Albert-Ludwigs-Universität Freiburg

Vorlesungsskript der zweistündigen Grundvorlesung

Stochastik I

Angelika Rohde

Wintersemester 2024/25

Inhaltsverzeichnis

1 Wahrscheinlichkeit

1.1 Axiome von Kolmogorov

Definition 1.1 (Diskreter Wahrscheinlichkeitsraum und diskreter Maßraum). *Gegeben sei eine nicht-leere Menge Ω (genannt Grundraum oder Stichprobenraum). Eine Abbildung $\mu : \mathcal{P}(\Omega) \to [0, \infty]$ heißt diskretes Maß, falls folgende Eigenschaften erfüllt sind:*

(i) *Nicht-Negativität: $\forall\, A \subset \Omega : \mu(A) \geq 0$.*

(ii) *Nulltreue: $\mu(\varnothing) = 0$.*

(iii) *σ-Additivität: Für $A_i \subset \Omega \;\forall\, i \in \mathbb{N}$ mit $A_i \cap A_j = \varnothing \;\forall\, i, j \in \mathbb{N}, i \neq j$ gilt:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

(iv) *Diskretheit: Es gibt eine abzählbare Teilmenge $\Omega_0 \subset \Omega$ mit $\mu(\Omega_0^C) = 0$.*

Das Tripel $(\Omega, \mathcal{P}(\Omega), \mu)$ nennt man einen diskreten Maßraum. Die Abbildung μ heißt diskretes Wahrscheinlichkeitsmaß, falls $\mu(\Omega) = 1$ ist. In diesem Fall nennt man das Tripel $(\Omega, \mathcal{P}(\Omega), \mu)$ auch diskreten Wahrscheinlichkeitsraum. Für $A \subset \Omega$ heißt dann $\mu(A)$ die Wahrscheinlichkeit von A.

Bemerkung 1.2. *Ein (diskretes) Wahrscheinlichkeitsmaß wird typischerweise mit $\mathbb{P}$ bezeichnet (wobei der Buchstabe P für „Probability“ steht).*

Übungsaufgabe 1.3. *Beweisen Sie: Eine Abbildung $\mu : \mathcal{P}(\Omega) \to [0, \infty]$ mit (i), (iii) und $\mu(\Omega) = 1$ erfüllt auch die Eigenschaft (ii). N.b.: Für abzählbares $\Omega \neq \varnothing$ werden die Eigenschaften (i), (iii) und $\mu(\Omega) = 1$ auch Axiome von Kolmogorov genannt (1933).*

Bemerkung 1.4. *Die Menge Ω_0 aus Definition 1.1 ist nicht eindeutig. Jede abzählbare Menge $\Omega_0 \subset \Omega$ mit $\mu(\Omega_0^C) = 0$ heißt Träger von μ. Lässt man aus einem Träger alle Elemente ω weg, die $\mu(\{\omega\}) = 0$ erfüllen, so bleibt ein Träger mit $\mu(\{\omega\}) > 0 \;\forall\, \omega \in \Omega_0$ übrig. Dieses Ω_0 nennt man dann Träger im strengen Sinne.*

Konvention 1.5. *Ist in einer endlichen oder unendlichen Summe von Zahlen aus $[0, \infty]$ mindestens ein Summand ∞, so wird die Summe gleich ∞ gesetzt.*

Notation 1.6. *Sind $A_i \in \mathcal{P}(\Omega)$ für $i \in \mathbb{N}$ paarweise disjunkte Mengen (d.h. $A_i \cap A_j = \varnothing \;\forall\, i, j \in \mathbb{N}, i \neq j$), so schreibt man auch:*

$$\sum_{i=1}^{\infty} A_i \quad \text{für} \quad \bigcup_{i=1}^{\infty} A_i$$

Lemma 1.7 (Regel von De Morgan). *Seien $\Omega \neq \emptyset$ eine beliebige Grundmenge und I eine beliebige Indexmenge. Für alle $i \in I$ sei $A_i \in \mathcal{P}(\Omega)$. Dann gilt:*

$$\left(\bigcap_{i\in I} A_i\right)^C = \bigcup_{i\in I} A_i^C \text{ und } \left(\bigcup_{i\in I} A_i\right)^C = \bigcap_{i\in I} A_i^C$$

Beweis. Übungsblatt 1, Aufgabe 2. □

Lemma 1.8. *Gegeben seien ein Grundraum $\Omega \neq \emptyset$, ein diskretes Maß $\mu : \mathcal{P}(\Omega) \to [0,\infty]$ und ein diskretes Wahrscheinlichkeitsmaß $\mathbb{P} : \mathcal{P}(\Omega) \to [0,\infty]$. Dann gelten folgende Aussagen:*

(i) *Endliche Additivität: Ist $n \in \mathbb{N}$ gegeben und sind $A_1, \ldots, A_n$ paarweise disjunkte Teilmengen von Ω (diese bezeichnet man gerne als „Ereignisse“), so gilt:*

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

(ii) *Isotonie: Ist $A \subset B$ für $A, B \in \mathcal{P}(\Omega)$, so ist $\mu(A) \leq \mu(B)$. Speziell folgt hieraus für das Wahrscheinlichkeitsmaß $\mathbb{P}$, dass $\forall\, A \in \mathcal{P}(\Omega)$ gilt, dass $0 \leq \mathbb{P}(A) \leq 1$ ist.*

(iii) *Subtraktivität: Ist $\mu(A) < \infty$ für ein $A \in \mathcal{P}(\Omega)$, so gilt die folgende Implikation $\forall\, B \in \mathcal{P}(\Omega)$:*

$$A \subset B \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$$

Wegen $\mathbb{P}(\Omega) = 1$ gilt diese Implikation für Wahrscheinlichkeitsmaße immer.

