Stochastik für Mathematiker Teil 2: Wahrscheinlichkeitstheorie

Sommersemester 2017

Prof. Dr. Hans-Jörg Starkloff

Stand: 12. Juli 2017

0 Organisatorisches

Informationen

- Vorlesung:

Dienstag	7:30-9:00	PRÜ-1103
Mittwoch ungerade Wochen	14:00-15:30	PRÜ-1103

- Lesender: Prof. Dr. Hans-Jörg Starkloff
- Kontaktdaten: *http://www.mathe.tu-freiberg.de/sto/mitarbeiter/hans-joerg-starkloff*
- Website zur Vorlesung: *http://www.mathe.tu-freiberg.de/wt*
- Übung: Mittwoch, 9:15-10:45, PRÜ-1103
- Übungsleiter: Dr. Udo Lorz
- Kontaktdaten: *http://www.mathe.tu-freiberg.de/dek/mitarbeiter/udo-lorz*
- Website zur Übung: *http://www.mathe.tu-freiberg.de/wt-ue*

Literatur

- Bauer, H.: *Wahrscheinlichkeitstheorie.* Verlag Walter de Gruyter, 5. Auflage, 2002. (E-Book TUBAF, 2011)
- Hesse, C.: *Wahrscheinlichkeitstheorie – Eine Einführung mit Beispielen und Anwendungen.* Vieweg-Teubner-Verlag, 2. Auflage, 2009. (1. Auflage TUBAF)
- Klenke, A.: *Wahrscheinlichkeitstheorie.* Springer-Verlag, 3. Auflage, 2013. (E-Book TUBAF)
- Brokate, M., Henze, N., Hettlich, F., Meister, A., Schranz-Kirlinger,G., Sonar,T.: *Grundwissen Mathematikstudium – Höhere Analysis, Numerik und Stochastik.* Springer-Verlag, 2016, Kapitel 19-24. (E-Book TUBAF)

1 Grundbegriffe

1.1 Ereignisse und Wahrscheinlichkeiten

Anmerkungen 1.1.

- Die Wahrscheinlichkeitstheorie liefert Methoden zur mathematischen Untersuchung von Versuchen, Beobachtungen, Situationen, etc., bei denen man die Ergebnisse nicht genau (vorher-)berechnen kann, aber eine Quantifizierung der vorhandenen Unsicherheit möglich und nützlich ist.
- Eine Möglichkeit dieser Unsicherheitsquantifizierung besteht in der Nutzung stochastischer Modelle (oft auch in Verbindung mit statistischen Modellen). Dabei werden Wahrscheinlichkeiten als Maßzahlen für die Chancen des Eintretens von bestimmten zufälligen Ereignissen und darauf aufbauende Konzepte genutzt.
- Der Wahrscheinlichkeitsbegriff ist dabei aus dem Begriff der relativen Häufigkeit für das Eintreten von zufälligen Ereignissen bei unabhängigen und gleichartigen Wiederholungen von Zufallsversuchen entstanden.

Anmerkungen 1.1 ((Fortsetzung))**.**

- Die Einbettung der Wahrscheinlichkeitstheorie in die Mathematik erfolgt durch ihre Grundlegung mit Hilfe der Maß- und Integrationstheorie und durch vielfältige Verflechtungen mit anderen mathematischen Teilgebieten.
- Dies wurde erstmalig umfassend und konsequent 1933 in dem Buch von A. N. Kolmogorow (Kolmogoroff, Kolmogorov) „Grundbegriffe der Wahrscheinlichkeitsrechnung“ realisiert.
- Das mathematische Modell eines zufälligen Versuchs ist ein **Wahrscheinlichkeitsraum**.

Definitionen 1.2. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum.

- Die Elemente $\omega \in \Omega$ werden **Elementarereignisse** genannt.
- Die messbaren Mengen $A \in \mathcal{F}$ heißen **(zufällige) Ereignisse**.
- Ein Ereignis $A \in \mathcal{F}$ tritt ein, falls eines seiner Elementarereignisse $\omega \in A$ eintritt.
- $\mathrm{P}(A) \in [0, 1]$ ist die **Wahrscheinlichkeit** für das Eintreten des Ereignisses $A \in \mathcal{F}$.
- Das Ereignis $\varnothing \in \mathcal{F}$ wird **unmögliches Ereignis** genannt. Es gilt $\mathrm{P}(\varnothing) = 0$.
- Das Ereignis $\Omega \in \mathcal{F}$ heißt **sicheres Ereignis**. Es gilt $\mathrm{P}(\Omega) = 1$.

Beispiel 1.3 ((Einmaliges Würfeln))**.**

$$\Omega = \{1, 2, 3, 4, 5, 6\}\,, \qquad \mathcal{F} = \mathcal{P}(\Omega)$$

Das Ereignis $A = \{2, 4, 6\}$ tritt ein, wenn eine gerade Zahl gewürfelt wird.

Beispiel 1.4 ((Außentemperatur in °C))**.**

$$\Omega = [-100, 100], \qquad \mathcal{F} = \mathcal{B} \cap \Omega$$

Das Ereignis $A = (16, 22)$ tritt ein, wenn eine (angenehme) Temperatur zwischen 16°C und 22°C herrscht.

Definitionen 1.5. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$.

- Das Ereignis $A^c \in \mathcal{F}$ wird das zum Ereignis $A \in \mathcal{F}$ **komplementäre** (oder auch **entgegengesetzte**) **Ereignis** genannt. Es tritt genau dann ein, wenn A nicht eintritt.
- Das Ereignis $A \cup B$ tritt ein, wenn das Ereignis A oder das Ereignis B oder beide Ereignisse A und B eintreten. Man sagt, das Ereignis A **oder** B tritt ein.
- Das Ereignis $A \cap B$ tritt ein, wenn sowohl das Ereignis A als auch das Ereignis B eintritt. Man sagt, das Ereignis A **und** B tritt ein.
- Es gelte $A \cap B = \varnothing$. Dann heißen die Ereignisse A und B **unvereinbar**, denn sie können nicht zugleich eintreten.
- Es gelte $A \subseteq B$. Dann sagt man **aus** A **folgt** B oder **das Ereignis** A **zieht das Ereignis** B **nach sich**.

Beispiele 1.6 ((Einmaliges Würfeln, vgl. Beispiel 1.3))**.**

- Das komplementäre Ereignis $A^c = \{1, 3, 5\}$ besteht darin, dass eine ungerade Zahl gewürfelt wird.
- Sei $B = \{1, 2\}$. Dann folgt

  $$A \cup B = \{1, 2, 4, 6\} \quad \text{und} \quad A \cap B = \{2\}\,.$$

  Das Ereignis A oder B tritt ein, wenn eine der Zahlen $1, 2, 4$ oder 6 gewürfelt wird und das Ereignis A und B tritt ein, wenn eine Zwei gewürfelt wird.
- Sei $C = \{4\}$. Dann gilt

  $$\{4\} = C \subseteq A = \{2, 4, 6\}\,.$$

  Aus dem Ereignis, dass eine Vier gewürfelt wird, folgt, dass eine gerade Zahl gewürfelt wird.

Beispiele 1.7 ((Außentemperatur in °C, vgl. Beispiel 1.4))**.**
Sei $B = [-100, 0)$, also das Ereignis Frost.

- Dann ist $A \cup B = (16, 22) \cup [-100, 0)$ das Ereignis, dass eine angenehme Temperatur oder Frost herrscht.
- Da $A \cap B = \varnothing$, sind die Ereignisse angenehme Temperatur und Frost miteinander unvereinbar.

Anmerkungen 1.8.

- Jede Zuordnung von Wahrscheinlichkeiten zu den zufälligen Ereignissen, die den Axiomen der Maßtheorie (siehe unten) genügt, ist mathematisch korrekt, aber nicht immer ein gutes mathematisches Modell für reale Versuche oder Situationen.
- Erinnerung an die Axiome:

$$\mathrm{P}(\varnothing) = 0\,; \quad \mathrm{P}(\Omega) = 1\,;$$
$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i) \text{ für } A_i \in \mathcal{F}, A_i \cap A_j = \varnothing \ (i \neq j, i, j \in \mathbb{N})\,.$$

- Einige Folgerungen:

$$\mathrm{P}(A^c) = 1 - \mathrm{P}(A)\ ;\ A \subseteq B \ \Rightarrow\ \mathrm{P}(A) \leqslant \mathrm{P}(B)\,;\ 0 \leqslant \mathrm{P}(A) \leqslant 1\,;$$
$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)\ .$$

Anmerkungen 1.9.

- Nicht immer, z.B. bei komplexen stochastischen Modellen, wird ein Wahrscheinlichkeitsraum explizit konstruiert oder angegeben.
- Ausgangspunkt für theoretische Überlegungen kann auch die Menge der zufälligen Ereignisse zu einem Zufallsversuch sein. Die grundlegende wesentliche Eigenschaft von zufälligen Ereignissen besteht darin, dass man nach der Durchführung des Zufallsversuches eindeutig entscheiden kann, ob das zufällige Ereignis eingetreten ist oder nicht. Ausgestattet mit den „besonderen“ zufälligen Ereignissen $(\varnothing, \Omega)$ und den Operationen (Komplementbildung, Vereinigung und Durchschnitt) besitzt die Menge der zufälligen Ereignisse zu einem Zufallsversuch die Struktur einer Booleschen Algebra.
- Nach einem Satz von Stone kann man zu jeder Ereignisalgebra eine isomorphe Mengenalgebra finden, folglich kann man zufällige Ereignisse immer durch Mengen repräsentieren, so wie das in unseren Axiomen geschehen ist.

Beispiel 1.10 ((Einmaliges Würfeln mit Gewinnbeträgen)).

$$\Omega = \{1, 2, 3, 4, 5, 6\}\,, \qquad \mathcal{F} = \mathcal{P}(\Omega)$$

Augenzahl	1	2	3	4	5	6
Gewinn Spieler 1	1		3			6
Gewinn Spieler 2		2		4	4	

1.2 Klassische Wahrscheinlichkeitsräume

Definition 1.11. Ein Wahrscheinlicheitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ heißt

- **klassischer Wahrscheinlichkeitsraum** bzw.
- **Laplacescher Wahrscheinlichkeitsraum**,

falls

(1) $0 < \operatorname{card}(\Omega) < \infty$,

(2) $\mathcal{F} = \mathcal{P}(\Omega)$ und

(3) für alle $\omega \in \Omega$ gilt $\mathrm{P}(\{\omega\}) = \frac{1}{\operatorname{card}(\Omega)}$.

Aufgabe 1.12. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein klassischer Wahrscheinlichkeitsraum. Zeigen Sie, dass dann für jedes Ereignis $A \subseteq \Omega$ gilt

$$\mathrm{P}(A) = \frac{\operatorname{card}(A)}{\operatorname{card}(\Omega)} = \frac{\text{„Anzahl der für das Ereignis } A \text{ günstigen Fälle“}}{\text{„Anzahl der möglichen Fälle“}}\,.$$

Anmerkung 1.13. Die Anzahl der für das Ereignis A günstigen Fälle kann in der Regel durch kombinatorische Überlegungen ermittelt werden.

1.3 Bedingte Wahrscheinlichkeiten

Definition 1.14. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $B \in \mathcal{F}$ ein Ereignis mit $\mathrm{P}(B) > 0$. Dann wird

$$\mathrm{P}(A\,|B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \quad \text{für} \quad A \in \mathcal{F}$$

die **bedingte Wahrscheinlichkeit** des zufälligen Ereignisses A unter der Bedingung B bzw. kurz von A unter B bzw. von A gegeben B genannt.

Satz 1.15. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und B ein Ereignis mit $\mathrm{P}(B) > 0$.

(1) Dann ist $\mathrm{P}(\,\cdot\,|B)$ ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(\Omega, \mathcal{F})$.

(2) Insbesondere gilt die Additionsformel

$$\mathrm{P}(A_1 \cup A_2|B) = \mathrm{P}(A_1|B) + \mathrm{P}(A_2|B) - \mathrm{P}(A_1 \cap A_2|B)\,.$$

(3) Seien $A_1, \ldots, A_n \in \mathcal{F}$, $n = 2, 3, \ldots$, mit $\mathrm{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt die Multiplikationsformel

$$\begin{aligned}\mathrm{P}(A_1 \cap \ldots \cap A_n) &= \\ &= \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\mathrm{P}(A_3|A_1 \cap A_2) \ldots \mathrm{P}(A_n|A_1 \cap \ldots \cap A_{n-1})\,.\end{aligned}$$

Beispiel 1.16 ((Multiplikationsformel))**.** In einer Urne befinden sich zehn Kugeln, davon sieben rote und drei schwarze. Es werden vier Kugeln ohne Zurücklegen entnommen. Wie groß ist die Wahrscheinlichkeit für das Ereignis R, dass alle vier gezogenen Kugeln rot sind? Zur Lösung dieser Aufgabe definieren wir die Ereignisse

- R ... alle vier gezogenen Kugeln sind rot und
- R_i ... die i-te gezogene Kugel ist rot, $i = 1, 2, 3, 4$,

und erhalten

$$\begin{aligned}\mathrm{P}(R) &= \mathrm{P}(R_1 \cap R_2 \cap R_3 \cap R_4) \\ &= \mathrm{P}(R_1) \cdot \mathrm{P}(R_2|R_1) \cdot \mathrm{P}(R_3|R_1 \cap R_2) \cdot \mathrm{P}(R_4|R_1 \cap R_2 \cap R_3) \\ &= \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} \cdot \frac{4}{7} = \frac{1}{6} \mathrel{\widehat{=}} 16.67\%\,.\end{aligned}$$

Satz 1.17 ((Formel der totalen Wahrscheinlichkeit, Zerlegungssatz))**.** Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum, $\varnothing \neq I \subseteq \mathbb{N}$ eine nichtleere, höchstens abzählbare Indexmenge und $(B_n)_{n \in I} \subseteq \mathcal{F}$ eine höchstens abzählbare Zerlegung von Ω, d.h. $\bigcup_{n \in I} B_n = \Omega$ und $B_i \cap B_j = \varnothing$ für alle $i \neq j$ mit $\mathrm{P}(B_n) > 0$ für alle $n \in I$. Dann gilt für $A \in \mathcal{F}$

$$\mathrm{P}(A) = \sum_{n \in I} \mathrm{P}(A\,|B_n)\ \mathrm{P}(B_n)\,.$$

Beispiel 1.18. Drei verschiedene Firmen liefern gleichartige Teile an einen Automobilkonzern. Der Anteil der drei Firmen beträgt $60\%, 30\%$ bzw. 10% am gesamten Lieferumfang. Die Ausschussquoten der drei Zulieferer betragen $1\%, 2\%$ bzw. 3%. Wie groß ist die Ausschussquote insgesamt?
Zur Lösung dieser Aufgabe definieren wir folgende Ereignisse

- A ... Teil ist Ausschuss
- F_n ... Teil wurde von der n-ten Firma geliefert, $n = 1, 2, 3$.

$$\mathrm{P}(A) = \sum_{n=1}^{3} \mathrm{P}(A\,|F_n)\ \mathrm{P}(F_n) =$$

$$= 0.01 \cdot 0.6 + 0.02 \cdot 0.3 + 0.03 \cdot 0.1 = 0.015 \mathrel{\widehat{=}} 1.5\%$$

Aufgabe 1.19. Konstruieren Sie für das Beispiel 1.18 einen geeigneten Wahrscheinlichkeitsraum. Geben Sie die Ereignisse A und $F_n\,, n = 1, 2, 3\,,$ explizit an.

Satz 1.20 ((Bayessche Formel, Satz von Bayes))**.** Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum, $\varnothing \neq I \subseteq \mathbb{N}$ eine nichtleere, höchstens abzählbare Indexmenge und $(B_n)_{n\in I} \subseteq \mathcal{F}$ eine höchstens abzählbare Zerlegung von Ω, d.h. $\bigcup\limits_{n\in I} B_n = \Omega$ und $B_i \cap B_j = \varnothing$ für alle $i \neq j$, mit $\mathrm{P}(B_n) > 0$ für alle $n \in I$. Dann gilt für $A \in \mathcal{F}$ mit $\mathrm{P}(A) > 0$

$$\mathrm{P}(B_n|A) = \frac{\mathrm{P}(A\,|B_n)\ \mathrm{P}(B_n)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A\,|B_n)\ \mathrm{P}(B_n)}{\sum\limits_{i\in I} \mathrm{P}(A\,|B_i)\ \mathrm{P}(B_i)} \quad \text{für} \quad n \in I\,.$$

Anmerkung 1.21. Man nennt in diesem Zusammenhang auch

$\mathrm{P}\,(B_n)$ **a-priori-Wahrscheinlichkeiten** und
$\mathrm{P}\,(B_n|A)$ **a-posteriori-Wahrscheinlichkeiten**.

Beispiel 1.22 ((vgl. Beispiel 1.18))**.** Wie groß ist die bedingte Wahrscheinlichkeit dafür, dass ein Teil von der n-ten Firma geliefert wurde unter der Bedingung, dass es Ausschuss ist?

$$\mathrm{P}(F_n|A) = \frac{\mathrm{P}(A\,|F_n)\ \mathrm{P}(F_n)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A\,|F_n)\ \mathrm{P}(F_n)}{\sum\limits_{i=1}^{3} P(A\,|F_i)\ \mathrm{P}(F_i)}\,, \quad n = 1, 2, 3$$

- $\mathrm{P}(F_1|A) = \dfrac{0.01 \cdot 0.6}{0.015} = \dfrac{2}{5} \mathrel{\widehat{=}} 40\%$
- $\mathrm{P}(F_2|A) = \dfrac{0.02 \cdot 0.3}{0.015} = \dfrac{2}{5} \mathrel{\widehat{=}} 40\%$
- $\mathrm{P}(F_3|A) = \dfrac{0.03 \cdot 0.1}{0.015} = \dfrac{1}{5} \mathrel{\widehat{=}} 20\%$

1.4 Unabhängigkeit

Definition 1.23. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathcal{F}$ heißen **unabhängig** bezüglich des Wahrscheinlichkeitsmaßes P (oder **stochastisch unabhängig**), falls gilt

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

Behauptung 1.24. Für unabhängige Ereignisse A und B mit $\mathrm{P}(A) > 0$ bzw. $\mathrm{P}(B) > 0$ gilt

$$\mathrm{P}(B|A) = \mathrm{P}(B) \quad \text{bzw.} \quad \mathrm{P}(A|B) = \mathrm{P}(A)\,.$$

Satz 1.25. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$ zwei Ereignisse. Dann sind die folgenden Aussagen äquivalent

(1) A und B sind unabhängig,

(2) A und B^c sind unabhängig,

(3) A^c und B sind unabhängig,

(4) A^c und B^c sind unabhängig.

Definitionen 1.26. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $I \neq \varnothing$ eine beliebige Indexmenge.

- Eine **Familie** $(A_i)_{i\in I} \subseteq \mathcal{F}$ **von Ereignissen** heißt **vollständig unabhängig**, falls für jede nichtleere, endliche Teilmenge $I_n \subseteq I$ gilt

$$\mathrm{P}\left(\bigcap_{i\in I_n} A_i\right) = \prod_{i\in I_n} \mathrm{P}(A_i)\,.$$

- Ist $I = \{1, \ldots, n\}$ so werden die **Ereignisse** $A_1, \ldots, A_n \in \mathcal{F}$ **vollständig unabhängig** genannt.

- Für $I = \mathbb{N}$ spricht man von einer **vollständig unabhängigen Folge** $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{F}$ **von Ereignissen**.

- Gilt lediglich $\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i) \cdot \mathrm{P}(A_j)$ für alle $i, j \in I, i \neq j$, dann heißt die **Familie** $(A_i)_{i\in I} \subseteq \mathcal{F}$ **von Ereignissen paarweise unabhängig**.

Beispiel 1.27. Wir betrachten den zweimaligen Wurf mit einer fairen Münze und dabei die Ereignisse

- $A_1 \ldots$ beim ersten Wurf wird Zahl (Z) geworfen
- $A_2 \ldots$ beim zweiten Wurf wird Wappen (W) geworfen
- $A_3 \ldots$ beide Wurfergebnisse sind verschieden

Als Grundmenge wählen wir $\Omega = \{(W, W), (W, Z), (Z, W), (Z, Z)\}$. Jedes der Elementarereignisse hat die Wahrscheinlichkeit $1/4$. Folglich ist $\mathrm{P}(A_1) = \mathrm{P}(\{(Z, Z), (Z, W)\}) = 1/2$, $\mathrm{P}(A_2) = \mathrm{P}(\{(W, W), (Z, W)\}) = 1/2$ und $\mathrm{P}(A_3) = \mathrm{P}(\{(W, Z), (Z, W)\}) = 1/2$. Wie man leicht sieht, ist $\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i) \cdot \mathrm{P}(A_j) = 1/4$ für alle $i, j \in \{1, 2, 3\}, i \neq j$. Jedoch ist einerseits $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(\{(Z, W)\}) = 1/4$, aber andererseits $\mathrm{P}(A_1)\,\mathrm{P}(A_2)\,\mathrm{P}(A_3) = 1/8$. Damit sind die Ereignisse A_1, A_2, A_3 zwar paarweise unabhängig, jedoch nicht vollständig unabhängig.

Definition 1.28. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum, $I \neq \varnothing$ eine beliebige Indexmenge und für jedes $i \in I$ sei $\mathcal{M}_i \subseteq \mathcal{F}$ ein Mengensystem. Dann heißt die **Familie** $(\mathcal{M}_i)_{i \in I}$ **von Mengensystemen unabhängig**, falls für jede mögliche Auswahl von Ereignissen $A_i \in \mathcal{M}_i$ für alle $i \in I$ die Familie $(A_i)_{i \in I}$ von Ereignissen vollständig unabhängig ist, d.h. falls für jede nichtleere, endliche Teilmenge $\{i_1, ..., i_n\} \subseteq I$ und jedes Ereignis $A_{i_k} \in \mathcal{M}_{i_k}$ mit $k = 1, ..., n$ gilt

$$\mathrm{P}\left(A_{i_1} \cap ... \cap A_{i_n}\right) = \mathrm{P}(A_{i_1}) \cdot ... \cdot \mathrm{P}(A_{i_n}) .$$

Anmerkungen 1.29.

(1) Die Familie $(\mathcal{M}_i)_{i \in I}$ ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist.

(2) Die Unabhängigkeit bleibt bestehen, wenn man die Mengensysteme $\mathcal{M}_i$ verkleinert.

1.5 Zufallsvariable und Verteilung

Definition 1.30. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{F}')$ ein messbarer Raum. Dann wird die $(\mathcal{F}, \mathcal{F}')$-messbare Abbildung $X : \Omega \to \Omega'$ **Zufallsvariable** bzw. **zufälliges Element** aus $(\Omega', \mathcal{F}')$ genannt. Man spricht von einer Zufallsvariable X von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ in einen messbaren Raum $(\Omega', \mathcal{F}')$ und schreibt dafür

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X} (\Omega', \mathcal{F}') \quad \text{oder} \quad X : (\Omega, \mathcal{F}, \mathrm{P}) \to (\Omega', \mathcal{F}').$$

Anmerkung 1.31. Zufallsvariable dienen als mathematische Objekte zur Beschreibung bestimmter Eigenschaften der Ergebnisse eines Experimentes mit zufälligem Ausgang (Zufallsexperiment). Eine Zufallsvariable weist jedem möglichen Ergebnis eines Zufallsexperimentes einen gewissen Wert $\omega' \in \Omega'$ zu. Der Wert $\omega' = X(\omega)$ der Zufallsvariable ist zum Beispiel das zahlenmäßige zufällige Versuchsergebnis oder eine zahlenmäßige Charakteristik des zufälligen Versuchsergebnisses.

Definition 1.32. Sei X eine Zufallsvariable von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ in einen messbaren Raum $(\Omega', \mathcal{F}')$. Dann wird das durch X übertragene Wahrscheinlichkeitsmaß (Bildmaß)

$$\mathrm{P}_X = \mathrm{P} \circ X^{-1}$$

Verteilungsgesetz bzw. kurz **Verteilung** von X genannt. Das Verteilungsgesetz P_X ist ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(\Omega', \mathcal{F}')$. Für $B \in \mathcal{F}'$ gilt

$$\mathrm{P}_X(B) = \mathrm{P}(X^{-1}(B)) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \in B\}) =: \mathrm{P}(X \in B).$$

Beispiel 1.33 (und Definition)**.** Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und

$$X : \Omega \to \Omega \quad \text{mit} \quad X(\omega) = \omega$$

die identische Abbildung. Es ist also

$$(\Omega', \mathcal{F}') = (\Omega, \mathcal{F}) \quad \text{und} \quad \mathrm{P}_X = \mathrm{P}\,.$$

In diesem Fall nennt man die Zufallsvariable X **unmittelbar gegeben**.

Behauptung 1.34. Für jedes Wahrscheinlichkeitsmaß P auf einem messbaren Raum $(\Omega, \mathcal{F})$ gibt es eine Zufallsvariable mit Werten in $(\Omega, \mathcal{F})$, für die P das Verteilungsgesetz ist.

Anmerkung 1.35. Oft wird für eine Zufallsvariable X von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ in einen messbaren Raum $(\Omega', \mathcal{F}')$ nur das Verteilungsgesetz P_X angegeben ohne den ursprünglichen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und insbesondere das Wahrscheinlichkeitsmaß P näher zu beschreiben. Man spricht dann von einem **hypothetischen Wahrscheinlichkeitsraum** $(\Omega, \mathcal{F}, \mathrm{P})$. Die Existenz eines solchen Wahrscheinlichkeitsraumes ist durch eine unmittelbar gegebene Zufallsvariable stets gesichert.

1.6 Zufallsgrößen

Definition 1.36. Eine Zufallsvariable X von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ in den messbaren Raum $(\mathbb{R}, \mathcal{B})$ wird **Zufallsgröße** oder **reelle Zufallsvariable**, **reellwertige Zufallsvariable** bzw. **zufällige reelle Zahl** genannt. Das Verteilungsgesetz P_X von X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ und somit $(\mathbb{R}, \mathcal{B}, \mathrm{P}_X)$ ein Wahrscheinlichkeitsraum.

Definition 1.37. Sei X eine Zufallsgröße. Dann heißt die Funktion

$$F_X \colon \mathbb{R} \to [0,1]$$

mit

$$F_X(x) := \mathrm{P}_X((-\infty, x)) = \mathrm{P}(\{\omega \in \Omega \colon X(\omega) < x\}) \quad \text{für} \quad x \in \mathbb{R}$$

Verteilungsfunktion der Zufallsgröße X bzw. des Verteilungsgesetzes P_X.

Anmerkungen 1.38.

(1) An Stelle von $\mathrm{P}(\{\omega \in \Omega \colon X(\omega) < x\})$ ist die Kurzschreibweise $\mathrm{P}(X < x)$ gebräuchlich, also $F_X(x) = \mathrm{P}(X < x)$, analog auch für Wahrscheinlichkeiten für andere zufällige Ereignisse im Zusammenhang mit einer Zufallsgröße X.

(2) Mit Hilfe der Verteilungsfunktion lässt sich die Wahrscheinlichkeit typischer Ereignisse einfach berechnen

$$\begin{aligned} \mathrm{P}(X < b) &= F_X(b) \quad \text{für} \quad b \in \mathbb{R}, \\ \mathrm{P}(X \geqslant a) &= 1 - F_X(a) \quad \text{für} \quad a \in \mathbb{R}, \\ \mathrm{P}(a \leqslant X < b) &= F_X(b) - F_X(a) \quad \text{für} \quad a, b \in \mathbb{R}, a < b. \end{aligned}$$

(3) Neben der oben gegebenen Definition der Verteilungsfunktion einer Zufallsgröße ist eine weitere Variante (mit geringfügig anderen Eigenschaften) verbreitet:

$$\widetilde{F}_X(x) = \mathrm{P}\left(X \leqslant x\right), \qquad x \in \mathbb{R}\,.$$

Satz 1.39. Die Verteilungsfunktion F_X einer Zufallsgröße X hat folgende Eigenschaften

(V1) F_X ist monoton wachsend,

(V2) F_X ist linksseitig stetig,

(V3) $\lim\limits_{x\to-\infty} F_X(x) = 0$ und

(V4) $\lim\limits_{x\to+\infty} F_X(x) = 1$.

Satz 1.40. Sei X eine Zufallsgröße. Dann ist eine messbare Funktion $f\colon \mathbb{R} \to \mathbb{R}$ genau dann P_X-integrierbar, wenn $f \circ X$ P-integrierbar ist. In diesem Fall gilt

$$\int_{\mathbb{R}} f(x)\ \mathrm{P}_X(\mathrm{d}x) = \int_{\Omega} f(X(\omega))\ \mathrm{P}(\mathrm{d}\omega)\,.$$

Definition 1.41. Eine Zufallsgröße X heißt **diskret**, falls ihr Wertebereich $W_X \subseteq \mathbb{R}$ höchstens abzählbar ist.

Anmerkung 1.41. Ist eine Zufallsgröße X nur durch ihr Verteilungsgesetz (z.B. ihre Verteilungsfunktion) gegeben, nennt man sie oft diskret, wenn eine höchstens abzählbare Menge W_X mit der Eigenschaft

$$\mathrm{P}\,(X \in W_X) = 1$$

existiert.

Satz 1.42 (und Definition)**.** Das Verteilungsgesetz P_X einer diskreten Zufallsgröße X ist eindeutig durch die **möglichen Werte** $\{x_k\,,\colon\ k \in I \subseteq \mathbb{N}\} = W_X$ und die **(zugehörigen) Einzelwahrscheinlichkeiten**

$$p_k := \mathrm{P}(\{\omega \in \Omega\colon X(\omega) = x_k\}) =: \mathrm{P}(X = x_k)\,, \quad k \in I\,,$$

bestimmt. Es gilt

$$\sum_{k\in I} p_k = 1 \quad \text{und} \quad \mathrm{P}_X = \sum_{k\in I} p_k\,\delta_{x_k}\,.$$

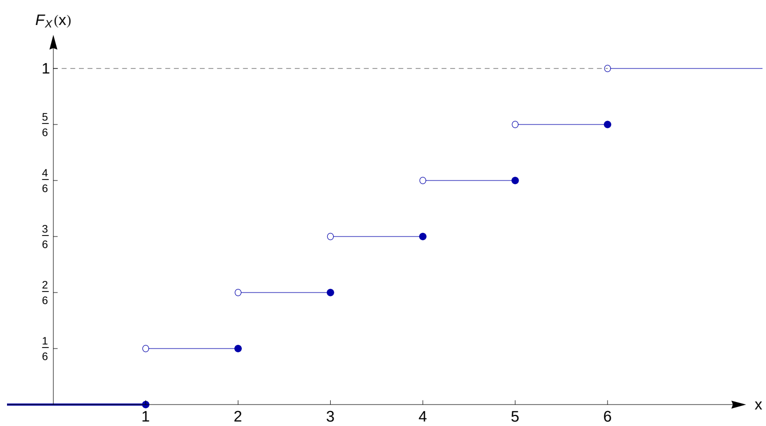
Die Verteilungsfunktion F_X ist eine stückweise konstante Funktion mit den Sprungstellen x_k und den Sprunghöhen $p_k, k \in I$. Sie hat die Darstellung

$$F_X(x) = \sum_{k\in I} p_k\, I_{(x_k,\infty)}(x)\,, \quad x \in \mathbb{R}\,.$$

Beispiel 1.43 ((Einmaliges Würfeln))**.** Die diskrete Zufallsgröße X beschreibe als unmittelbar gegebene Zufallsgröße das Ergebnis beim einmaligen Würfeln mit einem fairen Würfel.

$$\begin{aligned} W_X &= \{1,2,3,4,5,6\}\,, \\ x_k &= k\,, \quad p_k = \frac{1}{6}\,, \quad k = 1,\dots,6\,, \\ \mathrm{P}_X &= \frac{1}{6}\sum_{k=1}^{6}\delta_k\,, \\ F_X(x) &= \frac{1}{6}\sum_{k=1}^{6} I_{(k,\infty)}(x)\,, \quad x \in \mathbb{R}\,. \end{aligned}$$

Beispiel 1.43 ((Fortsetzung))**.** Grafische Darstellung der Verteilungsfunktion $F_X(x)$



Satz 1.44. Sei X eine diskrete Zufallsgröße. Dann ist eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ genau dann P_X-integrierbar, wenn $f \circ X$ P-integrierbar ist. In diesem Fall gilt

$$\int_\Omega f(X(\omega))\ \mathrm{P}(\mathrm{d}\omega) = \int_{\mathbb{R}} f(x)\ \mathrm{P}_X(\mathrm{d}x) = \sum_{k \in I} f(x_k)\, p_k\,.$$

Definition 1.45. Das Verteilungsgesetz P_X einer Zufallsgröße X sei absolut stetig bezüglich des Lebesgue-Borel-Maßes $\lambda\!\!\lambda$ auf dem messbaren Raum $(\mathbb{R}, \mathcal{B})$ $(\mathrm{P}_X \ll \lambda\!\!\lambda)$. Dann wird die Zufallgröße X **(absolut) stetig** genannt. Jede nichtnegative, messbare Funktion

$$f_X\colon \mathbb{R} \to [0,\infty] \quad \text{mit} \quad f_X(x) = \frac{\mathrm{d}\,\mathrm{P}_X}{\mathrm{d}\,\lambda\!\!\lambda}(x)\,,$$

die Radon-Nikodym-Ableitung von P_X bezüglich $\lambda\!\!\lambda$, wird **Verteilungsdichte** oder auch **Dichtefunktion** bzw. kurz **Dichte** der Zufallsgröße X bzw. des Verteilungsgesetzes P_X genannt.

