_2 \right] dx_1. \quad (4.4)$$

Iteratives Anwenden von (4.4) ergibt für

$$f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$$

$$x = (x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n), x_i \in \mathbb{R} \quad \forall 1 \leq i \leq n,$$

dass

$$\int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \left(\dots \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, \dots, x_n) dx_1 \right) dx_2 \right) \dots \right) dx_{n-1} \right] dx_n.$$

Beispiel 4.21

Sei $k = \ell = 1, n = 2$ und $f(x_1, x_2) = \lambda^2 \exp(-\lambda(x_1 + x_2)) \cdot \mathbf{1}_{[0, \infty)^2}((x_1, x_2))$ für $\lambda > 0$. Dann ist

$$\begin{aligned} \int_{\mathbb{R}^2} f(x) dx &= \int_0^\infty \int_0^\infty f(x_1, x_2) dx_1 dx_2 \\ &= \lambda^2 \int_0^\infty \left[\int_0^\infty \exp(-\lambda x_1) \exp(-\lambda x_2) dx_1 \right] dx_2 \\ &= \lambda^2 \int_0^\infty \exp(-\lambda x_2) \left[\int_0^\infty \exp(-\lambda x_1) dx_1 \right] dx_2 \\ &= \lambda^2 \int_0^\infty \exp(-\lambda x_2) \left[-\frac{1}{\lambda} \exp(-\lambda x_1) \right]_0^\infty dx_2 \\ &= \lambda \int_0^\infty \exp(-\lambda x_2) dx_2 = 1 \end{aligned}$$

wegen der Normierungsbedingung für die Exponentialverteilung. Also ist f eine Lebesgue-dichte auf \mathbb{R}^2 . Beachte: Es gilt, dass $f = f_{\text{Exp}}(\lambda) \cdot f_{\text{Exp}}(\lambda)$ ist.

Bemerkung 4.22

Unter Stetigkeits- und Kompaktheitsannahmen lassen sich mit dem Satz von Fubini Lebesgue-Integrale auf $\Omega \subseteq \mathbb{R}^n$ auf Riemann-Integrale auf \mathbb{R} zurückführen, vgl. Satz 1.18.(c) in Verbindung mit Bemerkung 1.19.(iii).

Definition 4.23

Seien $n \in \mathbb{N}$ und Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ gegeben, so dass für alle $1 \leq i \leq n$ der Ergebnisraum Ω_i eine Borel-Teilmenge von \mathbb{R} , $\mathcal{A}_i = \mathcal{B}(\Omega_i)$ und \mathbb{P}_i induziert ist von einer Lebesgue-dichte f_i auf $(\Omega_i, \mathcal{B}(\Omega_i))$.

Dann lässt sich der Produktraum

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$$

wie folgt definieren.

$$\begin{aligned} \Omega &= \prod_{i=1}^n \Omega_i, \\ \mathcal{A} &= \bigotimes_{i=1}^n \mathcal{A}_i = \{A \subseteq \Omega \mid A \in \mathcal{B}(\mathbb{R}^n)\}, \end{aligned}$$

und \mathbb{P} wird induziert von der Lebesgue-dichte $f = \prod_{i=1}^n f_i$.

Bemerkung 4.24

Nach Beispiel 1.5.(d) wird \mathcal{A} erzeugt durch das Mengensystem

$$\mathcal{F} = \{\pi_i^{-1}(A_i) : 1 \leq i \leq n, A_i \in \mathcal{A}_i\},$$

wobei $\pi_i : \Omega \rightarrow \Omega_i$ wie üblich die Projektion auf die i -te Koordinate bezeichnet.

Satz und Definition 4.25

Das Wahrscheinlichkeitsmaß \mathbb{P} aus Definition 4.23 heißt Produktmaß von $\mathbb{P}_1, \dots, \mathbb{P}_n$, in Zeichen:

$$\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i.$$

Es ist das eindeutig bestimmte Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) , für das gilt:

$$\forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n : \mathbb{P} \left(\prod_{i=1}^n A_i \right) = \prod_{i=1}^n \mathbb{P}_i(A_i).$$

Begründung (kein formaler Beweis):

Sei zur Vereinfachung der Notation $n = 2$. Der Satz von Fubini liefert

$$\begin{aligned}\mathbb{P}(A_1 \times A_2) &= \int_{A_1 \times A_2} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{A_1} \left[\int_{A_2} f(x_1, x_2) dx_2 \right] dx_1 \\ &= \int_{A_1} \left[\int_{A_2} f_1(x_1) f_2(x_2) dx_2 \right] dx_1 \\ &= \int_{A_1} f_1(x_1) \left[\int_{A_2} f_2(x_2) dx_2 \right] dx_1 \\ &= \int_{A_1} f_1(x_1) \mathbb{P}_2(A_2) dx_1 \\ &= \mathbb{P}_1(A_1) \mathbb{P}_2(A_2).\end{aligned}$$

Kapitel 5

Verteilungsfunktionen und Dichten, Transformationsformel

5.1 Verteilungsfunktionen und Dichten

Erinnerung 5.1

Für eine reellwertige Zufallsvariable $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt die Funktion $F_X : \mathbb{R} \rightarrow [0, 1]$ mit $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}^X([-\infty, x])$ für $x \in \mathbb{R}$ Verteilungsfunktion von X . Die Funktion F_X legt \mathbb{P}^X bereits fest.

Definition 5.2

Eine reellwertige Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt diskret, falls ihr Bild $\text{supp}(X) := X(\Omega) = \{X(\omega) : \omega \in \Omega\} \subset \mathbb{R}$ höchstens abzählbar ist (Träger von X). Wir nennen

$$f_X : \text{supp}(X) \rightarrow [0, 1], \quad x \mapsto \mathbb{P}^X(\{x\}) = \mathbb{P}(X = x)$$

die Zähldichte von X .

Korollar 5.3

Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine diskrete Zufallsvariable mit Zähldichte f_X und Verteilungsfunktion F_X .

Dann gilt:

$$F_X(x) = \sum_{y \in \text{supp}(X): y \leq x} f_X(y), \quad x \in \mathbb{R},$$

$$\text{so wie } f_X(x) = F_X(x) - F_X(x-), \quad x \in \text{supp}(X).$$

Hierbei (und in der Folge) bezeichnet $F_X(x-)$ den linksseitigen Grenzwert von F_X an der Stelle x . Wir setzen f_X auf ganz \mathbb{R} fort vermittels

$$f_X(x) := F_X(x) - F_X(x-) = 0, \quad x \in \mathbb{R} \setminus \text{supp}(X).$$

Beispiel 5.4 (Dirac-Verteilungen)

Für eine konstante Zufallsvariable X mit $X(\omega) = a \in \mathbb{R} \quad \forall \omega \in \Omega$ ist $\text{supp}(X) = \{a\}$. Es gilt $f_X(a) = 1$ und $f_X(x) = 0 \quad \forall x \in \mathbb{R} \setminus \{a\}$ sowie $F_X = 1_{[a, \infty)}$. Die Verteilung $\mathbb{P}^X =: \delta_a$ heißt Einpunktverteilung oder Dirac-Verteilung in $a \in \mathbb{R}$.

Definition 5.5

Eine reellwertige Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt stetig, falls $\text{supp}(X) = X(\Omega)$ eine Borelmenge auf \mathbb{R} ist und \mathbb{P}^X eine Lebesgue-dichte f_X besitzt. Das bedeutet, dass

$$\mathbb{P}(X \in A') = \mathbb{P}^X(A') = \int_{A'} f_X(x) dx$$

ist, $A' \in \mathcal{B}(\mathbb{R})$, wobei wir f_X auf ganz \mathbb{R} fortsetzen vermittels $f_X(x) = 0 \quad \forall x \in \mathbb{R} \setminus \text{supp}(X)$. Wir nennen f_X dann auch Dichte(-funktion) von X . Damit ist

$$F_X(x) = \int_{(-\infty, x]} f_X(y) dy.$$

Satz 5.6

Die Verteilungsfunktion F_X einer reellwertigen Zufallsvariablen X sei auf \mathbb{R} stetig und die Menge

$$D = \{x \in \mathbb{R} : F_X \text{ ist in } x \text{ nicht stetig differenzierbar}\} \text{ sei endlich.}$$

Dann ist die Funktion f_X , gegeben durch

$$f_X(x) = \begin{cases} \frac{d}{dx} F_X(x), & x \notin D \\ 0, & x \in D \end{cases}$$

die (kanonische) Dichte von X , und f_X ist höchstens auf D unstetig. Ändert man f_X auf D beliebig ab, so bleibt es eine Dichte von X .

Beweis:

Schreibe $D = \{d_1, \dots, d_K\}$, $K \geq 0$ und nehme o. B. d. A. an, dass $d_1 < d_2 < \dots < d_K$ gilt. Nach Voraussetzungen ist f_X höchstens auf D unstetig. Es genügt zu zeigen, dass

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(y) dy \tag{5.1}$$

für alle $x \in \mathbb{R}$ ist, vgl. Satz 1.10 und Erinnerung 5.1. Zum Nachweis von (5.1) sei $d_0 := -\infty$ und $d_{K+1} := +\infty$. Für jedes $0 \leq k \leq K$ ist F_X auf $I_k = (d_k, d_{k+1})$ stetig differenzierbar mit Ableitung f_X . Aus dem Hauptsatz der Differential- und Integralrechnung folgt daher

$$\forall a, b \in I_k : \int_a^b f_X(y) dy = F_X(b) - F_X(a).$$

Da F_X stetig ist folgt für $a \downarrow d_k$, dass

$$\int_{d_k}^b f_X(y)dy = F_X(b) - F_X(d_k) \quad \forall b \in I_k.$$

Ebenso erhält man für $b \uparrow d_{k+1}$, dass

$$\int_{d_k}^x f_X(y)dy = F_X(x) - F_X(d_k) \quad (5.2)$$

für alle $d_k < x \leq d_{k+1}$ gilt. Wir führen nun einen Induktionsbeweis, um für jedes $1 \leq \ell \leq K + 1$

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(y)dy \quad \forall x \leq d_\ell$$

zu zeigen.

$\ell = 1$:

Wegen $d_0 = -\infty$ und $F_X(-\infty) = 0$ folgt aus (5.2) mit $k = 0$, dass

$$F_X(x) = \int_{-\infty}^x f_X(y)dy$$

für alle $x \leq d_1$ ist.

$\ell \rightarrow \ell + 1$:

Für $d_\ell < x \leq d_{\ell+1}$ ergibt sich mit Induktionsvoraussetzung und (5.2), dass

$$\begin{aligned} \int_{-\infty}^x f_X(y)dy &= \int_{-\infty}^{d_\ell} f_X(y)dy + \int_{d_\ell}^x f_X(y)dy \\ &= F_X(d_\ell) + [F_X(x) - F_X(d_\ell)] \\ &= F_X(x) \text{ gilt. } \blacksquare \end{aligned}$$

5.2 Transformationsformel

Eine reellwertige Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ lässt sich unter Verwendung einer messbaren Abbildung $g : \text{supp}(X) \rightarrow \mathbb{R}$ „transformieren“ in eine neue Zufallsvariable $g(X) = g \circ X$.

Schema 5.7

$$\Omega \xrightarrow{X} \text{supp}(X) = X(\Omega) \xrightarrow{g} \mathbb{R}$$

Beispiel 5.8

- lineare Transformation: $g(X) = a + bX$
- Potenzen: $g(X) = X^k$

- *Absolutbetrag:* $g(X) = |X|$
- *Exponentialfunktion:* $g(X) = \exp(X)$
- *Logarithmus:* $g(X) = \log(X)$ für fast sicher positives X

Falls X eine stetige Zufallsvariable mit Dichte f_X ist, so kann die Dichte von $g(X)$ für spezielle Transformationen mit der Transformationsformel bestimmt werden.

Satz 5.9 (Transformationsformel)

Sei X eine stetige Zufallsvariable mit Träger $\text{supp}(X) = (a_X, b_X)$, Verteilungsfunktion F_X und Dichte $f_X : \text{supp}(X) \rightarrow [0, \infty)$. Wir nehmen an, dass die Menge D der Unstetigkeitsstellen von f_X endlich ist. Sei ferner $g : \text{supp}(X) \rightarrow \mathbb{R}$ eine stetig differenzierbare und streng monotone Funktion. Dann gilt:

- Ist g strikt isoton, so ergibt sich die Verteilungsfunktion von $g(X)$ als $F_{g(X)}(z) = F_X(g^{-1}(z))$, $z \in g(\text{supp}(X))$.
- Ist g strikt antiton, so ergibt sich die Verteilungsfunktion von $g(X)$ als $F_{g(X)}(z) = 1 - F_X(g^{-1}(z))$, $z \in g(\text{supp}(X))$.
- Ist $N = \{g' = 0\}$ endlich, so ist eine Dichte von $g(X)$ gegeben durch

$$f_{g(X)}(z) = \left\{ \begin{array}{ll} \frac{f_X(g^{-1}(z))}{|g'(g^{-1}(z))|}, & z \in g(\text{supp}(X)) \setminus g(N) \\ 0, & z \in g(N) \end{array} \right\}$$

und $f_{g(X)}$ ist höchstens auf der endlichen Menge $g(D \cup N)$ unstetig.

Beweis:

Definiere $Z := g(X)$ und $x = g^{-1}(z) \in \text{supp}(X)$ somit $z = g(x)$.

zu a):

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}(\{\omega \in \Omega : Z(\omega) \leq z\}) \\ &= \mathbb{P}(\{\omega \in \Omega : g(X(\omega)) \leq g(x)\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}), \end{aligned}$$

da g strikt isoton ist. Also ist

$$F_Z(z) = F_X(x) = F_X(g^{-1}(z)), \quad z \in g(\text{supp}(X)).$$

zu b):

Wegen strikter Antitonia von g ist hier

$$F_Z(z) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \geq x\}).$$

Da F_X in x stetig ist, gilt

$$F_Z(z) = \mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F_X(x) = 1 - F_X(g^{-1}(z)), \quad z \in g(\text{supp}(X)).$$

zu c):

Fall 1: g ist strikt isoton und somit $g' \geq 0$.

Dann ist $g(\text{supp}(X)) = (g(a_X), g(b_X))$ ein offenes Intervall und für $x \notin N$ ist $g'(x) > 0$.

Differenzieren in Teil a) liefert für $x = g^{-1}(z) \notin D \cup N$, dass

$$f_Z(z) = \frac{d}{dz} F_Z(z) = F'_X(g^{-1}(z)) \frac{d}{dz} g^{-1}(z).$$

Ferner ist bekannt, dass

$$\frac{d}{dz} g^{-1}(z) = \frac{1}{g'(g^{-1}(z))}$$

ist. Die Aussage folgt für $x \notin D \cup N$ aus $f_X(x) = F'_X(x)$, $x \notin D$, und $g'(g^{-1}(z)) = |g'(g^{-1}(z))|$.

Da $D \cup N$ endlich ist, können wir f_X durch $f_X(x) = 0 \quad \forall x \in D \cup N$ sowie f_Z durch $f_Z(z) = 0 \quad \forall z \in g(N)$ fortsetzen, was die Aussage impliziert.

Fall 2: g ist strikt antiton und somit $g'(x) \leq 0$.

Die Argumentation von Fall 1 kann analog geführt werden unter Beachtung von

$$f_Z(z) = F'_Z(z) = -F'_X(g^{-1}(z)) \frac{d}{dz} g^{-1}(z) \quad \text{und} \quad -g'(g^{-1}(z)) = |g'(g^{-1}(z))|. \quad \blacksquare$$

Bemerkung 5.10 (Satz 1.101 in Klenke (2008))

Satz 5.9 kann wie folgt verallgemeinert werden: Sei \mathbb{P} ein Wahrscheinlichkeitsmaß auf \mathbb{R}^d mit (stückweise) stetiger Dichte $f : \mathbb{R}^d \rightarrow [0, \infty)$. Sei $A \subseteq \mathbb{R}^d$ eine offene (oder abgeschlossene) Menge mit $\mathbb{P}(\mathbb{R}^d \setminus A) = 0$. Ferner sei $B \subseteq \mathbb{R}^d$ offen oder abgeschlossen sowie $g : A \rightarrow B$ bijektiv und stetig differenzierbar mit Ableitung g' . Dann hat das Wahrscheinlichkeitsmaß $\mathbb{P} \circ g^{-1}$ die Dichte

$$f_g(z) = \begin{cases} \frac{f(g^{-1}(z))}{|\det g'(g^{-1}(z))|}, & z \in B, \\ 0, & z \in \mathbb{R}^d \setminus B. \end{cases}$$

Anmerkung: g' ist die Jacobi-Matrix von g .

5.3 Zufallsvektoren

Sind bei einem Zufallsvorgang mehrere Merkmale gleichzeitig von Interesse, zwischen denen in aller Regel Zusammenhänge bestehen, so fasst man die entsprechenden (reellwertigen) Zufallsvariablen zu einem Zufallsvektor zusammen.

Beispiel 5.11

Bei einem (zufällig ausgewählten) neugeborenen Kind interessieren u. a.

$X_1 \hat{=} \text{Geburtsgewicht}$,

$X_2 \hat{=} \text{Geburtsgröße}$,

$X_3 \hat{=} \text{Schwangerschaftsdauer}$,

und diese drei Größen stehen erfahrungsgemäß in starkem Zusammenhang. Die Analyse des Zusammenhangs der drei Merkmale läuft auf die Untersuchung der Verteilung des Zufallsvektors $\mathbf{X} = (X_1, X_2, X_3)^\top$ hinaus.

Definition 5.12

Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $d \in \mathbb{N}$ und X_1, \dots, X_d mit $X_i : \Omega \rightarrow \mathbb{R} \quad \forall 1 \leq i \leq d$ reellwertige Zufallsvariablen. Die Abbildung

$$\mathbf{X} := (X_1, \dots, X_d)^\top : \Omega \rightarrow \mathbb{R}^d \text{ mit } \mathbf{X}(\omega) = (X_1(\omega), \dots, X_d(\omega))^\top, \quad \omega \in \Omega, \quad (5.3)$$

heißt Zufallsvektor. Die Bezeichnung (5.3) lässt sich auch schreiben als

$$\pi_i \circ \mathbf{X} = X_i, \quad 1 \leq i \leq d,$$

wobei $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ die Projektion auf die i -te Koordinate bezeichnet.

Der Zufallsvektor \mathbf{X} bildet messbar ab von $(\Omega, \mathcal{A}, \mathbb{P})$ nach $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Die Verteilung $\mathcal{L}(\mathbf{X})$ von \mathbf{X} heißt auch gemeinsame Verteilung von X_1, \dots, X_d . Für $1 \leq i \leq d$ heißt $\mathcal{L}(X_i)$ die Randverteilung von X_i . Die (gemeinsame) Verteilungsfunktion $F_{\mathbf{X}}$ des Zufallsvektor \mathbf{X} ist definiert vermittels

$$\forall \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d : F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathbb{P}\left(\bigcap_{i=1}^d \{X_i \leq x_i\}\right).$$

Sie legt die Verteilung $\mathcal{L}(\mathbf{X})$ fest, siehe Satz 1.10 in Verbindung mit Beispiel 1.5 (d).

Beispiel 5.13

(a) Seien X und Y zwei diskrete (reellwertige) Zufallsvariablen mit möglichen Werten x_1, \dots, x_m bzw. y_1, \dots, y_n . Dann ist $(X, Y)^\top$ ein bivariater diskreter Zufallsvektor. Bezeichne

$$\forall 1 \leq i \leq m : 1 \leq j \leq n : p_{ij} = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}^{(X,Y)}((x_i, y_j))$$

sowie

$$\forall 1 \leq i \leq m : p_{i\cdot} = \mathbb{P}(X = x_i) = \sum_{j=1}^n \mathbb{P}(X = x_i, Y = y_j) = \sum_{j=1}^n p_{ij},$$

$$\forall 1 \leq j \leq n : p_{\cdot j} = \mathbb{P}(Y = y_j) = \sum_{i=1}^m p_{ij}.$$

Dann lassen sich diese Größen tabellarisch wie folgt anordnen:

$i \setminus j$	1	2	\dots	n	Σ
1	p_{11}	p_{12}	\dots	p_{1n}	$p_{1\cdot}$
2	p_{21}	p_{22}	\dots	p_{2n}	$p_{2\cdot}$
\vdots					
m	p_{m1}	p_{m2}	\dots	p_{mn}	$p_{m\cdot}$
Σ	$p_{\cdot 1}$	$p_{\cdot 2}$	\dots	$p_{\cdot n}$	1

(b) Multinomialverteilung:

Bei einem Zufallsexperiment mit zugehörigem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ sei eine Zerlegung des Ergebnisraums Ω in d disjunkte Ereignisse (Klassen) gegeben, d. h. $\Omega = \bigcup_{i=1}^d A_i$ mit $A_i \cap A_j = \emptyset, i \neq j$. Bezeichne

$$p_i = \mathbb{P}(A_i) \in (0, 1), \quad \forall 1 \leq i \leq d, \text{ und } \mathbf{p} = (p_1, \dots, p_d)^\top.$$

Nun werden n unabhängige Wiederholungen dieses Zufallsexperimentes durchgeführt. Es bezeichne X_i die Anzahl an Wiederholungen, bei denen A_i eingetreten ist, $1 \leq i \leq d$, sowie \mathbf{X} den Zufallsvektor $(X_1, \dots, X_d)^\top$. Man beachte die Tatsache, dass $\sum_{i=1}^d X_i(\omega) \equiv n$ ist, für alle $\omega \in \Omega^n$. Die Verteilung von \mathbf{X} heißt Multinomialverteilung mit Parametern d, n und \mathbf{p} , in Zeichen $\mathcal{M}_d(n, \mathbf{p})$. Ihr Träger ist für $n \in \mathbb{N}$ gegeben als

$$\text{supp}(\mathbf{X}) = \{\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{N}_0^d : \sum_{i=1}^d x_i = n\}.$$

Die Zähldichte von \mathbf{X} ist gegeben durch

$$f_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}^{\otimes n}(\mathbf{X} = \mathbf{x}) = \binom{n}{x_1, \dots, x_d} \prod_{i=1}^d p_i^{x_i},$$

für alle $\mathbf{x} = (x_1, \dots, x_d)^\top \in \text{supp}(\mathbf{X})$.