(iv) *Komplementarität: Ist $\mu(A) < \infty$ für ein $A \in \mathcal{P}(\Omega)$, so folgt, dass $\mu(A^C) = \mu(\Omega) - \mu(A)$ ist. Insbesondere ist $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.*

(v) *Stetigkeit von unten: Ist $(A_i)_{i\in\mathbb{N}}$ eine aufsteigende Folge in $\mathcal{P}(\Omega)$ (d.h. $\forall\, i \in \mathbb{N}$ gilt $A_i \subset A_{i+1}$), so folgt:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mu(A_n)$$

(vi) *Stetigkeit von oben: Ist $(A_i)_{i\in\mathbb{N}}$ eine absteigende Folge in $\mathcal{P}(\Omega)$ (d.h. $\forall\, i \in \mathbb{N}$ gilt $A_i \supset A_{i+1}$), und ist $\mu(A_1) < \infty$, so folgt:*

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mu(A_n)$$

(*vii*) *<u>Subadditivität</u>: Für $n \in \mathbb{N}$ und $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$ gilt immer:*

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

(*viii*) *<u>σ-Subadditivität</u>: Ist $(A_i)_{i\in\mathbb{N}}$ eine Folge in $\mathcal{P}(\Omega)$, so gilt:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

Beweis. Wir weisen die Eigenschaften $(i) - (iv)$ nach. Die Eigenschaften $(v) - (viii)$ sind Gegenstand von Aufgabe 3 auf Übungsblatt 1.

(*i*) Mit $A_i := \varnothing \ \forall\, i > n$ gilt

$$\mu\left(\sum_{i=1}^{n} A_i\right) = \mu\left(\sum_{i=1}^{\infty} A_i\right) \overset{\substack{\sigma\text{-Additivität}\\ \downarrow}}{=} \sum_{i=1}^{\infty} \mu(A_i) \overset{\substack{\mu(\varnothing)=0\\ \downarrow}}{=} \sum_{i=1}^{n} \mu(A_i).$$

(*ii*) Sei $B = A \cup (B \setminus A)$. Es folgt aus (i), dass $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$ ist (denn $\mu(B \setminus A) \geq 0$).

(*iii*) $\mu(B) \overset{(ii)}{=\!=} \mu(A) + \mu(B \setminus A) \xRightarrow{\mu(A)<\infty} \mu(B \setminus A) = \mu(B) - \mu(A)$

(*iv*) folgt aus (iii) mit $B = \Omega$.

□

Diskrete Maße lassen sich leicht mithilfe von Zähldichten konstruieren.

Definition 1.9 (Zähldichte)**.** *Sei $\Omega \neq \varnothing$ ein Grundraum.*

(*i*) *Eine Abbildung $f : \Omega \to [0, \infty]$ heißt <u>Zähldichte</u>, wenn $T = \{\omega \in \Omega \mid f(\omega) > 0\}$ abzählbar ist. Gemäß*

$$\mu(A) := \sum_{\omega \in A\cap T} f(\omega) \text{ mit } \sum_{\omega\in\varnothing} f(\omega) := 0 \qquad (*)$$

wird offenbar ein diskretes Maß auf $(\Omega, \mathcal{P}(\Omega))$ definiert, wobei $\Omega_0 = T$ gesetzt werden kann. Gilt

$$\sum_{\omega\in T} f(\omega) = 1,$$

so ist das durch $(*)$ definierte Maß ein diskretes Wahrscheinlichkeitsmaß.

(*ii*) *Ist umgekehrt μ ein diskretes Maß auf $(\Omega, \mathcal{P}(\Omega))$, so ist die durch*

$$f : \Omega \to [0, \infty], \omega \mapsto \mu(\{\omega\})$$

definerte Funktion eine Zähldichte.

Beispiel 1.10. (*i*) *Seien $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $f : \Omega \to [0, \infty], \omega \mapsto \frac{1}{|\Omega|} = \frac{1}{6}$. Das zu f assoziierte Wahrscheinlichkeitsmaß $\mathbb{P}$ heißt Laplace-Verteilung auf $(\Omega, \mathcal{P}(\Omega))$. $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ist ein Modell für einen einmaligen Würfelwurf.*

(*ii*) *Betrachten wir nun einen einmaligen Münzwurf. Dafür verwenden wir die Grundmenge $\Omega = \{\text{Kopf}, \text{Zahl}\}$, versehen mit der Zähldichte $f(\text{Kopf}) = \frac{1}{2}$, $f(\text{Zahl}) = \frac{1}{2}$. Das zugehörigeWahrscheinlichkeitsmaß ist ebenfalls eine Laplace-Verteilung auf $(\Omega, \mathcal{P}(\Omega))$.*

1.2 Urnenmodelle und abgeleitete Verteilungen

Wir betrachten n Ziehungen aus einer Urne mit N Kugeln, die mit den Zahlen $1, \ldots, N$ durchnummeriert seien.