Anmerkungen 1.46.

(1) Die Dichtefunktion f_X einer stetigen Zufallsgröße X ist nur λ-fast überall eindeutig bestimmt.

(2) Sei $f\colon \mathbb{R} \to [0,\infty]$ eine nichtnegative, messbare Funktion mit

$$\int_{\mathbb{R}} f(x)\,\lambda(\mathrm{d}x) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1\,.$$

Dann ist $f\,\lambda$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R},\mathcal{B})$ und für die unmittelbar gegebene Zufallsgröße X gilt

$$\mathrm{P}_X = f\,\lambda\,,$$

d.h. X ist eine stetige Zufallsgröße mit der Dichtefunktion $f_X = f$.

Satz 1.47. Für eine stetige Zufallsgröße X mit der Dichtefunktion f_X gilt

(1)

$$\mathrm{P}_X(B) = \int_B f_X(x)\,\mathrm{d}x \quad \text{für} \quad B \in \mathcal{B}\,;$$

(2)

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \quad \text{für} \quad x \in \mathbb{R}\,;$$

(3) $\mathrm{P}(X = x) = 0$ für alle $x \in \mathbb{R}$;

(4) $f_X = F_X'$ λ-fast überall, falls F_X stetig differenzierbar ist.

Folgerung 1.48. Für eine stetige Zufallsgröße X mit der Dichtefunktion f_X gilt

$$\int_{\mathbb{R}} f_X(x)\,\lambda(dx) = \int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1\,,$$

$$\mathrm{P}(X < b) = \mathrm{P}(X \leqslant b) = F_X(b) = \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x \quad \text{für} \quad b \in \mathbb{R}\,,$$

$$\mathrm{P}(X \geqslant a) = \mathrm{P}(X > a) = 1 - F_X(a) = \int_{a}^{\infty} f_X(x)\,\mathrm{d}x \quad \text{für} \quad a \in \mathbb{R}\,,$$

$$\begin{aligned}\mathrm{P}(a \leqslant X < b) &= \mathrm{P}(a < X < b) = \mathrm{P}(a < X \leqslant b) = \mathrm{P}(a \leqslant X \leqslant b)\\ &= F_X(b) - F_X(a) = \int_a^b f_X(x)\,\mathrm{d}x \quad \text{für} \quad a,b \in \mathbb{R}\,, a < b\,.\end{aligned}$$

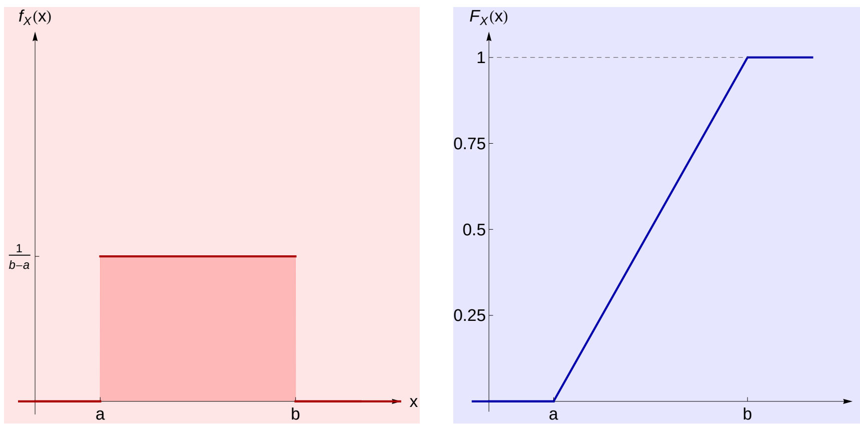
Beispiel 1.49. Sei X eine stetige Zufallsgröße mit der Dichtefunktion

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x) \quad \text{für} \quad -\infty < a < b < \infty\,.$$

Dann ist

$$F_X(x) = \begin{cases} 0 & \text{für} \quad x \leqslant a\,, \\ \frac{x-a}{b-a} & \text{für} \quad a < x \leqslant b\,, \\ 1 & \text{für} \quad x > b\,. \end{cases}$$

Beispiel 1.49 ((Fortsetzung))**.** Grafische Darstellung der Dichtefunktion $f_X(x)$ und der Verteilungsfunktion $F_X(x)$



Satz 1.50. Sei X eine stetige Zufallsgröße mit der Dichtefunktion f_X und $g\colon \mathbb{R} \to \mathbb{R}$ eine P_X-integrierbare Funktion. Dann ist $f_X \cdot g$ λ-integrierbar, und es gilt

$$\int_{\mathbb{R}} g(x)\ \mathrm{P}_X(\mathrm{d}x) = \int_{\mathbb{R}} g(x)\, f_X(x)\, \lambda(\mathrm{d}x) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, \mathrm{d}x.$$

1.7 Zufallsvektoren

Definition 1.51. Eine Zufallsvariable X von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ in den messbaren Raum $(\mathbb{R}^d, \mathcal{B}_d)$ mit $d \in \mathbb{N}$ wird **Zufallsvektor** bzw. **zufälliger, reeller Vektor** genannt. Das Verteilungsgesetz P_X von X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}_d)$ und somit $(\mathbb{R}^d, \mathcal{B}_d, \mathrm{P}_X)$ ein Wahrscheinlichkeitsraum.

Anmerkung 1.52 (und Definitionen)**.** Ein Zufallsvektor X kann auch als d-Tupel von Zufallsgrößen $X_i\,, i = 1, \ldots, d\,,$ aufgefasst werden

$$X = (X_1, \ldots, X_d)^T\,.$$

Für das Verteilungsgesetz P_X des Zufallsvektors X verwendet man auch die Schreibweisen

$$\mathrm{P}_X = \mathrm{P}_{X_1,\ldots,X_d} = \mathrm{P}_{(X_1,\ldots,X_d)}$$

und nennt $\mathrm{P}_{X_1,\ldots,X_d}$ die **gemeinsame Verteilung** der Zufallsgrößen $X_1,\ldots,X_d$. Die Verteilungsgesetze P_{X_k} der einzelnen Zufallsgrößen $X_k, k=1,\ldots,d$, werden **(eindimensionale) Randverteilungen** genannt.

Definition 1.53. Sei X ein Zufallsvektor. Dann heißt die Funktion

$$F_X\colon \mathbb{R}^d \to [0,1]$$

mit

$$\begin{aligned} F_X(\vec{x}) = F_{X_1,\ldots,X_d}(x_1,\ldots,x_d) &:= \mathrm{P}_X\left(\bigtimes_{i=1}^{d}(-\infty,x_i)\right) = \\ &= \mathrm{P}(\{\omega\in\Omega\colon X_1(\omega)<x_1,\ldots,X_d(\omega)<x_d\}) = \\ &= \mathrm{P}(X_1<x_1,\ldots,X_d<x_d) \quad \text{für} \quad \vec{x}=(x_1,\ldots,x_d)^T\in\mathbb{R}^d \end{aligned}$$

Verteilungsfunktion des Zufallsvektors X bzw. **gemeinsame Verteilungsfunktion** oder auch **Verbundverteilungsfunktion** der Zufallsgrößen $X_1,\ldots,X_d$.

Anmerkungen 1.54. Sei X ein Zufallsvektor.

(1) Dann gilt für die Randverteilung $\mathrm{P}_{X_k}, k=1,\ldots,d$,

$$\begin{aligned} \mathrm{P}_{X_k}(B) &= \mathrm{P}_{X_1,\ldots,X_d}(\mathbb{R}\times\ldots\times\mathbb{R}\times B\times\mathbb{R}\times\ldots\times\mathbb{R}) \\ &= \mathrm{P}_{X_1,\ldots,X_d}\left(\mathbb{R}^{k-1}\times B\times\mathbb{R}^{d-k}\right) \quad \text{für} \quad B\in\mathcal{B}. \end{aligned}$$

(2) Für die Verteilungsfunktion F_{X_k} der k-ten Randverteilung P_{X_k}, $k=1,\ldots,d$, gilt

$$\begin{aligned} F_{X_k}(x) &= \mathrm{P}_{X_1,\ldots,X_d}\left(\mathbb{R}^{k-1}\times(-\infty,x)\times\mathbb{R}^{d-k}\right) \\ &= \lim_{y\to\infty} F_{X_1,\ldots,X_d}(y,\ldots,y,x,y,\ldots,y) \quad \text{für} \quad x\in\mathbb{R}. \end{aligned}$$

Anmerkungen 1.54 ((Fortsetzung))**.**

(3) Sei $X=(X_1,X_2)^T$ ein zweidimensionaler Zufallsvektor und $-\infty<a_i<b_i<\infty, i=1,2$. Dann gilt

$$\begin{aligned} &\mathrm{P}(a_1\leqslant X_1<b_1, a_2\leqslant X_2<b_2) \\ =\ &\mathrm{P}_{X_1,X_2}([a_1,b_1)\times[a_2,b_2)) = \mathrm{P}_X([a_1,b_1)\times[a_2,b_2)) \\ =\ &\mathrm{P}_X((-\infty,b_1)\times(-\infty,b_2)) - \mathrm{P}_X((-\infty,a_1)\times(-\infty,b_2)) \\ &-\mathrm{P}_X((-\infty,b_1)\times(-\infty,a_2)) + \mathrm{P}_X((-\infty,a_1)\times(-\infty,a_2)) \\ =\ &F_X(b_1,b_2)-F_X(a_1,b_2)-F_X(b_1,a_2)+F_X(a_1,a_2). \end{aligned}$$

Satz 1.55. Sei X ein Zufallsvektor. Dann ist eine messbare Funktion $f\colon \mathbb{R}^d \to \mathbb{R}$ genau dann P_X-integrierbar, wenn $f \circ X$ P-integrierbar ist. In diesem Fall gilt

$$\int\limits_{\mathbb{R}^d} f(\vec{x})\ \mathrm{P}_X(\mathrm{d}\vec{x}) = \int\limits_{\Omega} f(X(\omega))\ \mathrm{P}(\mathrm{d}\omega)\,.$$

Definition 1.56. Das Verteilungsgesetz P_X des Zufallsvektors X sei absolut stetig bezüglich des Lebesgue-Borel-Maßes λ_d auf dem messbaren Raum $(\mathbb{R}^d, \mathcal{B}_d)$ $(\mathrm{P}_X \ll \lambda_d)$. Dann wird der Zufallsvektor X **(absolut) stetig** genannt. Jede nichtnegative, messbare Funktion

$$f_X\colon \mathbb{R}^d \to [0,\infty] \quad \text{mit} \quad f_X(\vec{x}) = \frac{\mathrm{d}\,\mathrm{P}_X}{\mathrm{d}\,\lambda_d}(\vec{x})\,,$$

die Radon-Nikodym-Ableitung von P_X bezüglich λ_d, wird **Verteilungsdichte** oder auch **Dichtefunktion** bzw. kurz **Dichte** des Zufallsvektors X bzw. des Verteilungsgesetzes P_X genannt.

Anmerkungen 1.57.

(1) Die Dichtefunktion f_X eines stetigen Zufallsvektors ist nur λ_d-fast überall eindeutig bestimmt.

(2) Sei $f\colon \mathbb{R}^d \to [0,\infty]$ eine nichtnegative, messbare Funktion mit

$$\int\limits_{\mathbb{R}^d} f(\vec{x})\,\lambda_d(\mathrm{d}\vec{x}) = \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} f(x_1,\dots,x_d)\,\mathrm{d}x_1 \dots \mathrm{d}x_d = 1\,.$$

Dann ist $f\,\lambda_d$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}_d)$ und für den unmittelbar gegebenen Zufallsvektor X gilt

$$\mathrm{P}_X = f\,\lambda_d\,,$$

d.h. X ist ein stetiger Zufallsvektor mit der Dichtefunktion $f_X = f$.

Satz 1.58. Für einen stetigen Zufallsvektor X mit der Dichtefunktion f_X gilt

(1)

$$\mathrm{P}_X(B) = \mathrm{P}(X \in B) = \int\limits_B f_X(\vec{x})\,\mathrm{d}\vec{x} = $$
$$= \int\limits_B \cdots \int f_X(x_1,\dots,x_d)\,\mathrm{d}x_1 \dots \mathrm{d}x_d \quad \text{für} \quad B \in \mathcal{B}_d$$

(2)

$$F_X(\vec{x}) = F_{X_1,\dots,X_d}(x_1,\dots,x_d) =$$
$$= \int\limits_{-\infty}^{x_1} \dots \int\limits_{-\infty}^{x_d} f_X(t_1,\dots,t_d)\,\mathrm{d}t_d \dots \mathrm{d}t_1, \quad \vec{x} = (x_1,\dots,x_d)^T \in \mathbb{R}^d$$

Satz 1.58 ((Fortsetzung))**.**

(3) Es gilt $\mathrm{P}_X(B) = 0$ für niederdimensionalen Mengen $B \in \mathcal{B}_d$, $\dim(B) \in \{0, \ldots, d-1\}$, insbesondere

$$\mathrm{P}(X = \vec{x}) = 0 \quad \text{für alle} \quad \vec{x} \in \mathbb{R}^d .$$

Folgerung 1.59. Sei $X = (X_1, X_2)^T$ ein zweidimensionaler, stetiger Zufallsvektor mit der Dichtefunktion f_X und $-\infty < a_i < b_i < \infty$, $i = 1, 2$. Dann gilt

$$\begin{aligned}
&\mathrm{P}(a_1 < X_1 < b_1, a_2 < X_2 < b_2)\\
=&\mathrm{P}(a_1 \leqslant X_1 < b_1, a_2 < X_2 < b_2)\\
&\vdots\\
=&\mathrm{P}(a_1 \leqslant X_1 \leqslant b_1, a_2 \leqslant X_2 \leqslant b_2)\\
=&F_X(b_1, b_2) - F_X(a_1, b_2) - F_X(b_1, a_2) + F_X(a_1, a_2)\\
=&\int_{a_1}^{b_1} \int_{a_2}^{b_2} f_X(x_1, x_2)\, \mathrm{d}x_2\, \mathrm{d}x_1 .
\end{aligned}$$

Anmerkungen 1.60. Sei X ein stetiger Zufallsvektor mit der Dichtefunktion f_X.

(1) Dann gilt für die Randverteilung P_{X_k}, $k = 1, \ldots, d$, und $B \in \mathcal{B}$

$$\mathrm{P}_{X_k}(B) = \mathrm{P}_{X_1,\ldots,X_d}\left(\mathbb{R}^{k-1} \times B \times \mathbb{R}^{d-k}\right) =$$
$$= \int_B \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} f_X(x_1, \ldots, x_d)\, \mathrm{d}x_1 \ldots \mathrm{d}x_{k-1}\, \mathrm{d}x_{k+1} \ldots \mathrm{d}x_d\, \mathrm{d}x_k .$$

(2) Die Komponenten X_k des Zufallsvektors X sind stetige Zufallsgrößen und für die Verteilungsdichte f_{X_k} der k-ten Randverteilung P_{X_k}, $k = 1, \ldots, d$, gilt

$$f_{X_k}(x_k) = \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} f_X(x_1, \ldots, x_d)\, \mathrm{d}x_1 \ldots \mathrm{d}x_{k-1}\, \mathrm{d}x_{k+1} \ldots \mathrm{d}x_d$$

für $x_k \in \mathbb{R}$ λ-fast überall.

Anmerkungen 1.60 ((Fortsetzung))**.**

(3) Die Umkehrung von (2) gilt nicht. Es gibt z.B. stetige Zufallsgrößen X_1, X_2, so dass $(X_1, X_2)^T$ kein stetiger Zufallsvektor ist.

Satz 1.61. Sei X ein stetiger Zufallsvektor mit der Dichtefunktion f_X und $g\colon \mathbb{R}^d \to \mathbb{R}$ eine P_X-integrierbare Funktion. Dann ist $f_X \cdot g$ λ_d-integrierbar, und es gilt

$$\int_{\mathbb{R}^d} g(\vec{x})\ \mathrm{P}_X(\mathrm{d}\vec{x}) = \int_{\mathbb{R}^d} g(\vec{x})\, f_X(\vec{x})\, \lambda_d(\mathrm{d}\vec{x}) =$$
$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_d)\, f_X(x_1, \ldots, x_d)\, \mathrm{d}x_1 \ldots \mathrm{d}x_d .$$

1.8 Unabhängigkeit von Zufallsvariablen

Definitionen 1.62. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $I \neq \varnothing$ eine beliebige Indexmenge. Für $i \in I$ sei $(\Omega_i', \mathcal{F}_i')$ ein messbarer Raum und X_i eine Zufallsvariable

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\Omega_i', \mathcal{F}_i'), \quad i \in I .$$

- Dann heißt die **Familie** $(X_i)_{i \in I}$ **von Zufallsvariablen unabhängig**, falls die Familie von Urbild-σ-Algebren $\big(X_i^{-1}(\mathcal{F}_i')\big)_{i \in I}$ unabhängig ist.
- Ist $I = \{1, \ldots, n\}$ so spricht man direkt von **unabhängigen Zufallsvariablen** $X_1, \ldots, X_n$.
- Für $I = \mathbb{N}$ spricht man von einer **unabhängigen Folge** $(X_i)_{i \in \mathbb{N}}$ **von Zufallsvariablen**.

Anmerkung 1.63.

(1) Eine Familie von Zufallsvariablen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist.

(2) Es genügt daher, Bedingungen für die Unabhängigkeit von Zufallsvariablen $X_1, \ldots, X_n$ zu finden. Sind diese für jede endliche Teilfamilie einer Familie von Zufallsvariablen erfüllt, so ist die gesamte Familie unabhängig.

Satz 1.64. Seien X_i Zufallsvariable von demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ in den messbaren Raum $(\Omega_i', \mathcal{F}_i')$

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\Omega_i', \mathcal{F}_i'), \quad i = 1, \ldots, n .$$

Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn

$$\mathrm{P}_{X_1, \ldots, X_n} = \bigotimes_{i=1}^{n} \mathrm{P}_{X_i} .$$

Folgerung 1.65. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $I \neq \varnothing$ eine beliebige Indexmenge. Für $i \in I$ sei $(\Omega_i', \mathcal{F}_i')$ ein messbarer Raum und X_i eine Zufallsvariable

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\Omega_i', \mathcal{F}_i') .$$

Die Familie $(X_i)_{i \in I}$ von Zufallsvariablen ist genau dann unabhängig, wenn für jede nichtleere, endliche Teilmenge $I_n = \{i_1, \ldots, i_n\} \subseteq I$ gilt

$$\mathrm{P}_{X_{i_1}, \ldots, X_{i_n}} = \bigotimes_{j=1}^{n} \mathrm{P}_{X_{i_j}} .$$

Satz 1.66. Seien X_i Zufallsgrößen auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\mathbb{R}, \mathcal{B}), \quad i = 1, \ldots, n.$$

Die Zufallsgrößen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn für alle $(x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

Folgerung 1.67. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $I \neq \varnothing$ eine beliebige Indexmenge. Für $i \in I$ sei X_i eine Zufallsgröße

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\mathbb{R}, \mathcal{B}).$$

Die Familie $(X_i)_{i \in I}$ von Zufallsgrößen ist genau dann unabhängig, wenn für jede nichtleere, endliche Teilmenge $I_n = \{i_1, \ldots, i_n\} \subseteq I$ und alle $(x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt

$$F_{X_{i_1},\ldots,X_{i_n}}(x_1, \ldots, x_n) = \prod_{j=1}^{n} F_{X_{i_j}}(x_j).$$

Satz 1.68. Seien X_i stetige Zufallsgrößen auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit den Dichtefunktionen f_{X_i}

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\mathbb{R}, \mathcal{B}), \quad i = 1, \ldots, n.$$

Die Zufallsgrößen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn für die gemeinsame Dichtefunktion $f_{X_1,\ldots,X_n}$ für λ_n-fast alle $(x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Folgerung 1.69. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $I \neq \varnothing$ eine beliebige Indexmenge. Für $i \in I$ sei X_i eine stetige Zufallsgröße

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\mathbb{R}, \mathcal{B}).$$

Die Familie $(X_i)_{i \in I}$ von Zufallsgrößen ist genau dann unabhängig, wenn für jede nichtleere, endliche Teilmenge $I_n = \{i_1, \ldots, i_n\} \subseteq I$ und λ_n-fast alle $(x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt

$$f_{X_{i_1},\ldots,X_{i_n}}(x_1, \ldots, x_n) = \prod_{j=1}^{n} f_{X_{i_j}}(x_j).$$

Satz 1.70. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum, $I \neq \varnothing$ eine beliebige Indexmenge und $(X_i)_{i\in I}$ eine unabhängige Familie von Zufallsvariablen

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{X_i} (\Omega_i', \mathcal{F}_i')\,, \quad i \in I\,.$$

Für $i \in I$ sei $(\Omega_i'', \mathcal{F}_i'')$ ein messbarer Raum und Y_i eine Zufallsvariable

$$(\Omega_i', \mathcal{F}_i', \mathrm{P}_{X_i}) \xrightarrow{Y_i} (\Omega_i'', \mathcal{F}_i'')\,.$$

Dann ist $(Y_i \circ X_i)_{i\in I}$ eine unabhängige Familie von Zufallsvariablen

$$(\Omega, \mathcal{F}, \mathrm{P}) \xrightarrow{Y_i \circ X_i} (\Omega_i'', \mathcal{F}_i'')\,, \quad i \in I\,.$$

1.9 Symmetrie

Definition 1.71. Die Zufallsgröße X bzw. deren Verteilungsgesetz P_X heißt **symmetrisch**, falls X und $-X$ dieselbe Verteilung besitzen, d.h.

$$\mathrm{P}_X = \mathrm{P}_{-X}\,.$$

Die Zufallsgröße X bzw. deren Verteilungsgesetz P_X heißt **symmetrisch bezüglich des Symmetriezentrums** $c \in \mathbb{R}$ bzw. kurz **symmetrisch bezüglich** $c \in \mathbb{R}$, falls $X - c$ symmetrisch ist.

Satz 1.72. Sei X eine stetige Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann sind die folgenden Aussagen äquivalent

(1) Die Verteilung von X ist symmetrisch.

(2) Für die Verteilungsfunktion F_X von X gilt

$$F_X(-x) = 1 - F_X(x) \quad \text{für alle} \quad x \in \mathbb{R}\,.$$

(3) Für die Verteilungsdichte f_X von X gilt

$$f_X(-x) = f_X(x) \quad \lambda\text{-fast überall}\,.$$

Beispiel 1.73. Die stetige Zufallsgröße X besitze die Verteilungsdichte

$$f_X(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}\,.$$

Offensichtlich ist die Verteilung von X symmetrisch.

Aufgabe 1.74.

(1) Zeigen Sie, dass eine stetige Zufallsgröße X genau dann symmetrisch bezüglich $c \in \mathbb{R}$ ist, wenn für die Verteilungsfunktion F_X von X gilt

$$F_X(c-x) = 1 - F_X(c+x) \quad \text{für alle} \quad x \in \mathbb{R}\,.$$

(2) Zeigen Sie, dass eine stetige Zufallsgröße X genau dann symmetrisch bezüglich $c \in \mathbb{R}$ ist, wenn für die Verteilungsdichte f_X von X gilt

$$f_X(c-x) = f_X(c+x) \quad \lambda\text{-fast überall}\,.$$

2 Kenngrößen für Zufallsgrößen

2.1 Erwartungswert

Definition 2.1. Sei X eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Falls X bezüglich P integrierbar ist, d.h.

$$\int_\Omega |X(\omega)|\ \mathrm{P}(\mathrm{d}\omega) < \infty$$

so sagt man, dass der Erwartungswert der Zufallsgröße X existiert und nennt die Größe

$$\mathrm{E}\,(X) := \int_\Omega X(\omega)\ \mathrm{P}(\mathrm{d}\omega) = \int_\Omega X\,\mathrm{d}\,\mathrm{P}$$

den **Erwartungswert** der Zufallsgröße X.
(**Bem.** Obige Bedingung wird auch durch $\mathrm{E}\,(|X|) < \infty$ ausgedrückt.)

Satz 2.2. Der Erwartungswert der Zufallsgröße X existiert genau dann, wenn

$$\int_\mathbb{R} |x|\ \mathrm{P}_X(\mathrm{d}x) < \infty.$$

Existiert der Erwartungswert $\mathrm{E}\,(X)$, so gilt

$$\mathrm{E}\,(X) = \int_\Omega X(\omega)\ \mathrm{P}(\mathrm{d}\omega) = \int_\mathbb{R} x\ \mathrm{P}_X(\mathrm{d}x).$$

Satz 2.3. Der Erwartungswert der diskreten Zufallsgröße X mit dem Wertebereich $W_X = \{x_k\}_{k\in I\subseteq\mathbb{N}}$ und den Einzelwahrscheinlichkeiten $\{p_k\}_{k\in I}$ existiert genau dann, wenn

$$\sum_{k\in I} |x_k|\,p_k < \infty.$$

Existiert der Erwartungswert $\mathrm{E}\,(X)$, so gilt

$$\mathrm{E}\,(X) = \sum_{k\in I} x_k\,p_k.$$

Beispiel 2.4 ((Einmaliges Würfeln))**.** Für die diskrete Zufallsgröße X aus Beispiel 1.43 mit

$$x_k = k, \quad p_k = \frac{1}{6}, \quad k = 1, \dots, 6,$$

gilt

$$\mathrm{E}(X) = \sum_{k=1}^{6} k \frac{1}{6} = \frac{7}{2} = 3.5$$

Satz 2.5. Der Erwartungswert der stetigen Zufallsgröße X mit der Dichtefunktion f_X existiert genau dann, wenn

$$\int_{\mathbb{R}} |x| f_X(x) \, \mathrm{d}x < \infty.$$

Existiert der Erwartungswert $\mathrm{E}(X)$, so gilt

$$\mathrm{E}(X) = \int_{\mathbb{R}} x \, f_X(x) \, \mathrm{d}x.$$

Beispiel 2.6. Für die stetige Zufallsgröße X aus Beispiel 1.49 mit der Dichtefunktion

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x) \quad \text{für} \quad -\infty < a < b < \infty$$

gilt

$$\mathrm{E}(X) = \int_{\mathbb{R}} x \, f_X(x) \, \mathrm{d}x = \frac{1}{b-a} \int_a^b x \, \mathrm{d}x = \frac{a+b}{2}.$$

Beispiel 2.7. Sei X eine stetige Zufallsgröße mit der Dichtefunktion

$$f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x) \quad \text{für} \quad \lambda > 0.$$

Dann ist

$$\mathrm{E}(X) = \int_{\mathbb{R}} x \, f_X(x) \, \mathrm{d}x = \int_0^\infty x \, \lambda e^{-\lambda x} \, \mathrm{d}x = \frac{1}{\lambda}.$$

Beispiel 2.8 ((Gegenbeispiel))**.** Die symmetrische, stetige Zufallsgröße X aus Beispiel 1.73 besitzt die Dichtefunktion

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

Der Erwartungswert $\mathrm{E}(X)$ existiert in diesem Fall nicht.

Anmerkung 2.9 ((Geometrische Interpretation des Erwartungswertes einer stetigen Zufallsgröße))**.** Sei X eine stetige Zufallsgröße mit der Dichtefunktion f_X. Dann ist $\mathrm{E}(X) \in \mathbb{R}$ gerade die x-Koordinate des Schwerpunktes der Fläche zwischen der x-Achse und dem Graph der Funktion $f_X \geqslant 0$.

Anmerkung 2.10 ((Bedeutung des Erwartungswertes))**.**

- wichtigste Kenngröße einer Zufallsgröße
- wichtigster Lageparameter einer Zufallsgröße
- durchschnittlicher Wert, Mittelwert einer Zufallsgröße
- Zentrum, Schwerpunkt der Verteilung einer Zufallsgröße
- Grundlage zur Definition weiterer Kenngrößen

Definition 2.11. Es sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und E eine Eigenschaft, so dass für alle Elementarereignisse $\omega \in \Omega$ definiert ist, ob ω diese Eigenschaft besitzt oder nicht. Dann sagt man, die Eigenschaft E gilt P**-fast sicher** (abgekürzt P-f.s.) auf Ω bzw. P-fast sicher alle $\omega \in \Omega$ besitzen die Eigenschaft E, falls es ein Ereignis $N \in \mathcal{F}$ mit $\mathrm{P}(N) = 0$ gibt, so dass

$$\{\omega \in \Omega \colon \omega \text{ besitzt Eigenschaft } E \text{ nicht}\} \subseteq N.$$

Ein Ereignis $A \in \mathcal{F}$ tritt P-fast sicher ein bzw. ist P-fast sicher, wenn $\mathrm{P}(A^c) = 0$ bzw. $\mathrm{P}(A) = 1$.

Definition 2.12. Sei X eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann spricht man von einer **nichtnegativen Zufallsgröße** X, wenn

$$X \geqslant 0 \quad \text{P-fast sicher.}$$

Aufgabe 2.13.

(1) Zeigen Sie, dass eine Zufallsgröße X genau dann nichtnegativ ist, wenn

$$F_X(x) = 0 \quad \text{für alle} \quad x \leqslant 0.$$

(2) Zeigen Sie, dass eine stetige Zufallsgröße X mit der Dichtefunktion f_X genau dann nichtnegativ ist, wenn

$$f_X \, I_{(-\infty,0)} = 0 \quad \lambda\text{-fast überall.}$$

Satz 2.14.

(1) Sei X eine Zufallsgröße, deren Erwartungswert existiert. Dann gilt

$$|\,\mathrm{E}\,(X)\,| \leqslant \mathrm{E}\,(|X|)\,.$$

(2) Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren Erwartungswert existiert, und es gelte $X \leqslant Y$ P-fast sicher. Dann gilt

$$\mathrm{E}\,(X) \leqslant \mathrm{E}\,(Y)\,.$$

(3) Seien $X_1, \ldots, X_n$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren Erwartungswert existiert, sowie $c_i \in \mathbb{R}, i = 1, \ldots, n$ reelle Zahlen. Dann existiert auch $\mathrm{E}\left(c_1 X_1 + \ldots + c_n X_n\right)$, und es gilt

$$\mathrm{E}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \,\mathrm{E}\left(X_i\right).$$

Satz 2.15.

(1) Für $c \in \mathbb{R}$ sei $X = c$ P-fast sicher, d.h. X ist eine P-fast sicher konstante Zufallsgröße. Dann existiert $\mathrm{E}(X)$, und es gilt

$$\mathrm{E}(X) = c.$$

(2) Sei $A \in \mathcal{F}$. Dann existiert der Erwartungswert der Zufallsgröße $X = I_A$, und es gilt

$$\mathrm{E}(X) = \mathrm{P}(A).$$

(3) Sei X eine Zufallsgröße, und es gelte $-\infty < a \leqslant X \leqslant b < \infty$ P-fast sicher. Dann existiert $\mathrm{E}(X)$, und es gilt

$$a \leqslant \mathrm{E}(X) \leqslant b.$$

(4) Sei X eine nichtnegative Zufallsgröße, deren Erwartungswert existiert. Dann gilt $\mathrm{E}(X) = 0$ genau dann, wenn $X = 0$ P-fast sicher.

Satz 2.16. Sei X eine nichtnegative Zufallsgröße, deren Erwartungswert existiert. Dann gilt

$$\mathrm{E}(X) = \int_0^\infty (1 - F_X(x))\,\mathrm{d}x.$$

Beispiel 2.17. Die Zufallsgröße X aus Beispiel 2.7 ist nichtnegativ und besitzt die Verteilungsfunktion

$$F_X(x) = \left(1 - e^{-\lambda x}\right) I_{(0,\infty)}(x) \quad \text{für} \quad \lambda > 0.$$

Man kann ihren Erwartungswert also auch gemäß

$$\mathrm{E}(X) = \int_0^\infty (1 - F_X(x))\,\mathrm{d}x = \int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

berechnen.

Satz 2.18 ((Multiplikationssatz))**.** Seien $X_1, \ldots, X_n$ unabhängige Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren Erwartungswert existiert. Dann existiert auch $\mathrm{E}\left(X_1 \cdot \ldots \cdot X_n\right)$, und es gilt

$$\mathrm{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathrm{E}\left(X_i\right).$$

2.2 Momente

Definition 2.19. Sei X eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Falls

$$\int\limits_{\Omega} |X(\omega)|^p \, \mathrm{P}(\mathrm{d}\omega) < \infty \quad \text{für} \quad p \in \mathbb{N},$$

so sagt man, dass das p-te Moment der Zufallsgröße X existiert und nennt die Größe

$$\mu_p := \mu_p(X) := \int\limits_{\Omega} X^p(\omega) \, \mathrm{P}(\mathrm{d}\omega) = \int\limits_{\Omega} X^p \, \mathrm{d}\,\mathrm{P}\,.$$

das **(gewöhnliche) Moment der Ordnung** p bzw. das p**-te (gewöhnliche) Moment** der Zufallsgröße X. Für $p = 1$ schreibt man kurz

$$\mu := \mu(X) := \mu_1(X).$$

Anmerkung 2.20. Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann existiert der Erwartungswert der Zufallsgröße $Y = X^p$, und es gilt

$$\mathrm{E}\,(Y) = \mathrm{E}\,(X^p) = \mu_p(X).$$

Für $p = 1$ gilt insbesondere

$$\mathrm{E}\,(X) = \mu(X).$$

Folglich kann man einerseits den Erwartungswert der Zufallsgröße als spezielles Moment und andererseits die Momente als Erwartungswerte spezieller Funktionen von der Zufallsgröße auffassen.