(c) Multivariate Normalverteilungen:

Für einen Vektor $\mu \in \mathbb{R}^d$ und eine symmetrische, positiv definite $(d \times d)$ -Matrix Σ ist die Funktion $f : \mathbb{R}^d \rightarrow (0, \infty)$ mit

$$f(\mathbf{x}) = (2\pi)^{-d/2} |\det(\Sigma)|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma (\mathbf{x} - \mu)\right), \mathbf{x} \in \mathbb{R}^d,$$

eine Lebesguedichte auf \mathbb{R}^d . Die durch f induzierte Verteilung \mathbb{P}_f auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ heißt d -dimensionale (multivariate) Normalverteilung mit Parametern $\mu = (\mu_1, \dots, \mu_d)^\top$ und Σ , in Zeichen: $\mathcal{N}_d(\mu, \Sigma)$. Besitzt ein Zufallsvektor $\mathbf{X} = (X_1, \dots, X_d)^\top$ mit Werten in \mathbb{R}^d die Verteilung $\mathcal{N}_d(\mu, \Sigma)$, so lässt sich zeigen, dass $\forall 1 \leq i \leq d : \mathcal{L}(X_i) = \mathcal{N}(\mu_i, \sigma_i^2)$, wobei $\sigma_i^2 = \Sigma_{ii} > 0$ das i -te Diagonalelement von Σ bezeichnet, vgl. Beispiel 1.22 (c).

Definition und Satz 5.14

Seien X und Y zwei reellwertige, stetige Zufallsvariablen auf dem selben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Die Lebesguedichte des Zufallsvektors (X, Y) sei $f_{(X,Y)} : \mathbb{R}^2 \rightarrow [0, \infty)$.

Dann ist f_Y , gegeben durch

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx$$

eine Randdichte von Y und f_X mit

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy$$

eine Randdichte von X .

Für festes $x \in \mathbb{R}$ nennen wir $f_{Y|X=x}$, gegeben durch

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}, y \in \mathbb{R},$$

bedingte Dichte von Y bezüglich $X = x$, wobei $0/0 = 0$ gesetzt wird.

Für $x \in \mathbb{R}$ mit $f_X(x) > 0$ heißt die Mengenfunktion

$$\mathcal{B}(\mathbb{R}) \ni B \mapsto \mathbb{P}(Y \in B | X = x) := \int_B f_{Y|X=x}(y) dy$$

bedingte Verteilung von Y bezüglich $X = x$.

Es gelten die folgenden Rechenregeln für $A, B \in \mathcal{B}(\mathbb{R})$ und $C \in \mathcal{B}(\mathbb{R}^2)$.

$$(i) \mathbb{P}(X \in A, Y \in B) = \int_A \mathbb{P}(Y \in B | X = x) f_X(x) dx.$$

$$(ii) \mathbb{P}(Y \in B) = \int_{-\infty}^{\infty} \mathbb{P}(Y \in B | X = x) f_X(x) dx.$$

$$(iii) \mathbb{P}((X, Y) \in C) = \int_{-\infty}^{\infty} \mathbb{P}(Y \in C(x) | X = x) f_X(x) dx,$$

wobei $C(x) = \{y \in \mathbb{R} | (x, y) \in C\}$ den x -Schnitt von C bezeichnet.

Beweis: Definition von Lebesguedichten und Satz von Fubini. ■

Kapitel 6

Stochastische Unabhängigkeit von Zufallsvariablen

Definition 6.1

Für eine beliebige Indexmenge $I \neq \emptyset$ heißt eine Familie von Zufallsvariablen $(X_i)_{i \in I}$ mit $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\Omega'_i, \mathcal{A}'_i)$ stochastisch unabhängig, falls für jede nicht-leere endliche Teilmenge $K \subseteq I$ die Teilfamilie $(X_k)_{k \in K}$ stochastisch unabhängig ist in dem Sinne, dass

$$(U_X) \quad \mathbb{P}(\forall k \in K : X_k \in B_k) = \prod_{k \in K} \mathbb{P}(X_k \in B_k) \text{ für alle } B_k \in \mathcal{A}'_k, k \in K, \text{ gilt.}$$

Anmerkung: Alle X_i sind auf dem selben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert, können aber unterschiedliche Wertebereiche Ω'_i besitzen.

Eigenschaft (U_X) besagt, dass die gemeinsame Verteilung der $(X_k)_{k \in K}$ das Produktmaß der (Rand-)Verteilungen der X_k mit $k \in K$ ist.

Beispiel 6.2

Seien $n \in \mathbb{N}$, $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ Wahrscheinlichkeitsräume für $1 \leq i \leq n$ und $(\Omega, \mathcal{A}, \mathbb{P}) = \otimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ der zugehörige Produktraum. Dann sind die Projektionen $(\pi_i)_{1 \leq i \leq n}$ mit

$$\begin{aligned} \pi_i : \Omega &\rightarrow \Omega_i, \\ \omega = (\omega_1, \dots, \omega_n) &\mapsto \omega_i \end{aligned}$$

stochastisch unabhängig, vgl. Abschnitte 4.3 und 4.4.

Satz 6.3

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Für $d \in \mathbb{N}$ seien X_1, \dots, X_d mit $X_i : \Omega \rightarrow \mathbb{R}$ reellwertige Zufallsvariablen mit (Rand-)Verteilungsfunktion F_{X_i} von X_i für $1 \leq i \leq d$. Ferner bezeichne $F_{\mathbf{X}}$ die (gemeinsame) Verteilungsfunktion des Zufallsvektors $\mathbf{X} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow$

$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ mit $\mathbf{X} = (X_1, \dots, X_d)^\top$. Dann sind X_1, \dots, X_d genau dann stochastisch unabhängig, wenn

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^d F_{X_i}(x_i) \quad \text{für alle } \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d \text{ gilt.}$$

Beweis:

Die Aussage folgt aus Beispiel 1.5 (d) zusammen mit dem folgenden Hilfsresultat, das analog zu Satz 1.10 ist.

Hilfssatz:

Unter den Voraussetzungen von Definition 6.1 sei $\mathcal{E}_i \subseteq \mathcal{A}_i$ ein \cap -stabiler Erzeuger von \mathcal{A}_i für alle $i \in I$. Ist dann die Familie $(X_i^{-1}(\mathcal{E}_i))_{i \in I}$ stochastisch unabhängig in dem Sinne, dass für jede endliche Teilmenge $K \subseteq I$ und jede Wahl von $E_k \in X_k^{-1}(\mathcal{E}_k)$, $k \in K$, die Ereignisse $(E_k)_{k \in K}$ stochastisch unabhängig sind, so ist die Familie $(X_i)_{i \in I}$ stochastisch unabhängig.

Der Beweis des Hilfssatzes findet sich in Klenke (2008), siehe Satz 2.16 dort. ■

Korollar 6.4 (Diskrete Zufallsvariablen)

Unter den Voraussetzungen von Definition 6.1 gelte o. B. d. A. $\Omega'_i = \text{supp}(X_i) \quad \forall i \in I = \{1, \dots, d\}$. Ferner seien alle $X_i, i \in I$, diskrete Zufallsvariablen mit Zähldichte f_{X_i} von X_i .

Wir definieren $\mathbf{X} = (X_1, \dots, X_d)^\top$ mit Zähldichte $f_{\mathbf{X}}$, gegeben durch

$$f_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 = x_1, \dots, X_d = x_d), \mathbf{x} = (x_1, \dots, x_d)^\top \in \Omega' := \prod_{i=1}^d \Omega'_i.$$

Dann sind X_1, \dots, X_d genau dann stochastisch unabhängig, falls

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^d f_{X_i}(x_i) \quad \mathbf{x} \in \Omega' \quad \text{gilt.} \quad (6.1)$$

Beweis:

$(U_X) \Rightarrow (6.1)$: Wähle $B_i = \{x_i\}$.

$(6.1) \Rightarrow (U_X)$: Wie beachten, dass (6.1) äquivalenterweise geschrieben werden kann als

$$\forall \mathbf{x} \in \Omega' : \mathbb{P}^{\mathbf{X}}(\{\mathbf{x}\}) = \prod_{i=1}^d \mathbb{P}^{X_i}(\{x_i\}).$$

Dies charakterisiert nach Definition 4.16 aber gerade das Produktmaß der Verteilungen der $(X_i)_{1 \leq i \leq d}$, und Korollar 4.17 liefert das Gewünschte. ■

Beispiel 6.5

Es seien $X_1 \sim \text{Bin}(n, p)$ und $X_2 \sim \text{Bin}(m, p)$ zwei stochastisch unabhängige, jeweils binomialverteilte Zufallsvariablen. Dann ist $S := X_1 + X_2 \sim \text{Bin}(n + m, p)$, denn wir rechnen für $0 \leq k \leq m + n$ wie folgt.

$$\begin{aligned}
\mathbb{P}(S = k) &= \mathbb{P}(X_1 + X_2 = k) \\
&= \mathbb{P}((X_1, X_2) \in \{(i_1, i_2) : i_1 + i_2 = k\}) \\
&= \sum_{i_1, i_2: i_1+i_2=k} \mathbb{P}(X_1 = i_1, X_2 = i_2) \\
&= \sum_{\ell=0}^k \mathbb{P}(X_1 = \ell, X_2 = k - \ell) \\
&= \sum_{\ell=0}^k \mathbb{P}(X_1 = \ell) \mathbb{P}(X_2 = k - \ell) \\
&= \sum_{\ell=0}^k \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \binom{m}{k-\ell} p^{k-\ell} (1-p)^{m-k+\ell} \\
&= \sum_{\ell=0}^k \binom{n}{\ell} \binom{m}{k-\ell} p^k (1-p)^{n+m-k} = \binom{n+m}{k} p^k (1-p)^{n+m-k}.
\end{aligned}$$

Dabei folgt die Gleichheit $\binom{n+m}{k} = \sum_{\ell=0}^k \binom{n}{\ell} \binom{m}{k-\ell}$ aus Additions- und Multiplikationsregel der Kombinatorik (Anzahl an Möglichkeiten, k Objekte aus $n+m$ Objekten auszuwählen), indem man alle Möglichkeiten betrachtet, wie viele der k auszuwählenden Objekte aus den ersten n Objekten ausgewählt werden, worauf sich der Index ℓ bezieht.

Korollar 6.6 (Stetige Zufallsvariablen)

Stetige, reellwertige Zufallsvariablen $(X_i)_{1 \leq i \leq d}$ mit Lebesguedichte f_{X_i} von X_i sind genau dann stochastisch unabhängig, wenn das Produkt $\prod_{i=1}^d f_{X_i} : \mathbb{R}^d \rightarrow [0, \infty)$ dieser Lebesguedichten eine Lebesguedichte des Zufallsvektors $\mathbf{X} = (X_1, \dots, X_d)^\top$ ist.

Beweis:

Nach Definition 4.23 induziert $\prod_{i=1}^d f_{X_i}$ das Produktmaß der Verteilungen der $(X_i)_{1 \leq i \leq d}$. Damit liefert Satz 4.25 das Gewünschte. ■

Beispiel 6.7

Seien $\alpha, r, s > 0$ und X, Y stochastisch unabhängige Zufallsvariablen mit $X \sim \Gamma_{\alpha, r}$ und $Y \sim \Gamma_{\alpha, s}$, vgl. Übungsaufgabe 11.b). Dann sind $S := X + Y$ und $R := \frac{X}{X+Y}$ stochastisch unabhängig mit $S \sim \Gamma_{\alpha, r+s}$ und $R \sim \text{Beta}(r, s)$, so dass

$$\begin{aligned}
f_R(x) &= [B(r, s)]^{-1} x^{r-1} (1-x)^{s-1}, \quad x \in (0, 1), \text{ wobei} \\
B(a, b) &= \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad (\text{Euler'sche Beta-Funktion}).
\end{aligned}$$

Beweis: Übungsaufgabe

Das folgende einfache Korollar über stochastische Unabhängigkeit von Zufallsvektoren geben wir ohne Beweis an.

Korollar 6.8

Für $1 \leq i \leq m$ sei $\mathbf{X}_i : \Omega \rightarrow \mathbb{R}^{d_i}$ ein Zufallsvektor mit Werten in $(\mathbb{R}^{d_i}, \mathcal{B}(\mathbb{R}^{d_i}))$.

- a) Bezeichnet $F_{\mathbf{X}_i}$ die (gemeinsame) Verteilungsfunktion von $\mathbf{X}_i, 1 \leq i \leq m$, und $F_{\mathbf{X}}$ die Verteilungsfunktion von $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_m^\top)^\top$ mit Werten in \mathbb{R}^d für $d = \sum_{i=1}^m d_i$, so sind $\mathbf{X}_1, \dots, \mathbf{X}_m$ genau dann stochastisch unabhängig, wenn für alle $\mathbf{x}_1 \in \mathbb{R}^{d_1}, \dots, \mathbf{x}_m \in \mathbb{R}^{d_m}$ gilt

$$F_{\mathbf{X}}((\mathbf{x}_1^\top, \dots, \mathbf{x}_m^\top)^\top) = \prod_{i=1}^m F_{\mathbf{X}_i}(\mathbf{x}_i).$$

- b) Sind $\mathbf{X}_1, \dots, \mathbf{X}_m$ stochastisch unabhängig und $g_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d'_i}$ messbare Abbildungen für $1 \leq i \leq m$, so sind auch $g_1(\mathbf{X}_1), \dots, g_m(\mathbf{X}_m)$ stochastisch unabhängig.

Kapitel 7

Faltungen von Verteilungen

Definition 7.1

Sind X und Y zwei stochastisch unabhängige, (jeweils) \mathbb{R}^d -wertige Zufallsvariablen, die auf dem selben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert sind, so heißt die Verteilung $\mathcal{L}(X + Y)$ ihrer Summe die Faltung der beiden Verteilungen $\mathcal{L}(X)$ und $\mathcal{L}(Y)$, in Zeichen:

$$\mathcal{L}(X) * \mathcal{L}(Y) := \mathcal{L}(X + Y).$$

7.1 Faltungen diskreter Verteilungen

Korollar 7.2

Es seien $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ zwei stochastisch unabhängige, diskrete Zufallsvariablen mit den (höchstens abzählbaren) Trägern

$$\text{supp}(X) = X(\Omega) \quad \text{und} \quad \text{supp}(Y) = Y(\Omega).$$

Dann hat die Summe $S := X + Y$ den höchstens abzählbaren Träger

$$\text{supp}(S) = \{x + y : x \in \text{supp}(X), y \in \text{supp}(Y)\}.$$

Für $s \in \text{supp}(S)$ ist die Elementarwahrscheinlichkeit bezüglich \mathbb{P}^S gegeben durch

$$\begin{aligned} \mathbb{P}(S = s) &= \mathbb{P}(X + Y = s) \\ &= \sum_{\substack{x \in \text{supp}(X): \\ s-x \in \text{supp}(Y)}} \mathbb{P}(X = x) \mathbb{P}(Y = s - x) \\ &= \sum_{\substack{y \in \text{supp}(Y): \\ s-y \in \text{supp}(X)}} \mathbb{P}(Y = y) \mathbb{P}(X = s - y). \end{aligned}$$

Unter Verwendung der Zähldichten f_X, f_Y und f_S lässt sich schreiben

$$f_S(s) = \sum_{\substack{x \in \text{supp}(X): \\ s-x \in \text{supp}(Y)}} f_X(x) f_Y(s-x) = \sum_{\substack{y \in \text{supp}(Y): \\ s-y \in \text{supp}(X)}} f_Y(y) f_X(s-y), \quad s \in \text{supp}(S).$$

Beispiel 7.3

(a) Faltungen von Bernoulliverteilungen

Die Binomialverteilung $\text{Bin}(n, p)$ ist die n -fache Faltung der Bernoulliverteilung $\text{Bernoulli}(p)$, d. h.

$$\text{Bin}(n, p) = \text{Bernoulli}(p) * \cdots * \text{Bernoulli}(p) \quad (n \text{ Faktoren}).$$

Anders ausgedrückt: Sind X_1, \dots, X_n stochastisch unabhängige Indikatorvariablen mit

$$\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0) \quad \forall 1 \leq i \leq n, \quad \text{so ist} \quad \mathcal{L}\left(\sum_{i=1}^n X_i\right) = \text{Bin}(n, p).$$

Beweis:

Folgt aus Beispiel 6.5.

(b) Faltungen von Multinomialverteilungen

Die Multinomialverteilung $\mathcal{M}_d(n, \mathbf{p})$ (siehe Beispiel 5.13.(b)) ist die n -fache Faltung von $\mathcal{M}_d(1, \mathbf{p})$. Allgemeiner gilt:

$$\mathcal{M}_d(n_1, \mathbf{p}) * \mathcal{M}_d(n_2, \mathbf{p}) = \mathcal{M}_d(n_1 + n_2, \mathbf{p}), \quad \mathbf{p} = (p_1, \dots, p_d)^\top.$$

Beweis:

Seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ stochastisch unabhängige, (jeweils) \mathbb{R}^d -wertige Zufallsvektoren mit $\mathbf{X}_i \sim \mathcal{M}_d(1, \mathbf{p}) \quad \forall 1 \leq i \leq n$. Es genügt zu zeigen, dass $\mathcal{L}(\sum_{i=1}^n \mathbf{X}_i) = \mathcal{M}_d(n, \mathbf{p})$ ist.

Dazu führen wir einen Induktionsbeweis.

Induktionsanfang ($n = 1$): Nichts zu zeigen.

Induktionsschritt ($n \rightarrow n + 1$):

Nach Induktionsvoraussetzung ist $\mathbf{S}_n := \sum_{i=1}^n \mathbf{X}_i \sim \mathcal{M}_d(n, \mathbf{p})$ mit Zähldichte f_{S_n} , gegeben durch

$$f_{S_n}(\mathbf{s}) = \binom{n}{s_1, \dots, s_d} \prod_{j=1}^d p_j^{s_j}, \quad \mathbf{s} = (s_1, \dots, s_d)^\top \in \text{supp}(\mathbf{S}_n).$$

Ferner ist \mathbf{S}_n stochastisch unabhängig von $\mathbf{X}_{n+1} \sim \mathcal{M}_d(1, \mathbf{p})$, vgl. Korollar 6.8.b). Wir beachten, dass

$$\text{supp}(\mathbf{S}_n) = \{\mathbf{s} = (s_1, \dots, s_d)^\top \in \mathbb{N}_0^d : \sum_{i=1}^d s_i = n\}$$

sowie

$$\text{supp}(\mathbf{X}_{n+1}) = \{\mathbf{e} = (e_1, \dots, e_d)^\top : e_i \in \{0, 1\} \quad \forall 1 \leq i \leq d, \quad \sum_{i=1}^d e_i = 1\}$$

(Menge der d Einheitsvektoren im \mathbb{R}^d) ist.

Damit ist

$$\text{supp}(\mathbf{S}_{n+1}) = \{\mathbf{s} + \mathbf{e} | \mathbf{s} \in \text{supp}(\mathbf{S}_n), \mathbf{e} \in \text{supp}(\mathbf{X}_{n+1})\}.$$

Sei $\mathbf{e}^{(i)}$ der i -te Einheitsvektor im \mathbb{R}^d , $1 \leq i \leq d$. Dann gilt nach Faltungsformel für $\mathbf{z} \in \text{supp}(\mathbf{S}_{n+1})$ sowie mit $\delta_{ij} = 1$, falls $i = j$ und $\delta_{ij} = 0$, falls $i \neq j$, dass

$$\begin{aligned} \mathbb{P}(\mathbf{S}_{n+1} = \mathbf{z}) &= \sum_{i=1}^d \mathbb{P}(\mathbf{X}_{n+1} = \mathbf{e}^{(i)}) \cdot \mathbb{P}(\mathbf{S}_n = \mathbf{z} - \mathbf{e}^{(i)}) \\ &= \sum_{i=1}^d p_i \left[n! \prod_{j=1}^d \frac{1}{(z_j - \delta_{ij})!} p_j^{z_j - \delta_{ij}} \right] \\ &= \sum_{i=1}^d p_i \left[n! \frac{1}{(z_i - 1)!} p_i^{z_i - 1} \times \prod_{\substack{j=1 \\ j \neq i}}^d \frac{1}{z_j!} p_j^{z_j} \right] \\ &= \sum_{i=1}^d p_i \left[n! \frac{z_i}{z_i!} p_i^{z_i - 1} \times \prod_{\substack{j=1 \\ j \neq i}}^d \frac{1}{z_j!} p_j^{z_j} \right] \\ &= n! \left[\sum_{i=1}^d z_i \right] \left[\prod_{j=1}^d \frac{1}{z_j!} p_j^{z_j} \right]. \end{aligned} \tag{7.1}$$

Da $\mathbf{z} \in \text{supp}(\mathbf{S}_{n+1})$ ist gilt $\sum_{i=1}^d z_i = n + 1$. Damit ist die rechte Seite von (7.1) gleich

$$(n + 1)! \prod_{j=1}^d \frac{1}{z_j!} p_j^{z_j} = \binom{n + 1}{z_1, \dots, z_d} \prod_{j=1}^d p_j^{z_j} = f_{\mathbf{S}_{n+1}}(\mathbf{z}),$$

was nach Beispiel 5.13.(b) zu zeigen war. ■

(c) Faltungen von Poisson-Verteilungen

Die Faltung von Poisson-Verteilungen ist wieder eine Poisson-Verteilung:

$$\text{Poisson}(\lambda_1) * \text{Poisson}(\lambda_2) = \text{Poisson}(\lambda_1 + \lambda_2).$$

Anders ausgedrückt: Sind X_1 und X_2 zwei stochastisch unabhängige Zufallsvariablen mit $\mathcal{L}(X_i) = \text{Poisson}(\lambda_i)$, $i = 1, 2$, so ist $\mathcal{L}(X_1 + X_2) = \text{Poisson}(\lambda_1 + \lambda_2)$.