Beispiel 1.11 (6 aus 49)**.** *Wieviele Möglichkeiten gibt es, aus 49 Zahlen 6 (unterschiedliche) Zahlen auszuwählen? Für die erste Zahl gibt es 49 Möglichkeiten; für die zweite dann nur noch 48, da eine Zahl dann ja bereits gewählt wurde; für die dritte gibt es nur noch 47 Möglichkeiten, usw.; wenn wir diese Anzahl noch durch die Anzahl der möglichen Permutationen eine 6-elementigen Menge teilen, ergibt sich schlussendlich folgende Formel:*

$$\frac{49 \cdot 48 \cdot \ldots \cdot 44}{6 \cdot \ldots \cdot 1} = \binom{49}{6} = 13.983.816$$

Dies ist genau die Anzahl aller 6-elementigen Teilmengen von $\{1, \ldots, 49\}$. Allgemeiner ist $\binom{n}{k}$ genau die Anzahl aller k-elementigen Teilmengen der Menge $\{1, \ldots, n\}$. Dieses Beispiel entspricht dem Ziehen ohne Reihenfolge und ohne Zurücklegen von 6 Kugeln aus einer Urne mit (ursprünglich) 49 Kugeln.

Satz 1.12 (Kombinatorik)**.** *Wir betrachten das Ziehen einer Stichprobe vom Umfang n aus einer N-elementigen Menge. Es gilt:*

	Mit Zurücklegen	Ohne Zurücklegen
Mit Reihenfolge	N^n	$N \cdot (N-1) \cdot \ldots \cdot (N-n+1) = \frac{N!}{(N-n)!}$
Ohne Reihenfolge	$\binom{N+n-1}{n}$	$\binom{N}{n}$

Ohne Beweis.

Beispiel 1.13 (Ziehen mit Zurücklegen, Binomialverteilung)**.** *Wir betrachten eine Urne mit N Kugeln; R davon seien rot, und $N-R$ weiß. Aus dieser Urne ziehen wir n-mal hintereinander eine Kugel, wobei wir nach jedem Zug die jeweils gezogene Kugel zurücklegen. Die Kugeln seien von $1,\dots,N$ durchnummeriert – o.B.d.A. seien die ersten R Kugeln rot. Wir betrachten dazu den Stichprobenraum*

$$\Omega = \big\{(\omega_1,\dots,\omega_n) \mid 1 \leq \omega_i \leq N\big\}$$

(ω_1 bezeichnet dabei die Nummer der ersten gezogenen Kugel, ω_2 die der zweiten, usw.) mit Laplace-Verteilung, d.h. die Wahrscheinlichkeit jedes n-Tupels ist gleich.

Frage: *Wie groß ist die Wahrscheinlichkeit, dass genau r rote Kugeln in der Stichprobe sind? Dies entspricht dem Ereignis $E_r = \{(\omega_1,\dots,\omega_n) : |\{i : \omega_i \leq R\}| = r\}$.*

Um die Mächtigkeit von E_r zu berechnen, schreiben wir E_r als Vereinigung disjunkter Ereignisse E_I, wobei $I \subseteq \{1,\dots,n\}$ die Nummern der Ziehungen enthält, bei denen genau eine rote Kugel gezogen wird. Die Idee ist, dass wir von diesen Ereignissen die Wahrscheinlichkeit leichter ermitteln können. Wir setzen also

$$E_I = \{(\omega_1,\dots,\omega_n) \mid \omega_i \in \{1,\dots,R\} \text{ für } i \in I \text{ und } \omega_i \in \{R+1,\dots,N\} \text{ für } i \notin I\}$$

und

$$E_r = \bigcup_{\substack{I\subseteq\{1,\dots,n\}\\ |I|=r}} E_I.$$

Da es genau $\binom{n}{r}$ Teilmengen $I \subseteq \{1,\dots,n\}$ mit Kardinalität r gibt, und da E_I für alle solchen I dieselbe Kardinalität hat, folgt:

$$|E_r| = \binom{n}{r} \cdot |E_{\{1,\dots,r\}}| = \binom{n}{r} \cdot R^r \cdot (N-R)^{n-r}$$

Mit der Laplace-Verteilung und der Tatsache, dass $|\Omega| = N^n$ ist, ergibt sich:

$$\mathbb{P}(E_r) = \frac{|E_r|}{|\Omega|} = \binom{n}{r} \cdot \left(\frac{R}{N}\right)^r \cdot \left(1 - \frac{R}{N}\right)^{n-r}$$

Da $E_0,\dots,E_n$ eine disjunkte Zerlegung des Ergebnisraumes Ω ist (denn irgendeine Anzahl roter Kugeln muss man ja gezogen haben, und diese Anzahl liegt bei n Kugeln zwischen 0 und n), wird durch

$$p(r) := \mathbb{P}(E_r)\ \forall\, r \in \{0,\dots,n\}$$

eine Zahldichte eines Wahrscheinlichkeitsmaßes auf $(\{0,\dots,n\}, \mathcal{P}(\{0,\dots,n\}))$ definiert, *denn es gilt*

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\sum_{i=0}^{n} E_i\right) = \sum_{i=0}^{n} \mathbb{P}(E_i) = \sum_{i=0}^{n} p(i).$$

Das Wahrscheinlichkeitsmaß heißt Binomialverteilung (mit Parametern n und $\frac{R}{N}$).