Satz 2.21. Das p-te Moment der Zufallsgröße X existiert genau dann, wenn

$$\int\limits_{\mathbb{R}} |x|^p \, \mathrm{P}_X(\mathrm{d}x) < \infty \quad \text{für} \quad p \in \mathbb{N}.$$

Existiert das p-te Moment $\mu_p(X)$, so gilt

$$\mu_p(X) = \mathrm{E}\,(X^p) = \int\limits_{\Omega} X^p(\omega) \, \mathrm{P}(\mathrm{d}\omega) = \int\limits_{\mathbb{R}} x^p \, \mathrm{P}_X(\mathrm{d}x).$$

Satz 2.22. Das p-te Moment der diskreten Zufallsgröße X mit dem Wertebereich $W_X = \{x_k\}_{k \in I \subseteq \mathbb{N}}$ und

$$\sum_{k \in I} |x_k|^p \, p_k < \infty \quad \text{für} \quad p \in \mathbb{N}.$$

Existiert das p-te Moment $\mu_p(X)$, so gilt

$$\mu_p(X) = \sum_{k \in I} x_k^p \, p_k\,.$$

Beispiel 2.23 ((Einmaliges Würfeln))**.** Für die diskrete Zufallsgröße X aus Beispiel 1.43 mit

$$x_k = k, \quad p_k = \frac{1}{6}, \quad k = 1, \ldots, 6,$$

gilt

$$\mu_p(X) = \mathrm{E}\left(X^p\right) = \sum_{k=1}^{6} k^p \, \frac{1}{6} = \frac{1}{6}\left(1 + 2^p + 3^p + 4^p + 5^p + 6^p\right).$$

Für $p = 2$ ergibt sich

$$\mu_2(X) = \mathrm{E}\left(X^2\right) = \frac{1}{6} \sum_{k=1}^{6} k^2 = \frac{91}{6}.$$

Satz 2.24. Das p-te Moment der stetigen Zufallsgröße X mit der Dichtefunktion f_X existiert genau dann, wenn

$$\int_{\mathbb{R}} |x|^p \, f_X(x) \, \mathrm{d}x < \infty \quad \text{für} \quad p \in \mathbb{N}.$$

Existiert das p-te Moment $\mu_p(X)$, so gilt

$$\mu_p(X) = \int_{\mathbb{R}} x^p \, f_X(x) \, \mathrm{d}x.$$

Beispiel 2.25. Für die stetige Zufallsgröße X mit der Dichtefunktion

$$f_X(x) = \lambda \, \mathrm{e}^{-\lambda x} \, I_{[0,\infty)}(x) \quad \text{für} \quad \lambda > 0$$

gilt

$$\mu_p(X) = \frac{p!}{\lambda^p} \quad \text{für} \quad p \in \mathbb{N}.$$

Satz 2.26. Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X), p = 2, 3, \ldots,$ existiert.

(1) Dann existieren auch alle Momente $\mu_q(X), q = 1, \ldots, p-1$, niedrigerer Ordnung und somit insbesondere $\mathrm{E}\,(X)$.

(2) Dann gilt

$$\int_{\Omega} |X - c|^p \, \mathrm{d}\,\mathrm{P} < \infty \quad \text{für alle} \quad c \in \mathbb{R}.$$

Definition 2.27. Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann heißt

$$\nu_p := \nu_p(X) := \int_{\Omega} (X - \mathrm{E}\,(X))^p \, \mathrm{d}\,\mathrm{P} \quad \text{für} \quad p \in \mathbb{N}$$

das **zentrale Moment der Ordnung** p bzw. das **p-te zentrale Moment** der Zufallsgröße X.

Anmerkung 2.28 (und Definition)**.** Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann existiert das p-te Moment der **zentrierten Zufallsgröße** $Y = X - \mathrm{E}(X)$, und es gilt

$$\mu_p(Y) = \nu_p(X) = \mathrm{E}\left((X - \mathrm{E}(X))^p\right).$$

Das p-te zentrale Moment $\nu_p(X)$ der Zufallsgröße X ist also gerade das (gewöhnliche) p-te Moment der zentrierten Zufallsgröße $X - \mathrm{E}(X)$ und kann zugleich als Erwartungswert spezieller Funktionen von der Zufallsgröße aufgefasst werden.

Satz 2.29. Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann existiert für $a, b \in \mathbb{R}$ auch $\nu_p(aX + b)$, und es gilt

$$\nu_p(aX + b) = a^p \nu_p(X).$$

Folgerung 2.30 (und Definition)**.** Zentrale Momente einer Zufallsgröße sind invariant gegenüber Verschiebungen (**translationsinvariant**).

Satz 2.31. Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann gilt für $p = 2, 3, 4$

$$\nu_2 = \mu_2 - \mu_1^2, \qquad \nu_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3,$$
$$\nu_4 = \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4,$$

und allgemein

$$\nu_p = \mu_p + \sum_{q=2}^{p-1} \binom{p}{q} \mu_q (-\mu)^{p-q} + (-1)^{p-1}(p-1)\mu^p \quad \text{für} \quad p \in \mathbb{N}.$$

Satz 2.32. Sei X eine Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann gilt für $p = 2, 3, 4$

$$\mu_2 = \nu_2 + \mu_1^2, \qquad \mu_3 = \nu_3 + 3\nu_2\mu_1 + \mu_1^3,$$
$$\mu_4 = \nu_4 + 4\nu_3\mu_1 + 6\nu_2\mu_1^2 + \mu_1^4,$$

und allgemein

$$\mu_p = \nu_p + \sum_{q=2}^{p-1} \binom{p}{q} \nu_q \mu^{p-q} + \mu^p \quad \text{für} \quad p \in \mathbb{N}.$$

Beispiel 2.33. Für die Zufallsgröße X mit der Dichtefunktion

$$f_X(x) = \lambda\, \mathrm{e}^{-\lambda x}\, I_{[0,\infty)}(x) \quad \text{für} \quad \lambda > 0$$

gilt

$$\nu_2(X) = \frac{1}{\lambda^2}, \qquad \nu_3(X) = \frac{2}{\lambda^3}, \qquad \nu_4(X) = \frac{9}{\lambda^4}.$$

Satz 2.34. Sei X eine symmetrische Zufallsgröße, deren p-tes Moment $\mu_p(X)$, $p \in \mathbb{N}$, existiert. Dann gilt

(1)
$$\mu_p(X) = 0, \quad \text{falls } p \text{ ungerade;}$$

(2)
$$\nu_p(X) = \mu_p(X).$$

Anmerkungen 2.35.

- Neben den gewöhnlichen und zentralen Momenten spielen auch absolute Momente $\mathrm{E}(|X|^p)$ bzw. absolute zentrale Momente $\mathrm{E}(|X - \mathrm{E}(X)|^p)$ eine Rolle, zum Teil auch nicht nur für Werte $p \in \mathbb{N}$.

- Für eine nichtnegative Zufallsgröße X mit existierendem p-ten Moment ($p \geqslant 1$) gilt
$$\mu_p(X) = \mathrm{E}(X^p) = p \int_0^\infty \mathrm{P}(X \geqslant t) t^{p-1} \, \mathrm{d}t\,.$$

2.3 Variabilitätskenngrößen

Definition 2.36. Sei X eine Zufallsgröße, deren zweites Moment existiert. Das zweite zentrale Moment $\nu_2(X)$ der Zufallsgröße X wird **Varianz** der Zufallsgröße X genannt und man schreibt dafür
$$\sigma^2 := \sigma^2(X) := \mathrm{Var}(X) := \nu_2(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right) \geqslant 0.$$

Anmerkung 2.37 ((Bedeutung der Varianz))**.**

- mittlere quadratische Abweichung einer Zufallsgröße von ihrem Erwartungswert
- Je größer die Varianz, desto weiter sind die Werte der Zufallsgröße X um $\mathrm{E}(X)$ herum verteilt.
- Je kleiner die Varianz, desto stärker konzentriert sich die Verteilung um den Erwartungswert herum.
- beschreibt die Variabilität, das Streuverhalten einer Zufallsgröße
- wichtigste Variabilitätskenngröße, wichtigster Streuungsparameter

Beispiel 2.38. Für die stetige Zufallsgröße X aus Beispiel 2.6 mit der Dichtefunktion
$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x) \quad \text{für} \quad -\infty < a < b < \infty$$
gilt
$$\mathrm{Var}(X) = \int\limits_{\mathbb{R}} \left(x - \frac{a+b}{2}\right)^2 f_X(x) \, \mathrm{d}x = \frac{(a-b)^2}{12}.$$

Satz 2.39.

(1) Für $c \in \mathbb{R}$ sei $X = c$ P-fast sicher. Dann existiert $\mathrm{Var}(X)$, und es gilt

$$\mathrm{Var}(X) = 0.$$

(2) Sei X eine Zufallsgröße, deren zweites Moment existiert. Dann gilt

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2 .$$

Insbesondere folgt

$$\mathrm{E}(X)^2 \leqslant \mathrm{E}\left(X^2\right).$$

(3) Sei X eine Zufallsgröße, deren zweites Moment existiert. Dann gilt für $a, b \in \mathbb{R}$

$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X) .$$

Somit ist die Varianz eine translationsinvariante Kenngröße.

Definition 2.40.

(1) Sei X eine Zufallsgröße, deren zweites Moment existiert. Dann heißt

$$\sigma := \sigma(X) := \mathrm{sd}(X) := \sqrt{\mathrm{Var}(X)} = \sqrt{\mathrm{E}\left((X - \mathrm{E}(X))^2\right)} \geqslant 0$$

Standardabweichung der Zufallsgröße X.

(2) Für eine Zufallsgröße X mit existierendem zweiten Moment und $\mathrm{Var}(X) > 0$ ist

$$\tilde{X} := \frac{X - \mathrm{E}(X)}{\mathrm{sd}(X)}$$

die **Standardisierung von** X.

Anmerkungen 2.41.

(1) Die Maßeinheit der Varianz einer Zufallsgröße X ist das Quadrat der Maßeinheit von X. Diese Maßeinheit ist häufig schwer oder nicht zu interpretieren.

(2) Die Standardabweichung einer Zufallsgröße X hat hingegen dieselbe Maßeinheit wie X und ist deshalb einfacher zu interpretieren.

(3) Die Standardabweichung einer Zufallsgröße ist eine translationsinvariante Kenngröße.

(4) Für $a, b \in \mathbb{R}$ gilt $\mathrm{sd}(aX + b) = |a| \,\mathrm{sd}(X)$.

(5) Für die Standardisierung $\tilde{X}$ von X gelten

$$\mathrm{E}\left(\tilde{X}\right) = 0, \qquad \mathrm{Var}\left(\tilde{X}\right) = 1, \qquad \mathrm{sd}\left(\tilde{X}\right) = 1.$$

Beispiel 2.42. Für die stetige Zufallsgröße X aus Beispiel 2.38 mit der Dichtefunktion

$$f_X(x) = \frac{1}{b-a}\, I_{[a,b]}(x) \quad \text{für} \quad -\infty < a < b < \infty$$

gilt

$$\mathrm{sd}\,(X) = \frac{b-a}{2\sqrt{3}}.$$

Definition 2.43. Sei X eine Zufallsgröße, deren zweites Moment existiert, und es gelte $\mathrm{E}\,(X) \neq 0$. Dann heißt

$$\mathrm{cv}\,(X) := \frac{\mathrm{sd}\,(X)}{\mathrm{E}\,(X)}$$

(Pearsonscher) Variationskoeffizient der Zufallsgröße X.

Anmerkungen 2.44.

(1) Der Variationskoeffizient ist eine dimensionslose Kenngröße.

(2) Für $a > 0$ gilt

$$\mathrm{cv}\,(aX) = \mathrm{cv}\,(X)\,.$$

Der Variationskoeffizient ist folglich eine **skaleninvariante** Kenngröße.

(3) Teilweise wird der Variationskoeffizient nur für nichtnegative Zufallsgrößen oder auch durch die Beziehung

$$\mathrm{cv}\,(X) := \frac{\mathrm{sd}\,(X)}{|\mathrm{E}\,(X)|} \geqslant 0$$

definiert.

Anmerkung 2.45 ((Bedeutung des Variationskoeffizienten))**.**

- Als relative Kenngröße ist der Variationskoeffizient besonders zum Vergleich der Variabilität verschiedener Zufallsgrößen bzw. Verteilungen geeignet.

Beispiel 2.46. Für die Zufallsgröße X mit der Dichtefunktion

$$f_X(x) = \lambda\, \mathrm{e}^{-\lambda x}\, I_{[0,\infty)}(x) \quad \text{für} \quad \lambda > 0$$

gilt (vgl. Beispiele 2.7 und 2.33)

$$\mathrm{cv}\,(X) \equiv 1 \quad \text{für alle } \lambda > 0.$$

Der Variationskoeffizient ist folglich unabhängig vom Parameter $\lambda > 0$.

2.4 Schiefe und Wölbung

Definition 2.47. Sei X eine Zufallsgröße, deren drittes Moment existiert mit $\mathrm{Var}\,(X) > 0$. Dann heißt

$$\gamma\,(X) := \frac{\nu_3(X)}{\sigma^3(X)} = \frac{\nu_3(X)}{\nu_2^{3/2}(X)}$$

(Charliersche) Schiefe der Zufallsgröße X.

Satz 2.48. Sei X eine Zufallsgröße, deren drittes Moment existiert mit $\mathrm{Var}\,(X) > 0$.

(1) Dann existiert für $a \neq 0$ und $b \in \mathbb{R}$ auch $\gamma\,(aX + b)$, und es gilt

$$\gamma\,(aX + b) = \mathrm{sign}\,(a)\,\gamma\,(X)\,.$$

(2) Ist $\tilde{X}$ die Standardisierung der Zufallsgröße X, dann gilt

$$\gamma\,(X) = \mu_3\left(\tilde{X}\right)\,.$$

Anmerkungen 2.49.

(1) Die Schiefe ist eine dimensionslose Kenngröße.

(2) Die Schiefe ist eine translationsinvariante Kenngröße.

(3) Die Schiefe ist eine skaleninvariante Kenngröße.

Anmerkung 2.50 ((Bedeutung der Schiefe))**.**

- Die Schiefe beschreibt die Asymmetrie einer Verteilung.
- Symmetrische Verteilungen haben die Schiefe 0.
- Eine rechtsschiefe (auch: linkssteile) Verteilung hat eine positive Schiefe.
- Hingegen ist die Schiefe einer linksschiefen (auch: rechtssteilen) Verteilung negativ.

Beispiel 2.51. Für die Zufallsgröße X aus Beispiel 2.33 mit der Dichtefunktion

$$f_X(x) = \lambda\,\mathrm{e}^{-\lambda x}\,I_{[0,\infty)}(x) \quad \text{für} \quad \lambda > 0$$

gilt

$$\gamma\,(X) \equiv 2 \quad \text{für alle } \lambda > 0.$$

Die Schiefe ist somit unabhängig vom Parameter $\lambda > 0$.

Definition 2.52. Sei X eine Zufallsgröße, deren viertes Moment existiert mit $\mathrm{Var}\,(X) > 0$. Dann heißt

$$\kappa(X) := \frac{\nu_4(X)}{\sigma^4(X)} = \frac{\nu_4(X)}{\nu_2^2(X)}$$

Wölbung oder **Kurtosis** der Zufallsgröße X.
Die Größe

$$\varepsilon(X) := \kappa(X) - 3$$

heißt **Exzess** der Zufallsgröße X.

Satz 2.53. Sei X eine Zufallsgröße, deren viertes Moment existiert mit $\mathrm{Var}\,(X) > 0$.

(1) Dann existiert für $a \neq 0$ und $b \in \mathbb{R}$ auch $\kappa\,(aX + b)$, und es gelten

$$\kappa\,(aX + b) = \kappa\,(X) \quad \text{und} \quad \varepsilon\,(aX + b) = \varepsilon\,(X)\ .$$

(2) Ist $\tilde{X}$ die Standardisierung der Zufallsgröße X, dann gilt

$$\gamma\,(X) = \mu_4\left(\tilde{X}\right)\ .$$

Anmerkungen 2.54.

(1) Wölbung und Exzess sind dimensionslose, translationsinvariante und skaleninvariante Kenngrößen.

(2) Der Exzess einer Zufallsgröße beschreibt die Abweichung der Wölbung im Vergleich zu einer Normalverteilung mit demselben Erwartungswert und derselben Varianz. So werden Dichtefunktionen unterteilt in

- normalgipflig ($\varepsilon(X) = 0$);
- steilgipflig ($\varepsilon(X) > 0$) bzw.
- flachgipflig ($\varepsilon(X) < 0$).

2.5 Kovarianz und Korrelationskoeffizient

Definition 2.55. Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Dann heißt

$$\mathrm{Cov}\,(X, Y) := \mathrm{E}\,((X - \mathrm{E}\,(X))(Y - \mathrm{E}\,(Y)))$$

Kovarianz der Zufallsgrößen X und Y.

Satz 2.56. Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert.

(1) Offensichtlich ist
$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X).$$

(2) Es gilt
$$\operatorname{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y).$$

(3) Für $a, b, c, d \in \mathbb{R}$ gilt
$$\operatorname{Cov}(aX + b, cY + d) = ac \operatorname{Cov}(X, Y).$$

Insbesondere ist die Kovarianz eine translationsinvariante Kenngröße.

(4) Speziell gilt
$$\operatorname{Cov}(X, X) = \operatorname{Var}(X).$$

Satz 2.57. Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert.

(1) Dann existiert auch $\operatorname{Var}(X + Y)$, und es gilt
$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Cov}(X, Y).$$

(2) Sind die Zufallsgrößen X und Y unabhängig, dann gilt
$$\operatorname{Cov}(X, Y) = 0 \quad \text{und} \quad \operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

Aufgabe 2.58. Seien X_i und $Y_i, i = 1, 2$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Zeigen Sie, dass dann gilt
$$\begin{aligned}\operatorname{Cov}(X_1 + X_2, Y_1 + Y_2) &= \\ &= \operatorname{Cov}(X_1, Y_1) + \operatorname{Cov}(X_1, Y_2) + \operatorname{Cov}(X_2, Y_1) + \operatorname{Cov}(X_2, Y_2).\end{aligned}$$

Definitionen 2.59.

- Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Weiterhin sei $\operatorname{Var}(X) > 0$ und $\operatorname{Var}(Y) > 0$. Dann heißt
$$\varrho := \varrho(X, Y) := \operatorname{Corr}(X, Y) := \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\operatorname{Cov}(X, Y)}{\operatorname{sd}(X)\operatorname{sd}(Y)}$$
  - **(Pearsonscher) Korrelationskoeffizient** bzw.
  - **(gewöhnlicher) Korrelationskoeffizient**

der Zufallsgrößen X und Y.

- Ist der Korrelationskoeffizient positiv (negativ), so nennt man die Zufallsgrößen X und Y **positiv (negativ) korreliert**.
- Für $\operatorname{Corr}(X, Y) = 0$ heißen die Zufallsgrößen X und Y **unkorreliert**.

Satz 2.60. Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Weiterhin sei $\operatorname{Var}(X) > 0$ und $\operatorname{Var}(Y) > 0$.

(1) Offensichtlich ist $\operatorname{Corr}(X, Y) = \operatorname{Corr}(Y, X)$.

(2) Für $a \neq 0, b, c \neq 0, d \in \mathbb{R}$ gilt

$$\operatorname{Corr}(aX + b, cY + d) = \operatorname{sign}(ac) \operatorname{Corr}(X, Y) .$$

(3) Speziell gilt für $a \neq 0, b \in \mathbb{R}$, dass $\operatorname{Corr}(X, aX + b) = \operatorname{sign}(a)$.

(4) Ist $\tilde{X}$ bzw. $\tilde{Y}$ die Standardisierung von X bzw. Y, dann gilt

$$\operatorname{Corr}(X, Y) = \mathrm{E}\left(\tilde{X} \cdot \tilde{Y}\right) .$$

Anmerkungen 2.61.

(1) Die Maßeinheit der Kovarianz $\operatorname{Cov}(X, Y)$ zweier Zufallsgrößen X und Y ist das Produkt der Maßeinheiten von X und Y. Diese Maßeinheit ist häufig schwer oder nicht zu interpretieren.

(2) Der Korrelationskoeffizient ist hingegen eine dimensionslose Kenngröße.

(3) Der Korrelationskoeffizient ist eine translationsinvariante Kenngröße.

(4) Der Korrelationskoeffizient ist eine skaleninvariante Kenngröße.

Satz 2.62. Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Weiterhin sei $\operatorname{Var}(X) > 0$ und $\operatorname{Var}(Y) > 0$.

(1) Es gilt

$$-1 \leqslant \operatorname{Corr}(X, Y) \leqslant 1.$$

(2) Es gilt $|\operatorname{Corr}(X, Y)| = 1$ genau dann, wenn Konstanten $a \neq 0$ und $b \in \mathbb{R}$ existieren mit $Y = aX + b$ P-f.s.

Anmerkung 2.63 ((Bedeutung des Korrelationskoeffizienten))**.**

- Der Korrelationskoeffizient ist ein Maß für die Stärke des linearen Zusammenhangs zwischen zwei Zufallsgrößen.

Satz 2.64. Seien X und Y Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Weiterhin sei $\mathrm{Var}\,(X) > 0$ und $\mathrm{Var}\,(Y) > 0$. Sind die Zufallsgrößen X und Y unabhängig, dann gilt

$$\mathrm{Corr}\,(X, Y) = 0.$$

Beispiel 2.65. Sei X eine diskrete Zufallsgröße mit dem Wertebereich $W_X = \{-1, 0, 1\}$und den Einzelwahrscheinlichkeiten

$$\mathrm{P}(X = -1) = \mathrm{P}(X = 0) = \mathrm{P}(X = 1) = \frac{1}{3}.$$

Offensichtlich sind die Zufallsgrößen X und $Y = X^2$ abhängig. Jedoch ist $\mathrm{Corr}\,(X, Y) = 0$.

Definition 2.66. Sei $X = (X_1, \ldots, X_d)^T$ ein Zufallsvektor und die Erwartungswerte $\mathrm{E}\,(X_i)$ der einzelnen Komponenten $X_i, i = 1, \ldots, d$, mögen existieren. Dann versteht man unter dem **Erwartungswert** $\mathrm{E}\,(X)$ des Zufallsvektors X (auch: **Erwartungswertvektor**) den Vektor der Erwartungswerte

$$(\mathrm{E}\,(X_1)\,, \ldots, \mathrm{E}\,(X_d))^T.$$

Definition 2.67. Sei $X = (X_1, \ldots, X_d)^T$ ein Zufallsvektor und die zweiten Momente der einzelnen Komponenten $X_i, i = 1, \ldots, d$, mögen existieren. Dann heißt die Matrix

$$\Sigma = (\sigma_{i,j})_{i,j=1}^d \quad \text{mit} \quad \sigma_{i,j} = \mathrm{Cov}\,(X_i, X_j)\,, \quad i, j = 1, \ldots, d,$$

Kovarianzmatrix des Zufallsvektors X.

Satz 2.68 (und Definition)**.** Sei $X = (X_1, \ldots, X_d)^T$ ein Zufallsvektor und die zweiten Momente der einzelnen Komponenten $X_i, i = 1, \ldots, d$, mögen existieren. Dann gilt

(1) Die Kovarianzmatrix ist eine reelle symmetrische Matrix.

(2) Die Kovarianzmatrix ist **positiv semidefinit**, d.h.

$$\vec{x}^T\, \Sigma\, \vec{x} \geqslant 0 \quad \text{für alle} \quad \vec{x} \in \mathbb{R}^d.$$

Beispiel 2.69. Seien X_1 und X_2 Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, deren zweites Moment existiert. Weiterhin sei $\sigma_1^2 = \mathrm{Var}\,(X_1) > 0$ und $\sigma_2^2 = \mathrm{Var}\,(X_2) > 0$, sowie $\varrho = \mathrm{Corr}\,(X_1, X_2)$.
Dann hat die Kovarianzmatrix Σ die Gestalt

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \varrho\,\sigma_1\sigma_2 \\ \varrho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Sind X_1 und X_2 unkorreliert, d.h. $\varrho = 0$, dann ist die Kovarianzmatrix Σ eine Diagonalmatrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

3 Transformationen von Zufallsgrößen

Problem
Seien $X_1, \dots, X_d$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit bekanntem Verteilungsgesetz $\mathrm{P}_{X_1,\dots,X_d}$ und $g\colon \mathbb{R}^d \to \mathbb{R}$ eine messbare Funktion. Welche Verteilung hat dann die Zufallsgröße

$$Y = g(X_1, \dots, X_d)\,?$$

Gesucht ist also

$$\mathrm{P}_Y \quad \text{oder} \quad F_Y$$

bzw. speziell

$$f_Y \quad \text{oder} \quad y_k \in W_Y \quad \text{und} \quad \mathrm{P}(Y = y_k)\,.$$

3.1 Die Verteilungsfunktion unter Transformationen

Satz 3.1. Seien $X_1, \dots, X_d$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $g\colon \mathbb{R}^d \to \mathbb{R}$ eine messbare Funktion. Dann gilt für die Zufallsgröße $Y = g(X_1, \dots, X_d)$

$$\mathrm{P}_Y = \mathrm{P}_{X_1,\dots,X_d} \circ g^{-1} = \mathrm{P} \circ (X_1, \dots, X_d)^{-1} \circ g^{-1}$$

und somit

$$F_Y(y) = \mathrm{P}_{X_1,\dots,X_d}(g^{-1}((-\infty, y))) = \mathrm{P}((X_1, \dots, X_d)^{-1}(g^{-1}((-\infty, y))))\,.$$

Beispiel 3.2. Sei X eine stetige Zufallsgröße. Dann gilt für die Zufallsgröße $Y = X^2$

$$F_Y(y) = \begin{cases} F_X\left(+\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) & \text{für} \quad y > 0\,, \\ 0 & \text{für} \quad y \leqslant 0\,. \end{cases}$$

Satz 3.3. Sei X eine Zufallsgröße und $g\colon \mathbb{R} \to \mathbb{R}$ eine streng monoton wachsende stetige Funktion mit

$$m = \inf_{x\in\mathbb{R}} g(x) \quad \text{und} \quad M = \sup_{x\in\mathbb{R}} g(x).$$

Dann gilt für die Zufallsgröße $Y = g(X)$

$$F_Y(y) = \begin{cases} 0 & \text{für} \quad y \leqslant m\,, \\ F_X\left(g^{-1}(y)\right) & \text{für} \quad m < y < M\,, \\ 1 & \text{für} \quad y \geqslant M\,. \end{cases}$$

Beispiele 3.4. Sei X eine Zufallsgröße. Dann gilt

(1)

$$F_{aX+b}(y) = F_X\left(\frac{y-b}{a}\right) \quad \text{für} \quad a > 0\,;$$

(2)

$$F_{\mathrm{e}^X}(y) = \begin{cases} 0 & \text{für} \quad y \leqslant 0\,, \\ F_X\left(\ln y\right) & \text{für} \quad y > 0\,. \end{cases}$$

3.2 Transformationssatz für Dichten

Satz 3.5 ((Transformationssatz für Dichten))**.** Sei $X = (X_1, \ldots, X_d)^T$ ein Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit der Dichtefunktion f_X bezüglich des Lebesgue-Borel-Maßes auf $(\mathbb{R}^d, \mathcal{B}_d)$.

- Für eine offene Menge $D \subseteq \mathbb{R}^d$ gelte
$$\int\limits_D f_X(\vec{x})\,\mathrm{d}\vec{x} = 1\,.$$
- Die Abbildung $g\colon D \to \mathbb{R}^d$ sei injektiv und stetig.
- Die inverse Abbildung $g^{-1}\colon W_g \to \mathbb{R}^d$ sei stetig partiell differenzierbar.

Satz 3.5 ((Fortsetzung))**.**

- Für die Determinante der Jacobi-Matrix gelte
$$J_{g^{-1}}(\vec{y}) = \begin{vmatrix} \frac{\partial g_1^{-1}}{\partial y_1}(\vec{y}) & \ldots & \frac{\partial g_1^{-1}}{\partial y_d}(\vec{y}) \\ \vdots & & \vdots \\ \frac{\partial g_d^{-1}}{\partial y_1}(\vec{y}) & \ldots & \frac{\partial g_d^{-1}}{\partial y_d}(\vec{y}) \end{vmatrix} \neq 0 \quad \text{für alle} \quad \vec{y} \in W_g\,.$$

Dann ist $Y = g(X)$ ebenfalls ein stetiger Zufallsvektor und besitzt die Dichtefunktion
$$f_Y(\vec{y}) = f_X(g^{-1}(\vec{y}))\; |J_{g^{-1}}(\vec{y})|\; I_{W_g}(\vec{y})\,.$$

Anmerkung 3.6. Ist die Abbildung $g\colon D \to \mathbb{R}^d$ stetig differenzierbar, so kann in der Behauptung von Satz 3.5 $J_{g^{-1}}(\vec{y})$ durch $\frac{1}{J_g(g^{-1}(\vec{y}))}$ ersetzt werden, wobei
$$J_g(\vec{x}) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1}(\vec{x}) & \ldots & \frac{\partial g_1}{\partial x_d}(\vec{x}) \\ \vdots & & \vdots \\ \frac{\partial g_d}{\partial x_1}(\vec{x}) & \ldots & \frac{\partial g_d}{\partial x_d}(\vec{x}) \end{vmatrix} \neq 0 \quad \text{für alle} \quad \vec{x} \in D\,.$$

Beispiel 3.7. Sei $X = (X_1, X_2)^T$ ein stetiger Zufallsvektor mit der Dichtefunktion f_{X_1,X_2} . Die Abbildung $g\colon \mathbb{R}^2 \to \mathbb{R}^2$ sei gegeben durch
$$g_1(x_1, x_2) = x_1 + x_2 \quad \text{und} \quad g_2(x_1, x_2) = x_2\,.$$
Dann besitzt der Zufallsvektor
$$Y = (Y_1, Y_2)^T = (g_1(X_1, X_2), g_2(X_1, X_2))^T = (X_1 + X_2, X_2)^T$$
die Dichtefunktion
$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(y_1 - y_2, y_2)\,.$$
Insbesondere ergibt sich für die Randverteilung von $Y_1 = X_1 + X_2$
$$f_{X_1+X_2}(y_1) = \int\limits_{\mathbb{R}} f_{X_1,X_2}(y_1 - y_2, y_2)\,\mathrm{d}y_2\,.$$

3.3 Summen unabhängiger Zufallsgrößen

Satz 3.8 (und Definition)**.** Seien P_1 und P_2 zwei Wahrscheinlichkeitsmaße auf dem messbaren Raum $(\mathbb{R}, \mathcal{B})$. Dann wird durch

$$\begin{aligned}(\mathrm{P}_1 * \mathrm{P}_2)(B) &:= \int_{\mathbb{R}} \int_{\mathbb{R}} I_B(x_1 + x_2)\ \mathrm{P}_1(\mathrm{d}x_1)\, \mathrm{P}_2(\mathrm{d}x_2) \quad \text{für} \quad B \in \mathcal{B} \\ &= \int_{\mathbb{R}} \mathrm{P}_1(B - x_2)\ \mathrm{P}_2(\mathrm{d}x_2) \\ &= \int_{\mathbb{R}} \mathrm{P}_2(B - x_1)\ \mathrm{P}_1(\mathrm{d}x_1)\end{aligned}$$

ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ definiert. Dieses Maß $\mathrm{P}_1 * \mathrm{P}_2$ wird die **Faltung** der Wahrscheinlichkeitsmaße P_1 und P_2 genannt.

Beispiel 3.9. Für die DIRAC-Maße δ_{x_1} und δ_{x_2} auf $(\mathbb{R}, \mathcal{B})$, $x_1, x_2 \in \mathbb{R}$, gilt

$$\delta_{x_1} * \delta_{x_2} = \delta_{x_1 + x_2}\,.$$

Anmerkung 3.10. Offensichtlich lässt sich die Definition der Faltung auf endliche viele Wahrscheinlichkeitsmaße $\mathrm{P}_1, \ldots, \mathrm{P}_n$ verallgemeinern

$$(\mathrm{P}_1 * \ldots * \mathrm{P}_n)(B) := \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} I_B(x_1 + \ldots + x_n)\ \mathrm{P}_1(\mathrm{d}x_1) \ldots \mathrm{P}_n(\mathrm{d}x_n)$$

für $B \in \mathcal{B}$.
Wie man leicht sieht, ist die Faltungsoperation sowohl assoziativ als auch kommutativ.