Beweis:

Sei $S := X_1 + X_2$ mit $\text{supp}(S) = \mathbb{N}_0$. Dann gilt nach Faltungsformel für $s \in \mathbb{N}_0$, dass

$$\begin{aligned}
 f_S(s) &= \mathbb{P}(S = s) = \mathbb{P}(X_1 + X_2 = s) \\
 &= \sum_{x \in \mathbb{N}_0: s-x \in \mathbb{N}_0} f_{X_1}(x) f_{X_2}(s-x) \\
 &= \sum_{x=0}^s f_{X_1}(x) f_{X_2}(s-x) \\
 &= \sum_{x=0}^s \frac{\lambda_1^x}{x!} \exp(-\lambda_1) \frac{\lambda_2^{s-x}}{(s-x)!} \exp(-\lambda_2) \\
 &= \exp(-(\lambda_1 + \lambda_2)) \sum_{x=0}^s \frac{1}{x!(s-x)!} \lambda_1^x \lambda_2^{s-x} \\
 &= \frac{1}{s!} \exp(-(\lambda_1 + \lambda_2)) \sum_{x=0}^s \binom{s}{x} \lambda_1^x \lambda_2^{s-x} \\
 &= \frac{(\lambda_1 + \lambda_2)^s}{s!} \exp(-(\lambda_1 + \lambda_2))
 \end{aligned}$$

nach binomischem Lehrsatz. ■

7.2 Faltungen stetiger Verteilungen mit Lebesguedichten

Satz 7.4

Seien X und Y zwei stochastisch unabhängige, (jeweils) reellwertige Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Trägern $\text{supp}(X) \subseteq \mathbb{R}$ und $\text{supp}(Y) \subseteq \mathbb{R}$. Wir nehmen an, dass diese Träger Borelmengen sind und Lebesguedichten f_X von X sowie f_Y von Y existieren, die wir auf ganz \mathbb{R} vermittels

$$f_X(x) = 0 \quad \forall x \in \mathbb{R} \setminus \text{supp}(X),$$

$$f_Y(y) = 0 \quad \forall y \in \mathbb{R} \setminus \text{supp}(Y)$$

fortsetzen. Ferner bezeichnen wir mit $f_{(X,Y)} = f_X \cdot f_Y$ die bivariate Lebesguedichte des Zufallsvektors (X, Y) . Dann besitzt die Summe $S := X + Y$ die Lebesguedichte f_S , gegeben durch

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x) dx = \int_{-\infty}^{\infty} f_Y(y) f_X(s-y) dy, \quad s \in \mathbb{R}.$$

Beweis:

Wir berechnen zunächst die Verteilungsfunktion F_S von S . Es gilt für $s \in \mathbb{R}$, dass

$$\begin{aligned}
 F_S(s) &= \mathbb{P}(X + Y \leq s) \\
 &= \int_{\{(x,y)^\top \in \mathbb{R}^2: x+y \leq s\}} f_{(X,Y)}(x,y) d(x,y) \\
 &= \int_{\{(x,y)^\top \in \mathbb{R}^2: y \leq s-x\}} f_{(X,Y)}(x,y) d(x,y) \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{s-x} f_{(X,Y)}(x,y) dy \right] dx. \tag{7.2}
 \end{aligned}$$

Wir substituieren

$$u := x + y \Leftrightarrow y = u - x$$

$$y = s - x \Rightarrow u = s$$

$$\frac{dy}{du} = 1 \Rightarrow dy = du$$

und erhalten für die rechte Seite von (7.2) unter Verwendung des Satzes von Fubini

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^s f_{(X,Y)}(x, u-x) du \right] dx = \int_{-\infty}^s \left[\int_{-\infty}^{\infty} f_{(X,Y)}(x, u-x) dx \right] du. \tag{7.3}$$

Nun beachten wir, dass $f_{(X,Y)} = f_X \cdot f_Y$ wegen der stochastischen Unabhängigkeit von X und Y gilt und erhalten

$$F_S(s) = \int_{-\infty}^s f_S(u) du \quad \text{mit} \quad f_S(u) = \int_{-\infty}^{\infty} f_X(x) f_Y(u-x) dx,$$

was die erste angegebene Darstellung von f_S zeigt. Die zweite angegebene Darstellung folgt analog durch Vertauschen der Rollen von x and y . ■

Anmerkung:

Bis zur Darstellung (7.3) bleibt die Rechnung auch ohne die Voraussetzung der stochastischen Unabhängigkeit von X und Y richtig; vgl. Satz 7.7 im nächsten Abschnitt.

Beispiel 7.5

(a) Faltungen von Normalverteilungen

Es gilt, dass $\mathcal{N}(\mu_1, \sigma_1^2) * \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ ist. Iterativ angewendet bedeutet dies: Sind X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen mit $\mathcal{L}(X_i) = \mathcal{N}(\mu_i, \sigma_i^2)$ für alle $1 \leq i \leq n$, so ist

$$\mathcal{L}\left(\sum_{i=1}^n X_i\right) = \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Beweis:

Wir betrachten zunächst den Spezialfall $\mu_1 = \mu_2 = 0$. Die Lebesguedichte f_i von $\mathcal{N}(0, \sigma_i^2)$ ist bekanntlich gegeben durch

$$f_i(x) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma_i^2}\right), \quad x \in \mathbb{R}, \quad i = 1, 2.$$

Nach der Faltungsformel für Dichten (d. h., Satz 7.4) ist demnach die Dichte f_S der Faltung gegeben durch

$$f_S(s) = \int_{-\infty}^{\infty} f_1(x) f_2(s-x) dx = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \left[\frac{x^2}{\sigma_1^2} + \frac{(s-x)^2}{\sigma_2^2} \right]\right) dx. \quad (7.4)$$

Wir definieren $\sigma^2 = \sigma_1^2 + \sigma_2^2$ und substituieren

$$u := \frac{\sigma x}{\sigma_1\sigma_2} - \frac{\sigma_1 s}{\sigma\sigma_2} \Rightarrow \frac{du}{dx} = \frac{\sigma}{\sigma_1\sigma_2} \Rightarrow dx = \frac{\sigma_1\sigma_2}{\sigma} du.$$

Ferner beachten wir, dass

$$\begin{aligned} \frac{x^2}{\sigma_1^2} + \frac{(s-x)^2}{\sigma_2^2} &= x^2 \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right] - \frac{2sx}{\sigma_2^2} + \frac{s^2}{\sigma_2^2} \\ &= x^2 \left[\frac{\sigma^2}{\sigma_1^2\sigma_2^2} \right] - \frac{2sx}{\sigma_2^2} + \frac{s^2}{\sigma_2^2} \\ &= x^2 \left[\frac{\sigma^2}{\sigma_1^2\sigma_2^2} \right] - \frac{2sx}{\sigma_2^2} + s^2 \left[\frac{\sigma_1^2}{\sigma^2\sigma_2^2} + \frac{1}{\sigma_2^2} \right] \\ &= x^2 \left[\frac{\sigma^2}{\sigma_1^2\sigma_2^2} \right] - \frac{2sx}{\sigma_2^2} + \frac{\sigma_1^2 s^2}{\sigma^2\sigma_2^2} + \frac{s^2}{\sigma_2^2} \\ &= u^2 + \frac{s^2}{\sigma^2} \text{ ist.} \end{aligned}$$

Damit ist die rechte Seite von (7.4) gleich

$$\begin{aligned} \frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \left[u^2 + \frac{s^2}{\sigma^2} \right]\right) du &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{s^2}{\sigma^2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right) du \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{s^2}{\sigma^2}\right) \end{aligned}$$

wegen der Normierungsbedingung der $\mathcal{N}(0, 1)$ -Verteilung. Damit ist der Spezialfall gezeigt.

Im allgemeinen Fall sind $U_i := X_i - \mu_i$ stochastisch unabhängig und gemäß Übungsaufgabe 25.c) gilt $\mathcal{L}(U_i) = \mathcal{N}(0, \sigma_i^2)$, $i = 1, 2$. Aus dem Spezialfall folgt daher, dass

$$U_1 + U_2 = (X_1 + X_2) - (\mu_1 + \mu_2) \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$$

und erneute Anwendung von Übungsaufgabe 25.c) liefert

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2). \quad \blacksquare$$

(b) Faltungen von Gammaverteilungen

Nach Beispiel 6.7 ist

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

Da $\Gamma_{\lambda,1} = \text{Exp}(\lambda)$ ist, folgt daraus: Sind X_1, \dots, X_n stochastisch unabhängig und identisch verteilt mit $\mathcal{L}(X_1) = \text{Exp}(\lambda)$, so ist $\mathcal{L}(\sum_{i=1}^n X_i) = \Gamma_{\lambda,n}$. Diese Verteilung heißt auch Erlang-Verteilung mit Parametern λ und n .

Lemma 7.6

Sei Γ , gegeben durch $\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt, x > 0$, die Euler'sche Gammafunktion.

Dann gilt:

- a) $\Gamma(x + 1) = x\Gamma(x), \quad x > 0.$
- b) $\Gamma(1) = 1.$
- c) $\Gamma(n) = (n - 1)!, \quad n \in \mathbb{N}.$
- d) $\Gamma(\frac{1}{2}) = \sqrt{\pi}.$

Beweis:

Zur Übung.

7.3 Ergebnisse für nicht notwendigerweise stochastisch unabhängige Zufallsvariablen

Satz 7.7

Sei $(X, Y)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^2$ ein stetig verteilter, bivariater Zufallsvektor mit Lebesgue-dichte $f_{(X,Y)} : \mathbb{R}^2 \rightarrow [0, \infty)$, die wir wie üblich durch $f_{(X,Y)}(x, y) = 0 \quad \forall (x, y)^\top \in \mathbb{R}^2 \setminus \text{supp}((X, Y)^\top)$ fortsetzen. Es wird nicht vorausgesetzt, dass $X \perp\!\!\!\perp Y$ gilt.

Dann gilt für $z, u \in \mathbb{R}$:

a)

$$F_{X+Y}(z) = \mathbb{P}(X + Y \leq z) = \int_{-\infty}^z f_{X+Y}(u) du \quad \text{mit}$$

$$f_{X+Y}(u) = \int_{-\infty}^\infty f_{(X,Y)}(x, u - x) dx = \int_{-\infty}^\infty f_{(X,Y)}(u - y, y) dy.$$

b)

$$F_{X-Y}(z) = \mathbb{P}(X - Y \leq z) = \int_{-\infty}^z f_{X-Y}(u) du \quad \text{mit}$$

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, x-u) dx = \int_{-\infty}^{\infty} f_{(X,Y)}(u+y, y) dy.$$

c)

$$F_{X \cdot Y}(z) = \mathbb{P}(X \cdot Y \leq z) = \int_{-\infty}^z f_{X \cdot Y}(u) du \quad \text{mit}$$

$$f_{X \cdot Y}(u) = \int_{\mathbb{R} \setminus \{0\}} |x|^{-1} f_{(X,Y)}\left(x, \frac{u}{x}\right) dx = \int_{\mathbb{R} \setminus \{0\}} |y|^{-1} f_{(X,Y)}\left(\frac{u}{y}, y\right) dy.$$

d)

$$F_{X/Y}(z) = \mathbb{P}\left(\frac{X}{Y} \leq z\right) = \int_{-\infty}^z f_{X/Y}(u) du \quad \text{mit}$$

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} |y| f_{(X,Y)}(uy, y) dy.$$

Beweis:

Übungsaufgabe.

Kapitel 8

Momente von Verteilungen, Integralungleichungen

8.1 Der Erwartungswert

8.1.1 Erwartungswert diskreter Verteilungen

Beispiel 8.1

a) Einfacher Würfelwurf

Wir betrachten den einfachen Würfelwurf mit einem homogenen Würfel und stellen die Frage: „Was würfelt man im Mittel?“

Eine plausible Antwort erscheint der Mittelwert der möglichen Werte zu sein:

$$3,5 = \frac{21}{6} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{\sum_{k \in \text{supp } X} k}{|\text{supp } X|}, \quad (8.1)$$

wobei X die Zufallsvariable bezeichne, die das Ergebnis des Würfelwurfs repräsentiert.

b) Zahlenlotto „6 aus 49“

Wir stellen beim Zahlenlotto „6 aus 49“ die analoge Frage: „Wie viele Richtige hat man im Mittel?“

Anwendung der Formel in (8.1) mit $\text{supp}(X) = \{0, 1, 2, 3, 4, 5, 6\}$ würde das Ergebnis 3 liefern. Dieses ist aber unplausibel, da es unserer Alltagserfahrung widerspricht.

Die Beobachtung, dass der relevante Unterschied zwischen den beiden Beispielen die unterschiedliche Verteilung ist (Gleichverteilung bzw. hypergeometrische Verteilung) motiviert die folgende Definition.

Definition 8.2 (Erwartungswert diskreter Zufallsvariablen)

Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable mit (höchstens abzählbarem) Träger $\text{supp}(X) =$

$X(\Omega) \subset \mathbb{R}$ und Zähldichte f_X . Falls

$$\sum_{x \in \text{supp}(X)} |x| \cdot \mathbb{P}(X = x) < \infty, \quad (8.2)$$

so ist der Erwartungswert von X definiert als

$$\mathbb{E}[X] := \sum_{x \in \text{supp}(X)} x \cdot \mathbb{P}(X = x) = \sum_{x \in \text{supp}(X)} x \cdot f_X(x) =: \mathbb{E}[\mathbb{P}^X]. \quad (8.3)$$

Zusatz: Für nicht-negatives X , d.h. $X(\Omega) \subseteq [0, \infty)$, definiert man $\mathbb{E}[X] \in [0, \infty]$ auch dann noch durch (8.3), falls (8.2) nicht gilt. In letzterem Falle ist dann $\mathbb{E}[X] = \infty$.

Beispiel 8.3

a) Indikatorvariablen

Sei $A \in \mathcal{A}$ ein Ereignis und $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$ die (messbare) Indikatorfunktion von A . Dann ist

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

b) Diskrete Gleichverteilung

Sei $m = |\text{supp}(X)| < \infty$ und $f_X(x) = \frac{1}{m}$ für alle $x \in \text{supp}(X)$. Dann ist

$$\mathbb{E}[X] = \sum_{x \in \text{supp}(X)} x \cdot f_X(x) = \frac{\sum_{x \in \text{supp}(X)} x}{m},$$

vgl. Beispiel 8.1.a).

c) Binomialverteilung

Sei $X \sim \text{Bin}(n, p)$ mit $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ für $0 \leq k \leq n$. Dann ist

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}, \end{aligned}$$

da $n-k = (n-1) - (k-1)$ ist.

Mit der Indextransformation $\ell = k-1 \Leftrightarrow k = \ell+1$ ergibt sich $k=1 \Rightarrow \ell=0$ und $k=n \Rightarrow \ell=n-1$ und somit

$$\mathbb{E}[X] = np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^\ell (1-p)^{(n-1)-\ell} = np$$

wegen der Normierungsbedingung für die $\text{Bin}(n-1, p)$ -Verteilung.

d) Hypergeometrische Verteilung

Wir verallgemeinern Beispiel 8.1.b) wie folgt: Seien $N, M, n \in \mathbb{N}$ vorgegebene Zahlen mit der Eigenschaft $0 < p := \frac{M}{N} < 1$. Diese können wie folgt interpretiert werden:

$$\begin{aligned} N &\hat{=} \text{Anzahl von Objekten,} \\ M &\hat{=} \text{Anzahl markierter Objekte,} \\ n &\hat{=} \text{Stichprobenumfang.} \end{aligned}$$

Sei $\Omega = \{A \subseteq \{1, \dots, N\} : |A| = n\}$, $\mathcal{A} = 2^\Omega$ und \mathbb{P} die diskrete Gleichverteilung auf (Ω, \mathcal{A}) mit $\mathbb{P}(\{\omega\}) = \frac{1}{\binom{N}{n}}$ für alle $\omega \in \Omega$.¹

Sei $X :=$ „Anzahl markierter Objekte unter den n ausgewählten Objekten“ eine Zufallsvariable mit

$$\text{supp}(X) = \{m \in \{0, \dots, n\} : n - (N - M) \leq m \leq M\}.$$

Dann ist für $m \in \text{supp}(X)$

$$f_X(m) = \mathbb{P}(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}.$$

Wir nennen $\mathcal{L}(X)$ die hypergeometrische Verteilung mit Parametern n, M, N , in Zeichen $\mathcal{H}(n, M, N)$.

Beim Zahlenlotto „6 aus 49“ wie in Beispiel 8.1 b) betrachtet gilt $X \sim \mathcal{H}(6, 6, 49)$.

Es gilt, dass $\mathbb{E}[\mathcal{H}(n, M, N)] = n \frac{M}{N} = np$ ist. Beim Zahlenlotto „6 aus 49“ ist demnach $\mathbb{E}[X] = \frac{36}{49} \approx 0.735 < 3$.

Beweis:

$$\mathbb{E}[X] = \sum_{m \in \text{supp}(X)} m \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}.$$

Unter Beachtung von $\binom{n}{k} = 0$ für beliebige Zahlen $n, k \in \mathbb{N}$ mit $k > n$ können wir die Summation erweitern und erhalten

$$\begin{aligned} \mathbb{E}[X] &= \sum_{m=0}^n m \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} \\ &= n \frac{M}{N} \sum_{m=1}^n \frac{\binom{M-1}{m-1} \binom{N-M}{n-m}}{\binom{N-1}{n-1}}. \end{aligned}$$

Mit der Indextransformation $\ell = m - 1 \Leftrightarrow m = \ell + 1$ ergibt sich $m = 1 \Rightarrow \ell = 0$ und $m = n \Rightarrow \ell = n - 1$ und somit

$$\mathbb{E}[X] = n \frac{M}{N} \sum_{\ell=0}^{n-1} \frac{\binom{M-1}{\ell} \binom{(N-1)-(M-1)}{(n-1)-\ell}}{\binom{N-1}{n-1}} = n \frac{M}{N} = np$$

wegen der Normierungsbedingung für die $\mathcal{H}(n-1, M-1, N-1)$ -Verteilung. ■

¹Dabei ist $\binom{N}{n}$ die Anzahl der Möglichkeiten, n Objekte aus N Objekten ohne Zurücklegen auszuwählen, vgl. Kapitel 2.

8.1.2 Erwartungswert stetiger Verteilungen

Aus Analogiegründen hinsichtlich Zähl- und Lebesguedichten (vgl. Bemerkung 1.21) ist die folgende Definition ein sinnvolles Analogon zu Definition 8.2 für den stetigen Fall.

Definition 8.4 (Erwartungswert stetiger Zufallsvariablen)

Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ eine stetige Zufallsvariable mit Lebesguedichte f_X , die wir wie üblich durch $f_X(x) = 0$ für $x \in \mathbb{R} \setminus \text{supp}(X)$ fortsetzen.

Falls

$$\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty \quad (8.4)$$

ist, so definieren wir $\mathbb{E}[X] = \mathbb{E}[\mathbb{P}^X] := \int_{-\infty}^{\infty} x \cdot f_X(x) dx$.

Gilt $X(\Omega) \subseteq [0, \infty)$ und ist die absolute Integrierbarkeitsbedingung (8.4) verletzt, so setzen wir $\mathbb{E}[X] = \infty$.

Beispiel 8.5

a) Stetige Gleichverteilung, $\text{UNI}[a, b]$

Sei X gleichverteilt auf dem Intervall $[a, b] \subset \mathbb{R}$ mit Lebesguedichte f_X , gegeben durch $f_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x)$, $x \in \mathbb{R}$, in Zeichen: $X \sim \text{UNI}[a, b]$.

Dann gilt

$$\begin{aligned} \mathbb{E}[X] &= \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) \\ &= \frac{1}{2(b-a)} (a+b)(b-a) = \frac{a+b}{2}. \end{aligned}$$

b) Exponentialverteilung, $\text{Exp}(\lambda)$

Sei X exponentialverteilt mit Intensitätsparameter $\lambda > 0$ mit Lebesguedichte f_X , gegeben durch $f_X(x) = \lambda \exp(-\lambda x) \mathbf{1}_{[0,\infty)}(x)$.

Dann gilt $\mathbb{E}[X] = \lambda \int_0^{\infty} x \exp(-\lambda x) dx$.

Wir setzen $g(x) = x$, $h'(x) = \exp(-\lambda x)$, so dass $g'(x) = 1$, $h(x) = -\frac{1}{\lambda} \exp(-\lambda x)$, und erhalten durch partielle Integration, dass

$$\begin{aligned} \mathbb{E}[X] &= \lambda \left\{ \left[-\frac{x}{\lambda} \exp(-\lambda x) \right]_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} \exp(-\lambda x) dx \right\} \\ &= \lambda \left\{ 0 + \frac{1}{\lambda} \left[-\frac{1}{\lambda} \exp(-\lambda x) \right]_0^{\infty} \right\} = \lambda \left\{ \frac{1}{\lambda^2} \right\} = \frac{1}{\lambda}. \end{aligned}$$

c) Normalverteilung $\mathcal{N}(\mu, \sigma^2)$

Sei X normalverteilt auf \mathbb{R} mit Parametern μ und σ^2 und mit Lebesguedichte f_X , gegeben

durch $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$, $x \in \mathbb{R}$. Dann gilt

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx.$$

Mit der Substitution

$$u = \frac{x-\mu}{\sigma} \Leftrightarrow x = \sigma u + \mu$$

$$\frac{du}{dx} = \frac{1}{\sigma} \Rightarrow dx = \sigma du$$

ergibt sich

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\sigma u + \mu) \exp\left(-\frac{1}{2}u^2\right) \sigma du \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \exp\left(-\frac{1}{2}u^2\right) du + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2\right) du \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[-\exp\left(-\frac{1}{2}u^2\right) \right]_{-\infty}^{\infty} + \mu = \mu \end{aligned}$$

wegen der Normierungsbedingung für die $\mathcal{N}(0, 1)$ -Verteilung.