2 Unabhängigkeit, Produkträume, bedingte Wahrscheinlichkeit

2.1 Stochastische Unabhängigkeit

Definition 2.1 (Stochastische Unabhängigkeit)**.** *Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Dann heißen zwei Ereignisse A und B stochastisch unabhängig, wenn gilt:*

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

Beispiel 2.2. (*i*) *Wir werfen einen unverfälschten Würfel und definieren folgende Ereignisse:*

- *A: „Die Augenzahl ist gerade“.*
- *B: „Die Augenzahl ist durch 3 teilbar“.*

Nach obiger Definition sind diese Ereignisse stochastisch unabhängig, denn es gilt:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2} \ ; \ \mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{1}{3} \ ; \ \mathbb{P}(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$$

(*ii*) *Wir ziehen zwei mal mit Zurücklegen aus einer Urne mit 3 roten und 5 weißen Kugeln, und betrachten folgende Ereignisse:*

- *A: „Die erste gezogene Kugel ist rot“.*
- *B: „Die zweite gezogene Kugel ist weiß“.*

Diese Ereignisse sind stochastisch unabhängig, denn es gilt:

$$\mathbb{P}(A) = \frac{3 \cdot 8}{8^2} = \frac{3}{8} \ ; \ \mathbb{P}(B) = \frac{8 \cdot 5}{8^2} = \frac{5}{8} \ ; \ \mathbb{P}(B) = \frac{15}{8^2} = \frac{3}{8} \cdot \frac{5}{8}$$

Definition 2.3. *Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Die Ereignisse $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$ heißen stochastisch unabhängig, wenn $\forall\, k \in \{1, \ldots, n\}$ und jede Auswahl von Indizes $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ gilt:*

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

Bemerkung. *Es folgt sofort, dass Teilfamilien von unabhängigen Ereignissen wieder unabhängig sind.*

Beispiel 2.4. *Wir werfen eine faire Münze zweimal. Wir betrachten also den folgenden Grundraum:*

$$\Omega = \{(K,K),(K,Z),(Z,K),(Z,Z)\}.$$

Setze:

- $A := \{(K,K),(K,Z)\} =$ *„Der erste Wurf ergibt Kopf"* $\rightsquigarrow \mathbb{P}(A) = 1/2$
- $B := \{(K,K),(Z,K)\} =$ *„Der zweite Wurf ergibt Kopf"* $\rightsquigarrow \mathbb{P}(B) = 1/2$
- $A := \{(K,Z),(Z,K)\} =$ *„Es wird genau einmal Kopf geworfen"* $\rightsquigarrow \mathbb{P}(C) = 1/2.$

Wir fragen uns, ob diese Ereignisse unabhängig sind. Dazu berechnen wir:

- $\mathbb{P}(A \cap B) = 1/4 = \mathbb{P}(A) \cdot \mathbb{P}(B) \rightsquigarrow A$ *und* B *sind unabhängig.*
- $\mathbb{P}(B \cap C) = 1/4 = \mathbb{P}(B) \cdot \mathbb{P}(C) \rightsquigarrow B$ *und* C *sind unabhängig.*
- $\mathbb{P}(A \cap C) = 1/4 = \mathbb{P}(A) \cdot \mathbb{P}(C) \rightsquigarrow A$ *und* C *sind unabhängig.*

Allerdings gilt: $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\varnothing) = 0 \neq \frac{1}{8} = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$. *Also sind zwar die Ereignisse* A, B *und* C *paarweise unabhängig, nicht aber das Gesamtsystem. Dies zeigt uns sofort, dass es nicht ausreicht, paarweise Unabhängigkeit zu überprüfen, um die Unabhängigkeit eines Systems nachzuweisen.*

Satz 2.5. *Sei* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum. Seien die Ereignisse* $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$ *stochastisch unabhängig. Dann sind auch die Ereignisse* $B_1, \ldots, B_n$ *mit* $B_i \in \{A_i, A_i^C\}\ \forall\, i \in \{1, \ldots, n\}$ *unabhängig.*

Beweis. Wir beweisen die Aussage mittels vollständiger Induktion nach n.

Induktionsanfang für $n = 2$*:* Es gilt:

$$A_1 = (A_1 \cap A_2) + (A_1 \cap A_2^C)$$

Dabei bezeichnet „+" eine disjunkte Vereinigung von Mengen.

$$\Rightarrow \mathbb{P}(A_1) = \underbrace{\mathbb{P}(A_1 \cap A_2)}_{=\mathbb{P}(A_1)\cdot\mathbb{P}(A_2)} + \mathbb{P}(A_1 \cap A_2^C)$$

$$\Rightarrow \mathbb{P}(A_1 \cap A_2^C) = \mathbb{P}(A_1) \cdot (1 - \mathbb{P}(A_2)) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2^C)$$

Damit sind die Ereignisse A_1 und A_2^C stochastisch unabhängig. Analog für A_1^C und A_2 bzw. A_1^C und A_2^C.

Induktionsschritt: Die Behauptung sei für $n-1$ richtig. Seien $A_1, \ldots, A_n$ stochastisch unabhängig. Sei $\{i_1, \ldots, i_m\} \subset \{1, \ldots, n\}$ mit $m := n-1$. Nach Induktionsvoraussetzung sind dann die Ereignisse $B_{i_1}, \ldots, B_{i_m}$ unabhängig (da dies $n-1$ Ereignisse sind). Es bleibt folgende Gleichung zu zeigen:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} \mathbb{P}(B_i) \tag{$*$}$$

Es seien o.B.d.A. $B_1 = A_1^C, \ldots, B_k = A_k^C, B_{k+1} = A_{k+1}, \ldots, B_n = A_n$. Wir zeigen die Aussage $(*)$ per Induktion über k.