Satz 3.11. Seien X_1 und X_2 unabhängige Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann gilt

$$\mathrm{P}_{X_1 + X_2} = \mathrm{P}_{X_1} * \mathrm{P}_{X_2}$$

und insbesondere

$$F_{X_1 + X_2}(x) = \int_{\mathbb{R}} F_{X_1}(x - x_2)\ \mathrm{P}_{X_2}(\mathrm{d}x_2) = \int_{\mathbb{R}} F_{X_2}(x - x_1)\ \mathrm{P}_{X_1}(\mathrm{d}x_1)\,.$$

Satz 3.12. Es seien X_1 und X_2 unabhängige stetige Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit den Dichtefunktionen f_{X_1} bzw. f_{X_2}. Dann ist $X_1 + X_2$ ebenfalls eine stetige Zufallsgröße und besitzt die Dichtefunktion

$$\begin{aligned}f_{X_1 + X_2}(x) &= \int_{\mathbb{R}} f_{X_1}(x - x_2) f_{X_2}(x_2)\, \mathrm{d}x_2 \\ &= \int_{\mathbb{R}} f_{X_2}(x - x_1) f_{X_1}(x_1)\, \mathrm{d}x_1\,.\end{aligned}$$

Beispiel 3.13. Die Zufallsgrößen X_1 und X_2 seien unabhängig und mögen dieselbe Verteilungsfunktion

$$F_X(x) = \left(1 - \mathrm{e}^{-\lambda x}\right) I_{(0,\infty)}(x)\,, \quad x \in \mathbb{R}, \lambda > 0\,,$$

(vgl. Beispiel 2.17) besitzen. Dann gilt

$$F_{X_1+X_2}(x) = \left[1 - (1 + \lambda x)\,\mathrm{e}^{-\lambda x}\right] I_{(0,\infty)}(x)$$

und

$$f_{X_1+X_2}(x) = \lambda^2 x\,\mathrm{e}^{-\lambda x}\, I_{[0,\infty)}(x)\,.$$

3.4 Minimum und Maximum unabhängiger Zufallsgrößen

Satz 3.14. Seien $X_1, \ldots, X_d$ unabhängige Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit der Verteilungsfunktion $F_{X_i}(x)$, $i = 1, \ldots, d$. Dann gilt für die Zufallsgrößen

$$X_{max} = \max\{X_1, \ldots, X_d\} \quad \text{bzw.} \quad X_{min} = \min\{X_1, \ldots, X_d\}$$

$$F_{X_{max}}(x) = \prod_{i=1}^{d} F_{X_i}(x) \quad \text{bzw.} \quad F_{X_{min}}(x) = 1 - \prod_{i=1}^{d} \left(1 - F_{X_i}(x)\right)\,.$$

Definitionen 3.15. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum und $I \neq \varnothing$ eine beliebige Indexmenge. Für $i \in I$ sei X_i eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.

- Dann heißt die **Familie** $(X_i)_{i\in I}$ **von Zufallsgrößen identisch verteilt**, falls $\mathrm{P}_{X_i} = \mathrm{P}_{X_j}$ für alle $i, j \in I$.
- Ist $I = \{1, \ldots, n\}$ so werden die **Zufallsgrößen** $X_1, \ldots, X_n$ **identisch verteilt** genannt.
- Für $I = \mathbb{N}$ spricht man von einer **identisch verteilten Folge** $(X_i)_{i\in\mathbb{N}}$ **von Zufallsgrößen**.
- Ist eine Familie $(X_i)_{i\in I}$ von identisch verteilten Zufallsgrößen zugleich auch noch eine Familie von unabhängigen Zufallsgrößen, so spricht man von einer **unabhängigen, identisch verteilten Familie von Zufallsgrößen** und verwendet dafür die Abkürzung **i.i.d.** („independent and identically distributed“).

Folgerung 3.16. Seien $X_1, \ldots, X_d$ i.i.d. Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit der Verteilungsfunktion F_X. Dann gilt für die Zufallsgrößen

$$X_{max} = \max\{X_1, \ldots, X_d\} \quad \text{bzw.} \quad X_{min} = \min\{X_1, \ldots, X_d\}$$

$$F_{X_{max}}(x) = \left(F_X(x)\right)^d \quad \text{bzw.} \quad F_{X_{min}}(x) = 1 - \left(1 - F_X(x)\right)^d\,.$$

Sind die Zufallsgrößen X_i außerdem stetig und die Verteilungsfunktion F_X stetig differenzierbar, so gilt für die Dichtefunktionen

$$f_{X_{max}}(x) = d\,(F_X(x))^{d-1}\,f_X(x)$$

bzw.

$$f_{X_{min}}(x) = d\,(1 - F_X(x))^{d-1}\,f_X(x)\,,$$

wobei $f_X = F_X'$ λ-fast überall.

Beispiel 3.17. Die i.i.d. Zufallsgrößen $X_1, \ldots, X_d$ mögen die Verteilungsfunktion

$$F_X(x) = \left(1 - \mathrm{e}^{-\lambda x}\right) I_{(0,\infty)}(x)\,, \quad x \in \mathbb{R}\,,\ \lambda > 0\,,$$

besitzen (vgl. Beispiel 2.17). Dann gilt

$$f_{X_{max}}(x) = \lambda d \left(1 - \mathrm{e}^{-\lambda x}\right)^{d-1} \mathrm{e}^{-\lambda x}\, I_{[0,\infty)}(x)$$

bzw.

$$f_{X_{min}}(x) = \lambda d\,\mathrm{e}^{-\lambda d x}\, I_{[0,\infty)}(x)\,,$$

d.h. das Minimum X_{min} hat dieselbe Art von Verteilung wie die Zufallsgrößen $X_1, \ldots, X_d$, jedoch nicht mit dem Parameter $\lambda > 0$, sondern mit dem Parameter λd.

Anmerkung 3.18 ((Anwendung in der Zuverlässigkeitstheorie))**.**

(1) Ausfallwahrscheinlichkeit von in Reihe geschalteten Bauelementen
→ Minimum

(2) Ausfallwahrscheinlichkeit von parallel geschalteten Bauelementen
→ Maximum

(3) Berechnung der Ausfallwahrscheinlichkeit von komplexen Systemen
→ Kombination beider Situationen

4 Ausgewählte diskrete Verteilungen

Bestimmungsstücke und Kenngrößen einer diskreten Zufallsgröße

Sei X eine diskrete Zufallsgröße mit dem Wertebereich $W_X = \{x_k\}_{k\in I}$ und den Einzelwahrscheinlichkeiten $(p_k)_{k\in I}$ mit einer Indexmenge $\varnothing \neq I \subseteq \mathbb{N}$. Dann gilt (vgl. die Sätze 1.42 und 2.22)

$$\sum_{k\in I} p_k = 1\,, \quad \mathrm{P}_X = \sum_{k\in I} p_k\,\delta_{x_k}\,, \quad F_X(x) = \sum_{k\in I} p_k\, I_{(x_k,\infty)}(x)\,, \quad x \in \mathbb{R}\,,$$

$$\mu_p(X) = \sum_{k\in I} x_k^p\, p_k \quad \text{für} \quad p \in \mathbb{N}\,,$$

falls $\mu_p(X)$ existiert. Insbesondere ist (vgl. die Sätze 2.3 und 1.44)

$$\mathrm{E}\,(X) = \sum_{k\in I} x_k\, p_k \quad \text{und} \quad \mathrm{Var}\,(X) = \sum_{k\in I} (x_k - \mathrm{E}\,(X))^2\, p_k\,.$$

4.1 Einpunktverteilung

Definition 4.1. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Einpunktverteilung** im Punkt $c \in \mathbb{R}$, falls $X = c$ P-fast sicher.
Man spricht in diesem Fall auch von einer entarteten Verteilung im Punkt $c \in \mathbb{R}$.

Bestimmungsstücke und Kenngrößen
Die Zufallsgröße X besitze eine Einpunktverteilung im Punkt $c \in \mathbb{R}$. Dann gilt (vgl. Sätze 2.15 (1) und 2.39 (1))

$$W_X = \{c\}\,, \qquad p = \mathrm{P}(X = c) = 1\,,$$

$$\mathrm{P}_X = \delta_c\,, \qquad F_X(x) = I_{(c,\infty)}(x)\,,$$

$$\mathrm{E}\,(X) = c\,, \qquad \mathrm{Var}\,(X) = 0\,,$$

$$\mu_p(X) = c^p \quad \text{für} \quad p \in \mathbb{N}\,.$$

4.2 Zweipunktverteilung

Definition 4.2. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Zweipunktverteilung**, falls X nur zwei verschiedene Werte $x_1 \in \mathbb{R}$ und $x_2 \in \mathbb{R}$ annehmen kann.

Modellvorstellung
Die Zweipunktverteilung wird als Modell für Zufallsexperimente benutzt, bei denen es nur zwei mögliche Versuchsausgänge gibt.

Bestimmungsstücke und Kenngrößen
Die Zufallsgröße X besitze eine Zweipunktverteilung. Dann gilt

$$W_X = \{x_1, x_2\}\,, \qquad p_k = \mathrm{P}(X = x_k)\,, \quad k = 1, 2\,, \quad p_1 + p_2 = 1\,,$$

$$\mathrm{P}_X = p_1\delta_{x_1} + p_2\delta_{x_2}\,, \qquad F_X(x) = p_1 I_{(x_1,\infty)}(x) + p_2 I_{(x_2,\infty)}(x)\,,$$

$$\mathrm{E}\,(X) = p_1 x_1 + p_2 x_2\,, \qquad \mathrm{Var}\,(X) = (x_1 - x_2)^2 p_1 p_2\,,$$

$$\mu_p(X) = p_1 x_1^p + p_2 x_2^p \quad \text{für} \quad p \in \mathbb{N}\,.$$

Für eine eigentliche Zweipunktverteilung gelten $0 < p_1 < 1\,, 0 < p_2 < 1$.

4.3 Bernoulli-Verteilung

Definition 4.3. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Bernoulli-Verteilung** mit dem Parameter $p \in [0,1]$, falls X nur die Werte 0 und 1 annehmen kann mit $\mathrm{P}(X = 1) = p$.
Hierfür schreibt man auch

$$X \sim \mathsf{B}(p) \,.$$

Anmerkung
Die Bernoulli-Verteilung ist eine spezielle Zweipunktverteilung. Es gilt

$$x_1 = 0\,, \quad x_2 = 1\,, \quad p_1 = 1 - p \quad \text{und} \quad p_2 = p\,.$$

Modellvorstellung und Definitionen

- Die Bernoulli-Verteilung wird für die Beschreibung von Zufallsexperimenten benutzt, bei denen das Eintreten eines festen Ereignisses $A \in \mathcal{F}$ mit $\mathrm{P}(A) = p \in [0,1]$ untersucht wird. Dabei wird die Zufallsgröße $X = I_A$ betrachtet.
  - Das Ereignis A tritt ein (Erfolg): $X = 1$.
  - Das Ereignis A tritt nicht ein (Misserfolg): $X = 0$.
- Diese Art von Experiment nennt man **Bernoulli-Experiment**.
- Die Wahrscheinlichkeit $p = \mathrm{P}(X = 1) = \mathrm{P}(A)$ wird **Erfolgswahrscheinlichkeit** genannt.

Beispiele

- Auftreten von Zahl beim Wurf einer idealen Münze $(p = 0.5)$.
- Beim Werfen eines Kronverschlusses fällt der Verschluss auf die glatte Seite $(p \approx 0.6)$.
- Ein zufällig heruntergefallenes Marmeladenbrötchen landet auf der Marmeladenseite (gefühlt: $p = 1$).
- Ziehen eines Loses auf dem Rummel (Niete oder Gewinn).
- Ein zufällig aus einem Warenposten ausgewähltes Werkstück ist Ausschuss.

Bestimmungsstücke und Kenngrößen
Es sei $X \sim \mathsf{B}(p)$ eine mit dem Parameter $p \in [0,1]$ Bernoulli-verteilte Zufallsgröße. Dann gilt

$$W_X = \{0, 1\}\,, \qquad p = \mathrm{P}(X = 1)\,, \qquad 1 - p = \mathrm{P}(X = 0)\,,$$

$$\mathrm{P}_X = (1-p)\delta_0 + p\delta_1\,, \qquad F_X(x) = \begin{cases} 0 & \text{für} \quad x \leqslant 0\,, \\ 1-p & \text{für} \quad 0 < x \leqslant 1\,, \\ 1 & \text{für} \quad x > 1, \end{cases}$$

$$\mathrm{E}\,(X) = p\,, \qquad \mathrm{Var}\,(X) = p(1-p)\,,$$

$$\mu_q(X) = p \quad \text{für} \quad q \in \mathbb{N}\,.$$

4.4 Gleichverteilung

Definition 4.4. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **diskrete Gleichverteilung**, falls X endlich viele Werte $x_1, \ldots, x_n \in \mathbb{R}\,, x_i \neq x_j\,, i, j \in \{1, \ldots, n\}$, jeweils mit der gleichen Wahrscheinlichkeit annimmt. Hierfür schreibt man auch

$$X \sim \mathsf{U}\{x_1, \ldots, x_n\}\ .$$

Modellvorstellung
Bei einem Zufallsexperiment sind endlich viele Versuchsausgänge möglich. Jeder Versuchsausgang hat dieselbe Wahrscheinlichkeit.

Beispiel

- Einmaliges Würfeln mit einem idealen Würfel.

Bestimmungsstücke und Kenngrößen
Es sei $X \sim \mathsf{U}\{x_1, \ldots, x_n\}$ eine diskret gleichverteilte Zufallsgröße. Dann gilt

$$W_X = \{x_1, \ldots, x_n\}\,, \qquad p_k = \mathrm{P}(X = x_k) = \frac{1}{n}\,, \quad k = 1, \ldots, n\,,$$

$$\mathrm{P}_X = \frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}\,, \qquad F_X(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(x_k,\infty)}(x)\,,$$

$$\mathrm{E}\,(X) = \frac{1}{n}\sum_{k=1}^{n} x_k\,, \qquad \mu_p(X) = \frac{1}{n}\sum_{k=1}^{n} x_k^p \quad \text{für} \quad p \in \mathbb{N}\,.$$

4.5 Hypergeometrische Verteilung

Definition 4.5. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **hypergeometrische Verteilung** mit den Parametern $N \in \mathbb{N}, K \in \{0, \ldots, N\}$ und $n \in \{1, \ldots, N\}$, falls X die Werte $k \in \{k_{min}, \ldots, k_{max}\} \subseteq \{0, \ldots, n\}$ mit den Einzelwahrscheinlichkeiten

$$p_k = \mathrm{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

annimmt, wobei $k_{min} = \max\{0, n + K - N\}$ und $k_{max} = \min\{K, n\}$. Hierfür schreibt man auch

$$X \sim \mathsf{Hyp}\,(n, K, N)\ .$$

Modellvorstellung: Urnenmodell ohne Zurücklegen

- In einem nicht einsehbaren Gefäß (Urne) befinden sich $N \in \mathbb{N}$ gleichartige Objekte (z.B. Kugeln).
- Davon besitzen K, $K = 0, \ldots, N,$ dieser Objekte eine spezielle Eigenschaft (z.B. rote Kugeln), die sie von den anderen $N - K$ Objekten (z.B. weiße Kugeln) unterscheiden.
- Die Objekte in der Urne sind gut durchmischt, so dass jedes Objekt die gleiche Chance hat, entnommen zu werden.
- Nun werden nacheinander insgesamt n Objekte, $n \in \{1, \ldots, N\}$, entnommen (ohne das jeweils gezogene Objekt vor der nächsten Entnahme wieder in die Urne zurückzulegen) und gezählt, wieviele der entnommenen Objekte die spezielle Eigenschaft aufweisen.

Beispiel (Lotto 6 aus 49)
Von den $N = 49$ Kugeln besitzen $K = 6$ die spezielle Eigenschaft, dass sie die von einem Spieler getippten Zahlen sind. Es werden $n = 6$ Kugeln gezogen. Folglich gibt es $\binom{49}{6} = 13\,983\,816$ mögliche Ziehungsergebnisse.

- Es gibt nur einen Sechser, dessen Wahrscheinlichkeit ist $1/13\,983\,816 = 0.00000007151$.

Die weiteren Wahrscheinlichkeiten sind:

- Fünfer: 0.00001845,
- Vierer: 0.00096862 und
- Dreier: 0.017650404.

Anwendung

- statistische Qualitätskontrolle (Stichprobenziehung ohne Zurücklegen, z.B. bei zerstörender Prüfung von Werkstücken)

Bestimmungsstücke und Kenngrößen
Es sei $X \sim \mathsf{Hyp}\,(n, K, N)$ eine mit den Parametern $N \in \mathbb{N}\,, K \in \{0, \ldots, N\}$ und $n \in \{1, \ldots, N\}$ hypergeometrisch verteilte Zufallsgröße. Dann gilt

$$W_X = \{\max\{0, n + K - N\}, \ldots, \min\{K, n\}\}\,,$$

$$p_k = \mathrm{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}\,, \quad k \in W_X\,,$$

$$\mathrm{E}\,(X) = n\,\frac{K}{N}\,, \qquad \mathrm{Var}\,(X) = n\,\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}\,.$$

4.6 Binomialverteilung

Definition 4.6. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Binomialverteilung** mit den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$, falls X die Werte $k \in \{0, \ldots, n\}$ mit den Einzelwahrscheinlichkeiten

$$p_k = \mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

annimmt. Hierfür schreibt man auch

$$X \sim \mathsf{B}\,(n; p)\ .$$

Modellvorstellung: Urnenmodell mit Zurücklegen

- In einer Urne befinden sich $N \in \mathbb{N}$ gleichartige Objekte (z.B. Kugeln).
- Darunter befinden sich Objekte mit einer speziellen Eigenschaft (z.B. rote Kugeln), die sie von den anderen Objekten (z.B. weiße Kugeln) unterscheiden.
- Der Anteil der Objekte mit der speziellen Eigenschaft sei $p \cdot 100\%$, $p \in [0, 1]$.
- Die Objekte in der Urne sind gut durchmischt, so dass jedes Objekt die gleiche Chance hat, entnommen zu werden.
- Nun wird ein Objekt aus der Urne entnommen und registiert, um welche Art von Objekt es sich handelt.
- Anschließend wird das Objekt wieder in die Urne zurückgelegt und die Objekte wieder gut durchmischt.
- Dieser Vorgang wird n-mal nacheinander ausgeführt und gezählt, wie oft ein Objekt mit der speziellen Eigenschaft gezogen wurde.

Anmerkungen und Definition

(1) Die Binomialverteilung beschreibt die n-malige unabhängige Durchführung eines Bernoulli-Experiments.

(2) Diese Vorgehensweise nennt man auch **Bernoulli-Schema**.

(3) Die binomialverteilte Zufallsgröße X gibt an, wieviele Erfolge bei der n-maligen unabhängigen Durchführung eines Bernoulli-Experiments eingetreten sind, d.h. wie häufig dabei das betrachtete Ereignis A eingetreten ist.

Anwendung

- statistische Qualitätskontrolle (Stichprobenziehung mit Zurücklegen, z.B. bei zerstörungsfreier Prüfung von Werkstücken)

Bestimmungsstücke und Kenngrößen

Es sei $X \sim \mathsf{B}\,(n;p)$ eine mit den Parametern $n \in \mathbb{N}$ und $p \in [0,1]$ binomialverteilte Zufallsgröße. Dann gilt

$$W_X = \{0,\dots,n\}\,, \quad p_k = \mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}\,, k = 0,\dots,n\,,$$

$$\mathrm{P}_X = \sum_{k=0}^{n} p_k \delta_k\,, \qquad F_X(x) = \sum_{k=0}^{n} p_k I_{(k,\infty)}(x)\,,$$

$$\mathrm{E}\,(X) = np\,, \qquad \mathrm{Var}\,(X) = np(1-p)\,, \qquad \gamma\,(X) = \frac{1-2p}{\sqrt{np(1-p)}}\,.$$

Spezialfall

Die Bernoulli-Verteilung ist eine spezielle Binomialverteilung. Es gilt

$$\mathsf{B}\,(p) = \mathsf{B}\,(1;p) \quad \text{für} \quad p \in [0,1]\,.$$

Satz 4.6.1

Seien $X_i \sim \mathsf{B}\,(p)\,, i = 1,\dots,n\,,$ i.i.d. Bernoulli-verteilte Zufallsgrößen mit dem Parameter $p \in [0,1]$. Dann gilt

$$X_1 + \ldots + X_n \sim \mathsf{B}\,(n;p)\ .$$

Satz 4.6.2

Seien $X_1 \sim \mathsf{B}\,(n_1;p)$ und $X_2 \sim \mathsf{B}\,(n_2;p)$ zwei unabhängige, binomialverteilte Zufallsgrößen. Dann gilt

$$X_1 + X_2 \sim \mathsf{B}\,(n_1+n_2;p)\ .$$

Vergleich von hypergeometrischer Verteilung und Binomialverteilung
Es sei $X \sim \mathsf{Hyp}\,(n, K, N)$ und $Y \sim \mathsf{B}\left(n; \frac{K}{N}\right)$. Dann gilt einerseits

$$\mathrm{E}\,(X) = \mathrm{E}\,(Y) = n\,\frac{K}{N}$$

und andererseits

$$\mathrm{Var}\,(X) = n\,\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1} \leqslant n\,\frac{K}{N}\left(1-\frac{K}{N}\right) = \mathrm{Var}\,(Y)\ .$$

Satz 4.6.3
Für $N \to \infty$ und $K \to \infty$ mit $\frac{K}{N} \to p \in [0,1]$ gilt

$$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \to \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für} \quad k = 0, \ldots, n\,, n \in \mathbb{N}\,,$$

d.h. die Einzelwahrscheinlichkeiten der hypergeometrischen Verteilung konvergieren unter den oben genannten Bedingungen gegen die Einzelwahrscheinlichkeiten der Binomialverteilung.

Anmerkung
Für große Werte von $N \in \mathbb{N}$ kann die hypergeometrische Verteilung $\mathsf{Hyp}\,(n, K, N)$ durch die Binomialverteilung $\mathsf{B}\,(n; p)$ mit $p = \frac{K}{N}$ approximiert werden.

4.7 Geometrische Verteilung

Definition 4.7. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **geometrische Verteilung** mit dem Parameter $p \in (0,1)$, falls X die Werte $k \in \mathbb{N}$ mit den Einzelwahrscheinlichkeiten

$$p_k = \mathrm{P}(X = k) = p(1-p)^{k-1}$$

annimmt. Hierfür schreibt man auch

$$X \sim \mathsf{Geo}\,(p)\ .$$

Modellvorstellung
Die geometrische Verteilung beschreibt die Anzahl der Versuche im Bernoulli-Schema bis zum ersten Erfolg, d.h. wieviele Versuche durchgeführt werden müssen, bis zum ersten Mal das betrachtete Ereignis A, der Erfolg, eintritt.

Anwendung

- Versicherungsmathematik (Risikobestimmung)

Bestimmungsstücke und Kenngrößen
Es sei $X \sim \mathsf{Geo}\,(p)$ eine mit dem Parameter $p \in (0,1)$ geometrisch verteilte Zufallsgröße. Dann gilt

$$W_X = \mathbb{N}\,, \qquad p_k = \mathrm{P}(X=k) = p(1-p)^{k-1} \quad \text{für} \quad k \in \mathbb{N}\,,$$

$$\mathrm{P}_X = \sum_{k=1}^{\infty} p_k \delta_k\,, \qquad F_X(x) = \sum_{k=1}^{\infty} p_k I_{(k,\infty)}(x)\,,$$

$$\mathrm{E}\,(X) = \frac{1}{p}\,, \qquad \mathrm{Var}\,(X) = \frac{1-p}{p^2}\,, \qquad \gamma\,(X) = \frac{2-p}{\sqrt{1-p}} > 0\,.$$

Satz und Definition
Es sei $X \sim \mathsf{Geo}\,(p)$ eine mit dem Parameter $p \in (0,1)$ geometrisch verteilte Zufallsgröße. Dann gilt

$$\mathrm{P}(X \geqslant n+k \,|\, X > n) = \mathrm{P}(X \geqslant k) \quad \text{für alle} \quad n,k \in \mathbb{N}\,.$$

Diese Eigenschaft der geometrischen Verteilung wird **Gedächtnislosigkeit** genannt.

Anmerkungen

- Von der geometrischen Verteilung gibt es eine Variante, bei der nicht die **Anzahl der Versuche bis zum ersten Erfolg** im Bernoulli-Schema, sondern die **Anzahl der Misserfolge vor dem ersten Erfolg** gezählt werden.
- Dies führt zu einer Verschiebung des Wertebereiches der Zufallsgröße und natürlich auch anderen Einzelwahrscheinlichkeiten.
- Folglich unterscheiden sich auch Momente und damit im Zusammenhang stehende Kenngrößen der beiden Varianten dieser Verteilung.
- In einigen Lehrbüchern bzw. Formelsammlungen wird beim Übergang von der Betrachtung der Anzahl der Versuche zur Anzahl der Misserfolge auch von der Erfolgswahrscheinlichkeit zur Misserfolgswahrscheinlichkeit übergegangen, d.h. p und $1-p$ wechseln die Plätze.

4.8 Negative Binomialverteilung

Definition 4.8. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **negative Binomialverteilung** mit den Parametern $n \in \mathbb{N}$ und $p \in (0,1)$, falls X die Werte $k \in \{n, n+1, \ldots\}$ mit den Einzelwahrscheinlichkeiten

$$p_k = \mathrm{P}(X=k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}$$

annimmt. Hierfür schreibt man auch

$$X \sim \mathsf{NB}\,(n;p)\,.$$

Modellvorstellung
Die negative Binomialverteilung beschreibt die Anzahl der erforderlichen Versuche im BERNOULLI-Schema bis zum n-ten Erfolg.

Anwendung

- Versicherungsmathematik (Schadenzahlverteilung)

Bestimmungsstücke und Kenngrößen
Es sei $X \sim \mathsf{NB}\,(n;p)$ eine mit den Parametern $n \in \mathbb{N}$ und $p \in (0,1)$ negativ binomialverteilte Zufallsgröße. Dann gilt

$$W_X = \{n, n+1, \ldots\}\,, \quad p_k = \binom{k-1}{n-1} p^n (1-p)^{k-n}\,, \quad k = n, n+1, \ldots\,,$$

$$\mathrm{P}_X = \sum_{k=n}^{\infty} p_k \delta_k\,, \qquad F_X(x) = \sum_{k=n}^{\infty} p_k I_{(k,\infty)}(x)\,,$$

$$\mathrm{E}\,(X) = \frac{n}{p}\,, \qquad \mathrm{Var}\,(X) = n\,\frac{1-p}{p^2}\,, \qquad \gamma\,(X) = \frac{2-p}{\sqrt{n(1-p)}} > 0\,.$$

Aufgabe 4.8.1
Seien $X_i \sim \mathsf{Geo}\,(p)$, $i = 1, \ldots, n$, i.i.d. geometrisch verteilte Zufallsgrößen mit dem Parameter $p \in (0,1)$. Zeigen Sie, dass dann gilt

$$X_1 + \ldots + X_n \sim \mathsf{NB}\,(n;p)\ .$$

Spezialfall
Die geometrische Verteilung ist eine spezielle negative Binomialverteilung. Es gilt

$$\mathsf{Geo}\,(p) = \mathsf{NB}\,(1;p) \quad \text{für} \quad p \in (0,1)\,.$$

Aufgabe 4.8.2
Seien $X_1 \sim \mathsf{NB}\,(n_1;p)$ und $X_2 \sim \mathsf{NB}\,(n_2;p)$ zwei unabhängige, negativ binomialverteilte Zufallsgrößen. Zeigen Sie, dass dann gilt

$$X_1 + X_2 \sim \mathsf{NB}\,(n_1 + n_2;p)\ .$$

Anmerkung

- Von der negativen Binomialverteilung gibt es – in Analogie zur geometrischen Verteilung – eine Variante, bei der nicht die **Anzahl der Versuche bis zum n-ten Erfolg** im Bernoulli-Schema, sondern die **Anzahl der Misserfolge vor dem n-ten Erfolg** gezählt werden.
- Dies führt analog zu einer Verschiebung des Wertebereiches der Zufallsgröße und ebenso anderen Einzelwahrscheinlichkeiten.
- Folglich unterscheiden sich auch Momente und damit im Zusammenhang stehende Kenngrößen der beiden Varianten dieser Verteilung.
- In einigen Lehrbüchern bzw. Formelsammlungen wird beim Übergang von der Betrachtung der Anzahl der Versuche zur Anzahl der Misserfolge auch von der Erfolgswahrscheinlichkeit zur Misserfolgswahrscheinlichkeit übergegangen, d.h. p und $1-p$ wechseln die Plätze.

4.9 Poisson-Verteilung

Definition 4.9. Eine Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Poisson-Verteilung** mit dem Parameter $\lambda > 0$, falls X die Werte $k \in \{0, 1, 2, \ldots\}$ mit den Einzelwahrscheinlichkeiten

$$p_k = \mathrm{P}(X = k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$$

annimmt. Hierfür schreibt man auch

$$X \sim \Pi\,(\lambda)\ .$$

Modellvorstellung und Definition

- Die Poisson-Verteilung beschreibt die Anzahl von Ereignissen, die mit konstanter Rate und unabhängig voneinander in einem festen Zeitintervall oder räumlichen Gebiet eintreten.
- Sie beschreibt näherungsweise die Anzahl der Erfolge bei sehr vielen Versuchen im Bernoulli-Schema, wenn die Erfolgswahrscheinlichkeit im einzelnen Bernoulli-Experiment sehr klein ist.
- Die Poisson-Verteilung wird deshalb auch die *Verteilung der seltenen Ereignisse* genannt.

Anwendungen

- Physik (radioaktiver Zerfall, Anzahl emittierter Teilchen)
- Versicherungsmathematik (Schadenzahlverteilung)
- Operations Research (Verteilung der Anzahl der Ankünfte bei Warteschlangenmodellen)

Bestimmungsstücke und Kenngrößen

Es sei $X \sim \Pi(\lambda)$ eine mit dem Parameter $\lambda > 0$ POISSON-verteilte Zufallsgröße. Dann gilt

$$W_X = \{0, 1, \ldots\}, \qquad p_k = \mathrm{P}(X = k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}, \quad k = 0, 1, \ldots,$$

$$\mathrm{P}_X = \sum_{k=0}^{\infty} p_k \delta_k, \qquad F_X(x) = \sum_{k=0}^{\infty} p_k I_{(k,\infty)}(x),$$

$$\mathrm{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda, \qquad \gamma(X) = \frac{1}{\sqrt{\lambda}} > 0.$$

Aufgabe

Seien $X_1 \sim \Pi(\lambda_1)$ und $X_2 \sim \Pi(\lambda_2)$ zwei unabhängige, POISSON-verteilte Zufallsgrößen. Zeigen Sie, dass dann gilt

$$X_1 + X_2 \sim \Pi(\lambda_1 + \lambda_2).$$

Satz 4.9.1

Seien $X_1 \sim \Pi(\lambda_1)$ und $X_2 \sim \Pi(\lambda_2)$ zwei unabhängige, POISSON-verteilte Zufallsgrößen. Dann gilt

$$\mathrm{P}(X_1 = k \,|\, X_1 + X_2 = n) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für alle} \quad k = 0, \ldots, n$$

mit

$$p = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Satz 4.9.2

Für $p_n \in [0,1]$ mit $\lim\limits_{n\to\infty} n p_n = \lambda > 0$ gilt

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} \quad \text{für} \quad k = 0, 1, \ldots,$$

d.h. die Einzelwahrscheinlichkeiten der Binomialverteilung konvergieren unter den oben genannten Bedingungen gegen die Einzelwahrscheinlichkeiten der POISSON-Verteilung.

Anmerkungen

(1) Für große Werte von $n \in \mathbb{N}$ kann die Binomialverteilung $\mathsf{B}(n;p)$ durch die Poisson-Verteilung $\Pi(\lambda)$ mit $\lambda = np$ approximiert werden.

(2) Für große Werte von $N \in \mathbb{N}$ und $n \leqslant N$ kann die hypergeometrische Verteilung $\mathsf{Hyp}(n, K, N)$ durch die Poisson-Verteilung $\Pi(\lambda)$ mit $\lambda = n\frac{K}{N}$ approximiert werden.

5 Ausgewählte stetige Verteilungen

Bestimmungsstücke und Kenngrößen
Sei X eine stetige Zufallsgröße mit der Dichtefunktion f_X. Dann gilt
(vgl. die Sätze 1.47 (2) und 2.24)

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \quad \text{für} \quad x \in \mathbb{R},$$

$$\mu_p(X) = \int_{\mathbb{R}} x^p f_X(x)\,\mathrm{d}x \quad \text{für} \quad p \in \mathbb{N},$$

falls $\mu_p(X)$ existiert. Insbesondere ist (vgl. die Sätze 2.5 und 1.50)

$$\mathrm{E}(X) = \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x \quad \text{und} \quad \mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mathrm{E}(X))^2 f_X(x)\,\mathrm{d}x\,.$$

5.1 Gleichverteilung

Definition 5.1. Eine stetige Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **stetige Gleichverteilung** auf dem Intervall $[a, b]$, $-\infty < a < b < +\infty$, falls ihre Dichtefunktion durch

$$f_X(x) = \frac{1}{b-a}\, I_{[a,b]}(x)\,, \quad x \in \mathbb{R},$$

gegeben ist. Hierfür schreibt man auch

$$X \sim \mathsf{U}[a, b]\,.$$

Für die Gleichverteilung auf einem Intervall ist auch die Bezeichnung **Rechteckverteilung** gebräuchlich.

Bedeutung der Parameter

- a und b sind Lageparameter bzw. Verschiebungsparameter.
- Sie begrenzen den Bereich, in dem die Wahrscheinlichkeitsmasse liegt.