8.1.3 Allgemeine Eigenschaften des Erwartungswertes

Lemma 8.6

Unter den Voraussetzungen von Definition 8.2 (diskrete Zufallsvariable X) existiere $\mathbb{E}[X]$ in \mathbb{R} .

Dann gilt

$$\begin{aligned} \mathbb{E}[X] &= \int_0^{\infty} [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx \\ &= \int_0^{\infty} \mathbb{P}(X > x) dx - \int_{-\infty}^0 \mathbb{P}(X \leq x) dx. \end{aligned} \tag{8.5}$$

Beweis: Für $x \in \mathbb{R}$ ist, weil X diskret ist,

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \in \text{supp}(X): y \leq x} f_X(y) \text{ und}$$

$$1 - F_X(x) = \mathbb{P}(X > x) = \sum_{y \in \text{supp}(X): y > x} f_X(y).$$

Damit ist (8.5) gleich

$$\int_0^{\infty} \sum_{y \in \text{supp}(X): y > x} f_X(y) dx - \int_{-\infty}^0 \sum_{y \in \text{supp}(X): y \leq x} f_X(y) dx =: I_1 - I_2.$$

Wir berechnen zunächst I_2 . Seien dazu

$$y_{[1]} < y_{[2]} < \cdots < y_{[j]} < y_{[j+1]} < \cdots < 0$$

die geordneten Elemente von $\text{supp}(X) \cap \{y \in \mathbb{R} : y < 0\}$. Wir erhalten, dass

$$\begin{aligned} I_2 &= \int_{y_{[1]}}^{y_{[2]}} f_X(y_{[1]}) \, dx + \int_{y_{[2]}}^{y_{[3]}} \{f_X(y_{[1]}) + f_X(y_{[2]})\} \, dx \\ &\quad + \int_{y_{[3]}}^{y_{[4]}} \{f_X(y_{[1]}) + f_X(y_{[2]}) + f_X(y_{[3]})\} \, dx + \cdots \\ &\quad + \int_{y_{[j]}}^{y_{[j+1]}} \sum_{\ell=1}^j f_X(y_{[\ell]}) \, dx + \cdots \\ &= (y_{[2]} - y_{[1]}) f_X(y_{[1]}) + (y_{[3]} - y_{[2]}) [f_X(y_{[1]}) + f_X(y_{[2]})] \\ &\quad + (y_{[4]} - y_{[3]}) [f_X(y_{[1]}) + f_X(y_{[2]}) + f_X(y_{[3]})] + \cdots \\ &\quad + (y_{[j+1]} - y_{[j]}) \sum_{\ell=1}^j f_X(y_{[\ell]}) + \cdots \\ &= - \sum_{y \in \text{supp}(X): y < 0} y f_X(y). \end{aligned}$$

Analog ist $I_1 = \sum_{y \in \text{supp}(X): y > 0} y f_X(y)$ und es folgt

$$I_1 - I_2 = \sum_{y \in \text{supp}(X)} y f_X(y) = \mathbb{E}[X].$$

■

Korollar 8.7 (zum Satz von Fubini)

Seien $a, b \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Dann ist (vgl. Abbildung 8.1)

$$\{(x, y) \in \mathbb{R}^2 : a \leq y \leq b \wedge a \leq x \leq y\} = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \wedge x \leq y \leq b\}.$$

Damit gilt für jede Lebesgue-integrierbare Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, dass

$$\int_a^b \int_a^y f(x, y) \, dx \, dy = \int_a^b \int_x^b f(x, y) \, dy \, dx.$$

Wir können damit nun das Analogon von Lemma 8.6 für stetige Zufallsvariablen beweisen.

Lemma 8.8

Unter den Voraussetzungen von Definition 8.4 (stetige Zufallsvariable X) existiere $\mathbb{E}[X]$ in \mathbb{R} .

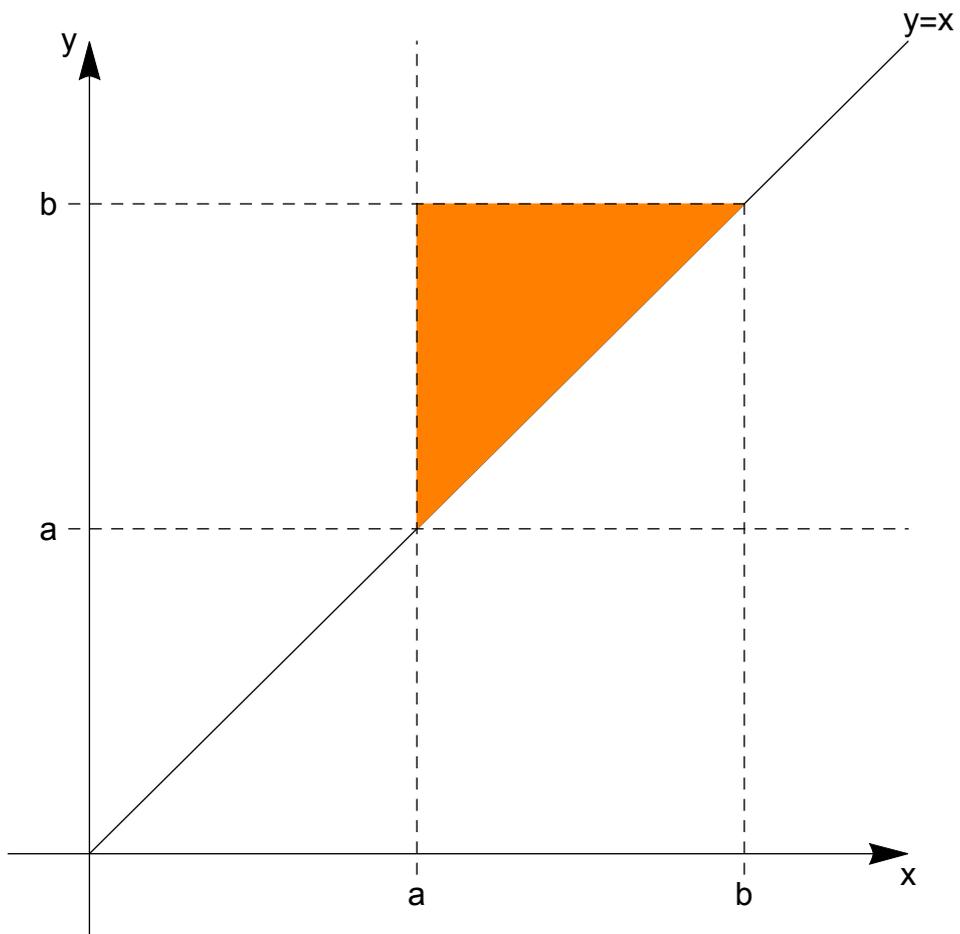


Abbildung 8.1: Illustration der Menge aus Korollar 8.7.

Dann gilt

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx \\ &= \int_0^\infty \mathbb{P}(X > x) dx - \int_{-\infty}^0 \mathbb{P}(X \leq x) dx \\ &= \int_0^\infty x f_X(x) dx + \int_{-\infty}^0 x f_X(x) dx.\end{aligned}$$

Beweis: Wir berechnen $\int_0^\infty x f_X(x) dx$. Beachte dazu, dass $x = \int_0^x 1 dy$ ist. Also folgt mit Korollar 8.7:

$$\begin{aligned}\int_0^\infty x f_X(x) dx &= \int_0^\infty \int_0^x f_X(x) dy dx = \int_0^\infty \int_y^\infty f_X(x) dx dy \\ &= \int_0^\infty \mathbb{P}(X > y) dy = \int_0^\infty \mathbb{P}(X > x) dx.\end{aligned}$$

Berechne nun $\int_{-\infty}^0 \mathbb{P}(X \leq x) dx = \int_{-\infty}^0 \int_{-\infty}^x f_X(y) dy dx$.

Korollar 8.7 liefert, dass

$$\begin{aligned}\int_{-\infty}^0 \int_{-\infty}^x f_X(y) dy dx &= \int_{-\infty}^0 \int_y^0 f_X(y) dx dy = \int_{-\infty}^0 [x f_X(y)]_{x=y}^0 dy \\ &= \int_{-\infty}^0 -y f_X(y) dy = - \int_{-\infty}^0 x f_X(x) dx,\end{aligned}$$

was den Beweis komplettiert. ■

Definition 8.9

Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ eine (beliebige) reellwertige Zufallsvariable mit Verteilungsfunktion F_X .

Falls die beiden Integrale

$$\begin{aligned}I_1 &= \int_0^\infty [1 - F_X(x)] dx = \int_0^\infty \mathbb{P}(X > x) dx \text{ und} \\ I_2 &= \int_{-\infty}^0 F_X(x) dx = \int_{-\infty}^0 \mathbb{P}(X \leq x) dx\end{aligned}$$

jeweils in \mathbb{R} existieren, so definieren wir

$$\mathbb{E}[X] =: \int_{\Omega} X d\mathbb{P} = I_1 - I_2.$$

Satz 8.10 (Rechenregeln für Erwartungswerte)

Seien $X, Y, (X_n)_{n \in \mathbb{N}}, (Y_n)_{n \in \mathbb{N}} : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen, deren (jeweilige) Erwartungswerte in \mathbb{R} existieren. Dann gilt:

a) Monotonie: Ist $X \leq Y$, so ist $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

b) Linearität: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ für beliebige $a, b \in \mathbb{R}$.

c) σ -Additivität bzw. monotone Konvergenz:

Sind alle $X_n \geq 0$ und ist $X = \sum_{n \geq 1} X_n$, so gilt $\mathbb{E}[X] = \sum_{n \geq 1} \mathbb{E}[X_n]$.

Wenn $Y_n \uparrow Y$ für $n \rightarrow \infty$, so folgt $\mathbb{E}[Y] = \lim_{n \rightarrow \infty} \mathbb{E}[Y_n]$.

d) Produktregel bei stochastischer Unabhängigkeit

Sind X und Y stochastisch unabhängig, so existiert der Erwartungswert von XY und es gilt $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Beweis: Lässt sich mit Transformations- und Faltungsformeln aus Definition 8.9 (ggfs. mit Hilfe von allgemeinen Grenzwertsätzen aus der Maßtheorie) folgern; Spezialfälle als Übungsaufgabe. ■

8.2 Momente und Varianz

Satz 8.11 (Transformationsatz für Erwartungswerte, siehe Maßtheorie-Vorlesung (MTV))

Unter den Voraussetzungen von Definition 8.9 sei $g : \text{supp}(X) \rightarrow \mathbb{R}$ so, dass $\mathbb{E}[g(X)]$ existiert.

Dann gilt

$$\begin{aligned} \mathbb{E}[g(X)] &= \int_{\Omega} g(X) \, d\mathbb{P} = \int_{\mathbb{R}} g \, d\mathbb{P}^X \\ &= \begin{cases} \sum_{x \in \text{supp}(X)} g(x) f_X(x), & X \text{ diskret mit Zähldichte } f_X, \\ \int_{\mathbb{R}} g(x) f_X(x) \, dx, & X \text{ stetig mit Lebesgue-dichte } f_X \text{ auf } \mathbb{R}. \end{cases} \end{aligned}$$

Anmerkung: Im Rahmen der Maßtheorie ist $\int_{\mathbb{R}} g \, d\mathbb{P}^X$ auch für Verteilungen \mathbb{P}^X erklärt, die weder diskret noch stetig sind.

Definition 8.12

Unter den Voraussetzungen von Definition 8.9 existiere $\mathbb{E}[X^k]$ für alle $1 \leq k \leq K$, $K \in \mathbb{N}$.

Dann heißt

a) $m_k(X) := \mathbb{E}[X^k]$ k -tes Moment von X .

b) $\mu_k(X) := \mathbb{E}[(X - \mathbb{E}[X])^k]$ k -tes zentrales Moment von X .

c) $M_k(X) := \mathbb{E}[|X|^k]$ k -tes absolutes Moment von X .

d) Ist $K \geq 2$, so heißt $\mu_2(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] =: \text{Var}(X)$ die Varianz von X und $SD(X) := \sqrt{\text{Var}(X)}$ die Standardabweichung von X .

e) Ist $K \geq 3$, $\mathbb{E}[X] =: \mu$ und $0 < \text{Var}(X) =: \sigma^2$, so heißt

$$m_3\left(\frac{X - \mu}{\sigma}\right) = \sigma^{-3} \mu_3(X)$$

die Schiefe von X .

f) Ist $K \geq 4$, $\mathbb{E}[X] =: \mu$ und $0 < \text{Var}(X) =: \sigma^2$, so heißt

$$m_4\left(\frac{X - \mu}{\sigma}\right) = \sigma^{-4} \mu_4(X)$$

die Wölbung (Kurtosis) von X und $m_4\left(\frac{X - \mu}{\sigma}\right) - 3$ die Exzess-Kurtosis von X .

Satz 8.13 (Verschiebungssatz)

Sei X eine Zufallsvariable mit endlicher Varianz. Dann ist

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &=: \mathbb{E}[X^2] - \mathbb{E}^2[X]. \end{aligned}$$

Beweis:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[X^2] - 2\mathbb{E}[X\mathbb{E}[X]] + \mathbb{E}^2[X] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}^2[X] + \mathbb{E}^2[X] \\ &= \mathbb{E}[X^2] - \mathbb{E}^2[X]. \end{aligned}$$



Beispiel 8.14

a) Bernoulli-Verteilung, Bernoulli(p)

Sei $X \sim \text{Bernoulli}(p)$. Dann ist $X^2 = X$. Damit ist $\mathbb{E}[X^2] = \mathbb{E}[X] = p$ (vgl. Beispiel 8.3.c) und folglich nach Verschiebungssatz $\text{Var}(X) = p - p^2 = p(1 - p)$.

b) Exponentialverteilung, Exp(λ)

Die Exponentialverteilung $\text{Exp}(\lambda)$ besitzt Momente beliebiger Ordnung und es gilt

$$m_k(\text{Exp}(\lambda)) = \frac{k!}{\lambda^k}, \quad k \in \mathbb{N}.$$

Beweis:

$$m_k(\text{Exp}(\lambda)) = \lambda \int_0^\infty x^k \exp(-\lambda x) dx.$$

Mit der Substitution

$$u = \lambda x \Leftrightarrow x = \frac{u}{\lambda}$$
$$\frac{du}{dx} = \lambda \Rightarrow dx = \frac{du}{\lambda}$$

ergibt sich

$$m_k(\text{Exp}(\lambda)) = \int_0^\infty \left(\frac{u}{\lambda}\right)^k \exp(-u) du$$
$$= \frac{1}{\lambda^k} \Gamma(k+1) = \frac{k!}{\lambda^k}.$$

■

c) Normalverteilung, $\mathcal{N}(\mu, \sigma^2)$

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Wir berechnen

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

unter Verwendung der Substitution

$$u = \frac{x-\mu}{\sigma} \Leftrightarrow x = \sigma u + \mu$$
$$\frac{du}{dx} = \frac{1}{\sigma} \Rightarrow dx = \sigma du.$$

Es ergibt sich (unter Verwendung der Normierungsbedingung für $X \sim \mathcal{N}(0, 1)$ sowie $\mathbb{E}[X] = 0$), dass

$$\begin{aligned} \mathbb{E}[X^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu)^2 \exp\left(-\frac{1}{2}u^2\right) du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 u^2 \exp\left(-\frac{u^2}{2}\right) du \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 2\sigma u \mu \exp\left(-\frac{u^2}{2}\right) du \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu^2 \exp\left(-\frac{u^2}{2}\right) du \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 \exp\left(-\frac{u^2}{2}\right) du + \mu^2. \end{aligned} \tag{8.6}$$

Sei $I := \int_{-\infty}^{\infty} u^2 \exp\left(-\frac{u^2}{2}\right) du$. Wir setzen

$$g(u) = u \Rightarrow g'(u) = 1,$$

$$h'(u) = u \exp\left(-\frac{u^2}{2}\right) \Rightarrow h(u) = -\exp\left(-\frac{u^2}{2}\right)$$

und somit ergibt sich durch partielle Integration

$$I = \left[-u \exp\left(-\frac{u^2}{2}\right) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) du = 0 + \sqrt{2\pi}.$$

Daher ist die rechte Seite von (8.6) gleich $\sigma^2 + \mu^2 = \mathbb{E}[X^2]$. Folglich ergibt sich $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \sigma^2$, vgl. Beispiel 8.5.c.

d) Poisson-Verteilung, Poisson(λ)

Sei $X \sim \text{Poisson}(\lambda)$.

Dann ist (unter Verwendung der Normierungseigenschaft von Poisson(λ))

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} k \exp(-\lambda) = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} k \exp(-\lambda) \\ &= \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell+1}}{(\ell+1)!} (\ell+1) \exp(-\lambda) \\ &= \lambda \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} \exp(-\lambda) = \lambda. \end{aligned}$$

Ferner ist

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} k^2 \exp(-\lambda) = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} k^2 \exp(-\lambda) \\ &= \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell+1}}{(\ell+1)!} (\ell+1)^2 \exp(-\lambda) \\ &= \lambda \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} (\ell+1) \exp(-\lambda) \\ &= \lambda [\mathbb{E}[X] + 1] = \lambda^2 + \lambda. \end{aligned}$$

Insgesamt ergibt sich also $\text{Var}(X) = \mathbb{E}[X] = \lambda$.

Satz 8.15 (Rechenregeln für die Varianz)

a) Die Dirac-Verteilung δ_a besitzt die Varianz 0 für beliebiges $a \in \mathbb{R}$.

b) Seien $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ zwei Zufallsvariablen mit endlichen Varianzen. Dann ist

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])].$$

c) Unter den Voraussetzungen von Teil b) ist $\text{Var}(a + bX) = b^2 \text{Var}(X)$ für beliebige reelle Konstanten a, b .

Beweis: Teil a) folgt unmittelbar aus der Tatsache, dass $\mathbb{P}(X = a) = 1$ ist, falls $X \sim \delta_a$, denn damit ist dann $\mathbb{E}[X] = a$ und $\mathbb{E}[X^2] = a^2$.

Zu Teil b) berechnen wir

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - \mathbb{E}^2[X + Y] \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}^2[X] - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}^2[Y] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\} \\ &= \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])]. \end{aligned}$$

Für Teil c) benutzen wir die Teile a) und b) und folgern, dass $\text{Var}(a + bX) = \text{Var}(bX)$ ist. Der Verschiebungssatz liefert dann

$$\text{Var}(bX) = \mathbb{E}[b^2 X^2] - \mathbb{E}^2[bX] = b^2 \mathbb{E}[X^2] - b^2 \mathbb{E}^2[X] = b^2 \text{Var}(X).$$

■

8.3 Momente von Zufallsvektoren

Für nicht-skalare zufällige Objekte (d.h. Zufallsvektoren, zufällige Matrizen, etc.) wird der Erwartungswert element-weise erklärt.

Definition 8.16 (Erwartungswert für Zufallsvektoren)

Sei $\mathbf{X} = (X_1, \dots, X_d)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ ein Zufallsvektor. Falls $\mathbb{E}[X_i]$ für alle $1 \leq i \leq d$ in \mathbb{R} existiert, so heißt

$$\mathbb{E}[\mathbf{X}] := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^\top \in \mathbb{R}^d$$

Erwartungswertvektor von \mathbf{X} .

Definition 8.17 (Kovarianz von Zufallsvariablen)

Seien $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ zwei Zufallsvariablen mit jeweils endlichen Varianzen.

Dann heißt

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Kovarianz von X und Y (vgl. den Beweis von Satz 8.15) und

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \in [-1, 1]$$

heißt Korrelationskoeffizient von X und Y .

Falls $\text{Cov}(X, Y) = \rho(X, Y) = 0$ gilt, so heißen X und Y unkorreliert.

Nebenbemerkung: Falls $\text{Var}(X) = 0$ oder $\text{Var}(Y) = 0$ gilt, so ist $\rho(X, Y)$ undefiniert (und bedeutungslos).

Satz 8.18 (Eigenschaften der Kovarianz)

Unter den Voraussetzungen von Definition 8.17 gilt:

a) $\text{Cov}(X, X) = \text{Var}(X)$

b) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (Symmetrie)

c) $\text{Cov}(a + X, b + Y) = \text{Cov}(X, Y)$ für alle reellen Konstanten a, b (Translationsinvarianz)

d) Bilinearität

Ist $Z : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ eine weitere Zufallsvariable mit endlicher Varianz, so ist

(i) $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$ für alle reellen Konstanten a, b .

(ii)

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

$$\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$$

e)

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y] + \text{Cov}(X, Y),$$

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \text{Cov}(X, Y).$$

f) $X \perp\!\!\!\perp Y$ (also X und Y stochastisch unabhängig) $\Rightarrow \text{Cov}(X, Y) = 0$.

Beweis: Folgt alles unmittelbar aus der Definition 8.17 und der Linearitätseigenschaft 8.10.b) des Erwartungswertes. Für Eigenschaft f) beachte die Produktregel bei Unabhängigkeit für den Erwartungswert, siehe Satz 8.10.d). ■

Gegenbeispiel 8.19

Die Umkehrung von Satz 8.18.f gilt im Allgemeinen nicht, d.h., aus der Unkorreliertheit von X und Y kann im Allgemein nicht auf $X \perp\!\!\!\perp Y$ geschlossen werden.