Induktionsanfang für $k = 1$: Mit endlicher Additivität erhalten wir:

$$\mathbb{P}\left(A_1^C \cap \bigcap_{i=2}^{n} A_i\right) + \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcap_{i=2}^{n} A_i\right)$$

$$\Rightarrow \mathbb{P}\left(A_1^C \cap \bigcap_{i=2}^{n} A_i\right) = \left(\prod_{i=2}^{n} \mathbb{P}(A_i)\right) \cdot (1 - \mathbb{P}(A_1)) = \mathbb{P}(A_1^C) \cdot \prod_{i=2}^{n} \mathbb{P}(A_i)$$

Induktionsschritt von $k-1$ nach k: Es gilt:

$$\begin{aligned}
&\mathbb{P}\left(\left(\bigcap_{i=1}^{k} A_i^C\right) \cap \left(\bigcap_{i=k+1}^{n} A_i\right)\right) + \mathbb{P}\left(\left(\bigcap_{i=1}^{k-1} A_i^C\right) \cap \left(\bigcap_{i=k}^{n} A_i\right)\right) \\
&\qquad = \mathbb{P}\left(\left(\bigcap_{i=1}^{k-1} A_i^C\right) \cap \left(\bigcap_{i=k+1}^{n} A_i\right)\right) \overset{\text{I.V.}}{=\!=} \left(\prod_{i=1}^{k-1} \mathbb{P}(A_i^C)\right) \cdot \left(\prod_{i=k+1}^{n} \mathbb{P}(A_i)\right)
\end{aligned}$$

Damit gilt analog zum Induktionsanfang:

$$\mathbb{P}\left(\left(\bigcap_{i=1}^{k} A_i^C\right) \cap \left(\bigcap_{i=k+1}^{n} A_i\right)\right) = \left(\prod_{i=1}^{k} \mathbb{P}(A_i^C)\right) \cdot \left(\prod_{i=k+1}^{n} \mathbb{P}(A_i)\right)$$

Insgesamt ergibt sich damit die Aussage $(*)$. $\square$

2.2 Produkträume und Produktexperimente

Motivation 2.6. *Zwei Studenten – Student 1 in Freiburg, Student 2 in Konstanz – führen gleichzeitig ein Experiment aus: Student 1 wirft eine (faire) Münze, Student 2 wirft einen (fairen) Würfel. Beide haben ihre eigenen Wahrscheinlichkeitsräume, $\Omega_1 = \{K, Z\}$ mit Laplace-Verteilung $\mathbb{P}_1$ auf $\mathcal{P}(\Omega_1)$ und $\Omega_2 = \{1, \ldots, 6\}$ mit Laplace-Verteilung $\mathbb{P}_2$ auf $\mathcal{P}(\Omega_2)$. Was die Wahrscheinlichkeit des Ereignisses A "Student 1 wirft Kopf, Student 2 erhält eine "6"" anbelangt, besteht das Problem , dass das Ereignis weder in $\mathbb{P}(\Omega_1)$ noch in $\mathbb{P}(\Omega_2)$ enthalten ist.*

Definition 2.7. *Es seien* $(\Omega_1, \mathcal{P}(\Omega_1), \mathbb{P}_1), \ldots, (\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)$ *diskrete Wahrscheinlichkeitsräume. Der* *Produktraum* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *ist der diskrete Wahrscheinlichkeitsraum mit*

$$\Omega := \Omega_1 \times \ldots \times \Omega_n = \{(\omega_1, \ldots, \omega_n) \mid \omega_i \in \Omega_i \ \forall\, i \in \{1, \ldots, n\}\}$$

und der Zähldichte

$$p((\omega_1, \ldots, \omega_n)) := \underbrace{p_1(\omega_1)}_{=\mathbb{P}_1(\{\omega_1\})} \cdot \ldots \cdot \underbrace{p_n(\omega_n)}_{=\mathbb{P}_n(\{\omega_n\})}$$

Das zu p gehörende Wahrscheinlichkeitsmaß $\mathbb{P}$ *auf* $(\Omega, \mathcal{P}(\Omega))$ *heißt* *Produkt-Wahrscheinlichkeitsmaß* *und wird oft mit* $\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n$ *bezeichnet.*

Satz 2.8. *Es seien* $(\Omega_1, \mathcal{P}(\Omega_1), \mathbb{P}_1), \ldots, (\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)$ *diskrete Wahrscheinlichkeitsräume und* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *der zugehörige Produktraum. Für* $i \in \{1, \ldots, n\}$ *seien* $A_i \subset \Omega$ *Ereignisse und* $A'_i := \{(\omega_1, \ldots, \omega_n) \in \Omega \mid \omega_i \in A_i\} = \Omega_1 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_n$. *Dann gilt* $\mathbb{P}(A'_i) = \mathbb{P}_i(A_i)$ *für alle* $i \in \{1, \ldots, n\}$ *und die Ereignisse* $A'_1, \ldots, A'_n$ *sind stochastisch unabhängig.*