Anmerkungen

(1) Die Gleichverteilung lässt sich auf einfache Weise von einem abgeschlossenen Intervall $[a, b]$ auf eine beliebige messbare Menge $B \in \mathcal{B}$ mit $0 < \lambda(B) < \infty$ (insbesondere z.B. halboffene und offene beschränkte Intervalle) verallgemeinern. Dazu verwendet man die Dichtefunktion

$$f_X(x) = \frac{1}{\lambda(B)} I_B(x) \quad \text{für} \quad x \in \mathbb{R}.$$

(2) Auch auf den mehrdimensionalen Fall eines gleichverteilten stetigen Zufallsvektors X bezüglich einer messbaren Menge $B \in \mathcal{B}_d$ mit $0 < \lambda_d(B) < \infty$ kann die Gleichverteilung verallgemeinert werden. Dazu verwendet man in analoger Weise die Dichtefunktion

$$f_X(\vec{x}) = \frac{1}{\lambda_d(B)} I_B(\vec{x}) \quad \text{für} \quad \vec{x} \in \mathbb{R}^d.$$

Anwendung

- geometrische Wahrscheinlichkeiten

Bestimmungsstücke und Kenngrößen
Sei $X \sim \mathsf{U}[a, b]$ eine auf dem Intervall $[a, b]$ gleichverteilte Zufallsgröße. Dann gilt (vgl. Beispiele 1.49, 2.6 und 2.38)

$$F_X(x) = \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,\infty)}(x), \quad x \in \mathbb{R},$$

$$\mathrm{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}, \qquad \gamma(X) = 0,$$

$$\mu_p(X) = \frac{b^{p+1} - a^{p+1}}{(b-a)(p+1)} \quad \text{für} \quad p \in \mathbb{N}.$$

Satz 5.1.1
Sei $X \sim \mathsf{U}[a, b]$ eine auf dem Intervall $[a, b]$ gleichverteilte Zufallsgröße. Dann gilt für $c, d \in \mathbb{R}, c \neq 0$,

$$cX + d \sim \begin{cases} \mathsf{U}[ac+d, bc+d] & \text{für} \quad c > 0, \\ \mathsf{U}[bc+d, ac+d] & \text{für} \quad c < 0, \end{cases}$$

d.h. die Familie der Gleichverteilungen auf abgeschlossenen beschränkten Intervallen ist abgeschlossen bezüglich (nicht entarteter) linearer Transformationen.

5.2 Cauchy-Verteilung

Definition 5.2. Eine stetige Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Cauchy-Verteilung** mit den Parametern $\mu \in \mathbb{R}$ und $\lambda > 0$, falls ihre Dichtefunktion durch

$$f_X(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x-\mu)^2}, \quad x \in \mathbb{R},$$

gegeben ist. Hierfür schreibt man auch

$$X \sim \mathsf{C}(\mu; \lambda).$$

Bedeutung der Parameter

- μ ist ein Lageparameter bzw. Verschiebungsparameter.
- λ ist ein Skalenparameter (Stauchung bzw. Streckung der x-Achse) bzw. Streuungsparameter.

Bestimmungsstücke und Kenngrößen
Sei $X \sim \mathsf{C}(\mu; \lambda)$ eine mit den Parametern $\mu \in \mathbb{R}$ und $\lambda > 0$ Cauchy-verteilte Zufallsgröße. Dann gilt

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\lambda}\right), \quad x \in \mathbb{R}.$$

Sämtliche Momente und davon abgeleitete Kenngrößen existieren nicht.

Aufgabe
Sei $X \sim \mathsf{C}(\mu; \lambda)$ eine mit den Parametern $\mu \in \mathbb{R}$ und $\lambda > 0$ Cauchy-verteilte Zufallsgröße. Zeigen Sie, dass dann für $a, b \in \mathbb{R}, a \neq 0$, gilt

$$aX + b \sim \mathsf{C}(a\mu + b; |a|\lambda),$$

d.h. die Familie der Cauchy-Verteilungen ist abgeschlossen bezüglich (nicht entarteter) linearer Transformationen.

Anmerkung
Mit Hilfe von charakteristischen Funktionen (später) kann folgendes leicht gezeigt werden.

Seien $X_1 \sim \mathsf{C}(\mu_1; \lambda_1)$ und $X_2 \sim \mathsf{C}(\mu_2; \lambda_2)$ unabhängige Zufallsgrößen. Dann gilt

$$X_1 + X_2 \sim \mathsf{C}(\mu_1 + \mu_2; \lambda_1 + \lambda_2),$$

d.h. die Familie der Cauchy-Verteilungen ist abgeschlossen bezüglich der Faltung, und

$$X_1 - X_2 \sim \mathsf{C}(\mu_1 - \mu_2; \lambda_1 + \lambda_2).$$

5.3 Normalverteilung

Definitionen 5.3. Eine stetige Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Normalverteilung** mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, falls ihre Dichtefunktion durch

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \,, \qquad x \in \mathbb{R} \,,$$

gegeben ist. Hierfür schreibt man auch

$$X \sim \mathsf{N}\left(\mu; \sigma^2\right) \,.$$

Der spezielle Fall $\mathsf{N}(0;1)$ wird **Standardnormalverteilung** genannt.

(Manchmal ist es günstig, die Einpunktverteilung im Punkt $\mu \in \mathbb{R}$ als Normalverteilung mit den Parametern μ und $\sigma^2 = 0$ zu betrachten, dies ist dann keine stetige Verteilung mehr.)

Bedeutung der Parameter

- μ ist ein Lageparameter bzw. Verschiebungsparameter.
- σ^2 ist ein Skalenparameter bzw. Streuungsparameter.

Anwendung

- Bei der summarischen Überlagerung vieler kleiner unabhängiger Einflüsse kann die daraus resultierende Größe annähernd durch eine normalverteilte Zufallsgröße modelliert werden.
- Beschreibung von Messfehlern in der Messtechnik.
- Zufällige Abweichung von Nennwerten, z.B. von Abmessungen von Werkstücken.
- Zentrale Bedeutung als Grenzverteilung bei Grenzwertsätzen für Summen von Zufallsgrößen.
- Zentrale Verteilung der parametrischen Statistik.

Bestimmungsstücke und Kenngrößen

Sei $X \sim \mathsf{N}(\mu; \sigma^2)$ eine mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ normalverteilte Zufallsgröße. Dann gilt

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{mit} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \, \mathrm{d}t \,, \quad x \in \mathbb{R} \,,$$

$$\mathrm{E}\,(X) = \mu \,, \qquad \mathrm{Var}\,(X) = \sigma^2 \,, \qquad \gamma\,(X) = 0,$$

$$\nu_{2p-1}(X) = 0\,, \qquad \nu_{2p}(X) = (2p-1)!!\,\sigma^{2p} \quad \text{für} \quad p \in \mathbb{N}\,.$$

Aufgabe 5.3.1
Sei $X \sim \mathsf{N}(\mu;\sigma^2)$ eine mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ normalverteilte Zufallsgröße. Zeigen Sie, dass dann für $a, b \in \mathbb{R}, a \neq 0,$ gilt

$$aX + b \sim \mathsf{N}\left(a\mu + b; a^2\sigma^2\right)\,,$$

d.h. die Familie der Normalverteilungen ist abgeschlossen bezüglich (nicht entarteter) linearer Transformationen.

Folgerung

(1) Ist $X \sim \mathsf{N}(0;1)$, so folgt für $\mu, \sigma \in \mathbb{R}, \sigma \neq 0,$ dass

$$\sigma X + \mu \sim \mathsf{N}\left(\mu;\sigma^2\right)\,.$$

(2) Ist $X \sim \mathsf{N}(\mu;\sigma^2)$, so ist

$$\frac{X-\mu}{\sigma} \sim \mathsf{N}(0;1)\,.$$

Satz 5.3.2
Seien $X_1 \sim \mathsf{N}(\mu_1;\sigma_1^2)$ und $X_2 \sim \mathsf{N}(\mu_2;\sigma_2^2)$ unabhängige Zufallsgrößen. Dann gilt

$$X_1 + X_2 \sim \mathsf{N}\left(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2\right)\,,$$

d.h. die Familie der Normalverteilungen ist abgeschlossen bezüglich der Faltung, und

$$X_1 - X_2 \sim \mathsf{N}\left(\mu_1 - \mu_2; \sigma_1^2 + \sigma_2^2\right)\,.$$

Satz 5.3.3
Sei $X \sim \mathsf{N}(0;1)$ eine standardnormalverteilte Zufallsgröße.

(1) Für beliebige $t > 0$ gilt

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{t} - \frac{1}{t^3}\right)\mathrm{e}^{-t^2/2} \leqslant 1 - \Phi(t) = \mathrm{P}(X > t) \leqslant \frac{1}{\sqrt{2\pi}}\frac{1}{t}\,\mathrm{e}^{-t^2/2}\,.$$

(2) Für beliebige beschränkte stetig differenzierbare Funktionen $g : \mathbb{R} \to \mathbb{R}$ gilt

$$\mathrm{E}\left(g'(X)\right) = \mathrm{E}\left(Xg(X)\right)\,.$$

5.4 Mehrdimensionale Normalverteilung

Definition 5.4. Ein d-dimensionaler Zufallsvektor $\vec{X} = (X_1, \dots, X_d)^T$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **d-dimensionale Standardnormalverteilung**, falls die Komponenten $X_i\,, i = 1, \dots, d\,,$ unabhängige standardnormalverteilte Zufallsgrößen sind. Man schreibt in diesem Fall

$$\vec{X} \sim \mathsf{N}\left(\vec{0}; \mathbf{I}_d\right) \,,$$

wobei $\mathbf{I}_d$ die d-dimensionale Einheitsmatrix bezeichnet.

Satz 5.4.1
Sei $\vec{X} \sim \mathsf{N}\left(\vec{0}; \mathbf{I}_d\right)$. Dann ist $\vec{X}$ ein stetiger Zufallsvektor mit Dichtefunktion

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^d}}\, \mathrm{e}^{-\frac{1}{2}\vec{x}^T\vec{x}} = \frac{1}{\sqrt{(2\pi)^d}}\, \mathrm{e}^{-\frac{1}{2}\|\vec{x}\|^2} \,, \quad \vec{x} \in \mathbb{R}^d \,,$$

($\|\cdot\|$ bezeichnet die Euklidische Norm in $\mathbb{R}^d$).

Außerdem gelten für den Erwartungswertvektor und die Kovarianzmatrix

$$\mathrm{E}\left(\vec{X}\right) = \vec{0} \,, \qquad \mathrm{Cov}\left(\vec{X}\right) = \mathbf{I}_d \,.$$

Definition 5.4.2
Ein n-dimensionaler Zufallsvektor $\vec{X} = (X_1, \dots, X_n)^T$ besitzt eine **Normalverteilung**, falls für einen d-dimensionalen standardnormalverteilten Zufallsvektor $\vec{Z}$, einen Vektor $\vec{\mu} \in \mathbb{R}^n$ und eine $n \times d-$Matrix $\mathbf{A}$ gilt

$$\vec{X} = \vec{\mu} + \mathbf{A}\vec{Z} \,.$$

Satz und Bezeichnung 5.4.3
Für den n-dimensionalen Zufallsvektor $\vec{X} = (X_1, \dots, X_n)^T$ aus Definition 5.4.2 gilt

$$\mathrm{E}\left(\vec{X}\right) = \vec{\mu} \,, \qquad \mathrm{Cov}\left(\vec{X}\right) = \mathbf{A}\mathbf{A}^T \,.$$

Man schreibt in diesem Fall

$$\vec{X} \sim \mathsf{N}\left(\vec{\mu}; \mathbf{A}\mathbf{A}^T\right) \,.$$

Definition und Satz 5.4.4
Gilt für den n-dimensionalen Zufallsvektor $\vec{X} = (X_1, \dots, X_n)^T$ aus Definition 5.4.2 $d = n\,,$ $\mathrm{Cov}\left(\vec{X}\right) = \mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}$ mit $\det(\mathbf{\Sigma}) \neq 0$, dann besitzt $\vec{X}$ eine **n-dimensionale nichtentartete Normalverteilung** $\mathsf{N}(\vec{\mu}; \mathbf{\Sigma})$ und $\vec{X}$ ist ein stetiger Zufallsvektor mit Dichtefunktion

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{\Sigma})}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{x} - \vec{\mu})\right\} \,, \quad \vec{x} \in \mathbb{R}^n \,.$$

$\mathbf{\Sigma}$ ist dann eine reelle, symmetrische und positiv definite Matrix.

Aufgabe 5.4.5
Sei $\vec{X} \sim \mathsf{N}(\vec{\mu}; \mathbf{\Sigma})$ ein mit den Parametern $\vec{\mu} \in \mathbb{R}^n$ und $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ normalverteilter n-dimensionaler Zufallsvektor, $\mathbf{A}_1 \in \mathbb{R}^{m\times n}$ und $\vec{b} \in \mathbb{R}^m$. Zeigen Sie, dass dann für den m-dimensionalen Zufallsvektor $\vec{Y} = \mathbf{A}_1\vec{X} + \vec{b}$ gilt

$$\vec{Y} \sim \mathsf{N}\left(\mathbf{A}_1\vec{\mu} + \vec{b}; \mathbf{A}_1\mathbf{\Sigma}\mathbf{A}_1^T\right).$$

Spezialfall $n = 2$
Sei $\vec{X} = (X_1, X_2)^T$ ein nichtentarteter zweidimensionaler normalverteilter Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit den Parametern $\vec{\mu} = (\mu_1, \mu_2)^T \in \mathbb{R}^2$ und $\mathbf{\Sigma} \in \mathbb{R}^{2\times 2}$. Die Kovarianzmatrix $\mathbf{\Sigma}$ hat die Gestalt (vgl. Beispiel 2.69)

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \varrho\,\sigma_1\sigma_2 \\ \varrho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \text{wobei}$$

$$\sigma_1^2 = \mathrm{Var}\,(X_1) > 0\,, \quad \sigma_2^2 = \mathrm{Var}\,(X_2) > 0 \quad \text{und}$$
$$\varrho = \mathrm{Corr}\,(X_1, X_2) \in (-1, +1)\,.$$

Dann ist

$$\mathbf{\Sigma}^{-1} = \frac{1}{\det(\mathbf{\Sigma})}\begin{pmatrix} \sigma_2^2 & -\varrho\,\sigma_1\sigma_2 \\ -\varrho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \quad \text{mit} \quad \det(\mathbf{\Sigma}) = \sigma_1^2\sigma_2^2\left(1 - \varrho^2\right).$$

Spezialfall $n = 2$ **(Fortsetzung)**
Hieraus folgt für $(x_1, x_2)^T \in \mathbb{R}^2$

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\Bigg\{ -\frac{1}{2(1-\varrho^2)} \times \\ \times \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\Bigg\} \end{aligned}$$

Herleitung: Für den Vorfaktor ergibt sich

$$\frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}}\,.$$

Für die Bestimmung des Arguments der Exponentialfunktion betrachtet man zunächst das Produkt

$$\Sigma^{-1}(\vec{x}-\vec{\mu}) = \frac{1}{\det(\Sigma)} \begin{pmatrix} \sigma_2^2 & -\varrho\,\sigma_1\sigma_2 \\ -\varrho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$= \frac{1}{\sigma_1^2\sigma_2^2\,(1-\varrho^2)} \begin{pmatrix} \sigma_2^2(x_1-\mu_1) - \varrho\,\sigma_1\sigma_2(x_2-\mu_2) \\ -\varrho\,\sigma_1\sigma_2(x_1-\mu_1) + \sigma_1^2(x_2-\mu_2) \end{pmatrix}$$

$$= \frac{1}{1-\varrho^2} \begin{pmatrix} \dfrac{x_1-\mu_1}{\sigma_1^2} - \varrho\,\dfrac{x_2-\mu_2}{\sigma_1\sigma_2} \\ -\varrho\,\dfrac{x_1-\mu_1}{\sigma_1\sigma_2} + \dfrac{x_2-\mu_2}{\sigma_2^2} \end{pmatrix}.$$

Somit ergibt sich insgesamt

$$-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu}) = -\frac{1}{2}\big((x_1-\mu_1) \quad (x_2-\mu_2)\big)\,\frac{1}{1-\varrho^2} \begin{pmatrix} \dfrac{x_1-\mu_1}{\sigma_1^2} - \varrho\,\dfrac{x_2-\mu_2}{\sigma_1\sigma_2} \\ -\varrho\,\dfrac{x_1-\mu_1}{\sigma_1\sigma_2} + \dfrac{x_2-\mu_2}{\sigma_2^2} \end{pmatrix} =$$

$$= -\frac{1}{2}\frac{1}{1-\varrho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} - \varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

$$= -\frac{1}{2(1-\varrho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right].$$

Satz 5.4.6
Sei $\vec{X} = (X_1, X_2)^T$ ein zweidimensionaler, normalverteilter Zufallsvektor mit den Parametern $\vec{\mu} = (\mu_1, \mu_2)^T \in \mathbb{R}^2$ und

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Dann gilt

(1) $X_i \sim \mathsf{N}\left(\mu_i; \sigma_i^2\right)$ für $i = 1, 2$.

(2) X_1 und X_2 sind genau dann unabhängig, wenn sie unkorreliert sind.

(3) Für beliebige $a_1, a_2 \in \mathbb{R}$ gilt

$$Y := a_1X_1 + a_2X_2 \sim \mathsf{N}\left(a_1\mu_1 + a_2\mu_2; a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_1\sigma_2\right).$$

Beweis. (zu 1) Für die Randverteilung von X_2 gilt

$$f_{X_2}(x_2) = \int\limits_{\mathbb{R}} f_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1 =$$

$$= \int\limits_{\mathbb{R}} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \mathrm{d}x_1.$$

Zwecks einer quadratischen Ergänzung wird der Ausdruck in eckigen Klammern im Argument der Exponentialfunktion entsprechend erweitert

$$\begin{aligned}
[\ldots] &= \frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \\
&= \frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\varrho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \varrho^2\,\frac{(x_2-\mu_2)^2}{\sigma_2^2} - \varrho^2\,\frac{(x_2-\mu_2)^2}{\sigma_2^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \\
&= \left(\frac{x_1-\mu_1}{\sigma_1} - \varrho\,\frac{x_2-\mu_2}{\sigma_2}\right)^2 + (1-\varrho^2)\,\frac{(x_2-\mu_2)^2}{\sigma_2^2}\,.
\end{aligned}$$

Somit erhält man

$$\begin{aligned}
&f_{X_2}(x_2) = \\
&= \int\limits_{\mathbb{R}} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1} - \varrho\,\frac{x_2-\mu_2}{\sigma_2}\right)^2 + (1-\varrho^2)\,\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \mathrm{d}x_1 \\
&= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\,\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\} \times \\
&\qquad \times \int\limits_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left(\frac{x_1-\left(\mu_1+\varrho\,\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)}{\sigma_1}\right)^2\right\} \mathrm{d}x_1 \\
&= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\,\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\} \int\limits_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\varrho^2}} \exp\left\{-\frac{\left(x_1-\left(\mu_1+\varrho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right)^2}{2\sigma_1^2(1-\varrho^2)}\right\} \mathrm{d}x_1
\end{aligned}$$

Der Integrand ist gerade die Dichtefunktion einer Normalverteilung mit den Parametern

$$\mu = \mu_1 + \varrho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2) \quad \text{und} \quad \sigma^2 = \sigma_1^2(1-\varrho^2)\,.$$

Deshalb ist der Wert des Integrals stets 1, und man erhält

$$f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\,\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\}\,,$$

d.h. $X_2 \sim \mathsf{N}\left(\mu_2;\sigma_2^2\right)$. Ganz analog kann man zeigen, dass die Randverteilung von X_1 ebenfalls eine Normalverteilung ist, und zwar $X_1 \sim \mathsf{N}\left(\mu_1;\sigma_1^2\right)$. □

5.5 Logarithmische Normalverteilung

Definition 5.5. Eine stetige Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **logarithmische Normalverteilung** bzw. **Log-Normalverteilung** mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, falls ihre Dichtefunktion durch

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x}\, \mathrm{e}^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}\, I_{(0,\infty)}(x)\,, \qquad x \in \mathbb{R}\,,$$

gegeben ist. Hierfür schreibt man auch

$$X \sim \mathsf{LogN}\left(\mu; \sigma^2\right)\,.$$

Satz 5.5.1
Sei $Y \sim \mathsf{N}(\mu; \sigma^2)$ eine normalverteilte Zufallsgröße. Dann gilt

$$X = \mathrm{e}^Y \sim \mathsf{LogN}\left(\mu; \sigma^2\right)\,.$$

Anwendung

- Bei der multiplikativen Überlagerung vieler (kleiner) unabhängiger Einflüsse kann die daraus resultierende Größe annähernd durch eine logarithmisch normalverteilte Zufallsgröße modelliert werden.
- Ingenieurwissenschaften (Korngrößenverteilung).
- Versicherungsmathematik (Schadenzahlverteilung, Schadenhöheverteilung).
- Finanzmathematik (Einkommensverteilung).

Bestimmungsstücke und Kenngrößen
Sei $X \sim \mathsf{LogN}(\mu; \sigma^2)$ eine mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ logarithmisch normalverteilte Zufallsgröße. Dann gilt für $x \in \mathbb{R}$

$$F_X(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) I_{(0,\infty)}(x) \quad \text{mit} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t\,,$$

$$\mathrm{E}\,(X) = \mathrm{e}^{\mu + \frac{\sigma^2}{2}}\,, \qquad \mathrm{Var}\,(X) = \left(\mathrm{e}^{\sigma^2} - 1\right) \mathrm{e}^{2\mu + \sigma^2}\,,$$

$$\gamma\,(X) = \sqrt{\mathrm{e}^{\sigma^2} - 1}\left(\mathrm{e}^{\sigma^2} + 2\right) > 0 \quad \text{und} \quad \mu_p(X) = \mathrm{e}^{p\mu + \frac{p^2\sigma^2}{2}} \qquad \text{für} \quad p \in \mathbb{N}\,.$$

Bedeutung der Parameter

- Wie man der Übersicht der Bestimmungsstücke und Kenngrößen entnehmen kann, treten die Parameter μ und σ^2 häufig gemeinsam in den Formeln für die Kenngrößen auf. Durch diese Kopplung ist es nicht möglich, eine klare Zuordnung etwa zu Lage- oder Streuungsparametern vorzunehmen.
- So wirkt der Parameter μ im Unterschied zur Normalverteilung hier nicht wie ein Lage- bzw. Verschiebungsparameter, sondern eher wie ein nichtlinearer Skalenparameter.
- Der Parameter σ^2 kann hingegen als Formparameter interpretiert werden, da er die Gestalt der Dichtefunktion wesentlich beeinflusst.

5.6 Exponentialverteilung

Definition 5.6. Eine stetige Zufallsgröße T auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Exponentialverteilung** mit dem Parameter $\lambda > 0$, falls ihre Dichtefunktion durch

$$f_T(t) = \lambda\, \mathrm{e}^{-\lambda t}\, I_{[0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

gegeben ist. Hierfür schreibt man auch

$$T \sim \mathsf{Exp}\,(\lambda)\ .$$

Bedeutung des Parameters

- λ ist ein Skalenparameter.

Bestimmungsstücke und Kenngrößen
Die Zufallsgröße T besitze eine Exponentialverteilung mit dem Parameter $\lambda > 0$. Dann gilt (vgl. Beispiele 2.17, 2.25, 2.7, 2.33, 2.51)

$$F_T(t) = \left(1 - \mathrm{e}^{-\lambda t}\right) I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

$$\mu_p(T) = \frac{p!}{\lambda^p} \quad \text{für} \quad p \in \mathbb{N}\,,$$

$$\mathrm{E}\,(T) = \frac{1}{\lambda}\,, \qquad \mathrm{Var}\,(T) = \frac{1}{\lambda^2}\,, \qquad \gamma\,(T) \equiv 2\,.$$

Satz 5.6.1
Es sei $T \sim \mathsf{Exp}\,(\lambda)$ eine mit dem Parameter $\lambda > 0$ exponentialverteilte Zufallsgröße. Dann gilt

$$\mathrm{P}(T \geqslant t_0 + t \,|\, T > t_0) = \mathrm{P}(T \geqslant t) \quad \text{für alle} \quad t_0, t > 0\,,$$

d.h. die Exponentialverteilung besitzt die Eigenschaft der Gedächtnislosigkeit.

Anmerkung 5.6.2

- Die Exponentialverteilung ist das stetige Analogon zur geometrischen Verteilung.

Anwendung

- Zuverlässigkeitstheorie (Beschreibung der Lebensdauer von nicht alternden technischen Systemen).
- Operations Research
  - Betriebsdauer zwischen Systemausfällen;
  - Zeit zwischen dem Eintreffen aufeinanderfolgender Kunden;
  - Zeit zwischen aufeinanderfolgender Telefonanrufen.
- Versicherungsmathematik (Verteilung der Schadenhöhe).

Satz 5.6.3
Seien $T_i \sim \mathsf{Exp}\,(\lambda)\,, i = 1, \ldots, d\,,$ i.i.d. exponentialverteilte Zufallsgrößen mit dem Parameter $\lambda > 0$. Dann gilt (vgl. Beispiel 3.17)

$$\min\{T_1, \ldots, T_d\} \sim \mathsf{Exp}\,(\lambda d)\ .$$

5.7 Gammaverteilung

Definition 5.7. Eine stetige Zufallsgröße T auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Gammaverteilung** mit den Parametern $\lambda > 0$ und $p > 0$, falls ihre Dichtefunktion durch

$$f_T(t) = \frac{\lambda^p}{\Gamma(p)}\, t^{p-1}\, \mathrm{e}^{-\lambda t}\, I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

gegeben ist. Hierfür schreibt man auch

$$T \sim \Gamma\,(\lambda; p)\ .$$

Definition Gammafunktion
Die **Gammafunktion** $\Gamma(x)$ ist für $x > 0$ durch das uneigentliche Integral erster Art

$$\Gamma(x) := \int\limits_0^\infty t^{x-1}\, \mathrm{e}^{-t}\ \mathrm{d}t$$

definiert. Die **unvollständige Gammafunktion der unteren Grenze** $\Gamma(x,y)$ ist gegeben durch

$$\Gamma(x,y) := \int\limits_y^\infty t^{x-1}\,\mathrm{e}^{-t}\,\mathrm{d}t \quad \text{für} \quad x>0 \quad \text{und} \quad y>0\,.$$

Eigenschaften der Gammafunktion

$$\lim_{y\downarrow 0}\Gamma(x,y) = \Gamma(x) \quad \text{für alle} \quad x>0\,.$$

$$\Gamma(x+1) = x\,\Gamma(x)\,, \quad \Gamma(1)=1 \quad \text{und} \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad \text{sowie}$$

$$\Gamma(n+1) = n! \quad \text{für} \quad n\in\mathbb{N}\,.$$

Bedeutung der Parameter

- λ ist ein Skalenparameter.
- p ist ein Struktur- bzw. Formparameter.

Bestimmungsstücke und Kenngrößen

Die Zufallsgröße T besitze eine Gammaverteilung mit den Parametern $\lambda>0$ und $p>0$. Dann gilt

$$F_T(t) = \left(1-\frac{\Gamma(p,\lambda t)}{\Gamma(p)}\right) I_{(0,\infty)}(t)\,, \qquad t\in\mathbb{R}\,,$$

$$\mathrm{E}\,(T) = \frac{p}{\lambda}\,, \qquad \mathrm{Var}\,(T) = \frac{p}{\lambda^2}\,, \qquad \gamma\,(T) = \frac{2}{\sqrt{p}} > 0\,,$$

$$\mu_q(T) = \frac{\Gamma(p+q)}{\lambda^q\,\Gamma(p)} = \frac{\prod\limits_{l=1}^{q}(p+l-1)}{\lambda^q} \quad \text{für} \quad q\in\mathbb{N}\,.$$

Anwendung

- Zuverlässigkeitstheorie (Beschreibung der Lebensdauer von technischen Systemen)

Spezialfall

Die Exponentialverteilung ist eine spezielle Gammaverteilung. Es gilt

$$\mathsf{Exp}\,(\lambda) = \Gamma\,(\lambda;1) \quad \text{für} \quad \lambda>0\,.$$

Anmerkung

Mit Hilfe der charakteristischen Funktionen kann leicht gezeigt werden:

Seien $T_1 \sim \Gamma\,(\lambda;p_1)$ und $T_2 \sim \Gamma\,(\lambda;p_2)$ unabhängige Zufallsgrößen. Dann gilt

$$T_1+T_2 \sim \Gamma\,(\lambda;p_1+p_2)\,,$$

d.h. die Familie der Gammaverteilungen ist abgeschlossen bezüglich der Faltung.

5.8 Erlang-Verteilung

Definition 5.8. Die Gammaverteilung mit den Parametern $\lambda > 0$ und $p = n \in \mathbb{N}$ wird **Erlang-Verteilung** mit den Parametern $\lambda > 0$ und $n \in \mathbb{N}$ genannt. Eine Erlang-verteilte Zufallsgröße T besitzt folglich die Dichtefunktion

$$f_T(t) = \frac{\lambda^n}{(n-1)!}\, t^{n-1}\, \mathrm{e}^{-\lambda t}\, I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,.$$

Hierfür schreibt man auch

$$T \sim \mathsf{Erl}\,(\lambda; n)\ .$$

Anmerkung
Seien $T_1 \sim \mathsf{Erl}\,(\lambda; n_1)$ und $T_2 \sim \mathsf{Erl}\,(\lambda; n_2)$ unabhängige Zufallsgrößen. Dann gilt

$$T_1 + T_2 \sim \mathsf{Erl}\,(\lambda; n_1 + n_2)\ ,$$

d.h. die Familie der Erlang-Verteilungen ist abgeschlossen bezüglich der Faltung.

Aufgabe
Seien $T_i \sim \mathsf{Exp}\,(\lambda)\,, i = 1, \ldots, n\,,$ i.i.d. exponentialverteilte Zufallsgrößen mit dem Parameter $\lambda > 0$. Zeigen Sie, dass dann gilt

$$T_1 + \ldots + T_n \sim \mathsf{Erl}\,(\lambda; n)\ ,$$

d.h. die Summe unabhängiger, exponentialverteilter Zufallsgrößen ist Erlang-verteilt.

5.9 Chi-Quadrat-Verteilung

Definition 5.9. Die Gammaverteilung mit den Parametern $\lambda = \frac{1}{2}$ und $p = \frac{n}{2}\,, n \in \mathbb{N}\,,$ wird **χ^2-Verteilung mit n Freiheitsgraden** genannt. Eine mit n Freiheitsgraden χ^2-verteilte Zufallsgröße T besitzt folglich die Dichtefunktion

$$f_T(t) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\, t^{\frac{n}{2}-1}\, \mathrm{e}^{-\frac{t}{2}}\, I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,.$$

Hierfür schreibt man auch

$$T \sim \chi^2_n\,.$$

Folgerung
Seien $T_1 \sim \chi^2_{n_1}$ und $T_2 \sim \chi^2_{n_2}$ unabhängige Zufallsgrößen. Dann gilt

$$T_1 + T_2 \sim \chi^2_{n_1+n_2}\,,$$

d.h. die Familie der χ^2-Verteilungen ist abgeschlossen bezüglich der Faltung.

Bestimmungsstücke und Kenngrößen
Die Zufallsgröße T besitze eine χ^2-Verteilung mit n Freiheitsgraden. Dann gilt

$$\mathrm{E}\,(T) = n\,, \qquad \mathrm{Var}\,(T) = 2n\,, \qquad \gamma\,(T) = \frac{2\sqrt{2}}{\sqrt{n}} > 0\,,$$

$$\mu_q(T) = \prod_{l=1}^{q}(n + 2(l-1)) \quad \text{für} \quad q \in \mathbb{N}\,.$$

Satz
Seien $X_i \sim \mathsf{N}\,(0;1)\,,\ i = 1,\ldots,n\,,$ i.i.d. standardnormalverteilte Zufallsgrößen. Dann gilt

$$X_1^2 + \ldots + X_n^2 \sim \chi_n^2\,,$$

d.h. die Summe der Quadrate unabhängiger, standardnormalverteilter Zufallsgrößen ist χ^2-verteilt mit n Freiheitsgraden.

5.10 Weibull-Verteilung

Definition 5.10. Eine stetige Zufallsgröße T auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ besitzt eine **Weibull-Verteilung** mit den Parametern $\lambda > 0$ und $p > 0$, falls ihre Dichtefunktion durch

$$f_T(t) = p\,\lambda^p\,t^{p-1}\,\mathrm{e}^{-(\lambda t)^p}\,I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

gegeben ist. Hierfür schreibt man auch

$$T \sim \mathsf{Wei}\,(\lambda; p)\ .$$

Bedeutung der Parameter

- λ ist ein Skalenparameter.
- p ist ein Struktur- bzw. Formparameter.

Bestimmungsstücke und Kenngrößen
Die Zufallsgröße T besitze eine Weibull-Verteilung mit den Parametern $\lambda > 0$ und $p > 0$. Dann gilt

$$F_T(t) = \left(1 - \mathrm{e}^{-(\lambda t)^p}\right) I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

$$\mathrm{E}\,(T) = \frac{\Gamma\left(\frac{1}{p}\right)}{\lambda\,p}\,, \qquad \mathrm{Var}\,(T) = \frac{2p\,\Gamma\left(\frac{2}{p}\right) - \Gamma\left(\frac{1}{p}\right)^2}{\lambda^2\,p^2}\,,$$

$$\mu_q(T) = \frac{\Gamma\left(\frac{p+q}{p}\right)}{\lambda^q} = \frac{q\,\Gamma\left(\frac{q}{p}\right)}{p\lambda^q} \quad \text{für} \quad q \in \mathbb{N}\,.$$

Anwendung

- Zuverlässigkeitstheorie (Beschreibung der Lebensdauer von technischen Systemen).
- Festigkeit spröder Materialien.
- Partikelgrößenverteilung in der mechanischen Verfahrenstechnik.
- Grenzverteilung geeignet normierter Minima
(„Verteilung des schwächsten Kettenglieds“).