Betrachte z.B. $X \sim \mathcal{N}(0, 1)$ und $Y = X^2$. Dann sind X und Y nicht stochastisch unabhängig, denn es gilt

$$\mathbb{P}(|X| < 1, X^2 > 1) = 0 \neq \mathbb{P}(|X| < 1) \cdot \mathbb{P}(X^2 > 1).$$

Nun ist aber $\mathbb{E}[X] = 0$ und $\mathbb{E}[X^2] = \text{Var}(X) = 1$ und somit $\text{Cov}(X, X^2) = \mathbb{E}[X(X^2 - 1)] = \mathbb{E}[X^3] - \mathbb{E}[X] = 0 - 0 = 0$.

Definition 8.20 (Kovarianzmatrix eines Zufallsvektors)

Sei $\mathbf{X} = (X_1, \dots, X_d)^\top : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ ein Zufallsvektor, so dass $\mathbb{E}[X_i^2] < \infty$ für alle $1 \leq i \leq d$ gilt. Dann heißt

$$\Sigma = (\sigma_{i,j})_{1 \leq i,j \leq d} := \mathbb{E} \left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top \right] \in \mathbb{R}^{d \times d}$$

Kovarianzmatrix von \mathbf{X} . Offenbar gilt $\sigma_{i,j} = \text{Cov}(X_i, X_j)$ für alle $1 \leq i, j \leq d$.

Satz 8.21

Unter den Voraussetzungen von Definition 8.20 gilt:

a) Σ ist positiv semi-definit, d.h. Σ ist symmetrisch und $\forall a = (a_1, \dots, a_d)^\top \in \mathbb{R}^d$ gilt

$$\sum_{i,j=1}^d a_i a_j \sigma_{i,j} = a^\top \Sigma a \geq 0.$$

b) Sei $A \in \mathbb{R}^{m \times d}$ eine deterministische Matrix und $\mathbf{Y} := A\mathbf{X}$. Dann ist \mathbf{Y} ein \mathbb{R}^m -wertiger Zufallsvektor mit Kovarianzmatrix $A\Sigma A^\top$.

Beweis: Zu Teil a) beachten wir, dass $\sum_{i,j=1}^d a_i a_j \sigma_{i,j} = \text{Var} \left(\sum_{i=1}^d a_i X_i \right)$ ist. Da Varianzen stets nicht-negativ sind folgt die Aussage. Teil b) ist zur Übung (lineare Algebra). ■

8.4 Integralungleichungen

Satz 8.22 (Markov-Ungleichung)

Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X : \Omega \rightarrow \mathbb{R}$ eine reellwertige Zufallsvariable, $h : \mathbb{R} \rightarrow [0, \infty)$ eine monoton wachsende, deterministische Funktion und a eine reelle Konstante mit $h(a) > 0$. Dann gilt

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[h(X)]}{h(a)}.$$

Beweis: Wegen $h(x) \geq 0$ für alle $x \in \mathbb{R}$ ist $\mathbb{E}[h(X)] \geq 0$. Damit gilt

$$\begin{aligned} \mathbb{E}[h(X)] &\geq \mathbb{E}[h(X) \cdot \mathbf{1}_{[a, \infty)}(X)] \\ &\geq h(a) \cdot \mathbb{E}[\mathbf{1}_{[a, \infty)}(X)] \\ &= h(a) \cdot \mathbb{P}(X \geq a). \end{aligned}$$

■

Korollar 8.23

a) Setzt man in Satz 8.22 $h = \text{id} \cdot \mathbf{1}_{[0, \infty)}$ und betrachtet statt X selbst die Zufallsvariable $|X|$, so ergibt sich für $a > 0$, dass

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}.$$

b) Setzt man in Satz 8.22 voraus, dass $\mathbb{E}[X^2] < \infty$ ist und betrachtet man $Y = |X - \mathbb{E}[X]|$ so erhält man für $a > 0$, dass

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

Diese Ungleichung ist als Chebyshev-Ungleichung bekannt.

Satz 8.24 (Jensen'sche Ungleichung)

Es sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$ eine Zufallsvariable mit in \mathbb{R} existierendem Erwartungswert und $h : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion, so dass $\mathbb{E}[h(X)]$ in \mathbb{R} existiert.

Dann ist $h(\mathbb{E}[X]) \leq \mathbb{E}[h(X)]$.

Beweis: Es bezeichne h'_+ die rechtsseitige Ableitung von h . Aus der Analysis ist bekannt, dass diese existiert und isoton (also monoton wachsend) ist.

Wir betrachten nun $f(x, y) := h(x) + h'_+(x)(y - x)$, die (rechtsseitige) Linearisierung von h in x . Dann gilt für alle $x, y \in \mathbb{R}$

$$h(y) \geq f(x, y) \tag{8.7}$$

und wegen $f(y, y) = h(y)$ folgt

$$h(y) = \sup_x f(x, y). \tag{8.8}$$

Mit $y = X(\omega)$ ergibt sich aus (8.7), dass für alle $x, X(\omega) \in \mathbb{R}$

$$h(X(\omega)) \geq f(x, X(\omega)) \text{ gilt.}$$

Unter Verwendung der Monotonie und Linearität des Erwartungswertes ergibt sich also

$$\begin{aligned} \mathbb{E}[h(X)] &\geq \sup_x \mathbb{E}[f(x, X)] \\ &= \sup_x [h(x) + h'_+(x)\{\mathbb{E}[X] - x\}] \\ &= \sup_x f(x, \mathbb{E}[X]). \end{aligned}$$

Anwendung von (8.8) mit $y = \mathbb{E}[X]$ liefert schließlich

$$\mathbb{E}[h(X)] \geq h(\mathbb{E}[X]).$$

■

Korollar 8.25

Sei $p \geq 1$ und X eine reellwertige Zufallsvariable, für die $\mathbb{E}[|X|^p]$ endlich ist. Dann ist

$$|\mathbb{E}[X]|^p \leq \mathbb{E}[|X|^p].$$

Insbesondere ergibt sich für $p = 2$, dass

$$\mathbb{E}^2[X] \leq \mathbb{E}[X^2] \iff \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] \geq 0.$$

Kapitel 9

Erzeugende Funktion, Laplacetransformierte, Charakteristische Funktion

Statt der Angabe von Zähldichten (diskrete Zufallsgrößen) oder Verteilungsdichten (stetiger Fall) ist es in manchen Fällen (Berechnung von Momenten, Herleitung von Faltungen) nützlicher, mit anderen Charakterisierungen von Wahrscheinlichkeitsverteilungen zu arbeiten. Insbesondere die charakteristische Funktion (Fourier-Transformierte) hat zentrale Bedeutung; mehr dazu im Abschnitt über Verteilungskonvergenz.

9.1 Erzeugende Funktion

Definition 9.1

Sei X eine Zufallsvariable mit Werten in \mathbb{N}_0 . Die Potenzreihe $G_X : [0, 1] \rightarrow [0, 1]$ mit

$$t \mapsto G_X(t) := \mathbb{E} [t^X] = \sum_{k=0}^{\infty} t^k \mathbb{P}(X = k)$$

heißt die erzeugende Funktion von X bzw. von \mathbb{P}^X (englisch: generating function).

Beispiel 9.2

a) Die Binomialverteilung $\text{Bin}(n, p)$ hat die erzeugende Funktion $t \mapsto (1 - p + pt)^n$ nach dem Binomischen Lehrsatz.

b) Die Poisson-Verteilung $\text{Poisson}(\lambda)$ hat die erzeugende Funktion

$$t \mapsto \sum_{k=0}^{\infty} t^k \exp(-\lambda) \frac{\lambda^k}{k!} = \exp(\lambda(t - 1)).$$

Satz 9.3 (Eigenschaften von G_X)

a) Eindeutigkeitssatz: Haben zwei Zufallsvariablen, jeweils mit Werten in \mathbb{N}_0 , die gleiche erzeugende Funktion, so haben sie die gleiche Verteilung.

Kurz: $G_X = G_Y \Rightarrow \mathbb{P}^X = \mathbb{P}^Y$.

b) Es gilt $\mathbb{P}(X = 0) = G_X(0) < G_X(t) < G_X(1) = 1 \quad \forall t \in (0, 1)$.

c) Die Funktion G_X ist stetig und in $(0, 1)$ unendlich oft stetig differenzierbar. Es gilt für $n \in \mathbb{N}$ und die n -te Ableitung $G_X^{(n)}$, dass

$$\lim_{t \nearrow 1} G_X^{(n)}(t) = \sum_{k=n}^{\infty} \mathbb{P}(X = k) \cdot \prod_{j=k-n+1}^k j,$$

wobei beide Seiten $+\infty$ sein können; d.h.

$$\lim_{t \nearrow 1} G_X'(t) = \mathbb{E}[X] \text{ und}$$

$$\lim_{t \nearrow 1} G_X^{(n)}(t) = \mathbb{E}[X(X-1) \times \dots \times (X-n+1)] \text{ ist das } n\text{-te faktorielle Moment von } X.$$

d) Ist Y eine weitere Zufallsvariable mit Werten in \mathbb{N}_0 , stochastisch unabhängig von X , so ist $t \mapsto G_X(t)G_Y(t)$ die erzeugende Funktion von $X + Y$, d.h. von der Faltung $\mathbb{P}^X * \mathbb{P}^Y$, kurz: $G_{X+Y} = G_X G_Y$.

e) Induktiv folgt für stochastisch unabhängige X_1, \dots, X_n , dass $G_{\sum_{i=1}^n X_i} = \prod_{i=1}^n G_{X_i}$ gilt.

Beweis:

zu a)-c): Analysis I, Eigenschaften von Potenzreihen, Koeffizientenvergleich.

zu d):

$$\begin{aligned} G_X(t)G_Y(t) &= \left(\sum_{k=0}^{\infty} \mathbb{P}(X = k)t^k \right) \left(\sum_{k=0}^{\infty} \mathbb{P}(Y = k)t^k \right) \\ &\stackrel{\text{Cauchy-Produkt-Formel}}{=} \sum_{k=0}^{\infty} t^k \left(\sum_{\ell=0}^k \mathbb{P}(X = \ell)\mathbb{P}(Y = k - \ell) \right) \\ &\stackrel{\text{stoch. Unabhängigkeit}}{=} \sum_{k=0}^{\infty} t^k \sum_{\ell=0}^k \mathbb{P}(X = \ell, Y = k - \ell) \\ &= \sum_{k=0}^{\infty} t^k \mathbb{P}(X + Y = k) = G_{X+Y}(t). \end{aligned}$$

■

Beispiel 9.4

a) Beispiel 9.2.a) zusammen mit 9.3.e) zeigt, dass die Summe von n stochastisch unabhängigen, identisch Bernoulli(p)-verteilter Indikatoren eine $B(n, p)$ -Verteilung besitzt.

Ist allgemeiner $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ und $X \perp\!\!\!\perp Y$, so ist $X + Y \sim \text{Bin}(n + m, p)$.

b) Ist $X \sim \text{Poisson}(\alpha)$, $Y \sim \text{Poisson}(\beta)$ und $X \perp\!\!\!\perp Y$, so folgt aus Beispiel 9.2.b), dass $X + Y \sim \text{Poisson}(\alpha + \beta)$, denn

$$\begin{aligned}\Rightarrow G_{X+Y}(t) &= G_X(t)G_Y(t) = \exp(\alpha(t-1))\exp(\beta(t-1)) \\ &= \exp((\alpha + \beta)(t-1)).\end{aligned}$$

Ferner gilt $\mathbb{E}[X] = \text{Var}(X) = \alpha$, denn

$$\left. \frac{d}{dt} G_X(t) \right|_{t=1-} = \alpha \exp(\alpha(t-1)) \Big|_{t=1-} = \alpha$$

und

$$\left. \frac{d^2}{dt^2} G_X(t) \right|_{t=1-} = \alpha^2 \exp(\alpha(t-1)) \Big|_{t=1-} = \alpha^2.$$

Folglich ist $\mathbb{E}[X] = \alpha$, $\mathbb{E}[X^2 - X] = \alpha^2$, $\mathbb{E}[X^2] = \alpha(\alpha + 1)$ und $\mathbb{E}[X^2] - \mathbb{E}^2[X] = \text{Var}(X) = \alpha$.

Für allgemeinere Verteilungen reellwertiger Zufallsvariablen, die auf $[0, \infty)$ konzentriert sind, empfiehlt sich häufig die Benutzung ihrer Laplace-Transformierten.

9.2 Laplace-Transformierte

Definition 9.5

Sei X eine reellwertige Zufallsvariable mit $\mathbb{P}^X([0, \infty)) = 1$. Dann heißt $L_X : [0, \infty) \rightarrow \mathbb{R}$, definiert durch

$$L_X(s) := \mathbb{E}[\exp(-sX)] = \int_{[0, \infty)} \exp(-sx) \mathbb{P}^X(dx)$$

für $s \in \mathbb{R}_{\geq 0}$, die Laplace-Transformierte von X (bzw. von \mathbb{P}^X oder F_X).

Ist X auf $[0, \infty)$ stetig verteilt mit Lebesgue-dichte f_X , so ist

$$L_X(s) = \int_0^\infty \exp(-sx) f_X(x) dx.$$

Satz 9.6 (Eigenschaften von L_X)

a) Wegen $0 \leq \exp(-sx) \leq 1$ für alle $x \geq 0$ und $s \geq 0$ existiert L_X auf $[0, \infty)$ und es gilt

$$0 \leq L_X(s) \leq 1 = L_X(0) \text{ sowie } \mathbb{P}(X = 0) = \lim_{s \rightarrow \infty} L_X(s).$$

b) Die Funktion L_X ist stetig auf $[0, \infty)$ und beliebig oft differenzierbar auf $(0, \infty)$ mit

$$\begin{aligned} L_X^{(k)}(s) &= (-1)^k \mathbb{E} \left[X^k \exp(-sX) \right], k \in \mathbb{N}_0, s > 0 \text{ und} \\ \mathbb{E} \left[X^k \right] &= \lim_{s \searrow 0} (-1)^k L_X^{(k)}(s), \end{aligned}$$

wobei beide Seiten $+\infty$ sein können.

c) Umkehrformel:

Sei $C(F) := \{t \in \mathbb{R} | F \text{ stetig in } t\}$ die Menge der Stetigkeitsstellen einer Verteilungsfunktion F auf \mathbb{R} . Es gilt:

$$\forall x \in C(F_X), x > 0 : F_X(x) = \lim_{n \rightarrow \infty} \sum_{k \leq nx} \frac{(-n)^k}{k!} L_X^{(k)}(n).$$

d) Eindeutigkeitssatz: \mathbb{P}^X ist durch L_X eindeutig bestimmt.

e) Ist Y eine weitere reellwertige Zufallsvariable mit $\mathbb{P}^Y([0, \infty)) = 1$, stochastisch unabhängig von X , so ist $L_{X+Y} = L_X \cdot L_Y$.

Beweis:

zu a):

$$\lim_{s \rightarrow \infty} \mathbb{E} [\exp(-sX)] = \mathbb{E} [\mathbf{1}_{\{X=0\}}] = \mathbb{P}(X = 0).$$

zu b):

$$\begin{aligned} \frac{d}{ds} L_X(s) &= \lim_{h \rightarrow 0} \frac{L_X(s+h) - L_X(s)}{h} \\ &= \lim_{h \rightarrow 0} h^{-1} \{ \mathbb{E} [\exp(-(s+h)X)] - \mathbb{E} [\exp(-sX)] \} \\ &= \lim_{h \rightarrow 0} \mathbb{E} [h^{-1} \{ \exp(-(s+h)X) - \exp(-sX) \}] \\ &\stackrel{\text{maj. Konvergenz}}{=} \mathbb{E} \left[\lim_{h \rightarrow 0} \frac{\exp(-(s+h)X) - \exp(-sX)}{h} \right] \\ &= \mathbb{E} \left[\frac{d}{ds} \exp(-sX) \right] \\ &= \mathbb{E} [-X \exp(-sX)] = -\mathbb{E} [X \exp(-sX)]. \end{aligned}$$

Induktion nach k liefert nun das Gewünschte.

zu c): Siehe Feller (1971), Abschnitt XIII.4.

zu d): Folgt aus c).

zu e): $\mathbb{E}[\exp(-s(X + Y))] \stackrel{\text{stoch. Unabhängigkeit}}{=} \mathbb{E}[e^{-sX}] \mathbb{E}[e^{-sY}]$.

■

Beispiel 9.7

a) Sei $X \sim \text{Exp}(\lambda)$, dann ergibt sich

$$\begin{aligned} L_X(s) &= \mathbb{E}[\exp(-sX)] = \int_0^\infty \exp(-sx) \lambda \exp(-\lambda x) dx \\ &= \lambda \int_0^\infty \exp(-(s + \lambda)x) dx = \frac{\lambda}{s + \lambda}. \\ \Rightarrow \mathbb{E}[X^k] &= (-1)^k \left. \frac{d^k}{ds^k} L_X(s) \right|_{s=0+} = (-1)^k (-1)^k \left. \frac{k! \lambda}{(s + \lambda)^{k+1}} \right|_{s=0+} = \frac{k!}{\lambda^k}. \end{aligned}$$

b) Die Erlang(λ, n)-Verteilung als n -fache Faltung von $\text{Exp}(\lambda)$ mit sich selbst hat die Laplace-Transformierte $s \mapsto \left(\frac{\lambda}{s + \lambda}\right)^n$. Ist $Y \sim \text{Erlang}(\lambda, n)$, so ergibt sich demnach

$$\begin{aligned} \mathbb{E}[Y] &= -\left. \frac{d}{ds} \left(\frac{\lambda}{s + \lambda}\right)^n \right|_{s=0+} = \left. \frac{n\lambda^n}{(s + \lambda)^{n+1}} \right|_{s=0+} = \frac{n}{\lambda}, \\ \mathbb{E}[Y^2] &= \left. \frac{d^2}{ds^2} \left(\frac{\lambda}{s + \lambda}\right)^n \right|_{s=0+} \\ &= \left. \frac{n(n + 1)\lambda^n}{(s + \lambda)^{n+2}} \right|_{s=0+} = \frac{n(n + 1)}{\lambda^2}, \\ \text{Var}(Y) &= \frac{n}{\lambda^2}. \end{aligned}$$

9.3 Charakteristische Funktion

Für eine reellwertige Zufallsvariable, deren Werte nicht auf $[0, \infty)$ eingeschränkt sind, existiert die Laplace-Transformierte häufig nur auf Teilbereichen des Trägers ihrer Verteilung. Einen Extremfall stellt die Cauchy-Verteilung dar, bei der die Laplace-Transformierte nur für $s = 0$ existiert. Folglich ist hier die Laplace-Transformierte nicht zur Charakterisierung der Verteilung geeignet. Zentrale Objekte der Wahrscheinlichkeitstheorie sind die charakteristischen Funktionen, die stets existieren.

Bezeichne dazu in diesem Abschnitt $i = \sqrt{-1}$ die imaginäre Einheit.

Definition 9.8

a) Sei μ ein endliches Maß auf \mathbb{R}^d für $d \in \mathbb{N}$. Die Abbildung $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$, definiert durch

$$\varphi_\mu(\mathbf{t}) := \int \exp(i\langle \mathbf{t}, \mathbf{x} \rangle) \mu(d\mathbf{x})$$

heißt Fourier - Transformierte von μ .

b) Sei $\mathbf{X} = (X_1, \dots, X_d)^\top$ ein Zufallsvektor mit (gemeinsamer) Verteilung $\mathbb{P}^{\mathbf{X}}$. Dann heißt $\varphi_{\mathbf{X}} := \varphi_{\mathbb{P}^{\mathbf{X}}}$ die charakteristische Funktion von \mathbf{X} .

c) Für eine komplexwertige Zufallsvariable Z mit Real- und Imaginärteilen $\operatorname{Re}(Z)$ und $\operatorname{Im}(Z)$ sei $\mathbb{E}[Z] := \mathbb{E}[\operatorname{Re}(Z)] + i\mathbb{E}[\operatorname{Im}(Z)]$, falls die Erwartungswerte von $\operatorname{Re}(Z)$ und $\operatorname{Im}(Z)$ (jeweils) existieren. Damit ist

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[\exp(i\langle \mathbf{t}, \mathbf{X} \rangle)], \mathbf{t} \in \mathbb{R}^d.$$

Man beachte dabei die Euler'sche Formel $\exp(i\vartheta) = \cos(\vartheta) + i \sin(\vartheta)$.

Wegen $|\exp(i\langle \mathbf{t}, \mathbf{x} \rangle)| = 1$ für alle $\mathbf{t}, \mathbf{x} \in \mathbb{R}^d$ existiert die charakteristische Funktion für alle $\mathbf{t} \in \mathbb{R}^d$.