Beweis. Es ist

$$\begin{aligned}
\mathbb{P}(A'_i) = \sum_{\omega \in A'_i} p(\omega) &= \sum_{\substack{(\omega_1, \ldots, \omega_n) \in \Omega \\ \omega_i \in A_i}} \prod_{i=1}^{n} p_i(\omega_i) \\
&= \underbrace{\left(\sum_{\omega_1 \in \Omega_1} p_1(\omega_1) \right)}_{=1} \cdot \ldots \cdot \left(\sum_{\omega_i \in A_i} p_i(\omega_i) \right) \cdot \ldots \cdot \underbrace{\left(\sum_{\omega_n \in \Omega_n} p_n(\omega_n) \right)}_{=1} \\
&= \sum_{\omega_i \in A_i} p_i(\omega_i) = \mathbb{P}_i(A_i),
\end{aligned}$$

und für die Indizes $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ gilt

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j=1}^{k} A'_{i_j} \right) &= \sum_{\substack{(\omega_1, \ldots, \omega_n) \in \Omega \\ \omega_{i_j} \in A_{i_j} \ \forall\, j \in \{1, \ldots, k\}}} \prod_{j=1}^{n} p_j(\omega_j) \\
&= \sum_{\substack{(\omega_{i_1}, \ldots, \omega_{i_k}) \in \Omega \\ \omega_{i_j} \in A_{i_j} \ \forall\, j \in \{1, \ldots, k\}}} \prod_{j=1}^{k} p_{i_j}(\omega_{i_j}) \\
&= \left(\sum_{\omega_{i_1} \in A_{i_1}} p_{i_1}(\omega_{i_1}) \right) \cdot \ldots \cdot \left(\sum_{\omega_{i_k} \in A_{i_k}} p_{i_k}(\omega_{i_k}) \right) = \prod_{j=1}^{k} \mathbb{P}_{i_j}(A_{i_j}) = \prod_{j=1}^{k} \mathbb{P}(A'_{i_j})
\end{aligned}$$

Damit sind die Ereignisse $A'_1, \ldots, A'_n$ unabhängig. $\square$

Bemerkung. *Mit der Identität* $\bigcap_{i=1}^n A_i' = A_1 \times \ldots \times A_n$ *bedeutet Satz 2.8, dass auf dem Produkt-Wahrscheinlichkeitsraum gilt:*

$$\mathbb{P}(A_1 \times \ldots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i).$$

Diese Identität ist Ausgangspunkt einer alternativen Definition des Produkt-Wahrscheinlichkeitsmaßes $\rightarrow$ *Maßtheorie (Analysis III).*

Beispiel 2.9 (Noch einmal die Studenten aus Freiburg and Konstanz)**.** *Mit der Notation aus Motivation 2.6, definiere* $\Omega := \Omega_1 \times \Omega_2$ *sowie das Produktmaß* $\mathbb{P} := \mathbb{P}_1 \otimes \mathbb{P}_2$ *auf* $\mathcal{P}(\Omega)$ *mit Komponenten* $\mathbb{P}_1$ *und* $\mathbb{P}_2$*. Das Ereignis* $\{(K, 6)\} =$ *"Student 1 wirft Kopf, Student 2 erhält eine "6""ist eine einelementige Teilmenge von* Ω *und*

$$\begin{aligned}
\mathbb{P}(\{K, 6\}) &= \mathbb{P}\Big(\big(\{K\} \times \Omega_2\big) \cap \big(\Omega_1 \times \{6\}\big)\Big) \\
&= \mathbb{P}(\{K\} \times \Omega_2) \cdot \mathbb{P}(\Omega_1 \times \{6\}) \\
&= \mathbb{P}_1(\{K\}) \cdot \mathbb{P}_2(\{6\}) = \frac{1}{2} \cdot \frac{1}{6}.
\end{aligned}$$

D.h., unter dem Produktmaß sind die Ereignisse in Experiment 1 unabhängig von denen in Experiment 2.

2.3 Bedingte Wahrscheinlichkeiten

Ziel: Beschreibung von Abhängigkeitseigenschaften von Ereignissen. Was ist die Wahrscheinlichkeit von einem Ereignis A, wenn wir wissen, dass ein anderes Ereignis B eingetreten ist?

Definition 2.10. *Sei* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum. Weiter seien* $A, B \in \mathcal{P}(\Omega)$ *Ereignisse mit* $\mathbb{P}(A) > 0$*. Dann wird die bedingte Wahrscheinlichkeit von* B *gegeben* A *durch*

$$\mathbb{P}(B \mid A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

definiert.

Beispiel 2.11. *Wir betrachten den zweimaligen Würfelwurf sowie die Ereignisse*

- $\Omega := \{1, \ldots, 6\}^2$
- $A := \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = 12\}$
- $B := \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 = 5\}$

Damit gilt:

$$\mathbb{P}(A) = \frac{1}{36} \;;\; \mathbb{P}(B) = \frac{1}{6} \;;\; \mathbb{P}(A \mid B) = \frac{P(\overbrace{A \cap B}^{=\varnothing})}{P(B)} = 0$$

Bemerkung 2.12. *Für unabhängige Ereignisse $A, B \in \mathcal{P}(\Omega)$ mit $\mathbb{P}(A) > 0$ gilt*

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

Anders gesagt: Sind die Ereignisse A und B unabhängig, so ändert das Wissen, dass das Ereignis A eingetreten ist, nichts an der Wahrscheinlichkeit, dass auch das Ereigniss B eintritt. Umgekehrt: Ist $\mathbb{P}(B \mid A) = \mathbb{P}(B)$, so folgt

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

d.h. A und B sind unabhängig.