Anmerkungen

(1) In einigen Lehrbüchern bzw. Formelsammlungen wird als Dichtefunktion der WEIBULL-Verteilung

$$f_T(t) = p\,\lambda\,t^{p-1}\,\mathrm{e}^{-\lambda t^p}\,I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

an Stelle von

$$f_T(t) = p\,\lambda^p\,t^{p-1}\,\mathrm{e}^{-(\lambda t)^p}\,I_{(0,\infty)}(t)\,, \qquad t \in \mathbb{R}\,,$$

angegeben. Dadurch erfolgt jedoch eine Kopplung der beiden Parameter λ und p, so dass λ nicht mehr als Skalenparameter interpretiert werden kann.

(2) Die WEIBULL-Verteilung wird in der Verfahrenstechnik auch als RRSB-Verteilung (nach ROSIN, RAMMLER, SPERLING und BENNET) bezeichnet.

Spezialfall
Die Exponentialverteilung ist eine spezielle WEIBULL-Verteilung. Es gilt

$$\mathsf{Exp}\,(\lambda) = \mathsf{Wei}\,(\lambda; 1) \quad \text{für} \quad \lambda > 0\,.$$

5.11 Rayleigh-Verteilung

Definition 5.11. Die WEIBULL-Verteilung mit den Parametern $\lambda = \frac{1}{\sqrt{2}\sigma}$, $\sigma > 0$, und $p = 2$ wird **Rayleigh-Verteilung** mit dem Parameter $\sigma^2 > 0$ genannt. Eine mit dem Parameter $\sigma^2 > 0$ RAYLEIGH-verteilte Zufallsgröße besitzt folglich die Dichtefunktion

$$f_T(t) = \frac{1}{\sigma^2}\,t\,\mathrm{e}^{-\frac{t^2}{2\sigma^2}}\,I_{(0,\infty)}(t)\,, \quad t \in \mathbb{R}\,.$$

Hierfür schreibt man auch

$$T \sim \mathsf{Ray}\left(\sigma^2\right)\,.$$

Bestimmungsstücke und Kenngrößen
Die Zufallsgröße T besitze eine RAYLEIGH-Verteilung mit dem Parameter $\sigma^2 > 0$. Dann gilt

$$F_T(t) = \left(1 - \mathrm{e}^{-\frac{t^2}{2\sigma^2}}\right) I_{(0,\infty)}(t)\,, \quad t \in \mathbb{R}\,,$$

$$\mathrm{E}\,(T) = \sqrt{\frac{\pi}{2}}\,\sigma\,, \qquad \mathrm{Var}\,(T) = \left(2 - \frac{\pi}{2}\right)\sigma^2\,.$$

Satz
Seien $X_1 \sim \mathsf{N}\,(0;\sigma^2)$ und $X_2 \sim \mathsf{N}\,(0;\sigma^2)$ unabhängige Zufallsgrößen. Dann gilt

$$\sqrt{X_1^2 + X_2^2} \sim \mathsf{Ray}\left(\sigma^2\right)\,.$$

6 Bedingte Verteilungen und bedingte Erwartungswerte

6.1 Zufälliges Ereignis mit positiver Wahrscheinlichkeit als Bedingung

Definition 6.1 ((und Behauptung))**.** Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $B \in \mathcal{F}$ ein zufälliges Ereignis mit $\mathrm{P}(B) > 0$.

Für $A \in \mathcal{F}$ ist $\mathrm{P}(A|B) := \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$ die **bedingte Wahrscheinlichkeit von** A **unter der Bedingung** B (auch: „**gegeben** B“).

Die Abbildung

$$\mathcal{F} \ni A \mapsto \mathrm{P}_B(A) = \mathrm{P}(A|B) \in [0;1] \subset \mathbb{R}$$

definiert ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F}, \mathrm{P})$, welches absolut stetig bezüglich P ist und die RADON-NIKODYM-Ableitung

$$\frac{\mathrm{d}\,\mathrm{P}_B}{\mathrm{d}P}(\omega) = \frac{1}{\mathrm{P}(B)} \cdot I_B(\omega)\,, \quad \omega \in \Omega\,,$$

besitzt.

Definition 6.2 ((und Behauptung))**.** Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, ein zufälliges Ereignis $B \in \mathcal{F}$ mit $\mathrm{P}(B) > 0$ und eine Zufallsvariable $X : \Omega \to \Omega'$ mit Werten im messbaren Raum $(\Omega', \mathcal{F}')$. Die Abbildung

$$\mathcal{F}' \ni A' \mapsto \mathrm{P}_{X|B}(A') := \frac{\mathrm{P}\left(\{X \in A'\} \cap B\right)}{\mathrm{P}(B)}$$

ist ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{F}')$, die **bedingte Verteilung der Zufallsvariable** X **unter der Bedingung** B bzw. kurz die **bedingte Verteilung von** X **unter** B.

Definitionen 6.3. Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, ein zufälliges Ereignis $B \in \mathcal{F}$ mit $\mathrm{P}(B) > 0$ und eine Zufallsgröße $X : \Omega \to \mathbb{R}$.

- Die Funktion
$$\mathbb{R} \ni x \mapsto F_{X|B}(x) := \mathrm{P}_{X|B}((-\infty; x)) = \frac{\mathrm{P}(\{X < x\} \cap B)}{\mathrm{P}(B)}, \quad x \in \mathbb{R},$$
ist die **bedingte Verteilungsfunktion der Zufallsgröße** X **unter der Bedingung** B bzw. kurz **bedingte Verteilungsfunktion von** X **unter** B.

- Falls existiert, ist
$$\mathrm{E}(X|B) := \int_{\mathbb{R}} x \, \mathrm{P}_{X|B}(\mathrm{d}x) \in \mathbb{R}$$
der **bedingte Erwartungswert der Zufallsgröße** X **unter der Bedingung** B.

Anmerkung 6.4. Mit den Bezeichnungen der vorigen Definitionen gilt:

Existiert $\mathrm{E}(X)$, dann existiert auch $\mathrm{E}(X|B)$ und es gilt
$$\mathrm{E}(X|B) = \frac{1}{\mathrm{P}(B)} \int_B X \, \mathrm{d}\,\mathrm{P} = \int_\Omega X \, \mathrm{d}\,\mathrm{P}_B \; .$$

Satz 6.5 ((und Definition))**.** Sei $(X, Y)^T$ ein stetiger Zufallsvektor mit gemeinsamer Dichtefunktion $f_{X,Y}$ und $B \in \mathcal{B}$ mit $\mathrm{P}(Y \in B) > 0$.

Dann ist die bedingte Verteilung $\mathrm{P}_{X|Y\in B}$ von X unter $Y \in B$ absolut stetig bezüglich $\lambda\!\!\lambda$ und die **bedingte Dichte von** X **unter** $Y \in B$ ist
$$f_{X|Y\in B}(x) = \frac{\mathrm{d}\,\mathrm{P}_{X|Y\in B}}{\mathrm{d}\lambda\!\!\lambda}(x) = \frac{1}{\mathrm{P}(Y \in B)} \int_B f_{X,Y}(x, y)\,\mathrm{d}y\,, \quad x \in \mathbb{R}.$$

6.2 Höchstens abzählbare Zerlegung als Bedingung

Anmerkung 6.6. In vielen Situationen kann (muss) man mit bedingenden Ereignissen rechnen, die aber noch nicht bekannt sind. Dann sind bedingte Wahrscheinlichkeiten und Erwartungswerte als Zufallsgrößen zu modellieren. Die wird hier zuerst für den (technisch einfacheren) Fall diskreter Zufallsgrößen behandelt.

Voraussetzung 6.7. Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und eine diskrete Zufallsgröße Y auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit Wertebereich $\{y_k\,;\, k \in J \subset \mathbb{N}\}$, so dass für alle $k \in J$ gilt: $\mathrm{P}(Y = y_k) > 0$.

Definitionen 6.8. Unter Voraussetzung 6.7 sei $A \in \mathcal{F}$ und X eine Zufallsgröße auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(|X|) < \infty$.

(i) Die **bedingte Wahrscheinlichkeit von A gegeben Y** (**bezüglich Y**) ist die Zufallsgröße $\mathrm{P}(A|Y)$ auf $(\Omega, \mathcal{F}, \mathrm{P})$, definiert durch

$$\mathrm{P}(A|Y)(\omega) = \mathrm{P}(A|\{Y = y_k\}) = \frac{\mathrm{P}(A \cap \{Y = y_k\})}{\mathrm{P}(Y = y_k)}$$

für $\omega \in \{Y = y_k\}, k \in J$.

(ii) Der **bedingte Erwartungswert von X gegeben Y** (oder **bezüglich Y**) ist die Zufallsgröße $\mathrm{E}(X|Y)$ auf $(\Omega, \mathcal{F}, \mathrm{P})$, definiert durch

$$\mathrm{E}(X|Y)(\omega) = \mathrm{E}(X|\{Y = y_k\}) = \frac{1}{\mathrm{P}(Y = y_k)} \int_{\{Y=y_k\}} X \,\mathrm{d}\,\mathrm{P}$$

für $\omega \in \{Y = y_k\}, k \in J$.

Anmerkung 6.9. Mit den Ereignissen $B_k := \{\omega \in \Omega : Y(\omega) = y_k\}, k \in J$, ist eine höchstens abzählbare Zerlegung des Grundraumes

$$\mathcal{Z} := \{B_k\,;\, k \in J\} \quad \text{mit} \quad \bigcup_{k \in J} B_k = \Omega\,;\; B_k \cap B_\ell = \varnothing\,, k, \ell \in J, k \neq \ell,$$

und die dazugehörige $\sigma-$Algebra

$$\mathcal{A} = \left\{ \bigcup_{j \in J_0} B_j\,;\, J_0 \subset J \right\} = \sigma(Y)$$

verbunden.

Definition 6.10. Unter Voraussetzung 6.7 definiert man mit den obigen Bezeichnungen die **bedingte Wahrscheinlichkeit von A bezüglich $\mathcal{Z}$** bzw. **bezüglich $\sigma(Y)$** als

$$\mathrm{P}(A|\mathcal{Z}) := \mathrm{P}(A|\mathcal{A}) := \mathrm{P}(A|\sigma(Y)) := \mathrm{P}(A|Y)\,.$$

Analog definiert man den **bedingten Erwartungswert von X bezüglich $\mathcal{Z}$** bzw. **bezüglich $\sigma(Y)$** als

$$\mathrm{E}(X|\mathcal{Z}) := \mathrm{E}(X|\mathcal{A}) := \mathrm{E}(X|\sigma(Y)) := \mathrm{E}(X|Y)\ .$$

Da dies eine Zufallsgröße ist, spricht man auch von der **bedingten Erwartung von X bezüglich $\mathcal{Z}$**.

Satz 6.11. Mit Definition 6.10 gilt für $A \in \mathcal{F}$ mit der Indikatorfunktion $I_A : \Omega \to \{0; 1\}$

$$\mathrm{P}(A|Y) = \mathrm{E}(I_A|Y)\ .$$

Anmerkung 6.12. Diese Beziehung ist die Entsprechung zu $\mathrm{P}(A) = \mathrm{E}(I_A)$ und zeigt, dass die Eigenschaften bedingter Wahrscheinlichkeiten aus denen bedingter Erwartungen abgeleitet werden können.

Der nachfolgende Satz ist der Schlüssel zur Definition bedingter Erwartungen (und Wahrscheinlichkeiten) bezüglich beliebiger Teil-σ-Algebren von $\mathcal{F}$.

Satz 6.13. Unter den Voraussetzungen zu Definition 6.10 gilt für eine Zufallsgröße X mit $\mathrm{E}(|X|) < \infty$:

(i) Die Zufallsgröße $\mathrm{E}(X|Y)$ ist eindeutig gekennzeichnet durch die beiden Bedingungen

(a) $\mathrm{E}(X|Y)$ ist messbar bezüglich $\mathcal{A} = \sigma(Y)$;

(b) für beliebige $B \in \mathcal{A} = \sigma(Y)$ gilt

$$\int_B \mathrm{E}(X|Y)\, \mathrm{d}\,\mathrm{P} = \int_B X \mathrm{d}\,\mathrm{P}\,.$$

(ii) Es existiert eine messbare Funktion $\psi : \mathbb{R} \to \mathbb{R}$, so dass gilt

$$\mathrm{E}(X|Y)(\omega) = \psi(Y(\omega))\,, \quad \omega \in \Omega\,.$$

Beispiel 6.14. Gegeben sei ein diskreter Zufallsvektor $(X,Y)^T$ mit

$$\mathrm{P}\left((X,Y)^T = (x_i, y_j)^T\right) = p_{ij}\,, \quad i = 1, \ldots, M\,; j = 1, \ldots, N\,;$$

wobei $p_{ij} > 0$ für beliebige $i = 1, \ldots, M\,; j = 1, \ldots, N$ gelte.

6.3 Teil-Sigma-Algebra als Bedingung

Anmerkung 6.15. Die Definition von bedingten Wahrscheinlichkeiten bzw. Erwartungen ist schwieriger, wenn als Bedingung $\{Y = y\}$ mit $\mathrm{P}(Y = y) = 0$ für ein $y \in \mathbb{R}$ berücksichtigt werden muss, z.B. für eine stetige Zufallsgröße Y. Dies ist notwendig und möglich mit Hilfe der Aussagen in (i) von Satz 6.13.

Beispiel 6.16. Gegeben sei eine Zufallsgröße $Y \sim U[0;1]$.
In Abhängigkeit der Realisierung y von Y werden n unabhängige Bernoulli-Experimente mit Erfolgswahrscheinlichkeit y durchgeführt, X sei die zufällige Anzahl der sich einstellenden Erfolge.
Dann sollte intuitiv gelten

$$\mathrm{P}(X = k|Y = y) = \binom{n}{k} y^k (1-y)^{n-k} \quad \text{für} \quad k = 0, \ldots, n\,.$$

Definition 6.17. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$, X und $X_{\mathcal{A}}$ Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}(|X|) < \infty$, sowie

(a) $X_{\mathcal{A}}$ ist $\mathcal{A}$-messbar und

(b) für beliebige $A \in \mathcal{A}$ gilt $\mathrm{E}(XI_A) = \mathrm{E}(X_{\mathcal{A}} I_A)$.

Dann heißt $X_{\mathcal{A}}$ **bedingte Erwartung von** X **gegeben** $\mathcal{A}$ (oder **bedingter Erwartungswert unter der Bedingung** $\mathcal{A}$, etc. und wird durch $\mathrm{E}(X|\mathcal{A})$ bezeichnet. Im Fall $\mathcal{A} = \sigma(Y)$ mit einer Zufallsvariable Y schreibt man auch $\mathrm{E}(X|Y)$.

Für $X = I_B$ mit $B \in \mathcal{F}$ heißt

$$\mathrm{E}(X|\mathcal{A}) = \mathrm{E}(I_B|\mathcal{A}) =: \mathrm{P}(B|\mathcal{A})$$

bedingte Wahrscheinlichkeit von B **unter der Bedingung** $\mathcal{A}$.

Satz 6.18 ((Korrektheit))**.** Erfüllen die Zufallsgrößen $X_{\mathcal{A}}$ und $\tilde{X}_{\mathcal{A}}$ die Bedingungen (i) und (ii) in Definition 6.17, dann gilt

$$X_{\mathcal{A}} = \tilde{X}_{\mathcal{A}} \quad \mathrm{P}-\text{f.s.}.$$

(Zufallsgrößen mit dieser Eigenschaft werden Versionen der bedingten Erwartung von X gegeben $\mathcal{A}$ genannt.)

Satz 6.19 ((Existenz))**.** Unter den Voraussetzungen von Definition 6.17 existiert $\mathrm{E}(X|\mathcal{A})$.

Satz 6.20. Sei $(\Omega, \mathcal{F}, \mathrm{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$, X_1 und X_2 Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit endlichem Erwartungswert mit $\mathrm{P}(X_1 = X_2) = 1$. Dann gilt

$$\mathrm{E}(X_1|\mathcal{A}) = \mathrm{E}(X_2|\mathcal{A}) \qquad \mathrm{P}-\text{f.s.}.$$

Satz 6.21 ((und Definition))**.** Unter den Voraussetzungen von Definition 6.17 sei $\mathcal{A} = \sigma(Y)$ für eine Zufallsgröße Y mit Verteilung P_Y. Dann existiert eine P_Y-f.s. eindeutig definierte messbare Funktion $g : \mathbb{R} \to \mathbb{R}$, so dass

$$\mathrm{E}(X|Y)(\omega) = g(Y(\omega)) \quad \mathrm{P}-\text{f.s.}$$

gilt. Eine solche Funktion wird bezeichnet mit $g(y) = \mathrm{E}(X|Y = y)$ und bedingter Erwartungswert von X unter der Bedingung $\{Y = y\}$ genannt.

Im Fall $X = I_B, B \in \mathcal{F}$, wird $g(Y) = \mathrm{E}(I_B|Y = y) = \mathrm{P}(B|Y = y)$ als bedingte Wahrscheinlichkeit von B unter der Bedingung $Y = y$ bezeichnet.

Satz 6.22 ((Faktorisierungslemma, Satz von Doob-Dynkin))**.** Seien Y, Z Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ und Z sei messbar bezüglich $\sigma(Y)$. Dann existiert eine messbare Funktion $g : \mathbb{R} \to \mathbb{R}$ mit $Z = g(Y)$ und die Funktionswerte von g sind für alle $y \in W_Y = \{Y(\omega) ; \omega \in \Omega\}$ eindeutig bestimmt.

Satz 6.23. Unter den Voraussetzungen zu Satz 6.21 wird eine Funktion $\mathbb{R} \ni y \mapsto g(y) = \mathrm{E}(X|Y = y)$ P_Y-f.s. eindeutig durch die Bedingung

$$\forall\, B \in \mathcal{B} : \int_{Y^{-1}(B)} X(\omega)\ \mathrm{P}(\mathrm{d}\omega) = \int_B g(y)\ \mathrm{P}_Y(\mathrm{d}y)$$

definiert. Im Fall $X = I_A, A \in \mathcal{F},$ lautet die Bedingung

$$\forall\, B \in \mathcal{B} : \mathrm{P}(A \cap \{Y \in B\}) = \int_B \mathrm{P}(A|Y = y)\ \mathrm{P}_Y(\mathrm{d}y).$$

Satz 6.24 ((und Definition))**.** Sei $(X,Y)^T$ ein stetiger Zufallsvektor mit gemeinsamer Dichtefunktion $f_{X,Y}$ und $h:\mathbb{R}\to\mathbb{R}$ eine messbare Funktion, so dass $\mathrm{E}\,(|h(X)|)<\infty$. Dann ist mit der **bedingten Dichte von X unter Y**

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)}\,, & \text{falls } f_Y(y) > 0\,; \\ 0\,, & \text{sonst} \end{cases}$$

die Zufallsgröße

$$\int_{\mathbb{R}} h(x) f_{X|Y}(x|Y)\,\mathrm{d}x$$

eine Version von $\mathrm{E}\,(h(X)|Y)$, der bedingten Erwartung von $h(X)$ unter Y, d.h.

$$\mathrm{E}\,(h(X)|Y=y) = \int_{\mathbb{R}} h(x) f_{X|Y}(x|y)\,\mathrm{d}x \quad \text{für } \mathrm{P}_Y\text{-fast alle } y\in\mathbb{R}\,.$$

Anmerkung 6.25. Die bedingte Dichte $f_{X|Y}(x|y)$ ergibt sich für stetige $f_{X,Y}$ durch einen Grenzprozess: für $A\in\mathcal{B}$ mit $\mathrm{P}(X\in A)>0$ und $y_0\in\mathbb{R}, h>0$, mit $\mathrm{P}(y_0\leqslant Y\leqslant y_0+h)>0$ gilt

$$\mathrm{P}(X\in A|y_0\leqslant Y\leqslant y_0+h) = \frac{\int_{y_0}^{y_0+h}\int_A f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y}{\int_{y_0}^{y_0+h}\int_{\mathbb{R}} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y} \to \frac{\int_A f_{X,Y}(x,y)\,\mathrm{d}x}{\int_{\mathbb{R}} f_{X,Y}(x,y)\,\mathrm{d}x} \quad (h\to 0)\,.$$

Satz 6.26. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},\mathrm{P})$ mit $\mathrm{E}\,(|X|)<\infty$ und $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$.

(i) $\forall\, c\in\mathbb{R}:\ \mathrm{E}\,(cX|\mathcal{A}) = c\,\mathrm{E}\,(X|\mathcal{A})$ P-f.s..

(ii) $\mathrm{E}\,(\mathrm{E}\,(X|\mathcal{A})) = \mathrm{E}\,(X)$.

(iii) Ist X $\mathcal{A}$-messbar, gilt $\mathrm{E}\,(X|\mathcal{A}) = X$ P-f.s..

(iv) Ist $\mathcal{A}_1$ eine Teil-σ-Algebra von $\mathcal{A}$ gilt die Turmeigenschaft

$$\mathrm{E}\,(\mathrm{E}\,(X|\mathcal{A})\,|\mathcal{A}_1) = \mathrm{E}\,(X|\mathcal{A}_1) \quad \mathrm{P}\text{-f.s.}.$$

(v) Ist Y eine weitere Zufallsgröße auf $(\Omega,\mathcal{F},\mathrm{P})$ mit $\mathrm{E}\,(|Y|)<\infty$, dann gilt

$$\mathrm{E}\,(X+Y|\mathcal{A}) = \mathrm{E}\,(X|\mathcal{A}) + \mathrm{E}\,(Y|\mathcal{A}) \quad \mathrm{P}\text{-f.s.}.$$

Satz 6.27. Seien $Y, X, X_i\,, i \in \mathbb{N}$, Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit endlichen Erwartungswerten sowie $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$.

(i) $Y \geqslant X \quad \mathrm{P}-\text{f.s.} \quad \Rightarrow \quad \mathrm{E}\,(Y|\mathcal{A}) \geqslant \mathrm{E}\,(X|\mathcal{A}) \quad \mathrm{P}-\text{f.s.}$

(ii) (Satz über monotone Konvergenz)

$$\forall\, i \in \mathbb{N} \,:\, X_{i+1} \geqslant X_i \quad \mathrm{P}-\text{f.s.}\,; \; X = \lim_{i\to\infty} X_i \quad \mathrm{P}-\text{f.s.}$$
$$\Rightarrow \quad \lim_{i\to\infty} \mathrm{E}\,(X_i|\mathcal{A}) = \mathrm{E}\,(X|\mathcal{A}) \qquad \mathrm{P}-\text{f.s.}$$

(iii) (Lemma von Fatou)

$$\forall\, i \in \mathbb{N} \,:\, X_i \geqslant 0 \quad \mathrm{P}-\text{f.s.}\,; \; \mathrm{E}\left(\liminf_{i\to\infty} X_i\right) < \infty$$
$$\Rightarrow \quad \mathrm{E}\left(\liminf_{i\to\infty} X_i|\mathcal{A}\right) \leqslant \liminf_{i\to\infty} \mathrm{E}\,(X_i|\mathcal{A}) \qquad \mathrm{P}-\text{f.s.}$$

Satz 6.27 ((Fortsetzung))**.** (iv) (Satz über dominierte Konvergenz)

$$\forall\, i \in \mathbb{N} \,:\, |X_i| \leqslant Y \quad \mathrm{P}-\text{f.s.}\,; \; X = \lim_{i\to\infty} X_i \quad \mathrm{P}-\text{f.s.}$$
$$\Rightarrow \quad \lim_{i\to\infty} \mathrm{E}\,(X_i|\mathcal{A}) = \mathrm{E}\,(X|\mathcal{A}) \qquad \mathrm{P}-\text{f.s.}$$

(v) (Ungleichung von Jensen)

$$g : \mathbb{R} \to \mathbb{R} \text{ konvex}, \; \mathrm{E}\,(|X|) + \mathrm{E}\,(|g(X)|) < \infty$$
$$\Rightarrow \quad \mathrm{E}\,(g(X)|\mathcal{A}) \geqslant g\,(\mathrm{E}\,(X|\mathcal{A})) \qquad \mathrm{P}-\text{f.s.}$$

Folgerung 6.28. Seien $X, X_i,\; i \in \mathbb{N}$, Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit endlichen Erwartungswerten sowie $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$.

(i) $X \geqslant 0 \quad \mathrm{P}-\text{f.s.} \quad \Rightarrow \quad \mathrm{E}\,(X|\mathcal{A}) \geqslant 0 \quad \mathrm{P}-\text{f.s.}$

(ii) Es gilt $|\mathrm{E}\,(X|\mathcal{A})| \leqslant \mathrm{E}\,\big(|X|\,\big|\mathcal{A}\big)$ $\mathrm{P}-$f.s. und allgemeiner für $p \geqslant 1$ $\mathrm{E}\,(|X|^p) < \infty \quad \Rightarrow$ $|\mathrm{E}\,(X|\mathcal{A})|^p \leqslant \mathrm{E}\,\big(|X|^p\,\big|\mathcal{A}\big)$ $\mathrm{P}-$f.s.

(iii) $\lim_{i\to\infty} \mathrm{E}\,(|X_i - X|) = 0 \quad \Rightarrow \quad \lim_{i\to\infty} \mathrm{E}\,(|\,\mathrm{E}\,(X_i|\mathcal{A}) - \mathrm{E}\,(X|\mathcal{A})\,|) = 0\,.$

Satz 6.29. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}\,(|X|) < \infty$ und $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$.

(i) Sind X und $\mathcal{A}$ unabhängig, dann gilt $\mathrm{E}\,(X|\mathcal{A}) = \mathrm{E}\,(X)$ $\mathrm{P}-$f.s. .

(ii) Für $B \in \mathcal{A}$ gilt $\mathrm{E}\,(XI_B|\mathcal{A}) = I_B\,\mathrm{E}\,(X|\mathcal{A})$ $\mathrm{P}-$f.s. .

(iii) Für eine $\mathcal{A}$-messbare Zufallsgröße Y mit $\mathrm{E}\,(|XY|) < \infty$ gilt

$$\mathrm{E}\,(XY|\mathcal{A}) = Y\,\mathrm{E}\,(X|\mathcal{A}) \qquad \mathrm{P}-\text{f.s.}\,.$$

(iv) Sei $g : \mathbb{R}^2 \to \mathbb{R}_+$ messbar, die Zufallsgröße Y sei $\mathcal{A}$-messbar, X sei unabhängig von $\mathcal{A}$ und es gelte $\mathrm{E}\,(g(X,Y)) < \infty$.

Es bezeichne für $y \in \mathbb{R}$ $\quad G(y) := \mathrm{E}\,(g(X,y))$.

Dann ist $G : \mathbb{R} \to \mathbb{R}_+$ eine messbare Funktion und es gilt

$$\mathrm{E}\,(g(X,Y)|\mathcal{A}) = G(Y) \quad \mathrm{P}-\text{f.s.}.$$

Satz 6.30. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}\,(|X|^2) < \infty$, $\mathcal{A}$ eine Teil-σ-Algebra von $\mathcal{F}$, $H = \mathrm{L}^2(\Omega, \mathcal{F}, \mathrm{P})$, $H_1 := \mathrm{L}^2(\Omega, \mathcal{A}, \mathrm{P}\,|_{\mathcal{A}})$ und $\pi_{\mathcal{A}} : H \to H_1$ der Operator der orthogonalen Projektion auf H_1.

Dann gilt

$$\mathrm{E}\,(X|\mathcal{A}) = \pi_{\mathcal{A}} X \qquad \mathrm{P}-\text{f.s.}$$

und insbesondere

$$\begin{aligned} &\mathrm{E}\left((X - \mathrm{E}\,(X|\mathcal{A}))^2\right) \\ &\quad = \inf\{\mathrm{E}\left((X-Y)^2\right) \,:\, Y \,:\, \Omega \to \mathbb{R},\ Y \text{ ist } \mathcal{A}-\text{messbar}\}. \end{aligned}$$

Satz 6.31. Sei $(X, X_1, \ldots, X_n)^T$ ein Gaussscher (d.h. normalverteilter) Zufallsvektor auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ und $\mathcal{A} = \sigma(X_1, \ldots, X_n)$.

Dann gilt

$$\mathrm{E}\,(X|\mathcal{A}) = a + b_1 X_1 + \ldots + b_n X_n \qquad \mathrm{P}-\text{f.s.},$$

wobei $(a, b_1, \ldots, b_n)$ eine (beliebige) Lösung des linearen Gleichungssystems

$$\begin{aligned} \mathrm{E}\,(X) &= a + b_1\,\mathrm{E}\,(X_1) + \ldots + b_n\,\mathrm{E}\,(X_n) \\ \mathrm{E}\,(XX_i) &= a\,\mathrm{E}\,(X_i) + b_1\,\mathrm{E}\,(X_1X_i) + \ldots + b_n\,\mathrm{E}\,(X_nX_i), \quad i = 1, \ldots, n, \end{aligned}$$

ist. Dieses lineare Gleichungssystem besitzt mindestens eine Lösung.

Folgerung 6.32. Sei $(X_1, X_2)^T$ ein normalverteilter Zufallsvektor mit Parametern $\vec{\mu} = (\mu_1, \mu_2)^T \in \mathbb{R}^2$, $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \varrho\,\sigma_1\sigma_2 \\ \varrho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ mit $\sigma_1^2\sigma_2^2 > 0$, $|\varrho| < 1$. Dann gelten für $x_1 \in \mathbb{R}, x_2 \in \mathbb{R}$

$$f_{X_2|X_1}(x_2|x_1) = \frac{1}{\sqrt{2\pi(1-\varrho^2)}\sigma_2} \exp\left\{ -\frac{\left(x_2 - \left(\mu_2 + \varrho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)\right)\right)^2}{2(1-\varrho^2)\sigma_2^2} \right\},$$

d.h. $X_2|X_1 = x_1 \sim \mathsf{N}\left(\mu_2 + \varrho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1); (1-\varrho^2)\sigma_2^2\right)$, sowie

$$\mathrm{E}\,(X_2|X_1) = a + bX_1 \quad \mathrm{P}-\text{f.s.} \quad \text{mit} \quad b = \varrho\frac{\sigma_2}{\sigma_1}, a = \mu_2 - b\mu_1,$$

d.h. $\mathrm{E}\,(X_2|X_1) \sim \mathsf{N}\,(\mu_2; \varrho^2\sigma_2^2)$, $\mathrm{E}\,((X_2 - \mathrm{E}\,(X_2|X_1))^2) = (1-\varrho^2)\sigma_2^2$.

7 Grenzwertsätze

7.1 Charakteristische Funktionen

Definition 7.1. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Die komplexwertige Funktion φ_X einer reellen Variablen mit

$$\varphi_X(t) := \mathrm{E}\left(\mathrm{e}^{\mathrm{i}tX}\right) = \int_\Omega \mathrm{e}^{\mathrm{i}tX}\,\mathrm{d}\,\mathrm{P} = \int_\Omega \cos(tX)\,\mathrm{d}\,\mathrm{P} + \mathrm{i}\int_\Omega \sin(tX)\,\mathrm{d}\,\mathrm{P}$$

für $t \in \mathbb{R}$ heißt **charakteristische Funktion der Zufallsgröße** X (bzw. **der Verteilung** P_X).

Anmerkung 7.2. $\varphi_X(t)$ existiert für beliebige $t \in \mathbb{R}$, da $|\,\mathrm{e}^{\mathrm{i}tx}\,| \leqslant 1$ gilt und die Funktion messbar ist.

Satz 7.3. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.

(i) Es gilt für $t \in \mathbb{R}$

$$\begin{aligned}\varphi_X(t) &= \int_\mathbb{R} \mathrm{e}^{\mathrm{i}tx}\ \mathrm{P}_X(\mathrm{d}x) = \int_\mathbb{R} \cos(tx)\ \mathrm{P}_X(\mathrm{d}x) + \mathrm{i}\int_\mathbb{R} \sin(tx)\ \mathrm{P}_X(\mathrm{d}x) \\ &= \int_\mathbb{R} \mathrm{e}^{\mathrm{i}tx}\ \mathrm{d}F_X(x) = \int_\mathbb{R} \cos(tx)\,\mathrm{d}F_X(x) + \mathrm{i}\int_\mathbb{R} \sin(tx)\,\mathrm{d}F_X(x).\end{aligned}$$

(ii) Für eine diskrete Zufallsgröße X mit möglichen Werten $(x_k\,; k \in J)$ und Einzelwahrscheinlichkeiten $(p_k\,; k \in J)$ gilt

$$\varphi_X(t) = \sum_{k\in J} \mathrm{e}^{\mathrm{i}tx_k}\, p_k = \sum_{k\in J} \cos(tx_k)p_k + \mathrm{i}\sum_{k\in J} \sin(tx_k)p_k\,.$$

(iii) Für eine stetige Zufallsgröße X mit Dichtefunktion f_X gilt

$$\varphi_X(t) = \int_\mathbb{R} \mathrm{e}^{\mathrm{i}tx}\, f_X(x)\,\mathrm{d}x = \int_\mathbb{R} \cos(tx) f_X(x)\,\mathrm{d}x + \mathrm{i}\int_\mathbb{R} \sin(tx) f_X(x)\,\mathrm{d}x.$$

Beispiele 7.4.

a) Besitzt die Zufallsgröße X eine Einpunktverteilung in $c \in \mathbb{R}$, dann gilt

$$\varphi_X(t) = \mathrm{e}^{\mathrm{i}tc} = \cos(tc) + \mathrm{i}\sin(tc)\,, \quad t \in \mathbb{R}\,.$$

b) Ist die Zufallsgröße X exponentialverteilt mit Parameter $\lambda > 0$, dann gilt

$$\varphi_X(t) = \frac{\lambda}{\lambda - \mathrm{i}t}\,, \quad t \in \mathbb{R}\,.$$

Satz 7.5. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit charakteristischer Funktion φ_X. Dann gelten:

(i) $\varphi_X(0) = 1$;

(ii) $|\varphi_X(t)| \leqslant 1, \quad t \in \mathbb{R}$;

(iii) $\varphi_X(-t) = \varphi_{-X}(t) = \overline{\varphi_X(t)}, \quad t \in \mathbb{R}$;

(iv) für $a, b \in \mathbb{R}$ ist $\varphi_{aX+b}(t) = \mathrm{e}^{\mathrm{i}tb}\varphi_X(at), \quad t \in \mathbb{R}$;

(v) φ_X ist gleichmäßig stetig auf $\mathbb{R}$;

(vi) für beliebige $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{R}$, $z_1, \ldots, z_n \in \mathbb{C}$ gilt

$$\sum_{k,\ell=1}^{n} \varphi_X(t_k - t_\ell) z_k \overline{z_\ell} \geqslant 0$$

(d.h., φ_X ist eine **positiv semidefinite** oder **nichtnegativ definite Funktion** auf $\mathbb{R}$).