Satz 9.9 (Eigenschaften der charakteristischen Funktion)

a) $\forall \mathbf{t} \in \mathbb{R}^d : |\varphi_{\mathbf{X}}(\mathbf{t})| \leq 1 = \varphi_{\mathbf{X}}(\mathbf{0})$.

b) Affine Transformationen: Sei \mathbf{X} eine Zufallsgröße mit Werten in \mathbb{R}^d und $\mathbf{Y} := A\mathbf{X} + \mathbf{b}$ mit $A \in \mathbb{R}^{m \times d}$ und $\mathbf{b} \in \mathbb{R}^m$, wobei $d, m \in \mathbb{N}$. Dann gilt $\varphi_{\mathbf{Y}}(\mathbf{u}) = \exp(i\langle \mathbf{u}, \mathbf{b} \rangle) \varphi_{\mathbf{X}}(A^\top \mathbf{u})$, $\mathbf{u} \in \mathbb{R}^m$. Ist speziell $d = m = 1$ und $A = a = -1$, $\mathbf{b} = 0$, so ergibt sich z. B.

$$\varphi_{-X}(u) = \varphi_X(-u) = \overline{\varphi_X(u)}$$

aufgrund der Symmetrieeigenschaften von Sinus und Cosinus.

c) $\mathbb{P}^X = \mathbb{P}^{-X}$ genau dann, wenn φ_X (rein) reellwertig ist.

d) Die Zufallsvariablen X_1, \dots, X_d sind genau dann stochastisch unabhängig, wenn $\forall \mathbf{u} = (u_1, \dots, u_d)^\top \in \mathbb{R}^d : \varphi_{\mathbf{X}}(\mathbf{u}) = \prod_{k=1}^d \varphi_{X_k}(u_k)$ gilt, $\mathbf{X} = (X_1, \dots, X_d)^\top$.

e) Faltungsformel: Sind \mathbf{X} und \mathbf{Y} stochastisch unabhängige Zufallsvektoren mit Werten in \mathbb{R}^d , so ist $\varphi_{\mathbf{X}+\mathbf{Y}} = \varphi_{\mathbf{X}} \cdot \varphi_{\mathbf{Y}}$.

Beweis:

zu a): $\mathbb{P}^{\mathbf{X}}(\mathbb{R}^d) = 1$.

zu b): Zur Übung (Lineare Algebra).

zu c): Folgt aus den Symmetrieeigenschaften von Sinus und Cosinus.

zu d): Folgt aus der Charakterisierung der stochastischen Unabhängigkeit über

$$\mathbb{E}[f(X_i)g(X_j)] = \mathbb{E}[f(X_i)] \mathbb{E}[g(X_j)]$$

für alle komplexwertigen, messbaren Funktionen f und g , Details z.B. in Kapitel 8 von Breiman (1992).

zu e): Die Beweisführung erfolgt analog zum Beweis für Laplace-Transformierte in Satz 9.6.e). ■

Es existieren eine ganze Reihe von “Umkehrformeln”, die es erlauben, Verteilungsfunktionen, Dichtefunktionen oder Wahrscheinlichkeitsfunktionen aus charakteristischen Funktionen zurückzugewinnen.

Satz 9.10

a) Diskrete Fourier-Inversionsformel:

Sei μ ein Wahrscheinlichkeitsmaß auf $\mathbb{Z}^d \Rightarrow \forall \mathbf{x} \in \mathbb{Z}^d$ gilt:

(i)

$$\mu(\{\mathbf{x}\}) = (2\pi)^{-d} \int_{[-\pi, \pi]^d} \exp(-i\langle \mathbf{t}, \mathbf{x} \rangle) \varphi_\mu(\mathbf{t}) d\mathbf{t},$$

(ii)

$$\sum_{\mathbf{x} \in \mathbb{Z}^d} \mu(\{\mathbf{x}\})^2 = (2\pi)^{-d} \int_{[-\pi, \pi]^d} |\varphi_\mu(\mathbf{t})|^2 d\mathbf{t} \quad (\text{Plancherel}).$$

b) Besitzt μ eine stetige und beschränkte λ^d -Dichte f , so gilt

$$f(\mathbf{x}) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-i\langle \mathbf{t}, \mathbf{x} \rangle) \varphi_\mu(\mathbf{t}) \lambda^d(d\mathbf{t}), \quad \mathbf{x} \in \mathbb{R}^d.$$

c) In Dimension $d = 1$ gilt

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\text{Im}(e^{-itx} \varphi_X(t))}{t} dt$$

für alle Stetigkeitspunkte x von F_X .

d) Chungs Inversionsformel (hier nur $d = 1$):

Falls $a < b$ und $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$, so folgt

$$F_X(b) - F_X(a) = \lim_{T \rightarrow \infty} \left\{ \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt \right\}.$$

Beweis:

zu a): Für Teil (i) folgen wir dem Beweis von Satz 15.10 in Klenke (2008). Wir rechnen für festes $\mathbf{x} \in \mathbb{Z}^d$, dass

$$\int_{[-\pi, \pi]^d} \exp(-i\langle \mathbf{t}, \mathbf{x} \rangle) \varphi_\mu(\mathbf{t}) d\mathbf{t} = \int_{[-\pi, \pi]^d} \exp(-i\langle \mathbf{t}, \mathbf{x} \rangle) \left[\lim_{n \rightarrow \infty} \sum_{|\mathbf{y}| \leq n} \exp(i\langle \mathbf{t}, \mathbf{y} \rangle) \mu(\{\mathbf{y}\}) \right] d\mathbf{t}.$$

Wegen des Satzes von der majorisierten Konvergenz aus der MTV dürfen wir Grenzwert und Integration vertauschen und erhalten

$$\begin{aligned} \int_{[-\pi, \pi]^d} \exp(-i\langle \mathbf{t}, \mathbf{x} \rangle) \varphi_\mu(\mathbf{t}) d\mathbf{t} &= \lim_{n \rightarrow \infty} \int_{[-\pi, \pi]^d} \exp(-i\langle \mathbf{t}, \mathbf{x} \rangle) \sum_{|\mathbf{y}| \leq n} \exp(i\langle \mathbf{t}, \mathbf{y} \rangle) \mu(\{\mathbf{y}\}) d\mathbf{t} \\ &= \sum_{\mathbf{y} \in \mathbb{Z}^d} \mu(\{\mathbf{y}\}) \int_{[-\pi, \pi]^d} \exp(i\langle \mathbf{t}, \mathbf{y} - \mathbf{x} \rangle) d\mathbf{t} \\ &= \mu(\{\mathbf{x}\}) (2\pi)^d \end{aligned}$$

wegen der (2π) -Periodizität von Sinus und Cosinus, was die Aussage impliziert.

Teil (ii) folgt aus Teil (i), siehe Übung 15.1.3 in Klenke (2008).

zu b): Wir beweisen hier nur den Fall $d = 1$. Sei dazu $x \in \mathbb{R}$ beliebig, aber fest. Wir beachten, dass

$$\begin{aligned} \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx) \varphi_\mu(t) dt &= \lim_{\varepsilon \rightarrow 0} \left\{ \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx - \varepsilon t^2/2) \varphi_\mu(t) dt \right\} \\ &=: \lim_{\varepsilon \rightarrow 0} I(\varepsilon). \end{aligned}$$

Wir berechnen nun $I(\varepsilon)$ und erhalten, dass

$$I(\varepsilon) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx - \varepsilon t^2/2) \left[\int_{\mathbb{R}} \exp(ity) f(y) dy \right] dt$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}^2} \exp(it(y-x) - \varepsilon t^2/2) f(y) dy dt \quad (9.1)$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \exp(it(y-x) - \varepsilon t^2/2) dt \right] f(y) dy \quad (9.2)$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left[\sqrt{\frac{2\pi}{\varepsilon}} \exp\left(-\frac{(x-y)^2}{2\varepsilon}\right) \right] f(y) dy \quad (9.3)$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\varepsilon}} \exp\left(-\frac{(x-y)^2}{2\varepsilon}\right) f(y) dy$$

$$= f_S(x),$$

wobei S die Summe ist aus $Y \sim \mu$ mit Lebesgue-dichte f und $X \sim \mathcal{N}(0, \varepsilon)$, siehe Formel für das Faltungsintegral aus Satz 7.4. Es gilt nun aber, dass $f_S(x) \rightarrow f(x)$ für $\varepsilon \rightarrow 0$, denn dann degeneriert die Verteilung von X zu δ_0 .

Dabei gelten (9.1) und (9.2) wegen des Satzes von Fubini und (9.3) folgt, wenn wir in dem folgenden Lemma $a = 0$, $b = i(x - y)$ und $c = \varepsilon/2$ setzen.

Lemma

Für $c > 0$ gilt, dass

$$\int_{\mathbb{R}} \exp(-a - bt - ct^2) dt = \sqrt{\frac{\pi}{c}} \exp\left(-a + \frac{b^2}{4c}\right).$$

Beweis: Sei $g(t) = \exp(-a - bt - ct^2)$. Wir rechnen

$$\begin{aligned} g(t) &= \exp\left(-c\left(t^2 + \frac{b}{c}t + \frac{a}{c}\right)\right) \\ &= \exp\left(-c\left\{\left(t + \frac{b}{2c}\right)^2 - \frac{b^2}{4c^2} + \frac{a}{c}\right\}\right) \\ &= \exp\left(-a + \frac{b^2}{4c}\right) \exp\left(-c\left(t + \frac{b}{2c}\right)^2\right) \\ &= \exp\left(-a + \frac{b^2}{4c}\right) \sqrt{\frac{\pi}{c}} \sqrt{\frac{c}{\pi}} \exp\left(-\frac{1}{2} \frac{\left(t + \frac{b}{2c}\right)^2}{1/(2c)}\right). \end{aligned}$$

Die Normierungsbedingung für die Normalverteilung mit Erwartungswert $-b/(2c)$ und Varianz $(2c)^{-1}$ liefert daher

$$\int_{\mathbb{R}} g(t) dt = \sqrt{\frac{\pi}{c}} \exp\left(-a + \frac{b^2}{4c}\right)$$

wie gewünscht. ■

zu c): Siehe Gil-Pelaez (1951).

zu d): Wir folgen Chung (2000) und benutzen das folgende Lemma aus der Analysis ohne Beweis.

Lemma

Für alle $\alpha \in \mathbb{R}$ gilt

$$\forall y \geq 0 : 0 \leq \operatorname{sgn}(\alpha) \int_0^y \frac{\sin(\alpha x)}{x} dx \leq \int_0^\pi \frac{\sin(x)}{x} dx < \infty, \tag{9.4}$$

$$\int_0^\infty \frac{\sin(\alpha x)}{x} dx = \frac{\pi}{2} \operatorname{sgn}(\alpha). \tag{9.5}$$

Sei nun \mathbb{P} ein Wahrscheinlichkeitsmaß auf \mathbb{R} . Wir zeigen, dass für $x_1 < x_2$ gilt:

$$\mathbb{P}((x_1, x_2)) + \frac{1}{2}\mathbb{P}(\{x_1\}) + \frac{1}{2}\mathbb{P}(\{x_2\}) = \lim_{T \rightarrow \infty} \left\{ \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi_{\mathbb{P}}(t) dt \right\}.$$

Beachte dazu, dass nach dem Satz von Fubini

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \left[\int_{-\infty}^{\infty} e^{itx} \mathbb{P}(dx) \right] dt &= \int_{-\infty}^{\infty} \left[\int_{-T}^T \frac{e^{it(x-x_1)} - e^{it(x-x_2)}}{2\pi it} dt \right] \mathbb{P}(dx) \\ &=: \int_{-\infty}^{\infty} I(T, x, x_1, x_2) \mathbb{P}(dx) \end{aligned} \quad (9.6)$$

gilt. Symmetrieeigenschaften von Sinus und Cosinus liefern, dass

$$I(T, x, x_1, x_2) = \frac{1}{\pi} \int_0^T \frac{\sin(t(x-x_1))}{t} dt - \frac{1}{\pi} \int_0^T \frac{\sin(t(x-x_2))}{t} dt.$$

Wir wenden Formel (9.5) an und folgern, dass

$$\lim_{T \rightarrow \infty} I(T, x, x_1, x_2) = \begin{cases} -\frac{1}{2} - (-\frac{1}{2}) = 0, & x < x_1, \\ 0 - (-\frac{1}{2}) = \frac{1}{2}, & x = x_1, \\ \frac{1}{2} - (-\frac{1}{2}) = 1, & x_1 < x < x_2, \\ \frac{1}{2} - 0 = \frac{1}{2}, & x = x_2, \\ \frac{1}{2} - \frac{1}{2} = 0, & x > x_2. \end{cases}$$

Damit ist schließlich

$$\begin{aligned} \lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} I(T, x, x_1, x_2) \mathbb{P}(dx) &= \left\{ \int_{(-\infty, x_1)} 0 + \int_{\{x_1\}} \frac{1}{2} + \int_{(x_1, x_2)} 1 + \int_{\{x_2\}} \frac{1}{2} + \int_{(x_2, \infty)} 0 \right\} \mathbb{P}(dx) \\ &= \frac{1}{2} \mathbb{P}(\{x_1\}) + \mathbb{P}((x_1, x_2)) + \frac{1}{2} \mathbb{P}(\{x_2\}) \end{aligned}$$

wie gewünscht. ■

Korollar 9.11 (Eindeutigkeitsatz)

Ein Wahrscheinlichkeitsmaß μ auf \mathbb{R}^d ist durch Angabe der charakteristischen Funktion φ_μ eindeutig festgelegt.

Satz 9.12 (Momentenberechnung)

Sei $\mathbf{X} = (X_1, \dots, X_d)^\top$ ein Zufallsvektor mit Werten in \mathbb{R}^d . Falls $\mathbb{E}[|\mathbf{X}|^m]$ für $m \in \mathbb{N}$ endlich ist, dann ist $\varphi_{\mathbf{X}}$ m -mal stetig partiell differenzierbar und es gilt für alle $\mathbf{t} \in \mathbb{R}^d$, dass

$$\frac{\partial^m}{\partial t_{j_1} \partial t_{j_2} \dots \partial t_{j_m}} \varphi_{\mathbf{X}}(\mathbf{t}) = i^m \mathbb{E}[X_{j_1} X_{j_2} \dots X_{j_m} \exp(i\langle \mathbf{t}, \mathbf{X} \rangle)].$$

Beweisskizze:

Nach Definition ist

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_{\mathbb{R}^d} \exp(i\langle \mathbf{t}, \mathbf{x} \rangle) \mathbb{P}_{\mathbf{X}}(d\mathbf{x}).$$

Es ist nun

$$\frac{\partial}{\partial t_{j_1}} \exp(i\langle \mathbf{t}, \mathbf{x} \rangle) = ix_{j_1} \exp(i\langle \mathbf{t}, \mathbf{x} \rangle) \quad (9.7)$$

sowie

$$\int_{\mathbb{R}^d} ix_{j_1} \exp(i\langle \mathbf{t}, \mathbf{x} \rangle) \mathbb{P}_{\mathbf{X}}(d\mathbf{x}) = i\mathbb{E}[X_{j_1} \exp(i\langle \mathbf{t}, \mathbf{X} \rangle)]. \quad (9.8)$$

Man zeigt nun (siehe, z. B., Jacod and Protter (2000), Theorem 13.2), dass Differenzieren unter dem Integral unter den gemachten Annahmen zulässig ist, und wendet (9.7) und (9.8) m -mal hintereinander an. ■

Beispiel 9.13 (Normalverteilungen)

a) Sei $X \sim \mathcal{N}(0, 1)$ im \mathbb{R}^1 . Dann ist

$$\begin{aligned} \varphi_X(t) = \mathbb{E}[\exp(itX)] &= \int_{\mathbb{R}} \cos(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &\quad + i \underbrace{\int_{\mathbb{R}} \sin(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx}_{=0, \text{ da Integrand ungerade Funktion}}. \end{aligned}$$

Folglich ergibt sich für die Ableitung, dass

$$\begin{aligned} \Rightarrow \varphi'_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} -x \sin(tx) \exp\left(-\frac{x^2}{2}\right) dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t \cos(tx) \exp\left(-\frac{x^2}{2}\right) dx \\ &= -t\varphi_X(t), \end{aligned}$$

wobei wir im mittleren Schritt partiell integriert haben mit $u'(x) = -x \exp(-x^2/2)$ und $v(x) = \sin(tx)$. Also ergibt sich insgesamt

$$\frac{\varphi'_X(t)}{\varphi_X(t)} = -t \quad \Rightarrow \quad \ln(\varphi_X(t)) = -\frac{t^2}{2} + C \quad \Rightarrow \quad \varphi_X(t) = \exp\left(-\frac{t^2}{2}\right) \exp(C).$$

Wegen $\varphi_X(0) = 1$ ist $C = 0$, also folgt schließlich $\varphi_X(t) = \exp(-t^2/2)$.

b) Sei $Y \sim \mathcal{N}(\mu, \sigma^2)$ im \mathbb{R}^1 . Dann ist $Y \stackrel{\mathcal{D}}{=} \sigma X + \mu$ mit $X \sim \mathcal{N}(0, 1)$. Damit gilt nach Satz 9.9.b), dass

$$\varphi_Y(t) = \exp(it\mu) \exp\left(-\frac{\sigma^2 t^2}{2}\right) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

c) Sei $\mathbf{X} = (X_1, \dots, X_d)^\top$ standardnormalverteilt im \mathbb{R}^d . Dann liefert Satz 9.9.d), dass

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^d \exp\left(-\frac{t_k^2}{2}\right) = \exp\left(-\frac{1}{2}|\mathbf{t}|^2\right).$$

d) Sei $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$ allgemein normalverteilt, $\mathbf{Y} \sim \mathcal{N}_m(\boldsymbol{\mu}, \Sigma)$.

Dann lässt sich $\Sigma = QQ^\top$ zerlegen und $\mathbf{Y} = Q\mathbf{X} + \boldsymbol{\mu}$ schreiben, wobei \mathbf{X} standardnormalverteilt ist. Somit gilt

$$\begin{aligned} \varphi_{\mathbf{Y}}(\mathbf{u}) &= \exp(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle) \exp\left(-\frac{1}{2}|\mathbf{Q}^\top \mathbf{u}|^2\right) = \exp(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle) \exp\left(-\frac{1}{2}\langle \mathbf{Q}^\top \mathbf{u}, \mathbf{Q}^\top \mathbf{u} \rangle\right) \\ &= \exp(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle) \exp\left(-\frac{1}{2}(\mathbf{Q}^\top \mathbf{u})^\top \mathbf{Q}^\top \mathbf{u}\right) = \exp(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle) \exp\left(-\frac{1}{2}\mathbf{u}^\top \mathbf{Q} \mathbf{Q}^\top \mathbf{u}\right) \\ &= \exp(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle) \exp\left(-\frac{1}{2}\mathbf{u}^\top \Sigma \mathbf{u}\right) = \exp\left(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle - \frac{1}{2}\mathbf{u}^\top \Sigma \mathbf{u}\right) \\ &= \exp\left(i\langle \mathbf{u}, \boldsymbol{\mu} \rangle - \frac{1}{2}\langle \mathbf{u}, \Sigma \mathbf{u} \rangle\right). \end{aligned}$$

Beispiel 9.14 (Weitere Beispiele (in $d = 1$))

a) Binomialverteilung: Sei $X \sim \text{Bin}(n, p)$, so gilt:

$$\begin{aligned} \varphi_X(t) &= \sum_{k=0}^n \exp(itk) p^k (1-p)^{n-k} \binom{n}{k} \\ &= \sum_{k=0}^n [\exp(it)p]^k (1-p)^{n-k} \binom{n}{k} \\ &\stackrel{\text{bin. Lehrsatz}}{=} [p \exp(it) + (1-p)]^n. \end{aligned}$$

b) Gammaverteilung: Sei $Y \sim \text{Gamma}(1, r)$, so gilt:

$$\begin{aligned} \varphi_Y(t) &= \int_0^\infty \exp(it y) \frac{y^{r-1}}{\Gamma(r)} e^{-y} dy \\ &= \int_0^\infty \frac{y^{r-1}}{\Gamma(r)} \exp(-y(1-it)) dy \\ &= (1-it)^{-r} \int_0^\infty \frac{(1-it)^r}{\Gamma(r)} y^{r-1} \exp(-y(1-it)) dy \\ &= (1-it)^{-r} \text{ wegen Normierungsbedingung von "Gamma}(1-it, r)". \end{aligned}$$

Sei nun $X \sim \text{Gamma}(\alpha, r)$, so gilt $X \stackrel{\mathcal{D}}{=} Y/\alpha$ und damit

$$\varphi_X(t) = \varphi_Y\left(\frac{t}{\alpha}\right) = \left(1 - \frac{it}{\alpha}\right)^{-r} = \left(\frac{\alpha}{\alpha - it}\right)^r.$$

c) Sei $X \sim \text{UNI}[a, b]$ (stetige Gleichverteilung auf dem Intervall $[a, b]$). Dann ist selbstverständ-

lich $\varphi_X(0) = 1$. Für $t \neq 0$ rechnen wir:

$$\begin{aligned}
 \varphi_X(t) &= \int_a^b \frac{\exp(itx)}{(b-a)} dx = \left[(it(b-a))^{-1} \exp(itx) \right]_a^b \\
 &= \frac{\exp(itb) - \exp(ita)}{it(b-a)}. \\
 a = -b \Rightarrow \varphi_X(t) &= \frac{\exp(itb) - \exp(-itb)}{2itb} \\
 &= \frac{\cos(tb) + i \sin(tb) - \cos(-tb) - i \sin(-tb)}{2itb} \\
 &= \frac{\sin(tb)}{tb}.
 \end{aligned}$$

d) Seien $(X_k)_{k \in \mathbb{N}}$ stochastisch unabhängig und identisch verteilt. Sei N eine weitere Zufallsvariable, stochastisch unabhängig von den X_k , mit Werten in \mathbb{N} . Sei $S := \sum_{k=1}^N X_k$.

$$\begin{aligned}
 \Rightarrow \varphi_S(t) &= \mathbb{E} \left[\exp \left(it \sum_{j=1}^N X_j \right) \right] \\
 &= \sum_{n \in \mathbb{N}} \mathbb{P}(N = n) \varphi_{X_1}^n(t) = \sum_n \mathbb{P}(N = n) \exp(n \ln \varphi_{X_1}(t)) \\
 &= \mathbb{E} [\exp(N \ln \varphi_{X_1}(t))] = \mathbb{E} [\exp(iN(-i) \ln \varphi_{X_1}(t))] \\
 &= \varphi_N(-i \ln \varphi_{X_1}(t)) \text{ bei entsprechendem Konvergenzradius in } \mathbb{C}.
 \end{aligned}$$

Kapitel 10

Folgen von Zufallsvariablen, Grenzwertsätze

In diesem Kapitel betrachten wir Folgen $(X_n)_{n \geq 1}$ von (reellwertigen) Zufallsvariablen mit $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $n \geq 1$, und beschreiben, in welchen Weisen die Folge $(X_n)_{n \geq 1}$ gegen einen Grenzwert, also eine Grenz-Zufallsvariable $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, konvergieren kann (für $n \rightarrow \infty$). Da X_n , $n \geq 1$, und X Funktionen sind, lassen sich (wie in der Funktionalanalysis) verschiedene Konvergenzarten unterscheiden, die in der Wahrscheinlichkeitstheorie mit besonderen Begriffen belegt werden. Es bestehen ferner Implikationsbeziehungen zwischen den Konvergenzarten, d.h., die “Stärke” der Konvergenz lässt sich unterscheiden.