Satz 7.6. Für unabhängige Zufallsgrößen X, Y auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit charakteristischen Funktionen φ_X, φ_Y gilt

(i) $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t), \qquad t \in \mathbb{R}$;

(ii) $\varphi_{X-Y}(t) = \varphi_X(t) \cdot \overline{\varphi_Y(t)}, \qquad t \in \mathbb{R}$;

(iii) sind die Zufallsgrößen $X_1, \ldots, X_n$ mit $n \in \mathbb{N}$ unabhängig und identisch verteilt, dann gilt für die Zufallsgröße $S_n := X_1 + \ldots + X_n$

$$\varphi_{S_n}(t) = \left(\varphi_{X_1}(t)\right)^n, \qquad t \in \mathbb{R}.$$

Satz 7.7. Die charakteristische Funktion einer Zufallsgröße bestimmt deren Verteilungsgesetz eindeutig, d.h. aus

$$\varphi_X(t) = \varphi_Y(t), \qquad t \in \mathbb{R},$$

für Zufallsgrößen X und Y folgt

$$\mathrm{P}_X = \mathrm{P}_Y .$$

(Die Zufallsgrößen dürfen prinzipiell auch auf unterschiedlichen Wahrscheinlichkeitsräumen definiert sein.)

Satz 7.8. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit charakteristischer Funktion φ_X, Verteilungsfunktion F_X und Verteilung P_X. Es gelten:

(i) $\displaystyle\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \varphi_X(t)\,\mathrm{d}t = \mathrm{P}_X\left((a;b)\right) + \frac{1}{2}\mathrm{P}_X\left(\{a,b\}\right)$;

(ii) sind a und b Stetigkeitsstellen von F_X, dann ist

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \varphi_X(t)\,\mathrm{d}t = F_X(b) - F_X(a)\,;$$

(iii) falls zusätzlich $\int_{\mathbb{R}} |\varphi_X(t)|\,\mathrm{d}t < \infty$, dann ist X eine stetige Zufallsgröße mit gleichmäßig stetiger Dichtefunktion f_X und

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx}\, \varphi_X(t)\,\mathrm{d}t\,, \qquad x \in \mathbb{R}\,.$$

Satz 7.9. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit charakteristischer Funktion φ_X. Dann ist X genau dann symmetrisch, wenn die charakteristische Funktion reellwertig ist, d.h. es gilt $\varphi_X(t) \in \mathbb{R}$ für beliebige $t \in \mathbb{R}$.

Beispiele 7.10. Für $t \in \mathbb{R}$ gilt für eine diskrete Zufallsgröße X

a) $X \sim \mathsf{U}\{x_1, \dots, x_n\} \quad\Rightarrow\quad \varphi_X(t) = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} \mathrm{e}^{\mathrm{i}tx_k}\,;$

b) $X \sim \mathsf{B}\,(n;p) \quad\Rightarrow\quad \varphi_X(t) = \left(1 - p\left(1 - \mathrm{e}^{\mathrm{i}t}\right)\right)^n;$

c) $X \sim \mathsf{Geo}\,(p) \quad\Rightarrow\quad \varphi_X(t) = \dfrac{p\,\mathrm{e}^{\mathrm{i}t}}{1 - (1-p)\,\mathrm{e}^{\mathrm{i}t}}\,;$

d) $X \sim \mathsf{Geo}_o\,(p) \quad\Rightarrow\quad \varphi_X(t) = \dfrac{p}{1 - (1-p)\,\mathrm{e}^{\mathrm{i}t}}\,;$

e) $X \sim \mathsf{NB}\,(n;p) \quad\Rightarrow\quad \varphi_X(t) = \left(\dfrac{p\,\mathrm{e}^{\mathrm{i}t}}{1 - (1-p)\,\mathrm{e}^{\mathrm{i}t}}\right)^n;$

f) $X \sim \Pi\,(\lambda) \quad\Rightarrow\quad \varphi_X(t) = \mathrm{e}^{\lambda\left(\mathrm{e}^{\mathrm{i}t} - 1\right)}\,.$

Beispiele 7.11. Für $t \in \mathbb{R}$ gilt für eine stetige Zufallsgröße X

a) $X \sim \mathsf{U}[a;b] \quad\Rightarrow\quad \varphi_X(t) = \dfrac{\mathrm{e}^{\mathrm{i}tb} - \mathrm{e}^{\mathrm{i}ta}}{(b-a)\mathrm{i}t}\,;$

b) $X \sim \mathsf{C}\,(\mu;\lambda) \quad\Rightarrow\quad \varphi_X(t) = \mathrm{e}^{\mathrm{i}\mu t - \lambda|t|}\,;$

c) $X \sim \mathsf{N}\left(\mu;\sigma^2\right) \quad\Rightarrow\quad \varphi_X(t) = \mathrm{e}^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2}\,;$

d) $X \sim \mathsf{Exp}\,(\lambda) \quad\Rightarrow\quad \varphi_X(t) = \dfrac{\lambda}{\lambda - \mathrm{i}t}\,;$

e) $X \sim \Gamma\,(\lambda;p) \quad\Rightarrow\quad \varphi_X(t) = \left(\dfrac{\lambda}{\lambda - \mathrm{i}t}\right)^p;$

f) $X \sim \mathsf{Erl}\,(\lambda; n) \quad \Rightarrow \quad \varphi_X(t) = \left(\dfrac{\lambda}{\lambda - \mathrm{i}t}\right)^n .$

Satz 7.12. Seien $X_1, \ldots, X_N$ unabhängige Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ und $S_N = X_1 + \ldots + X_N$. Dann gelten

(i) $X_i \sim \mathsf{B}\,(n_i; p) \quad \Rightarrow \quad S_N \sim \mathsf{B}\,(n_1 + \ldots + n_N; p)\,;$

(ii) $X_i \sim \mathsf{Geo}\,(p) \quad \Rightarrow \quad S_N \sim \mathsf{NB}\,(N; p)\,;$

(iii) $X_i \sim \mathsf{NB}\,(n_i; p) \quad \Rightarrow \quad S_N \sim \mathsf{NB}\,(n_1 + \ldots + n_N; p)\,;$

(iv) $X_i \sim \Pi\,(\lambda_i) \quad \Rightarrow \quad S_N \sim \Pi\,(\lambda_1 + \ldots + \lambda_N)\,;$

Satz 7.12 ((Fortsetzung))**.**

(v) $X_i \sim \mathsf{N}\,(\mu_i; \sigma_i^2) \quad \Rightarrow \quad S_N \sim \mathsf{N}\,(\mu_1 + \ldots + \mu_N; \sigma_1^2 + \ldots + \sigma_N^2)$ und

$X_1 - X_2 \sim \mathsf{N}\,(\mu_1 - \mu_2; \sigma_1^2 + \sigma_2^2)\,;$

(vi) $X_i \sim \Gamma\,(\lambda; p_i) \quad \Rightarrow \quad S_N \sim \Gamma\,(\lambda; p_1 + \ldots + p_N)\,;$

(vii) $X_i \sim \mathsf{Exp}\,(\lambda) \quad \Rightarrow \quad S_N \sim \mathsf{Erl}\,(\lambda; N)\,;$

(viii) $X_i \sim \mathsf{Erl}\,(\lambda; n_i) \quad \Rightarrow \quad S_N \sim \mathsf{Erl}\,(\lambda; n_1 + \ldots + n_N)\,;$

(ix) $X_i \sim \mathsf{C}\,(\mu_i; \lambda_i) \quad \Rightarrow \quad S_N \sim \mathsf{C}\,(\mu_1 + \ldots + \mu_N; \lambda_1 + \ldots + \lambda_N)$ und

$X_1 - X_2 \sim \mathsf{C}\,(\mu_1 - \mu_2; \lambda_1 + \lambda_2)\,.$

Satz 7.13. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$ mit charakteristischer Funktion φ_X.

(i) Falls $\mathrm{E}\,(|X|^p) < \infty$ mit $p \in \mathbb{N}$, dann ist

(a) φ_X p-fach differenzierbar auf $\mathbb{R}$ mit

$$\varphi_X^{(p)}(t) = \int_\Omega (\mathrm{i}X)^p \, \mathrm{e}^{\mathrm{i}tX} \, \mathrm{d}P = \int_{\mathbb{R}} (\mathrm{i}x)^p \, \mathrm{e}^{\mathrm{i}tx} \, \mathrm{P}_X(\mathrm{d}x)\,, \quad t \in \mathbb{R}\,;$$

(b) $\mu_p(X) = \mathrm{E}\,(X^p) = (-\mathrm{i})^p \varphi_X^{(p)}(0)\,;$

(c) $\varphi_X(t) = \displaystyle\sum_{q=0}^{p} \frac{(\mathrm{i}t)^q}{q!} \mu_q(X) + \frac{(\mathrm{i}t)^p}{p!} R_p(t)\,, \quad t \in \mathbb{R}\,,$

wobei $|R_p(t)| \leqslant 2\,\mathrm{E}\,(|X|^p)$ und $\lim\limits_{t \to 0} R_p(t) = 0$ gelten.

(ii) Falls $\varphi_X^{(2n)}(0)$ existiert für ein $n \in \mathbb{N}$, dann gilt $\mathrm{E}\,(|X|^{2n}) < \infty$.

Definition 7.14. Sei $\vec{X} = (X_1, \dots, X_n)^T$ ein Zufallsvektor auf $(\Omega, \mathcal{F}, \mathrm{P})$.
Die komplexwertige Funktion $\varphi_{\vec{X}}$ von n reellen Variablen $\vec{t} = (t_1, \dots, t_n) \in \mathbb{R}^n$ mit

$$\varphi_{\vec{X}}\left(\vec{t}\,\right) = \varphi_{\vec{X}}(t_1, \dots, t_n) := \mathrm{E}\left(\mathrm{e}^{\mathrm{i}(t_1 X_1 + \ldots + t_n X_n)}\right) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\vec{t}\vec{X}}\right)$$

heißt **charakteristische Funktion des Zufallsvektors** $\vec{X}$ (bzw. **der Verteilung** $\mathrm{P}_{\vec{X}}$).

Anmerkungen 7.15.

- Die Eigenschaften von charakteristischen Funktionen von Zufallsvektoren entsprechen zum großen Teil den Eigenschaften von charakteristischen Funktionen von Zufallsgrößen.
- Für einen Zufallsvektor $\vec{X} = (X_1, \dots, X_n)^T$ mit charakteristischer Funktion $\varphi_{\vec{X}}$ erhält man die charakteristische Funktion zum Beispiel von der ersten Komponente X_1, indem in $\varphi_{\vec{X}}$ alle Variablen außer der ersten Null gesetzt werden, d.h.

$$\varphi_{X_1}(t) = \varphi_{\vec{X}}(t, 0, \dots, 0)\,, \qquad t \in \mathbb{R}\,;$$

  entsprechende Aussagen gelten für die anderen Komponenten bzw. andere Teilvektoren.

Satz 7.16. Sei $\vec{X} = (X_1, \dots, X_n)^T$ ein Zufallsvektor auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit charakteristischer Funktion $\varphi_{\vec{X}}$.
Dann sind die Zufallsgrößen $X_1, \dots, X_n$ genau dann vollständig unabhängig, wenn gilt

$$\varphi_{\vec{X}}(t_1, t_2, \dots, t_n) = \varphi_{X_1}(t_1) \cdot \varphi_{X_2}(t_2) \cdot \ldots \cdot \varphi_{X_n}(t_n)\,.$$

Satz 7.17. Ein Zufallsvektor $\vec{X} = (X_1, \dots, X_n)^T$ auf $(\Omega, \mathcal{F}, \mathrm{P})$ ist genau dann normalverteilt mit Erwartungswertvektor $\vec{\mu} = (\mu_1, \dots, \mu_n)^T$ und Kovarianzmatrix $\mathbf{\Sigma}$, d.h. $\vec{X} \sim \mathsf{N}(\vec{\mu}; \mathbf{\Sigma})$, falls für $\vec{t} = (t_1, \dots, t_n) \in \mathbb{R}^n$ gilt

$$\varphi_{\vec{X}}(\vec{t}) = \mathrm{e}^{\mathrm{i}\,\vec{t}\vec{\mu} - \frac{1}{2}\vec{t}\,\mathbf{\Sigma}\,\vec{t}^T}\,.$$

Anmerkungen 7.18.

- Dies kann als alternative Definition von normalverteilten Zufallsvektoren genutzt werden.
- Sind die Komponenten unkorreliert, dann sind sie auch unabhängig.

7.2 Konvergenzbegriffe für Folgen von Zufallsgrößen

Definition 7.19. Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Man sagt, die Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ **konvergiert P-fast sicher** gegen X, falls

$$\mathrm{P}\left(\left\{\omega \in \Omega\colon \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

und schreibt dafür

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}\text{-f.s.}} X.$$

Anmerkung 7.20. Die P-fast sichere Konvergenz einer Folge von Zufallsgrößen ist ein Spezialfall der in der Maß- und Integrationstheorie eingeführten μ-fast überall Konvergenz einer Folge von messbaren Abbildungen für den Fall eines Wahrscheinlichkeitsmaßes $\mu = \mathrm{P}$.

Definition 7.21. Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, und es existiere $\mathrm{E}\left(X^p\right)$ bzw. $\mathrm{E}\left(X_n^p\right)$ für $1 \leqslant p < \infty$ und $n \in \mathbb{N}$. Man sagt, die Folge $(X_n)_{n\in\mathbb{N}}$ **konvergiert im p-ten Mittel** gegen X, falls

$$\lim_{n\to\infty} \int\limits_{\Omega} |X_n(\omega) - X(\omega)|^p \, \mathrm{P}(\mathrm{d}\omega) = 0$$

und schreibt dafür

$$X_n \xrightarrow[n\to\infty]{\mathcal{L}^p(\mathrm{P})} X \quad \text{bzw.} \quad X_n \xrightarrow[n\to\infty]{\mathcal{L}^p} X.$$

Anmerkung 7.22. Die Konvergenz im p-ten Mittel einer Folge von Zufallsgrößen ist ein Spezialfall der in der Maß- und Integrationstheorie eingeführten Konvergenz im p-ten Mittel einer Folge von messbaren Abbildungen für den Fall eines Wahrscheinlichkeitsmaßes $\mu = \mathrm{P}$.

Definition 7.23. Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Man sagt, die Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ **konvergiert in Wahrscheinlichkeit** oder **stochastisch** gegen X, wenn für alle $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} \mathrm{P}\left(\{\omega \in \Omega\colon |X_n(\omega) - X(\omega)| \geqslant \varepsilon\}\right) = 0$$

und schreibt dafür

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} X.$$

Anmerkung 7.24. Die Konvergenz in Wahrscheinlichkeit einer Folge von Zufallsgrößen ist ein Spezialfall der in der Maß- und Integrationstheorie eingeführten Konvergenz einer Folge von messbaren Abbildungen dem Maße nach für den Fall eines Wahrscheinlichkeitsmaßes $\mu = \mathrm{P}$.

Anmerkungen 7.25. Für die P-fast sichere Konvergenz, die Konvergenz im p-ten Mittel und die stochastische Konvergenz gelten:

- Sind zusätzlich X'_n, $n \in \mathbb{N}$, Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ und gilt $\mathrm{P}(X_n = X'_n) = 1\,, n \in \mathbb{N}$, dann folgen aus der Konvergenz der Zufallsgrößen X_n gegen X auch die entsprechende Konvergenz für $n \to \infty$ der Zufallsgrößen X'_n gegen eine Zufallsgröße X' auf $(\Omega, \mathcal{F}, \mathrm{P})$ und auch $\mathrm{P}(X' = X) = 1$.

  Insbesondere sind für diese Konvergenzarten die Grenzwerte P-fast sicher eindeutig bestimmt.

- Für diese Konvergenzarten können äquivalente Cauchy-Bedingungen formuliert werden.

- Diese Konvergenzarten sind verträglich mit der linearen Struktur, d.h. die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert genau dann gegen X, wenn die Folge der Zufallsgrößen $(X_n - X)_{n\in\mathbb{N}}$ gegen 0 konvergiert.

Definition 7.26. Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Man sagt, die Folge von Zufallsgrößen $(X_n)_{n\in\mathbb{N}}$ **konvergiert in Verteilung** gegen X bzw. die Folge von Verteilungsgesetzen $(\mathrm{P}_{X_n})_{n\in\mathbb{N}}$ **konvergiert schwach** gegen P_X, wenn für jede stetige beschränkte Funktion $g\colon \mathbb{R} \to \mathbb{R}$ gilt

$$\lim_{n\to\infty} \int_{\mathbb{R}} g\, d\,\mathrm{P}_{X_n} = \int_{\mathbb{R}} g\, d\,\mathrm{P}_X \quad \Big(\text{bzw.} \quad \lim_{n\to\infty} \mathrm{E}\left(g(X_n)\right) = \mathrm{E}\left(g(X)\right)\Big)$$

und schreibt dafür

$$X_n \xrightarrow[n\to\infty]{d} X \quad \text{bzw.} \quad \mathrm{P}_{X_n} \underset{n\to\infty}{\Longrightarrow} \mathrm{P}_X\,.$$

Anmerkung 7.27. Die Verteilungskonvergenz kann auch betrachtet werden, wenn die Zufallsgrößen $X_n\,, n \in \mathbb{N}$, auf unterschiedlichen Wahrscheinlichkeitsräumen definiert sind, da nur die Verteilungen im Bildraum wesentlich sind.

Satz 7.28. Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ genau dann in Verteilung gegen X, wenn

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

für alle Stetigkeitsstellen $x \in \mathbb{R}$ von $F_X(x)$.

Beispiel 7.29. Sei $(c_n)_{n\in\mathbb{N}}$ eine streng monoton wachsende Folge von reellen Zahlen mit

$$\lim_{n\to\infty} c_n = c \in \mathbb{R}$$

und es gelte für Zufallsgrößen X_n auf $(\Omega, \mathcal{F}, \mathrm{P})$: $P(X_n = c_n) = 1\,, n \in \mathbb{N}$.
Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ in Verteilung gegen eine Zufallsgröße X mit $\mathrm{P}(X = c) = 1$, aber für die entsprechenden Verteilungsfunktionen gilt $F_{X_n}(c) = 1\,,\; n \in \mathbb{N}$, $F_X(c) = 0$, d.h.

$$\lim_{n\to\infty} F_{X_n}(c) = 1 \neq F_X(c) = 0\,.$$

Satz 7.30 ((Stetigkeitssatz)). Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$.

(i) Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ genau dann in Verteilung gegen X, wenn gilt

$$\lim_{n\to\infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \text{für alle} \quad t \in \mathbb{R}.$$

(ii) Gilt für die Folge der charakteristischen Funktionen

$$\lim_{n\to\infty} \varphi_{X_n}(t) = \psi(t)\,, \qquad t \in \mathbb{R}\,,$$

wobei die Grenzfunktion ψ an der Stelle 0 stetig ist oder die Konvergenz gleichmäßig auf endlichen Teilintervallen von $\mathbb{R}$ erfolgt, dann ist ψ die charakteristische Funktion einer Zufallsgröße und die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen diese Zufallsgröße.

Beispiele 7.31.

(1) Satz 4.6.3 $\quad \mathsf{Hyp}\,(n, K, N) \xRightarrow[N\to\infty,\, K\to\infty,\, \frac{K}{N}\to p]{} \mathsf{B}\,(n; p).$

(2) Satz 4.9.2 $\quad \mathsf{B}\,(n; p_n) \xRightarrow[n\to\infty,\, np_n\to\lambda]{} \Pi\,(\lambda).$

Beispiel 7.32. Seien $T_i \sim \mathsf{Exp}\,(\lambda)$, $i = 1, \dots, n$, i.i.d. exponentialverteilte Zufallsgrößen mit dem Parameter $\lambda > 0$. Dann gilt

$$\min\{T_1, \dots, T_n\} \xrightarrow[n\to\infty]{d} 0$$

und

$$\mathrm{P}_{Y_n} \xRightarrow[n\to\infty]{} \mathsf{Exp}\,(\lambda) \quad \text{für} \quad Y_n = n \min\{T_1, \dots, T_n\}$$

sowie

$$\frac{1}{\ln n} \max\{T_1, \dots, T_n\} \xrightarrow[n\to\infty]{d} \frac{1}{\lambda}.$$

Aufgabe 7.33. Seien X_i, $i \in \mathbb{N}$, i.i.d. Zufallsgrößen mit der Verteilungsfunktion

$$F_X(x) = x^p I_{[0,1]}(x) + I_{(1,\infty)}(x) \quad \text{mit} \quad p > 0.$$

Zeigen Sie, dass dann gilt

$$\min\{X_1, \dots, X_n\} \xrightarrow[n\to\infty]{d} 0$$

und

$$\mathrm{P}_{Y_n} \xRightarrow[n\to\infty]{} \mathsf{Wei}\,(\lambda; p)$$

für

$$Y_n = \frac{\sqrt[p]{n}}{\lambda} \min\{X_1, \dots, X_n\} \quad \text{mit} \quad \lambda > 0.$$

Satz 7.34. Seien X, X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann folgt aus

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} X$$

stets

$$X_n \xrightarrow[n\to\infty]{d} X \quad \text{bzw.} \quad \mathrm{P}_{X_n} \underset{n\to\infty}{\Longrightarrow} \mathrm{P}_X\,.$$

Beispiel 7.35 ((Gegenbeispiel für die Umkehrung von Satz 7.34))**.** Es sei $X \sim \mathsf{B}\left(\frac{1}{2}\right)$ eine Bernoulli-verteilte Zufallsgröße und $X_n = 1 - X$ für alle $n \in \mathbb{N}$. Dann besitzen X und X_n dieselbe Verteilung und trivialerweise gilt

$$X_n \xrightarrow[n\to\infty]{d} X \quad \text{bzw.} \quad \mathrm{P}_{X_n} \underset{n\to\infty}{\Longrightarrow} \mathrm{P}_X\,.$$

Andererseits ist

$$|X(\omega) - X_n(\omega)| = 1 \quad \text{für alle} \quad \omega \in \Omega$$

und somit kann die Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsgrößen nicht in Wahrscheinlichkeit gegen X konvergieren.

Satz 7.36. Seien X_n, $n \in \mathbb{N}$, Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$, und es gelte

$$X_n \xrightarrow[n\to\infty]{d} c \quad \text{bzw.} \quad \mathrm{P}_{X_n} \underset{n\to\infty}{\Longrightarrow} \delta_c \quad \text{für ein} \quad c \in \mathbb{R}.$$

Dann folgt

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} c.$$

Übersicht: Zusammenhänge zwischen den Konvergenzbegriffen

$$\begin{tikzcd}
X_n \xrightarrow[n\to\infty]{\text{(gleichmäßig)}} X \arrow[d] & \\
X_n \xrightarrow[n\to\infty]{\text{(punktweise)}} X \arrow[d] & \\
X_n \xrightarrow[n\to\infty]{\mathrm{P}\text{-f.s.}} X \arrow[d, shift right] & X_n \xrightarrow[n\to\infty]{\mathcal{L}^p(\mathrm{P})} X \arrow[d, shift right] \\
X_n \xrightarrow[n\to\infty]{\mathrm{P}} X \arrow[u, shift right, "\text{Teilfolge}"'] \arrow[d, shift right] & X_n \xrightarrow[n\to\infty]{\mathrm{P}} X \arrow[u, shift right, "\lim\limits_{n\to\infty} \|X_n\|_p = \|X\|_p,\, 1\leqslant p<\infty"'] \\
X_n \xrightarrow[n\to\infty]{d} X \arrow[u, shift right, "X\equiv c"'] &
\end{tikzcd}$$

7.3 Gesetze der großen Zahlen

Anmerkung 7.37.

- Grenzwertsätze betreffen asymptotische Aussagen über Verteilungen einer Folge von Zufallsvariablen, hier über Partialsummen $S_n = X_1 + \ldots + X_n$ für eine Folge $(X_k)_{k\in\mathbb{N}}$ von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$.
- Man spricht von schwachen Gesetzen, wenn die stochastische Konvergenz vorliegt.
- Man spricht von starken Gesetzen, wenn die fast sichere Konvergenz vorliegt.
- Die Partialsummen S_n müssen noch geeignet normiert werden, damit eine Grenzwert (hier eine entartete Zufallsgröße, also eine reelle Konstante) existiert (z.B. wächst die Varianz mit größer werdendem n falls sie existiert und die Zufallsgrößen X_k unabhängig sind).

Satz 7.38 ((Schwaches Gesetz der großen Zahlen von Tschebyschew))**.** Sei $(X_k)_{k\in\mathbb{N}}$ eine i.i.d. Folge von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}\left(X_1^2\right) < \infty$ und für $n \in \mathbb{N}$ sei $S_n := \sum_{k=1}^{n} X_k$.

Dann gilt

$$\frac{1}{n} S_n - \mathrm{E}\left(X_1\right) \xrightarrow[n\to\infty]{\mathrm{P}} 0 \qquad \left(\text{bzw.} \quad \frac{1}{n} S_n \xrightarrow[n\to\infty]{\mathrm{P}} \mathrm{E}\left(X_1\right)\right).$$

Folgerung 7.39 ((Schwaches Gesetz der großen Zahlen von Bernoulli))**.** Sei $(X_k)_{k\in\mathbb{N}}$ eine i.i.d. Folge von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $X_k \sim \mathsf{B}(p)$ und für $n \in \mathbb{N}$ sei $S_n := \sum_{k=1}^{n} X_k$.

Dann gilt

$$\frac{1}{n} S_n \xrightarrow[n\to\infty]{\mathrm{P}} p\,.$$

Anmerkung 7.40. Diese Aussage liefert in der axiomatischen Wahrscheinlichkeitstheorie die theoretische Begründung für die Häufigkeitsinterpretation des Wahrscheinlichkeitsbegriffs.

Satz 7.41 ((Schwaches Gesetz der großen Zahlen))**.** Sei $(X_k)_{k\in\mathbb{N}}$ eine i.i.d. Folge von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}\left(|X_1|\right) < \infty$ und für $n \in \mathbb{N}$ sei $S_n := \sum_{k=1}^{n} X_k$.

Dann gilt

$$\frac{1}{n} S_n - \mathrm{E}\left(X_1\right) \xrightarrow[n\to\infty]{\mathrm{P}} 0 \qquad \left(\text{bzw.} \quad \frac{1}{n} S_n \xrightarrow[n\to\infty]{\mathrm{P}} \mathrm{E}\left(X_1\right)\right).$$

Satz 7.42 ((Starkes Gesetz der großen Zahlen von Kolmogorow)). Sei $(X_k)_{k\in\mathbb{N}}$ eine i.i.d. Folge von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ und für $n \in \mathbb{N}$ sei $S_n := \sum_{k=1}^{n} X_k$.

Dann konvergiert die Folge der Zufallsgrößen $\left(\frac{1}{n}S_n\right)_{n\in\mathbb{N}}$ genau dann fast sicher, falls $\mathrm{E}\left(|X_1|\right) < \infty$ ist und in diesem Fall gilt

$$\frac{1}{n}S_n - \mathrm{E}\left(X_1\right) \xrightarrow[n\to\infty]{\mathrm{P}\text{-f.s.}} 0 \qquad \left(\text{bzw.} \quad \frac{1}{n}S_n \xrightarrow[n\to\infty]{\mathrm{P}\text{-f.s.}} \mathrm{E}\left(X_1\right)\right).$$

7.4 Zentrale Grenzwertsätze

Anmerkung 7.43.

- Inhalt des zentralen Grenzwertsatzes ist die Verteilungskonvergenz von geeignet normierten Summen $S_n = X_1 + \ldots + X_n$ einer Folge $(X_k)_{k\in\mathbb{N}}$ von unabhängigen Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$.
- Hier wird nur der wichtigste grundlegende zentrale Grenzwertsatz behandelt, die Voraussetzungen in dem Satz können in verschiedener Hinsicht stark abgeschwächt werden.

Satz 7.44 ((Zentraler Grenzwertsatz von Lindeberg-Levy)). Sei $(X_k)_{k\in\mathbb{N}}$ eine i.i.d. Folge von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $\mathrm{E}\left(X_1^2\right) < \infty$, $\mathrm{Var}\left(X_1\right) > 0$ und für $n \in \mathbb{N}$ sei $S_n := \sum_{k=1}^{n} X_k$.

Dann gilt für die standardisierten Partialsummen $Z_n := \dfrac{S_n - n\,\mathrm{E}\left(X_1\right)}{\sqrt{n\,\mathrm{Var}\left(X_1\right)}}$

$$\mathrm{P}_{Z_n} \underset{n\to\infty}{\Longrightarrow} \mathsf{N}\left(0;1\right), \quad \text{d.h.}$$

$$\lim_{n\to\infty} \mathrm{P}(Z_n < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}}\, \mathrm{d}u, \qquad x \in \mathbb{R}.$$

Folgerung 7.45 ((Zentraler Grenzwertsatz von Moivre-Laplace)). Sei $(S_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathrm{P})$ mit $S_n \sim \mathsf{B}\left(n;p\right)$, $n \in \mathbb{N}, 0 < p < 1$.

Dann gilt für die standardisierten Partialsummen $Z_n := \dfrac{S_n - np}{\sqrt{np(1-p)}}$

$$\mathrm{P}_{Z_n} \underset{n\to\infty}{\Longrightarrow} \mathsf{N}\left(0;1\right), \quad \text{d.h.}$$

$$\lim_{n\to\infty} \mathrm{P}(Z_n < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}}\, \mathrm{d}u, \qquad x \in \mathbb{R}.$$

Beispiel 7.46. Seien $X_i \sim \mathsf{U}[-1,+1]$, $i \in \mathbb{N}$, i.i.d. Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathrm{P})$. Dann gilt

$$\mathrm{P}_{Z_n} \underset{n\to\infty}{\Longrightarrow} \mathsf{N}(0;1)$$

für

$$Z_n = \sqrt{\frac{3}{n}}\, S_n \,.$$

Satz 7.47 ((Berry-Esseen-Ungleichung))**.** Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von i.i.d. Zufallsgrößen, deren dritten Momente existieren, wobei $\operatorname{Var}(X_i) > 0$. Dann gilt

$$\sup_{x\in\mathbb{R}} |F_{Z_n}(x) - \Phi(x)| \leqslant \frac{1}{2}\, \frac{\mathrm{E}\left(|X_1 - \mathrm{E}(X_1)|^3\right)}{\sqrt{n}\operatorname{Var}(X_1)^{3/2}}$$

für

$$Z_n = \frac{S_n - n\,\mathrm{E}(X_1)}{\sqrt{n \operatorname{Var}(X_1)}} \,.$$

8 Elemente der mathematischen Statistik

8.1 Einige Grundbegriffe

Anmerkungen 8.1.

(i) In vielen praktischen Fällen sind Verteilungsfunktionen von Zufallsgrößen nicht (exakt) bekannt (nicht theoretisch berechenbar), aber es sind Daten verfügbar, die Informationen über die unbekannte Verteilungsfunktion liefern. Eine analoge Situation gilt auch für allgemeinere Zufallsvariable.

(ii) In der „klassischen“ („Fisherschen“) mathematischen Statistik geht man oft von einer Verteilungsfunktion einer Zufallsgröße mit unbekannten „exakten (wahren)“ Parametern aus.

(iii) In der Bayesschen Statistik werden alle unbekannten Größen, insbesondere auch unbekannte Parameter einer Verteilungsfunktion, als Zufallsgrößen mit einer Verteilung modelliert.

(iv) In gewisser Weise sind die Aufgabenstellungen der mathematischen Statistik invers zu den Aufgabenstellungen der Wahrscheinlichkeitstheorie.

Definition 8.2. Gegeben sei ein messbarer Raum $(\Omega, \mathcal{F})$, eine $\mathcal{F}-\mathcal{B}$-messbare Abbildung $X : \Omega \to \mathbb{R}$ und eine Familie $(\mathrm{P}_\theta)_{\theta\in\Theta}$, $\Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{F})$.

(i) Dann nennt man $(\Omega, \mathcal{F}, (\mathrm{P}_\theta)_{\theta\in\Theta})$ ein **parametrisches statistisches Modell für die Grundgesamtheit** X. Die Verteilungsfunktion von X unter dem Wahrscheinlichkeitsmaß P_θ $(\theta \in \Theta)$ werde mit $F_{X|\theta}$ bezeichnet.