Definition 10.1 (Konvergenzarten)

Sei $(X_n)_{n \geq 1}$ eine Folge von Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum, d. h., für alle $n \in \mathbb{N}$ ist $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine messbare Abbildung. Ferner sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine weitere (reellwertige) Zufallsvariable auf dem gleichen Wahrscheinlichkeitsraum wie $(X_n)_{n \geq 1}$.

a) Die Folge $(X_n)_{n \geq 1}$ konvergiert \mathbb{P} -fast sicher (mit Wahrscheinlichkeit 1) gegen X für $n \rightarrow \infty$

$$:\Leftrightarrow \mathbb{P} \left(\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right) = 1$$

$$\Leftrightarrow \mathbb{P} \left(\lim_{n \rightarrow \infty} X_n = X \right) = 1.$$

In Zeichen: $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$

b) Die Folge $(X_n)_{n \geq 1}$ konvergiert \mathbb{P} -stochastisch (in Wahrscheinlichkeit) gegen X für $n \rightarrow \infty$

$$:\Leftrightarrow \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P} (|X_n - X| > \varepsilon) = 0.$$

In Zeichen: $X_n \xrightarrow{\mathbb{P}} X$

c) Die Folge $(X_n)_{n \geq 1}$ konvergiert in Verteilung (schwach) gegen X für $n \rightarrow \infty$

$$:\Leftrightarrow \forall x \in C(F_X) : \lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

In Zeichen: $X_n \xrightarrow{D} X$ bzw. $\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X)$.

Beachte: Das Maß \mathbb{P} wird für die Definition der Verteilungskonvergenz nicht benötigt. Daher können die X_n und/oder X in dieser Definition sogar auf unterschiedlichen Wahrscheinlichkeitsräumen definiert sein. Eine exaktere Definition lautet daher:

Sei (Ω', d) ein metrischer Raum und \mathcal{A}' die von den offenen Kugeln in der Metrik d erzeugte σ -Algebra. Seien \mathbb{P} und $(\mathbb{P}_n)_{n \geq 1}$ Wahrscheinlichkeitsmaße auf dem Messraum (Ω', \mathcal{A}') . Dann konvergiert die Folge $(\mathbb{P}_n)_{n \geq 1}$ schwach gegen \mathbb{P} für $n \rightarrow \infty$

$$:\Leftrightarrow \forall f \in C_b(\Omega') : \lim_{n \rightarrow \infty} \int f d\mathbb{P}_n = \int f d\mathbb{P},$$

wobei $C_b(\Omega')$ die Menge aller stetigen und beschränkten Abbildungen $f : \Omega' \rightarrow \mathbb{R}$ bezeichnet.

d) Sei $p \geq 1$ und seien X, X_1, X_2, \dots reellwertige Zufallsvariablen mit in \mathbb{R} existierendem p -ten Moment. Dann konvergiert die Folge $(X_n)_{n \geq 1}$ im p -ten Mittel gegen X für $n \rightarrow \infty$

$$:\Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

In Zeichen: $X_n \xrightarrow{L_p} X$

Spezialfälle:

$p = 1$: Konvergenz im Mittel

$p = 2$: Konvergenz im quadratischen Mittel

Aus der Diskussion in Definition 10.1.c) über die Verteilungskonvergenz (schwache Konvergenz der Verteilungsgesetze) hat sich bereits ergeben, dass es unterschiedliche, äquivalente Charakterisierungen der vier in Definition 10.1 beschriebenen Konvergenzarten gibt. Dazu nun mehr.

Satz 10.2 (Alternative Charakterisierungen)

a)

$$X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X \Leftrightarrow \mathbb{P} \left(\liminf_{n \rightarrow \infty} (X_n - X) = \limsup_{n \rightarrow \infty} (X_n - X) = 0 \right) = 1$$

$$\Leftrightarrow \forall \omega \in \Omega \setminus N : \lim_{n \rightarrow \infty} (X_n(\omega) - X(\omega)) = 0,$$

wobei N eine \mathbb{P} -Nullmenge bezeichnet, d. h., $\mathbb{P}(N) = 0$.

Beachte: $Y_n := X_n - X \Rightarrow$

$$\left\{ \lim_{n \rightarrow \infty} Y_n = 0 \right\} = \bigcap_{m=1}^{\infty} \bigcup_{k=0}^{\infty} \bigcap_{n=k}^{\infty} \left\{ |Y_n| < \frac{1}{m} \right\}$$

und damit messbar!

b)

$$X_n \xrightarrow{\mathcal{D}} X \Leftrightarrow \forall f \in \mathcal{C}_b(\mathbb{R}) : \mathbb{E}[f(X_n)] = \int f d\mathcal{L}(X_n) \xrightarrow{n \rightarrow \infty} \int f d\mathcal{L}(X) = \mathbb{E}[f(X)].$$

Beweis:

zu a): Die Aussage ist unmittelbar klar.

zu b): Der Beweis macht von dem folgendem Hilfssatz Gebrauch, der Bezüge zwischen der Topologie und der Integrationstheorie auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ herstellt. Er ist Teil des sogenannten "Portmanteau Theorem" und findet sich z.B. in Ash (1972), Theorem 5.4.1 d) + e).

Hilfssatz 10.3 (ohne Beweis)

$$\begin{aligned} & \forall f \in \mathcal{C}_b(\mathbb{R}) : \mathbb{E}[f(X_n)] \xrightarrow{(n \rightarrow \infty)} \mathbb{E}[f(X)] \\ \Leftrightarrow & \liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(A) \geq \mathbb{P}^X(A) \text{ für alle offenen Teilmengen } A \text{ von } \mathbb{R} \\ \Leftrightarrow & \mathbb{P}^{X_n}(A) \rightarrow \mathbb{P}^X(A) \forall A \in \mathcal{B}(\mathbb{R}) \text{ mit } \mathbb{P}^X(\partial A) = 0 \text{ ("randlose Mengen").} \end{aligned}$$

Da $(-\infty, x]$ für $x \in C(F_X)$ eine randlose Menge ist, liefert die zweite Äquivalenz im Hilfsatz unmittelbar die " \Leftarrow "-Richtung der Aussage unter b).

Zum Beweis der " \Rightarrow "-Richtung zeigen wir:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \forall x \in C(F_X) \Rightarrow \forall A \subseteq \mathbb{R} \text{ offen: } \liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(A) \geq \mathbb{P}^X(A).$$

Sei dazu $A \subseteq \mathbb{R}$ offen beliebig ausgewählt. Wir schreiben A als disjunkte Vereinigung offener Intervalle I_1, I_2, \dots . Damit ergibt sich nach dem Lemma von Fatou

$$\liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(A) = \liminf_{n \rightarrow \infty} \sum_k \mathbb{P}^{X_n}(I_k) \geq \sum_k \liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(I_k). \quad (10.1)$$

Da F_X nur abzählbar viele Unstetigkeitsstellen besitzen kann, lässt sich für jede Konstante $\varepsilon > 0$ die folgende Konstruktion durchführen:

Für jedes k sei I'_k ein rechtsseitig abgeschlossenes Teilintervall von I_k , so dass

(1) alle Endpunkte der I'_k in $C(F_X)$ enthalten sind und

(2) $\forall k : \mathbb{P}^X(I'_k) \geq \mathbb{P}^X(I_k) - \varepsilon 2^{-k}$.

Da $X_n \xrightarrow{\mathcal{D}} X$, gilt nun

$$\liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(I_k) \geq \liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(I'_k) = \mathbb{P}^X(I'_k).$$

Folglich gilt für (10.1), dass

$$\liminf_{n \rightarrow \infty} \mathbb{P}^{X_n}(A) \geq \sum_k \mathbb{P}^X(I'_k) \geq \sum_k \mathbb{P}^X(I_k) - \varepsilon = \mathbb{P}^X(A) - \varepsilon,$$

wegen $\sum_{k=1}^{\infty} 2^{-k} = 1$. Da ε beliebig klein gewählt werden kann, ist hiermit alles gezeigt. ■

Satz 10.4 (Lévy'scher Stetigkeitssatz)

Es sei $(X_n)_{n \geq 1}$ eine Folge von Zufallsvariablen mit zugehörigen charakteristischen Funktionen $(\varphi_n)_{n \geq 1}$.

a) Falls $(X_n)_{n \geq 1}$ gegen eine Zufallsvariable X in Verteilung konvergiert, dann konvergiert $(\varphi_n)_{n \geq 1}$ gegen die charakteristische Funktion von X , und zwar gleichmäßig auf jedem endlichen Intervall.

b) Falls $(\varphi_n)_{n \geq 1}$ punktweise gegen eine Funktion φ konvergiert, deren Realteil im Punkte $(0, 1)$ stetig ist, dann gilt:

(i) φ ist eine charakteristische Funktion, und damit existiert (genau) eine Wahrscheinlichkeitsverteilung μ , deren charakteristische Funktion gerade φ ist.

(ii) $\mathcal{L}(X_n) \xrightarrow{w} \mu$ für $n \rightarrow \infty$.

Beweis zu Teil a). Sei $\mu := \mathbb{P}^X$, $\mu_n := \mathbb{P}^{X_n}$ und $\varphi = \varphi_X$. Zum Beweis von Teil a) beachten wir, dass für jedes feste t die Funktion $x \mapsto \exp(itx)$, $i = \sqrt{-1}$, beschränkt und stetig ist. Damit kann Satz 10.2.b) auf Real- und Imaginärteil angewendet werden. Daraus folgt die punktweise (bzgl. t) Konvergenz von φ_n gegen φ sofort. Ferner ist

$$|\varphi_n(t+h) - \varphi_n(t)| \leq \int |\exp(ihx) - 1| d\mu_n(x) \rightarrow \int |\exp(ihx) - 1| d\mu(x),$$

$n \rightarrow \infty$, für alle t und h . Das letzte Integral hängt nicht von t ab und strebt gegen 0 für $h \rightarrow 0$. Also ist $\{\varphi_n\}_{n \geq 1}$ gleichgradig stetig. Zusammen mit der punktweisen Konvergenz von φ_n gegen φ folgt die Aussage von Teil a). ■

Beweisskizze zu Teil b). Für den Beweis von Teil b) ist das folgende Lemma hilfreich.

Lemma

Für jedes Wahrscheinlichkeitsmaß μ auf \mathbb{R} mit zugehöriger charakteristischer Funktion φ gilt

$$\forall A > 0 : \mu([-2A, 2A]) \geq A \left| \int_{-\frac{1}{A}}^{\frac{1}{A}} \varphi(t) dt \right| - 1.$$

Beweis: In Analogie zu den Rechnungen im Beweis der Umkehrformel von Chung ist

$$\forall T > 0 : \frac{1}{2T} \int_{-T}^T \varphi(t) dt = \int_{-\infty}^{\infty} \frac{\sin(Tx)}{Tx} d\mu(x).$$

Nun ist aber $\frac{\sin(Tx)}{Tx} \leq 1$ für alle x und $|Tx|^{-1} \leq (2TA)^{-1}$ für $|x| > 2A$. Es folgt, dass

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\sin(Tx)}{Tx} d\mu(x) &\leq \mu([-2A, 2A]) + \frac{1}{2TA} \{1 - \mu([-2A, 2A])\} \\ &= \left(1 - \frac{1}{2TA}\right) \mu([-2A, 2A]) + \frac{1}{2TA}. \end{aligned}$$

Setzen wir $T = A^{-1}$, so ergibt sich

$$\left| \frac{A}{2} \int_{-\frac{1}{A}}^{\frac{1}{A}} \varphi(t) dt \right| \leq \frac{1}{2} \mu([-2A, 2A]) + \frac{1}{2},$$

und die Aussage folgt. ■

Um nun Teil b) des Satzes zu beweisen, beachten wir, dass für alle $n \geq 1$ und $\delta > 0$ nach Dreiecksungleichung gilt:

$$\left| \frac{1}{2\delta} \int_{-\delta}^{\delta} \varphi_n(t) dt \right| \geq \left| \frac{1}{2\delta} \int_{-\delta}^{\delta} \varphi(t) dt \right| - \frac{1}{2\delta} \int_{-\delta}^{\delta} |\varphi_n(t) - \varphi(t)| dt.$$

Wegen der angenommenen Stetigkeit von φ in $(0, 1)$ strebt der erste Summand gegen $1 = \varphi(0)$ für $\delta \downarrow 0$. Andererseits strebt für jedes feste $\delta > 0$ der zweite Summand gegen 0 für $n \rightarrow \infty$ wegen der angenommenen punktweisen Konvergenz von φ_n gegen φ . Insgesamt folgt damit

$$\forall \varepsilon > 0 \exists \delta(\varepsilon) \text{ und } n_0(\varepsilon) : \forall n \geq n_0 : \left| \frac{1}{2\delta} \int_{-\delta}^{\delta} \varphi_n(t) dt \right| \geq 1 - \varepsilon.$$

Wegen des vorigen Lemmas mit $A = 1/\delta$ folgt

$$\begin{aligned} \mu_n \left(\left[-\frac{2}{\delta}, \frac{2}{\delta} \right] \right) &\geq \frac{1}{\delta} \left| \int_{-\delta}^{\delta} \varphi_n(t) dt \right| - 1 \\ &\geq \frac{1}{\delta} (2\delta - 2\delta\varepsilon) - 1 = 2 - 2\varepsilon - 1 \\ &= 1 - 2\varepsilon \end{aligned}$$

für solche δ und n .

Nun besitzt aber jede Teilfolge von $\{\mu_n\}_{n \geq 1}$ eine konvergente Teilfolge $\{\mu_{n_k}\}_{k \geq 1}$.

Hinweis: Das wird hier nicht gezeigt, deswegen ist dies hier nur eine Beweisskizze!

Sei μ der Grenzwert einer solchen Teilfolge. Dann ist

$$\mu(\mathbb{R}) \geq \mu \left(\left[-\frac{2}{\delta}, \frac{2}{\delta} \right] \right) \geq 1 - 2\varepsilon.$$

Also ist jedes solche μ ein Wahrscheinlichkeitsmaß, da ε beliebig klein gewählt werden kann. Sei φ_μ die charakteristische Funktion von μ . Da $(\varphi_n)_{n \geq 1}$ punktweise gegen eine fest vorgegebene Funktion φ konvergent ist, muss $\varphi_\mu = \varphi$ gelten und μ damit eindeutig bestimmt sein, denn Teil a) des Satzes kann auf jede der konvergenten Teilfolgen angewendet werden.

Ein vollständiger Beweis zu Teil b) findet sich z. B. in Klenke (2008), siehe Satz 15.23 dort. ■

Anmerkung: Analoge Stetigkeitssätze gelten auch für erzeugende Funktionen und Laplace-Transformierte.

Satz 10.5 (Implikationsbeziehungen zwischen Konvergenzarten)

Es sei $(X_n)_{n \geq 1}$ eine Folge von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und X eine weitere Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Ferner sei $p \geq 1$ eine reelle Konstante.

(a) $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$

(b) $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$ impliziert die Existenz des p -ten Moments von X sowie $X_n \xrightarrow{L_p} X$ genau dann, wenn $\mathcal{H} := \{|X_n|^p : n \geq 1\}$ gleichgradig integrierbar ist, d.h., falls

$$\lim_{c \rightarrow \infty} \sup_{f \in \mathcal{H}} \int_{\{|f| \geq c\}} |f| d\mathbb{P} = 0.$$

(c) $X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{L_q} X, \forall 1 \leq q \leq p.$

(d) $X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$

(e) $X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathcal{D}} X.$

(f) Es ergibt sich damit die in Abbildung 10.1 dargestellte Grafik.

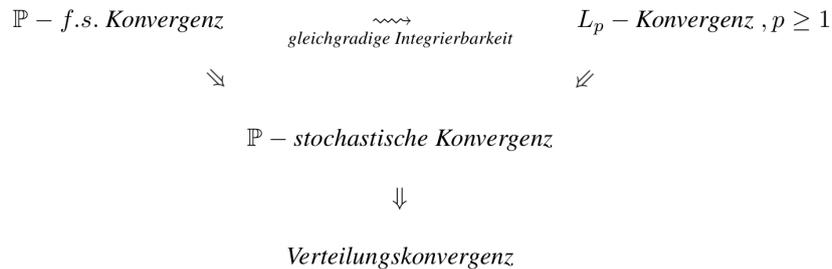


Abbildung 10.1: Zusammenhang von Konvergenzarten

Beweis:

zu a): Sei $\varepsilon > 0$ beliebig, aber fest. Definiere $A_n(\varepsilon) := \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\}$. Wir müssen zeigen, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n(\varepsilon)) = 1.$$

Sei dazu $A := \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$. Nach Voraussetzung dürfen wir annehmen, dass $\mathbb{P}(A) = 1$ ist. Nun ist aber für hinreichend großes n die Menge A eine Teilmenge von $A_n(\varepsilon)$, denn für alle $\omega \in A$ liegen alle bis auf endlich viele Folgenglieder $X_n(\omega)$ in jeder ε -Umgebung von $X(\omega)$. Dies impliziert die Aussage wegen der Monotonie von \mathbb{P} .

zu b): vgl. Abschnitt 6.2 in Klenke (2008).

zu c): Die Funktion g , definiert durch $g(t) := t^{\frac{p}{q}}$, ist konvex auf $\mathbb{R}_{\geq 0} \ni t$. Nach der Jensen'schen Ungleichung (Satz 8.24) gilt daher für alle $n \in \mathbb{N}$, dass

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}\left[|X_n - X|^{\frac{p}{q}}\right]^q \geq (\mathbb{E}[|X_n - X|^q])^{\frac{p}{q}}$$

und damit

$$(\mathbb{E}[|X_n - X|^p])^{\frac{1}{p}} \geq (\mathbb{E}[|X_n - X|^q])^{\frac{1}{q}}.$$

zu d): Wir wenden die Markov-Ungleichung (Satz 8.22) auf $Y_n := |X_n - X|$ mit $h(t) := t^p$ an und erhalten für jedes $\varepsilon > 0$, dass

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-p} \mathbb{E}[|X_n - X|^p].$$

zu e): Sei f eine gleichmäßig stetige, beschränkte Funktion auf \mathbb{R} und $\varepsilon > 0$ beliebig vorgegeben. Dann gibt es ein $\delta > 0$ mit der Eigenschaft

$$|x - y| \leq \delta \Rightarrow |f(x) - f(y)| < \varepsilon, \quad x, y \in \mathbb{R}.$$

Wir rechnen:

$$\begin{aligned} & \left| \int f(X_n) d\mathbb{P} - \int f(X) d\mathbb{P} \right| \leq \int |f(X_n) - f(X)| d\mathbb{P} \\ &= \int_{\{|X_n - X| \leq \delta\}} |f(X_n) - f(X)| d\mathbb{P} + \int_{\{|X_n - X| > \delta\}} |f(X_n) - f(X)| d\mathbb{P} \\ &\leq \varepsilon \mathbb{P}(|X_n - X| \leq \delta) + 2 \sup_{x \in \mathbb{R}} |f(x)| \cdot \mathbb{P}(|X_n - X| > \delta). \end{aligned}$$

Also gilt wegen $X_n \xrightarrow{\mathbb{P}} X$, dass

$$\limsup_{n \rightarrow \infty} \left| \int f(X_n) d\mathbb{P} - \int f(X) d\mathbb{P} \right| \leq \varepsilon$$

und damit

$$\int f(X_n) d\mathbb{P} \xrightarrow{n \rightarrow \infty} \int f(X) d\mathbb{P},$$

da ε beliebig gewählt wurde. Da aber nach Transformationssatz

$$\int f(X_n) d\mathbb{P} \xrightarrow{n \rightarrow \infty} \int f(X) d\mathbb{P} \iff \int f d\mathbb{P}^{X_n} \xrightarrow{n \rightarrow \infty} \int f d\mathbb{P}^X$$

gilt, ist hiermit alles gezeigt. ■

Bemerkung 10.6

Die Implikationen aus Satz 10.5 sind im Allgemeinen strikt, d.h., die Umkehrungen gelten allgemein nicht. Ein Beispiel für $X_n \xrightarrow{\mathcal{D}} X$, aber $X_n \not\xrightarrow{\mathbb{P}} X$ ist gegeben durch

$$X_n(\omega) = \mathbf{1}_{[0, \frac{1}{2}]}(\omega), n \geq 1, \text{ und } X(\omega) = \mathbf{1}_{(\frac{1}{2}, 1]}(\omega)$$

auf $([0, 1], \mathcal{B}([0, 1]), \text{UNI}[0, 1])$.

In dem Spezialfall, dass $X \equiv x_0$ \mathbb{P} -fast sicher konstant ist, gilt jedoch:

$$X_n \xrightarrow{\mathbb{P}} x_0 \Leftrightarrow X_n \xrightarrow{\mathcal{D}} X = x_0.$$

Beweis: siehe Bauer (1991), Beweis von Satz 5.1. ■

Ein für die mathematische Statistik ungemein wichtiger Satz beschließt den technischen Teil dieses Kapitels.