(ii) Die $\mathcal{F} - \mathcal{B}(\mathbb{R}^n)$-messbare Abbildung $X^{(n)} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ heißt **mathematische Stichprobe vom Umfang n aus der Grundgesamtheit** X, falls für beliebige $\theta \in \Theta$ die Zufallsgrößen $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit der Verteilungsfunktion $F_{X|\theta}$ sind.

(iii) Für jedes $\omega \in \Omega$ wird $x^{(n)} = (x_1, \ldots, x_n) = X^{(n)}(\omega) \in \mathbb{R}^n$ **Realisierung der mathematischen Stichprobe** oder **konkrete Stichprobe** vom Umfang n genannt.

Beispiel 8.3.

- Die Grundgesamtheit X beschreibe ein Bernoulli-Experiment.
- Wir betrachten ein Bernoulli-Schema, d.h. die n-malige, unabhängige Wiederholung eines Bernoulli-Experiments. Beschreibt X_i den Ausgang des i-ten Experiments, d.h. $X_i = 1$ bei Erfolg und $X_i = 0$ bei Misserfolg, so ist das n-Tupel $(X_1, \ldots, X_n)$ eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X.
- Jede Realisierung, d.h. jedes n-Tupel aus Nullen und Einsen, z.B. $x^{(5)} = (1, 0, 1, 1, 0)$, ist eine konkrete Stichprobe.

Definition 8.4. Sei $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus einer Grundgesamtheit X auf $(\Omega, \mathcal{F}, (\mathrm{P}_\theta)_{\theta\in\Theta})$.

(i) Für eine nicht von θ abhängige messbare Funktion $\tilde{T}_n : \mathbb{R}^n \to \mathbb{R}$ nennt man die messbare Abbildung

$$T_n = \tilde{T}_n(X^{(n)}) = \tilde{T}_n(X_1, \ldots, X_n) : \Omega \to \mathbb{R}$$

eine **Stichprobenfunktion**.

(ii) Für jede Realisierung $x^{(n)} = X^{(n)}(\omega)$, $\omega \in \Omega$, ist

$$T_n(\omega) = \tilde{T}_n(X_1(\omega), \ldots, X_n(\omega)) = \tilde{T}_n(x_1, \ldots, x_n) \in \mathbb{R}$$

eine Realisierung der Stichprobenfunktion T_n.

Beispiele 8.5 (und Definitionen)**.** Sei $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus einer Grundgesamtheit X. Dann nennt man die Stichprobenfunktion

(i)

$$\overline{X}_n = \overline{X}_n\left(X^{(n)}\right) := \frac{1}{n}(X_1 + \ldots + X_n)$$

Stichprobenmittelwert,

(ii) allgemeiner für $p \in \mathbb{N}$

$$M_n^p = M_n^p\left(X^{(n)}\right) := \frac{1}{n}(X_1^p + \ldots + X_n^p)$$

p**-tes Stichprobenmoment** und

(iii)

$$S_n^2 = S_n^2\left(X^{(n)}\right) := \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Stichprobenstreuung.

Beispiele 8.5 (und Definitionen (Fortsetzung))**.**

(iv) Für jedes $x \in \mathbb{R}$ definiert man die Stichprobenfunktion

$$\widehat{F}_n(x) = \left(\widehat{F}_n(x)\right)\left(X^{(n)}\right) := \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x)}(X_i)$$

und nennt die Abbildung $\widehat{F}_n(\cdot)\colon \mathbb{R} \to [0,1]$ **empirische Verteilungsfunktion**.

Anmerkungen 8.6.

(1) Für jedes $\omega \in \Omega$ und damit jede Realisierung $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$ der mathematischen Stichprobe $X^{(n)}$ ergibt sich für jedes $x \in \mathbb{R}$ eine Realisierung

$$\begin{aligned}\left(\widehat{F}_n(x)\right)(\omega) &= \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x)}(X_i(\omega)) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x)}(x_i) \\ &= \frac{1}{n}\,\mathrm{card}\left(\{i \in \{1,\ldots,n\}\colon x_i < x\}\right)\end{aligned}$$

der Stichprobenfunktion $\widehat{F}_n(x)$.

Die Funktion $\left(\widehat{F}_n(\cdot)\right)(\omega)\colon \mathbb{R} \to [0,1]$ besitzt für jedes $\omega \in \Omega$ die Eigenschaften einer Verteilungsfunktion.

Anmerkungen 8.6 ((Fortsetzung))**.**

(2) Für alle $x \in \mathbb{R}$ gilt

$$\begin{aligned}\mathrm{E}_\theta\left(\widehat{F}_n(x)\right) &= F_{X|\theta}(x) \qquad \text{und} \\ \mathrm{Var}_\theta\left(\widehat{F}_n(x)\right) &= \frac{1}{n}F_{X|\theta}(x)(1 - F_{X|\theta}(x))\,.\end{aligned}$$

(3) Aus dem starken Gesetz der großen Zahlen folgt

$$\widehat{F}_n(x) \xrightarrow[n\to\infty]{\mathrm{P}_\theta\text{-f.s.}} F_{X|\theta}(x) \quad \text{für jedes } \ x \in \mathbb{R}.$$

Satz 8.7 ((Hauptsatz der mathematischen Statistik, Satz von Glivenko-Cantelli))**.** Sei $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus einer Grundgesamtheit X, so gilt

$$\sup_{x\in\mathbb{R}} \left| \widehat{F}_n(x) - F_{X|\theta}(x) \right| \xrightarrow[n\to\infty]{\mathrm{P}_\theta\text{-f.s.}} 0\,.$$

Anmerkungen 8.8.

(i) Wichtige Aufgabenstellungen in der mathematischen Statistik sind unter anderem

- die Bestimmung unbekannter Parameter der Verteilung einer Grundgesamtheit, d.h. die Parameterschätzung, wobei hier Punkt- oder Bereichsschätzungen genutzt werden;
- die Überprüfung, ob bestimmte Hypothesen (Annahmen) über unbekannte Parameter (oder andere Eigenschaften) angenommen oder abgelehnt werden sollten; dies wird durch statistische Tests oder Signifikanztests realisiert.

(ii) Zu beachten ist, dass im Allgemeinen keine Aussagen mit 100%-iger Sicherheit getroffen werden können. Es müssen in der Regel Fehlerwahrscheinlichkeiten mit berücksichtigt werden.

8.2 Punktschätzungen von Parametern

Definitionen 8.9.

- Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X und $g\colon \Theta \to \mathbb{R}$ eine Funktion. Unter einer **Punktschätzung** bzw. kurz einem **Schätzer** für $g(\theta)$ versteht man eine Stichprobenfunktion $\widehat{G}_n = \widehat{G}_n\left(X^{(n)}\right)$.
- Für jedes $\omega \in \Omega$ und damit jede Realisierung $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$ der mathematischen Stichprobe $X^{(n)}$ ist

  $$\widehat{g}_n := \widehat{G}_n(\omega) = \widehat{G}_n(X_1(\omega), \ldots, X_n(\omega)) = \widehat{G}_n(x_1, \ldots, x_n) \in \mathbb{R}$$

  eine Realisierung der Schätzfunktion. Diese nennt man den **Schätzwert** für $g(\theta)$ anhand der konkreten Stichprobe $(x_1, \ldots, x_n)$.

Definitionen 8.10 ((Eigenschaften von Schätzern))**.** Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X und $g\colon \Theta \to \mathbb{R}$ eine Funktion.

- Ein Schätzer $\widehat{G}_n$ für $g(\theta)$ heißt **erwartungstreu**, wenn für alle $\theta \in \Theta$ $\mathrm{E}_\theta\left(\widehat{G}_n\right)$ existiert und gilt
$$\mathrm{E}_\theta\left(\widehat{G}_n\right) := \int\limits_\Omega \widehat{G}_n \,\mathrm{d}\,\mathrm{P}_\theta = g(\theta).$$
- Die Differenz
$$\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)$$
nennt man die **Verzerrung**, den **systematischen Fehler** oder auch den **Bias** des Schätzers $\widehat{G}_n$.

Definitionen 8.10 ((Eigenschaften von Schätzern, Fortsetzung))**.**

- Sind $\widehat{G}_n$ und $\widehat{G}_n^*$ zwei erwartungstreue Schätzer für $g(\theta)$, so heißt $\widehat{G}_n^*$ **besser** als $\widehat{G}_n$, falls für alle $\theta \in \Theta$ gilt
$$\mathrm{Var}_\theta\left(\widehat{G}_n^*\right) \leqslant \mathrm{Var}_\theta\left(\widehat{G}_n\right).$$
- Falls für alle $\theta \in \Theta$ gilt
$$\lim_{n\to\infty} \mathrm{E}_\theta\left(\widehat{G}_n\right) = g(\theta),$$
so nennt man die Folge $\left(\widehat{G}_n\right)_{n\in\mathbb{N}}$ von Schätzern für $g(\theta)$ **asymptotisch erwartungstreu**.

Definitionen 8.10 ((Eigenschaften von Schätzern, Fortsetzung))**.**

- Die Folge $\left(\widehat{G}_n\right)_{n\in\mathbb{N}}$ von Schätzern für $g(\theta)$ heißt **(schwach) konsistent**, falls für alle $\theta \in \Theta$ gilt
$$\widehat{G}_n \xrightarrow[n\to\infty]{\mathrm{P}_\theta} g(\theta).$$
- Die Folge $\left(\widehat{G}_n\right)_{n\in\mathbb{N}}$ von Schätzern für $g(\theta)$ heißt **stark konsistent**, falls für alle $\theta \in \Theta$ gilt
$$\widehat{G}_n \xrightarrow[n\to\infty]{\mathrm{P}_\theta\,-\text{f.s.}} g(\theta).$$

Beispiel 8.11 (und Definition)**.** Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X, und es existiere $\mathrm{E}_\theta(|X|^p)$, $p \in \mathbb{N}$, für alle $\theta \in \Theta$. Dann ist das p-te Stichprobenmoment
$$M_n^{(p)} = \frac{1}{n}\sum_{i=1}^n X_i^p$$

ein erwartungstreuer Schätzer für

$$g(\theta) = \mathrm{E}_\theta(X^p) = \int\limits_\Omega X^p \,\mathrm{d}\,\mathrm{P}_\theta$$

und die Folge $\left(M_n^{(p)}\right)_{n\in\mathbb{N}}$ ist stark konsistent.

Beispiel 8.12. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X, und es existiere $\mathrm{E}_\theta(X^2)$ für alle $\theta \in \Theta$. Dann ist die Stichprobenstreuung

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

ein erwartungstreuer Schätzer für

$$g(\theta) = \mathrm{Var}_\theta(X) = \int\limits_\Omega \left(X - \mathrm{E}_\theta(X)\right)^2 \,\mathrm{d}\,\mathrm{P}_\theta$$

und die Folge $(S_n^2)_{n\in\mathbb{N}}$ ist stark konsistent.

Definition 8.13. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X und $g\colon \Theta \to \mathbb{R}$ eine Funktion. Existiert für einen Schätzer $\widehat{G}_n$ das zweite Moment $\mathrm{E}_\theta\left(\widehat{G}_n^2\right)$ für alle $\theta \in \Theta$, so heißt

$$\mathrm{E}_\theta\left(\left(\widehat{G}_n - g(\theta)\right)^2\right)$$

mittlerer quadratischer Fehler des Schätzers $\widehat{G}_n$ für $g(\theta)$.

Man nennt den Schätzer $\widehat{G}_n^*$ den **gleichmäßig besten Schätzer** in einer gewissen Menge $\mathcal{G}$ von Schätzern für $g(\theta)$, wenn

$$\mathrm{E}_\theta\left(\left(\widehat{G}_n^* - g(\theta)\right)^2\right) \leqslant \mathrm{E}_\theta\left(\left(\widehat{G}_n - g(\theta)\right)^2\right)$$

gilt für alle $\widehat{G}_n \in \mathcal{G}$ und alle $\theta \in \Theta$.

Satz 8.14. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X und $g\colon \Theta \to \mathbb{R}$ eine Funktion. Existiert für einen Schätzer $\widehat{G}_n$ für $g(\theta)$ das zweite Moment $\mathrm{E}_\theta\left(\widehat{G}_n^2\right)$ für alle $\theta \in \Theta$, so gilt für den mittleren quadratischen Fehler

$$\mathrm{E}_\theta\left(\left(\widehat{G}_n - g(\theta)\right)^2\right) = \mathrm{Var}_\theta\left(\widehat{G}_n\right) + \left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)^2.$$

Beweis.

$$
\begin{aligned}
&\mathrm{E}_\theta\left(\left(\widehat{G}_n - g(\theta)\right)^2\right) = \mathrm{E}_\theta\left(\left(\widehat{G}_n - \mathrm{E}_\theta\left(\widehat{G}_n\right) + \mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)^2\right)\\
&= \mathrm{E}_\theta\left(\left(\widehat{G}_n - \mathrm{E}_\theta\left(\widehat{G}_n\right)\right)^2 + 2\left(\widehat{G}_n - \mathrm{E}_\theta\left(\widehat{G}_n\right)\right)\left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right) + \left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)^2\right)\\
&= \mathrm{E}_\theta\left(\left(\widehat{G}_n - \mathrm{E}_\theta\left(\widehat{G}_n\right)\right)^2\right) + 2\left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)\mathrm{E}_\theta\left(\widehat{G}_n - \mathrm{E}_\theta\left(\widehat{G}_n\right)\right) +\\
&\quad + \mathrm{E}_\theta\left(\left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)^2\right)\\
&= \mathrm{Var}\left(\widehat{G}_n\right) + 2\left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)\underbrace{\left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - \mathrm{E}_\theta\left(\mathrm{E}_\theta\left(\widehat{G}_n\right)\right)\right)}_{=0} + \left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)^2\\
&= \mathrm{Var}\left(\widehat{G}_n\right) + \left(\mathrm{E}_\theta\left(\widehat{G}_n\right) - g(\theta)\right)^2 .
\end{aligned}
$$

□

Beispiel 8.15. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X und es existiere $\mathrm{E}_\theta(X^2)$ für alle $\theta \in \Theta$.

Für jeden Vektor $\vec{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$ mit $a_1 + \ldots + a_n = 1$ ist der Schätzer $\widehat{G}_n^{\vec{a}} = \widehat{G}_n^{\vec{a}}\left(X^{(n)}\right) = a_1X_1 + \ldots + a_nX_n$ ein erwartungstreuer Schätzer für $g(\theta) = \mathrm{E}_\theta(X)$.

Es gilt

$$\mathrm{Var}_\theta\left(\widehat{G}_n^{\vec{a}}\right) = (a_1^2 + \ldots + a_n^2)\,\mathrm{Var}_\theta(X)\,.$$

Da $a_1^2 + \ldots + a_n^2$ unter der Nebenbedingung $a_1 + \ldots + a_n = 1$ minimal wird, wenn $a_i = \frac{1}{n}$ für alle $i = 1, \ldots, n$ gilt, ist $\overline{X}_n$ der beste lineare erwartungstreue Schätzer für $g(\theta) = \mathrm{E}_\theta(X)$ (BLUE: „best linear unbiased estimator").

Satz 8.16. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X und $g\colon \Theta \to \mathbb{R}$ eine Funktion. Für eine Folge $\left(\widehat{G}_n\right)_{n\in\mathbb{N}}$ von Schätzern für $g(\theta)$ existiere das zweite Moment $\mathrm{E}_\theta\left(\widehat{G}_n^2\right)$ für alle $\theta \in \Theta$ und alle $n \in \mathbb{N}$. Ist diese Folge asymptotisch erwartungstreu und gilt

$$\lim_{n\to\infty} \mathrm{Var}_\theta\left(\widehat{G}_n\right) = 0 \quad \text{für alle} \quad \theta \in \Theta,$$

so ist sie konsistent.

Definition 8.17 ((Momentenmethode))**.** Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X und es existiere $\mathrm{E}_\theta(X^p)$, $p \in \mathbb{N}$, für alle $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$.

Hat der Parameter θ_j, die j-te Komponente von θ, $j = 1, \ldots, d$, eine Darstellung der Form

$$\theta_j = h_j(\mu_1(X), \ldots, \mu_p(X))$$

mit einer stetigen Funktion $h_j\colon \mathbb{R}^p \to \mathbb{R}$, so nennt man

$$\widehat{\theta_j}_n^{MM} = h_j\left(M_n^{(1)}, \ldots, M_n^{(p)}\right)$$

den **Momentenschätzer** bzw. **den nach der Momentenmethode konstruierten Schätzer** für θ_j, $j = 1, \ldots, d$.

Beispiel 8.18. Gegeben sei eine exponentialverteilte Grundgesamtheit $X \sim \mathsf{Exp}(\lambda)$, $\theta = \lambda$, $\Theta = (0, \infty) \subseteq \mathbb{R}$.

$$\mu_1(X) = \frac{1}{\theta}.$$

Folglich erhält man die Darstellung

$$\theta = \frac{1}{\mu_1(X)}.$$

Also ist

$$\widehat{\theta}_n^{MM} = \frac{1}{M_n^{(1)}}$$

der Momentenschätzer für θ.

Beispiel 8.19. Gegeben sei eine normalverteilte Grundgesamtheit $X \sim \mathsf{N}(\mu; \sigma^2)$, $\theta = (\theta_1, \theta_2)^T = (\mu, \sigma^2)^T$, $\Theta = \mathbb{R} \times (0, \infty) \subseteq \mathbb{R}^2$. Dann gilt

$$\mu_1(X) = \theta_1 \quad \text{und} \quad \mu_2(X) = \theta_1^2 + \theta_2.$$

Folglich erhält man die Darstellungen

$$\theta_1 = \mu_1(X) \quad \text{und} \quad \theta_2 = \mu_2(X) - \mu_1(X)^2.$$

Also sind

$$\widehat{\theta_1}_n^{MM} = M_n^{(1)} \quad \text{und} \quad \widehat{\theta_2}_n^{MM} = M_n^{(2)} - \left(M_n^{(1)}\right)^2$$

die Momentenschätzer für θ_1 und θ_2.

Satz 8.20. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X und es existiere $\mathrm{E}_\theta(X^p)$, $p \in \mathbb{N}$, für alle $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$.

Dann ist die Folge $\left(\widehat{\theta_j}_n^{MM}\right)_{n\in\mathbb{N}}$ der Momentenschätzer für den Parameter θ_j, die j-te Komponente von θ, $j = 1, \ldots, d$, stark konsistent.

Folgerung 8.21. Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X, $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X und es existiere $\mathrm{E}_\theta(X^p)$, $p \in \mathbb{N}$, für alle $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$. Weiterhin sei $g\colon \Theta \to \mathbb{R}$ eine stetige Funktion.

Dann ist die Folge $\left(\widehat{G}_n\right)_{n\in\mathbb{N}}$ mit

$$\widehat{G}_n = g\left(\widehat{\theta_1}_n^{MM}, \ldots, \widehat{\theta_d}_n^{MM}\right)$$

eine stark konsistente Folge von Schätzern für $g(\theta)$.

Definitionen 8.22 ((Maximum-Likelihood-Methode))**.** Sei $(\Omega, \mathcal{F}, \{\mathrm{P}_\theta\}_{\theta\in\Theta})$ ein parametrisches statistisches Modell für die Grundgesamtheit X mit Wertebereich W_X und $X^{(n)}$ eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit X.

- Für den Fall, dass X eine diskrete Grundgesamtheit ist, definieren wir $L_n\colon \mathbb{R}^n \times \Theta \to [0, \infty)$ gemäß
$$L_n(x_1, \ldots, x_n, \theta) := \prod_{i=1}^{n} \mathrm{P}_\theta(X = x_i).$$
- Für den Fall, dass X eine stetige Grundgesamtheit mit der Dichtefunktion f_θ ist, d.h. $f_\theta = \frac{d\left(\mathrm{P}_\theta \circ X^{-1}\right)}{d\lambda}$, definieren wir $L_n\colon \mathbb{R}^n \times \Theta \to [0, \infty)$ gemäß
$$L_n(x_1, \ldots, x_n, \theta) := \prod_{i=1}^{n} f_\theta(x_i).$$

Die Funktion L_n wird **Likelihood-Funktion** genannt.

Definitionen 8.22 ((Maximum-Likelihood-Methode, Fortsetzung))**.** Die j-te Komponente $\widehat{\theta_j}_n^{ML}$, $j = 1, \ldots, d$, des d-dimensionalen zufälligen Vektors $T_n\left(X^{(n)}\right)$ mit $T_n\colon \mathbb{R}^n \to \mathbb{R}^d$ nennt man **Maximum-Likelihood-Schätzer** für den Parameter θ_j, die j-te Komponente von θ, wenn für alle $(x_1, \ldots, x_n) \in (W_X)^n$ gilt

$$T_n(x_1, \ldots, x_n) \in \arg\max_{\theta\in\Theta} L_n(x_1, \ldots, x_n, \theta).$$

Anmerkungen 8.23 (und Definition)**.**

(1) Für jedes $\omega \in \Omega$ und damit jede Realisierung $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$ der mathematischen Stichprobe $X^{(n)}$ liefert der Vektor der Maximum-Likelihood-Schätzer denjenigen Wert $\widehat{\theta} = T_n(x_1, \ldots, x_n)$ als Schätzwert für θ, der die Realisierung genau dieser Stichprobe am wahrscheinlichsten macht.

(2) Ist die Likelihood-Funktion differenzierbar, so erfüllt der Vektor der Maximum-Likelihood-Schätzer die sogenannten **Maximum-Likelihood-Gleichungen**

$$\frac{\partial \log L_n(x_1, \ldots, x_n, \theta_1, \ldots, \theta_d)}{\partial \theta_j} = 0 \quad \text{für} \quad j = 1, \ldots, d.$$

Beispiel 8.24. Gegeben sei eine normalverteilte Grundgesamtheit $X \sim \mathsf{N}(\mu;\sigma^2)$, $\theta = (\theta_1, \theta_2)^T = (\mu, \sigma^2)^T$, $\Theta = \mathbb{R} \times (0, \infty) \subseteq \mathbb{R}^2$. Dann sind

$$\widehat{\theta_1}_n^{ML} = \overline{X}_n \quad \text{und} \quad \widehat{\theta_2}_n^{ML} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

die Maximum-Likelihood-Schätzer für θ_1 und θ_2 und stimmen mit den Momentenschätzern überein. Folglich sind

$$\left(\widehat{\theta_1}_n^{ML}\right)_{n\in\mathbb{N}} \quad \text{und} \quad \left(\widehat{\theta_2}_n^{ML}\right)_{n\in\mathbb{N}}$$

stark konsistente Folgen von Schätzern. Während $\widehat{\theta_1}_n^{ML}$ erwartungstreu ist, ist $\widehat{\theta_2}_n^{ML} = \frac{n-1}{n} S_n^2$ nur asymptotisch erwartungstreu.

Herleitung: Es ist $X \sim \mathsf{N}(\mu;\sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Folglich ist $\theta = (\theta_1, \theta_2)^T = (\mu, \sigma^2)^T$ und $\Theta = \mathbb{R} \times (0, \infty)$. Die zugehörige Dichtefunktion ist

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2}\right\}.$$

Als Likelihood-Funktion ergibt sich

$$L_n(x_1, \ldots, x_n, \theta_1, \theta_2) = \prod_{i=1}^{n} f_\theta(x_i) = (2\pi\theta_2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2\right\}$$

bzw. als Log-Likelihood-Funktion

$$\ln L_n(x_1, \ldots, x_n, \theta_1, \theta_2) = -\frac{n}{2}\ln(2\pi\theta_2) - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2.$$

Aus

$$\frac{\partial}{\partial\theta_1}\ln L_n(x_1, \ldots, x_n, \theta_1, \theta_2) = \frac{1}{\theta_2}\sum_{i=1}^{n}(x_i - \theta_1) \overset{!}{=} 0$$

ergibt sich

$$\theta_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

als Lösung und somit ist

$$\widehat{\theta_1}_n^{ML} = \overline{X}_n = M_n^{(1)}$$

der Maximum-Likelihood-Schätzer für θ_1. Aus

$$\frac{\partial}{\partial\theta_2}\ln L_n(x_1, \ldots, x_n, \theta_1, \theta_2) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i - \theta_1)^2 \overset{!}{=} 0$$

ergibt sich

$$\theta_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta_1)^2$$

als Lösung und somit ist

$$\widehat{\theta_2}_n^{ML} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = M_n^{(2)} - \left(M_n^{(1)}\right)^2 = \frac{n-1}{n}\,S_n^2$$

der Maximum-Likelihood-Schätzer für θ_2.

Beide Konstruktionsmethoden liefern also in diesem Fall dasselbe Resultat.

Die starke Konsistenz der Folgen $\left(\widehat{\theta_1}_n^{ML}\right)_{n\in\mathbb{N}}$ und $\left(\widehat{\theta_2}_n^{ML}\right)_{n\in\mathbb{N}}$ folgt unmittelbar aus Beispiel 8.11.

Ebenfalls aus Beispiel 8.11 folgt die Erwartungstreue von $\widehat{\theta_1}_n^{ML}$.

Nach Beispiel 8.12 ist die Stichprobenstreuung S_n^2 eine erwartungstreuer Schätzer für $g(\theta) = \mathrm{Var}_\theta(X) = \theta_2$. Folglich gilt

$$\mathrm{E}_\theta\left(\widehat{\theta_2}_n^{ML}\right) = \mathrm{E}_\theta\left(\frac{n-1}{n}\,S_n^2\right) = \frac{n-1}{n}\,\mathrm{E}_\theta\left(S_n^2\right) = \frac{n-1}{n}\,\theta_2$$

und $\widehat{\theta_2}_n^{ML}$ ist ein asymptotisch erwartungstreuer Schätzer für θ_2.

8.3 Bereichsschätzungen von Parametern

Anmerkung 8.25. Die Genauigkeit einer Punktschätzung kann z.B. mit Hilfe der Varianz des Schätzers bzw. des mittleren quadratischen Fehlers quantifiziert werden.

Explizitere Aussagen dazu können getroffen werden, wenn statt einer Punktschätzung eine Bereichsschätzung durchgeführt wird.

Aus mathematischer Sicht ergibt sich dabei die im Allgemeinen nicht einfach zu lösende Aufgabe, Funktionen der mathematischen Stichprobe zu bestimmen, für die eine bekannte, von weiteren unbekannten Parametern nicht abhängige Verteilung berechenbar ist.

Definition 8.26. Ein **Konfidenzintervall** (auch **Vertrauensbereich**) ist ein in Lage und/oder Breite zufälliger Bereich, der den unbekannten (reellen) Parameter θ mit Wahrscheinlichkeit $1-\alpha$ überdeckt. Der Wert $1-\alpha$ wird **Konfidenzniveau** genannt.

Anmerkungen 8.27.

- Typische Werte für α sind 0.05, 0.1 oder 0.01.
- Man unterscheidet drei Arten von Konfidenzintervallen:
  - zentral oder beidseitig begrenzt $(\hat{\theta}_{n;\alpha/2} \leqslant \theta \leqslant \hat{\theta}_{n;1-\alpha/2})$;

– einseitig, oben begrenzt $(\theta \leqslant \hat{\theta}_{n;1-\alpha})$;

– einseitig, unten begrenzt $(\hat{\theta}_{n;\alpha} \leqslant \theta)$.

Anmerkung 8.28. Für normalverteilte Grundgesamtheiten spielen sowohl bei Konfidenzintervallen, als auch bei statistischen Tests neben der χ^2-Verteilung auch die t-Verteilung, die auch Student- oder Gosset-Verteilung genannt wird, eine große Rolle.
Die Dichtefunktion einer t-verteilten Zufallsgröße mit dem Parameter $n \in \mathbb{N}$ („Anzahl der Freiheitsgrade") ist

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

Sind $X_1, \ldots, X_n$ unabhängige und identisch normalverteilte Zufallsgrößen mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, dann ist die Zufallsgröße

$$Y := \frac{\overline{X}_n - \mu}{S_n}\sqrt{n}$$

t-verteilt mit $n-1$ Freiheitsgraden.

Satz 8.29. Für eine normalverteilte Grundgesamtheit mit unbekannten Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ sind die zweiseitigen Konfidenzintervalle

$$\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1;1-\alpha/2} \leqslant \mu \leqslant \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1;1-\alpha/2} \qquad \text{bzw.}$$

$$\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\alpha/2}} \leqslant \sigma^2 \leqslant \frac{(n-1)S_n^2}{\chi^2_{n-1;\alpha/2}}.$$

Dabei sind $t_{n-1;\gamma}$ bzw. $\chi^2_{n-1;\gamma}$ jeweils γ-Quantile der t- bzw. χ^2-Verteilung mit $n-1$ Freiheitsgraden.

8.4 Statistische Tests

Anmerkung 8.30. Ein statistischer Test ist ein Verfahren zur Überprüfung einer statistischen Hypothese (eine Behauptung über Eigenschaften von Zufallsgrößen, z.B. etwa über einen Parameter der Verteilung) auf der Basis einer Stichprobe.

Das Vorgehen dazu soll an einem kleinen Beispiel erläutert werden.

Beispiel 8.31 ((Abfüllmenge Waschmittelpackungen))**.** Bei einem Verbrauchertest für Waschmittel werde auch die Abfüllmenge kontrolliert. Dabei ergaben sich bei 10 zufällig ausgewählten 5 kg Packungen einer bestimmten Sorte folgende Abfüllmengen (in kg):

$$4.6\,,\ 4.95\,,\ 4.8\,,\ 4.9\,,\ 4.75\,,\ 5.05\,,\ 4.9\,,\ 5.1\,,\ 4.85\,,\ 4.95\,.$$

Ist auf der Basis dieser Beobachtungswerte die Auffassung vertretbar, dass die Packungen im Mittel weniger Waschmittel als angegeben enthalten?

- Wir modellieren die tatsächliche Abfüllmenge (in kg) einer Waschmittelpackung als Zufallsgröße X, die als normalverteilt mit unbekannten Parametern angenommen wird.
- Berechnete Schätzwerte für den Erwartungswert, die Standardabweichung und die Varianz der Merkmalsgröße sind:

$$\overline{x} = 4.885\,, \qquad s = 0.145\,, \qquad s^2 = 0.0211\,.$$

- Zu überprüfen ist eigentlich die Richtigkeit der Vermutung, dass der Erwartungswert kleiner ist als der Sollwert $\mu_0 = 5$. Dies kann aber nicht einfach aus der Tatsache

$$\overline{x} = 4.885 < 5 = \mu_0$$

gefolgert werden, da die auftretenden zufälligen Schwankungen berücksichtigt werden müssen.
- Aus mathematischen Gründen ist die direkte Überprüfung nicht möglich, deshalb wird geprüft, ob der Erwartungswert gleich dem Sollwert $\mu_0 = 5$ ist, oder ob man im Gegensatz dazu von einem kleineren Erwartungswert ausgehen kann (muss).

Testablauf Beispiel Waschmittelpackungen I

1. Aufstellen der Hypothesen $H_0: \ \mu = 5 \ \ H_1: \ \mu < 5$
2. Festlegen des Signifikanzniveaus α $\alpha = 0.05$
3. Auswahl der Testgröße (Prüfgröße) T

$$T = \frac{\overline{X} - \mu_0}{S}\sqrt{n}\,,$$

unter H_0 besitzt diese Zufallsgröße eine $t-$Verteilung mit $n-1$ Freiheitsgraden. Im Beispiel ist $t = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n} = \dfrac{4.885 - 5}{0.145}\sqrt{10} = -2.508\,.$

Testablauf Beispiel Waschmittelpackungen II

4. Festlegung des kritischen Bereiches K

$$K = \{t \in \mathbb{R} : t < t_{n-1;\alpha}\} = (-\infty; t_{9;0.05}) = (-\infty; -t_{9,0.95})$$
$$K = (-\infty, -1.833)\,.$$

5. Entscheidung $t = -2.508$, $K = (-\infty, -1.833)$ $\Rightarrow$ $t \in K$, die Hypothese H_0 wird abgelehnt (verworfen) und die Auffassung ist vertretbar, dass die Packungen im Mittel weniger Waschmittel als angegeben enthalten.

   Man kann auch sagen: die Abfüllmengen sind signifikant zu gering.

Ausblick: Module, die auf der Grundvorlesung Stochastik aufbauen

- Stochastische Finanzmarktmodelle
  - *Teil 1*, ungerade WS, 5.Mm+5.BWM
  - *Teil 2*, gerade SS, 6.Mm+6.BWM
- Angewandte Statistik
  - *Teil 1*, ungerade WS, 5.Mm+5.BWM
  - *Teil 2*, gerade SS, 6.Mm+6.BWM
- Finanz- und Versicherungsmathematik
  (gerade SS, 6.Mm+6.BWM+2.MWM)
- Stochastische Prozesse
  - *Stochastische Prozesse*, gerade WS, 5.Mm+5.BWM
  - *Stochastische Analysis*, ungerade SS, 6.Mm+6.BWM
- Zeitreihenanalyse in den Wirtschaftswissenschaften, SS, BWM

Ausblick: Module für die weitere Vertiefung

- Stochastische Geometrie und räumliche Statistik
  - *Stochastische Geometrie*, ungerade WS, 7.Mm
  - *Räumliche Statistik*, gerade SS, 8.Mm
- Theoretische Statistik
  - *Schätz- und Testtheorie*, gerade WS, 7.Mm+1.MWM
  - *Asymptotische und algorithmische Statistik*, ungerade SS, 8.Mm+2.MWM
- Statistische Analysemethoden für Mathematiker
  - *Multivariate Statistik*, gerade WS, 7.Mm+1.MWM
  - *Zeitreihenanalyse*, ungerade SS, 8.Mm+2.MWM
- Aktuelle Themen aus der Stochastik, WS, Mm

Inhaltsverzeichnis