Satz 10.7 (Satz von Cramér-Slutsky (Slutzky))

Seien $(X_n)_{n \geq 1}$ und $(Y_n)_{n \geq 1}$ zwei Folgen von Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine weitere Zufallsvariable.

a) $X_n \xrightarrow{\mathcal{D}} X$ und $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0 \Rightarrow Y_n \xrightarrow{\mathcal{D}} X$.

b) Sei $c \in \mathbb{R}$. $X_n \xrightarrow{\mathcal{D}} X$ und $Y_n \xrightarrow{\mathcal{D}} c \Rightarrow \left\{ \begin{array}{l} (i) \quad X_n + Y_n \xrightarrow{\mathcal{D}} X + c. \\ (ii) \quad X_n Y_n \xrightarrow{\mathcal{D}} cX. \end{array} \right\}$

Beweis:

zu a): Sei $f \in \mathcal{C}_b(\mathbb{R})$ mit Lipschitz-Konstante K . Dann ist

$$|f(x) - f(y)| \leq K|x - y| \wedge 2 \sup_{u \in \mathbb{R}} |f(u)|, \forall x, y \in \mathbb{R}.$$

Der Satz von der majorisierten Konvergenz liefert, dass

$$\limsup_{n \rightarrow \infty} \mathbb{E} [|f(X_n) - f(Y_n)|] = 0.$$

Also ergibt sich

$$\begin{aligned} & \limsup_{n \rightarrow \infty} |\mathbb{E} [f(Y_n)] - \mathbb{E} [f(X)]| \\ & \leq \limsup_{n \rightarrow \infty} |\mathbb{E} [f(X)] - \mathbb{E} [f(X_n)]| + \limsup_{n \rightarrow \infty} |\mathbb{E} [f(X_n) - f(Y_n)]| \\ & = 0. \end{aligned}$$

zu b): (i) Definiere $Z_n := X_n + c$ und $\tilde{Z}_n := X_n + Y_n$. Dann gilt $Z_n \xrightarrow{\mathcal{D}} X + c$ und $|Z_n - \tilde{Z}_n| \xrightarrow{\mathbb{P}} 0$, vgl. Bemerkung 10.6. Also kann Teil a) angewendet werden.

- (ii) Siehe Theorem 2.3.3 in Lehmann (1999); Beweis in Bickel and Doksum (1977) bzw. Cramér (1946). ■

“Stillschweigend” haben wir den folgenden Satz benutzt.

Satz 10.8 (Continuous Mapping Theorem, siehe Abschnitt 1.7 in Serfling (1980))

Sei $h : \mathbb{R} \rightarrow \mathbb{R}$ messbar und stetig $\Rightarrow [X_n \xrightarrow{D} X \Rightarrow h(X_n) \xrightarrow{D} h(X)]$.

Wir kommen nun zu Anwendungen der Konvergenztheorie für Folgen von Zufallsvariablen.

Satz 10.9 (Kolmogoroffsches 0 – 1 Gesetz)

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge stochastisch unabhängiger Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit beliebigen Wertebereichen. Dann gilt für jedes terminale (bzw. asymptotische) Ereignis, d.h., für jedes Ereignis $A \in \mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(\{X_m\} : m > n)$ entweder $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$. Dabei heißt \mathcal{T} terminale σ -Algebra. Sie enthält Ereignisse, über deren Eintreten nach der Beobachtung von endlich vielen der X_n noch keine Aussage getroffen werden kann.

Beweis: Sei $(\Omega'_k, \mathcal{A}'_k)$ der Wertebereich von $X_k, k \in \mathbb{N}$, und seien $n \in \mathbb{N}$ sowie $C_k \in \mathcal{A}'_k, k = 1, \dots, n$, beliebig ausgewählt. Definiere $C := \{X_1 \in C_1, \dots, X_n \in C_n\}$. Dann ist

$$\mathbf{1}_C = \prod_{k=1}^n \mathbf{1}_{C_k}(X_k)$$

stochastisch unabhängig von $\mathbf{1}_A$.

Ferner erzeugt das System aller Mengen C die Produkt- σ -Algebra $\bigotimes_{k \geq 1} \mathcal{A}'_k$ und deswegen ist $(X_k)_{k \geq 1} \perp \perp \mathbf{1}_A$. Insbesondere ist A als Element von $\bigcap_{n \geq 1} \sigma(\{X_m\} : m > n)$ damit stochastisch unabhängig von $A = \{\mathbf{1}_A = 1\}$, d.h., $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C) \Rightarrow \mathbb{P}(A) = [\mathbb{P}(A)]^2$. Die Gleichung $x = x^2$ hat aber nur die Lösungen 0 und 1. ■

Korollar 10.10

Es sei $(X_n)_{n \geq 1}$ eine Folge stochastisch unabhängiger, reellwertiger Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann sind $\liminf_{n \rightarrow \infty} X_n, \limsup_{n \rightarrow \infty} X_n$, sowie die Cesàro-Limiten $\liminf_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i$ und $\limsup_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i$ allesamt \mathbb{P} -fast sicher konstant.

Beweis: Korollar 2.39 in Klenke (2008). ■

Satz 10.11 (Lemma von Borel-Cantelli)

Sei $(A_k)_{k \geq 1}$ eine Folge von Ereignissen in einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und $A := \limsup_{k \rightarrow \infty} A_k = \{\omega \in \Omega : \omega \in A_k \text{ für unendlich viele } k\}$.

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

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

Beweis:

zu (a): Für alle $m \in \mathbb{N}$ ist $A \subseteq \bigcup_{k \geq m} A_k$ und daher $\mathbb{P}(A) \leq \sum_{k \geq m} \mathbb{P}(A_k)$. Falls nun $\sum_{k \geq 1} \mathbb{P}(A_k) < \infty$, so folgt $\lim_{m \rightarrow \infty} \sum_{k \geq m} \mathbb{P}(A_k) = 0$ und damit $\mathbb{P}(A) = 0$.

zu (b): Wir beachten, dass $A^c = \bigcup_{m \geq 1} \bigcap_{k \geq m} A_k^c$ ist. Es folgt

$$\begin{aligned} \mathbb{P}(A^c) &\leq \sum_{m \geq 1} \mathbb{P} \left(\bigcap_{k \geq m} A_k^c \right) = \sum_{m \geq 1} \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcap_{k=m}^n A_k^c \right) \\ &= \sum_{m \geq 1} \lim_{n \rightarrow \infty} \prod_{k=m}^n (1 - \mathbb{P}(A_k)) \end{aligned}$$

wegen der vorausgesetzten stochastischen Unabhängigkeit der $(A_k)_{k \geq 1}$. Anwendung der bekannten Abschätzung $1 - x \leq \exp(-x) \forall x \in [0, 1]$ ergibt, dass

$$\begin{aligned} \mathbb{P}(A^c) &\leq \sum_{m \geq 1} \lim_{n \rightarrow \infty} \exp \left(- \sum_{k=m}^n \mathbb{P}(A_k) \right) \\ &= \sum_{m \geq 1} 0 = 0. \end{aligned}$$

■

Eine zentrale Fragestellung in der (mathematischen) Statistik lautet: „Unter welchen Voraussetzungen konzentriert sich der arithmetische Mittelwert (das empirische Mittel) einer Folge $(X_n)_{n \geq 1}$ von Zufallsvariablen ‘hinreichend gut’ um die theoretischen Mittelwerte $\mathbb{E}[X_n]$ für $n \rightarrow \infty$?“

Die Beantwortung dieser Frage ist zentral zur Beurteilung der Qualität von Schätz- und Testverfahren. Das einfachste Beispiel ist vermutlich ein Bernoulli’sches Versuchsschema. Kann die Trefferwahrscheinlichkeit p aus einer ”langen“ Messreihe ”gut“ inferiert werden?

Wahrscheinlichkeitstheoretisch wird dieser Problemkreis mit den Gesetzen der großen Zahlen bearbeitet.

Satz 10.12 (Gesetze der großen Zahlen)

Es sei $(X_n)_{n \geq 1}$ eine Folge von integrierbaren, reellwertigen Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Sei

$$S_n := \sum_{i=1}^n (X_i - \mathbb{E}[X_i]).$$

Wir sagen, dass $(X_n)_{n \geq 1}$ dem schwachen bzw. starken Gesetz der großen Zahlen genügt, falls

$$n^{-1} S_n \xrightarrow{\mathbb{P}} 0 \text{ bzw. } n^{-1} S_n \xrightarrow{\mathbb{P}\text{-f.s.}} 0.$$

(a) $(X_n)_{n \geq 1}$ genügt dem schwachen Gesetz der großen Zahlen, falls die $(X_n)_{n \in \mathbb{N}}$ paarweise unkorreliert sind und

$$\lim_{n \rightarrow \infty} n^{-2} \sum_{i=1}^n \text{Var}(X_i) = 0$$

gilt.

(b) $(X_n)_{n \geq 1}$ genügt dem starken Gesetz der großen Zahlen, falls die $(X_n)_{n \in \mathbb{N}}$ identisch verteilt und paarweise stochastisch unabhängig sind.

Beweis:

zu (a): Offenbar besitzt X_n eine endliche Varianz für alle $n \in \mathbb{N}$. Ferner ist $\mathbb{E}[S_n] = 0$ und $\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i)$ (nach Bienaymé) für alle $n \in \mathbb{N}$.

$$\text{Also ist } \text{Var}(n^{-1}S_n) = n^{-2} \sum_{i=1}^n \text{Var}(X_i) =: \sigma_n^2.$$

Nach der Chebyshev-Ungleichung (Korollar 8.23.(b)) folgt, dass

$$\forall \varepsilon > 0 : \mathbb{P}(|n^{-1}S_n| \geq \varepsilon) \leq \varepsilon^{-2} \sigma_n^2.$$

Die Bedingung $\sigma_n^2 \xrightarrow[n \rightarrow \infty]{} 0$ impliziert die \mathbb{P} -stochastische Konvergenz von $n^{-1}S_n$.

zu (b): Etemadi (1981) benutzt das Lemma von Borel-Cantelli (Satz 10.11), den Satz von der monotonen Konvergenz und eine Abschneidetechnik, die ähnlich auch beim Zentralen Grenzwertsatz in der Version von Lindeberg/Feller (siehe unten) gebraucht wird.

■

Satz 10.13 (Zentraler Grenzwertsatz)

Sei $(X_n)_{n \geq 1}$ eine Folge (reellwertiger) stochastisch unabhängiger Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit endlichen zweiten Momenten und nicht-trivialer Varianz. O.B.d.A. sei $\mathbb{E}[X_k] = 0$ für alle $k \in \mathbb{N}$. Wir bezeichnen ferner für $k \in \mathbb{N}$ mit $\sigma_k^2 := \text{Var}(X_k) = \mathbb{E}[X_k^2] > 0$ die Varianz von X_k .

Sei $S_n := \sum_{j=1}^n X_j$. Beachte, dass $\text{Var}(S_n) = \sum_{k=1}^n \sigma_k^2$ gilt.

Wir sagen, dass für die Folge $(X_n)_{n \geq 1}$ ein Zentraler Grenzwertsatz gilt, falls

$$\mathcal{L} \left(\frac{S_n}{\sqrt{\text{Var}(S_n)}} \right) \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(0, 1).$$

Die folgenden drei Bedingungen sind jeweils hinreichend dafür, dass ein Zentraler Grenzwertsatz für $(X_n)_{n \geq 1}$ gilt:

(i) Alle X_k , $k \in \mathbb{N}$, haben die selbe Verteilung.

(ii) *Ljapunov-Bedingung:*

$$\exists \delta > 0 : \alpha_k := \mathbb{E} \left[|X_k^{2+\delta}| \right] < \infty \quad \forall k \in \mathbb{N} \quad \text{und} \quad \sum_{i=1}^n \alpha_i = o \left((\text{Var}(S_n))^{\frac{2+\delta}{2}} \right)$$

$$\Leftrightarrow \lim_{n \rightarrow \infty} (\text{Var}(S_n))^{-\frac{2+\delta}{2}} \sum_{j=1}^n \mathbb{E} \left[|X_j|^{2+\delta} \right] = 0.$$

(iii) *Lindeberg-Bedingung:*

$$\forall \varepsilon > 0 : [\text{Var}(S_n)]^{-1} \sum_{j=1}^n \int_{\{|y| \geq \varepsilon \sqrt{\text{Var}(S_n)}\}} y^2 F_j(dy) \xrightarrow{(n \rightarrow \infty)} 0,$$

wobei $F_j(x) = \mathbb{P}(X_j \leq x), j \in \mathbb{N}$.

Bemerkung 10.14

a) $(i) \Rightarrow (ii) \Rightarrow (iii)$.

b) Die Lindeberg-Bedingung stellt sicher, dass die individuellen Varianzen der X_k klein sind im Vergleich zu ihrer Summe, denn (iii) impliziert, dass für gegebenes $\delta > 0$ ein $N(\delta)$ existiert mit der Eigenschaft

$$\forall n > N(\delta) : \frac{\sigma_k}{\sqrt{\text{Var}(S_n)}} < \delta \quad \forall k = 1, \dots, n.$$

c) $(i) \Rightarrow (iii)$ ist leicht einzusehen. Sind $(X_n)_{n \geq 1}$ stochastisch unabhängig und identisch verteilt, so ist $\text{Var}(S_n) = n\sigma^2$ (mit $\sigma^2 = \text{Var}(X_1)$) und die linke Seite der Lindeberg-Bedingung wird zu $\sigma^{-2} \int_{\{|y| \geq \varepsilon \sqrt{n}\sigma\}} y^2 F(dy)$ mit $F(x) = \mathbb{P}(X_1 \leq x)$.

Da X_1 ein endliches zweites Moment sowie eine nicht-triviale Varianz besitzt und der Integrationsweg für $n \rightarrow \infty$ verschwindet, folgt die Gültigkeit der Lindeberg-Bedingung.

Beweis:

Beweis unter (i):

Sei φ die charakteristische Funktion von X_1/σ , wobei $\sigma^2 = \text{Var}(X_1)$. Wir müssen zeigen, dass

$$\mathcal{L} \left(\sum_{j=1}^n X_j / (\sqrt{n}\sigma) \right) \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(0, 1).$$

Für fixes n ist die charakteristische Funktion von $\sum_{j=1}^n X_j / (\sqrt{n}\sigma)$ gegeben durch $t \mapsto \varphi^n \left(\frac{t}{\sqrt{n}} \right)$.

Es bleibt nach Lévy'schem Stetigkeitssatz (Satz 10.4) zu zeigen:

$$\lim_{n \rightarrow \infty} \varphi^n \left(\frac{t}{\sqrt{n}} \right) = \exp \left(-\frac{t^2}{2} \right) \quad \text{punktweise für alle } t \in \mathbb{R}.$$

Da X_1 ein endliches zweites Moment besitzt, ist φ nach Satz 9.12 zweimal stetig differenzierbar. Wegen

$$\mathbb{E}[X_1] = 0 = \left. \frac{d}{dt} \varphi(t) \right|_{t=0} \quad \text{sowie} \quad \mathbb{E} \left[\left(\frac{X_1}{\sigma} \right)^2 \right] = 1 = - \left. \frac{d^2}{dt^2} \varphi(t) \right|_{t=0}$$

gilt somit für die Taylorentwicklung um 0, dass

$$\varphi \left(\frac{t}{\sqrt{n}} \right) = 1 + 0 - \frac{t^2}{2n} + o(n^{-1}).$$

$$\begin{aligned} \text{Damit ist} \quad \lim_{n \rightarrow \infty} \varphi^n \left(\frac{t}{\sqrt{n}} \right) &= \lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n} \right)^n \\ &= \exp \left(-\frac{t^2}{2} \right), \text{ da } \forall x \in \mathbb{R} : \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n = \exp(x). \end{aligned}$$

Der Beweis unter (iii), der die Aussage unter (ii) impliziert, wird ähnlich geführt und findet sich in Feller (1971), Theorem 1 in Abschnitt XV.6. ■

Zur Gewinnung von präzisen Aussagen in der Statistik ist es überdies nützlich, dass unter der Annahme der Existenz dritter Momente auch die (asymptotische) Größenordnung der Differenz der Verteilungsfunktion der standardisierten Summenstatistik und Φ (der Verteilungsfunktion von $\mathcal{N}(0, 1)$) angegeben werden kann.

Satz 10.15 (Satz von Berry und Esséen)

Unter den Voraussetzungen von Satz 10.13 sei F_n die Verteilungsfunktion von $S_n / \sqrt{\text{Var}(S_n)}$, $n \in \mathbb{N}$. Dann gilt

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6}{(\text{Var}(S_n))^{\frac{3}{2}}} \sum_{j=1}^n \mathbb{E}[|X_j|^3].$$

Sind $(X_n)_{n \geq 1}$ stochastisch unabhängig und identisch verteilt, so ergibt sich

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6}{\sqrt{n} (\text{Var}(X_1))^{\frac{3}{2}}} \mathbb{E}[|X_1|^3] \sim \frac{1}{\sqrt{n}}.$$

Beweis: Satz 4.2.10 in Gaenssler and Stute (1977). ■

Zum Abschluss dieses Kapitels nun noch der sogenannte "Hauptsatz der Statistik".

Satz 10.16 (Glivenko-Cantelli)

Sei $((X_{n1}, \dots, X_{nn}))_{n \in \mathbb{N}}$ ein Dreiecksschema von zeilenweise stochastisch unabhängigen Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Für jedes $n \in \mathbb{N}$ seien also X_{n1}, \dots, X_{nn} stochastisch unabhängig mit zugehörigen Verteilungsfunktionen F_{n1}, \dots, F_{nn} . Bezeichne $\bar{F}_n = n^{-1} \sum_{i=1}^n F_{ni}$, $n \in \mathbb{N}$.

Für jedes $n \in \mathbb{N}$ sei $G_n : \mathbb{R} \rightarrow [0, 1]$, definiert durch $G_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}_{(-\infty, t]}(X_{ni})$ für $t \in \mathbb{R}$, die sogenannte empirische Verteilungsfunktion von $(X_{nj})_{j=1, \dots, n}$.

Dann gilt:

$$\sup_{t \in \mathbb{R}} |G_n(t) - \bar{F}_n(t)| \xrightarrow{\mathbb{P}\text{-f.s.}} 0 \text{ für } n \rightarrow \infty.$$

Ist insbesondere $(X_n)_{n \geq 1}$ eine Folge von stochastisch unabhängigen und identisch verteilten Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Verteilungsfunktion F von X_1 , so gilt:

$$\sup_{t \in \mathbb{R}} |G_n(t) - F(t)| \xrightarrow{\mathbb{P}\text{-f.s.}} 0 \text{ für } n \rightarrow \infty.$$

Beweis: Theorem 3.2.1 in Shorack and Wellner (1986). ■

Bemerkung 10.17

Für jedes fixe $t \in \mathbb{R}$ folgt die (punktweise) \mathbb{P} -fast sichere Konvergenz bereits aus dem starken Gesetz der großen Zahlen, falls die $(X_{ni})_{i=1, \dots, n}$ stochastisch unabhängig und identisch verteilt sind. Der allgemeine Fall wird bewiesen unter Anwendung des Prinzips der Quantilstransformation und des Lemmas von Borel-Cantelli (Satz 10.11).

Tabellenverzeichnis

2.1	Tabelle zum Geburtstagsparadoxon	18
-----	--	----

Abbildungsverzeichnis

1.1	Berechnung von Wahrscheinlichkeiten mit Dichtefunktionen	13
8.1	Illustration der Menge aus Korollar 8.7.	64
10.1	Zusammenhang von Konvergenzarten	93

Literaturverzeichnis

- Ash, R. B. (1972). *Measure, integration, and functional analysis*. New York-London: Academic Press, Inc.
- Bauer, H. (1991). *Probability theory. (Wahrscheinlichkeitstheorie.) 4., völlig überarb. u. neugestaltete Aufl. des Werkes: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. Berlin etc.: Walter de Gruyter.
- Bickel, P. J. and K. A. Doksum (1977). *Mathematical statistics. Basic ideas and selected topics*. Holden-Day Series in Probability and Statistics. San Francisco etc.: Holden-Day, Inc.
- Breiman, L. (1992). *Probability*. Philadelphia, PA: SIAM.
- Chung, K. L. (2000). *A course in probability theory. 3rd ed.* Orlando, FL: Academic Press.
- Cramér, H. (1946). *Mathematical methods of statistics*. Princeton Mathematical series. Princeton N. J.: Princeton University Press.
- Etemadi, N. (1981). An elementary proof of the strong law of large numbers. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 55, 119–122.
- Feller, W. (1971). *An introduction to probability theory and its applications. Vol II* (2nd ed.). Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley and Sons, Inc.
- Feller, W. (1971). *An introduction to probability theory and its applications. Vol II. 2nd ed.* Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley and Sons, Inc.
- Forster, O. (2012). *Analysis 3: Maß- und Integrationstheorie, Integralsätze im \mathbb{R}^n und Anwendungen*. (7th revised ed.). Wiesbaden: Springer Spektrum.
- Gaenssler, P. and W. Stute (1977). *Wahrscheinlichkeitstheorie*. Hochschultext. Berlin-Heidelberg-New York: Springer-Verlag.
- Georgii, H.-O. (2007). *Stochastics. Introduction to probability theory and statistics. (Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik.) 3rd ed.* de Gruyter Lehrbuch. Berlin: de Gruyter.

- Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika* 38, 481–482.
- Jacod, J. and P. Protter (2000). *Probability essentials*. Berlin: Springer.
- Klenke, A. (2008). *Probability theory. (Wahrscheinlichkeitstheorie.) 2nd revised ed.* Berlin: Springer.
- Lehmann, E. L. (1999). *Elements of large-sample theory*. New York, NY: Springer.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley & Sons.
- Shorack, G. R. and J. A. Wellner (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics. New York, NY: Wiley.