Satz 2.13 (Satz von der totalen Wahrscheinlichkeit)**.** *Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Seien $B_1, \ldots, B_n \in \mathcal{P}(\Omega)$ mit folgenden Eigenschaften:*

- $\bigcup\limits_{i=1}^{n} B_i = \Omega$
- $B_i \cap B_j = \varnothing \;\forall\, i, j \in \{1, \ldots, n\}, i \neq j$
- $\mathbb{P}(B_j) > 0 \;\forall\, j \in \{1, \ldots, n\}.$

Eine solche Zerlegung des Raumes Ω wird auch Partition von Ω genannt. Dann gilt für $A \in \mathcal{P}(\Omega)$:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)$$

Beweis. Da auch die Mengen $A \cap B_i$, $i = 1, \ldots, n$, paarweise disjunkt sind, folgt

$$\begin{aligned} \mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(A \cap \bigcup_{j=1}^{n} B_j\right) &= \mathbb{P}\left(\bigcup_{j=1}^{n} (A \cap B_j)\right) \\ &= \sum_{j=1}^{n} \mathbb{P}(A \cap B_j) = \sum_{j=1}^{n} \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j) \end{aligned}$$

□

Satz 2.14 (Formel von Bayes)**.** *Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Sei $B_1, \ldots, B_n \in \mathcal{P}(\Omega)$ eine Partition von Ω wie oben, und sei $A \in \mathcal{P}(\Omega)$. Dann gilt $\forall\, i \in \{1, \ldots, n\}$:*

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j)}$$

Beweis. Es gilt $\forall\, i \in \{1, \ldots, n\}$:

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\mathbb{P}(A)} \overset{3.13}{=\!=} \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j)}.$$

□

Beispiel 2.15. *Gegeben seien 6 Urnen, die jeweils 5 Kugeln enthalten. Die erste Urne enthält 5 weiße Kugeln, die zweite Urne enthält 4 weiße und eine schwarze Kugel, usw. bis zur 6. Urne, die 5 schwarze Kugeln enthält:*

Urne 1	○ ○ ○ ○ ○
Urne 2	● ○ ○ ○ ○
Urne 3	● ● ○ ○ ○
Urne 4	● ● ● ○ ○
Urne 5	● ● ● ● ○
Urne 6	● ● ● ● ●

Experiment: Wir werfen zunächst einen unverfälschten Würfel und wählen diejenige Urne aus, die der geworfenen Augenzahl entspricht. Anschließend ziehen wir aus der ausgewählten Urne mit Zurücklegen Kugeln. Wir definieren folgende Ereignisse:

- $B_i :=$ *„Es wird aus der i-ten Urne gezogen“.*
- $A_1 :=$ *„Die erste gezogene Kugel ist schwarz“.*
- $A_2 :=$ *„Die zweite gezogene Kugel ist schwarz“.*

Ziel ist nun, die Wahrscheinlichkeiten $\mathbb{P}(A_1)$ und $\mathbb{P}(A_2 \mid A_1)$ zu ermitteln. Dafür halten wir zunächst fest, dass die Ereignisse $B_1, \ldots, B_6$ eine disjunkte Zerlegung des Ergebnisraumes bilden. Ferner gilt $\forall\, i \in \{1, \ldots, 6\}$:

$$\mathbb{P}(A_1 \mid B_i) = \frac{i-1}{5}$$

Damit gilt nach dem Satz der totalen Wahrscheinlichkeit:

$$\mathbb{P}(A_1) = \sum_{i=1}^{6} \underbrace{\mathbb{P}(A_1 \mid B_i)}_{=\frac{i-1}{5}} \cdot \underbrace{\mathbb{P}(B_i)}_{=\frac{1}{6}} = \frac{1}{30} \cdot \sum_{i=0}^{5} i = \frac{1}{2}$$

Bemerkung: Diese Aussage war bereits klar aus Symmetriegründen.
Außerdem gilt:

$$\mathbb{P}(A_2 \mid A_1) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} = 2\mathbb{P}(A_1 \cap A_2)$$

Nun ist

$$\mathbb{P}(A_1 \cap A_2 \mid B_i) = \left(\frac{i-1}{5}\right)^2$$

also gilt:

$$\mathbb{P}(A_1 \cap A_2) = \sum_{i=1}^{6} \underbrace{\mathbb{P}(A_1 \cap A_2 \mid B_i)}_{=\left(\frac{i-1}{5}\right)^2} \cdot \underbrace{\mathbb{P}(B_i)}_{=\frac{1}{6}} = \frac{1}{150} \cdot \sum_{i=0}^{5} i^2 = \frac{11}{30}$$

$$\Rightarrow \mathbb{P}(A_2 \mid A_1) = \frac{11}{15}.$$

3 Diskrete Zufallsvariablen

3.1 Abbildungen und Bildmaße

Beispiel 3.1. *Wir betrachten das dreimalige Werfen einer Münze mit dem Grundraum*

$$\Omega := \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{0,1\} \ \forall\, i \in \{1,2,3\}\}.$$

Wir nutzen die Laplace-Verteilung $\mathbb{P}$ *auf* $(\Omega, \mathcal{P}(\Omega))$*, d.h.* $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{8}$ *für* $A \in \mathcal{P}(\Omega)$*.*
Frage: Wie groß ist die Wahrscheinlichkeit, dass k-mal Zahl geworfen wird?
Dafür gibt es zwei mögliche Herangehensweisen:

(i) *Wir definieren das Ereignis*

$$A_k := \left\{(\omega_1, \omega_2, \omega_3) \in \Omega \;\middle|\; \sum_{i=1}^{3} \omega_i = k\right\} \quad \textit{und berechnen } \mathbb{P}(A_k) = \frac{|A_k|}{8}.$$

(ii) *Die Anzahl ist binomialverteilt, d.h. wir betrachten* $Bin(3, \frac{1}{2})$*. Diese Verteilung ist allerdings kein Wahrscheinlichkeitsmaß auf* $(\Omega, \mathcal{P}(\Omega))$*. Wir müssen also einen anderen Grundraum und ein anderes Wahrscheinlichkeitsmaß verwenden:*

$$\Omega^X = \{0,1,2,3\} \text{ und } \mathbb{P}^X(\{k\}) := \binom{3}{k}\left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{3-k} = \binom{3}{k}\frac{1}{8}, \; k = 0,1,2,3.$$

Den Raum $(\Omega^X, \mathcal{P}(\Omega^X), \mathbb{P}^X)$ *erhält man aus* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *mit folgender Abbildung:*

$$X : \Omega \to \mathbb{R}, \quad \omega = (\omega_1, \omega_2, \omega_3) \mapsto \sum_{i=1}^{3} \omega_i.$$

Dann ist $\Omega^X = \mathrm{Bild}(X) = \{x \in \mathbb{R} \mid \exists\, \omega \in \Omega : x = X(\omega)\}$*, und es gilt* $\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B))$ *für jede Menge* $B \in \mathcal{P}(\mathrm{Bild}(X))$*.*

Definition 3.2. *Sei* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum. Eine Abbildung* $X : \Omega \to D$ *mit* $D \subset \mathbb{R}$ *heißt* *diskrete Zufallsvariable. Die durch* X *induzierte Verteilung*

$$\mathbb{P}^X(A) := \mathbb{P}(X^{-1}(A)) \; \forall \, A \subset D$$

ist ein diskretes Wahrscheinlichkeitsmaß auf $(D, \mathcal{P}(D))$. *Die Funktion*

$$F_X(x) = \mathbb{P}^X((-\infty, x]) = \mathbb{P}(\underbrace{X^{-1}((-\infty, x])}_{=\{\omega | X(\omega) \leq x\}}) =: \mathbb{P}(X \leq x)$$

heißt *Verteilungsfunktion von* X.

$$\begin{tikzcd}
(\Omega, \mathcal{P}(\Omega), \mathbb{P}) \arrow[rr, bend left, "X"] & & (D, \mathcal{P}(D), \mathbb{P}^X)
\end{tikzcd}$$

Bemerkung 3.3. *(i) Man rechnet leicht nach, dass* $\mathbb{P}^X$ *tatsächlich (wie in obiger Definition behauptet) ein diskretes Wahrscheinlichkeitsmaß ist (da die Axiome von Kolmogorov gelten).*

(ii) Man rechnet ebenfalls leicht nach, dass F_X *monoton wachsend und rechtsseitig stetig ist, und folgende Limites hat:*

$$\lim_{x \to -\infty} F_X(x) = 0 \text{ und } \lim_{x \to \infty} F_X(x) = 1.$$

(iii) Da der Träger von $\mathbb{P}^X$ *in* Ω^X *enthalten ist und somit* $\Omega^X \subset D$ *gilt, kann man* $\mathbb{P}^X$ *gleichermaßen auf* $(\Omega^X, \mathcal{P}(\Omega^X))$ *definieren. Eine andere typische Wahl ist* $D = \mathbb{R}$.

Die Poisson-Verteilung

In der Praxis stellt sich mitunter die Frage, explizite Zahlenwerte der Verteilungsfunktion F_X auszurechnen. Dies kann bspw. bei der Binomialverteilung für große Werte von n sehr aufwendig sein, da die Binomialkoeffizienten für große n sehr groß werden können.

Satz 3.4 (Poisson-Grenzwertsatz). *Es seien* $X_n \sim Bin(n, p_n)$ *für* $n \in \mathbb{N}$ *diskrete Wahrscheinlichkeitsräume. Falls ein* $\lambda \in (0, \infty)$ *existiert, sodass* $np_n \xrightarrow{n \to \infty} \lambda$ *gilt, so gilt*

$$\lim_{n \to \infty} \underbrace{\mathbb{P}(X_n = k)}_{= \binom{n}{k} p_n^k (1 - p_n)^{n-k}} = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \forall \, k \in \mathbb{N}_0.$$

Die Abbildung $p_\lambda : \mathbb{N}_0 \to \mathbb{R}, \; k \mapsto e^{-\lambda} \frac{\lambda^k}{k!}$, *definiert die Zähldichte eines Wahrscheinlichkeitsmaßes auf* $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$. *Dieses heißt* *Poisson-Verteilung.*

Beweis. Für festes $k \in \mathbb{N}_0$ ist

$$\begin{aligned}
\mathbb{P}(X_n = k) &= \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} \\
&= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} \cdot \frac{1}{n^k} \cdot (np_n)^k \cdot \left(1 - \frac{np_n}{n}\right)^n \cdot (1-p_n)^{-k} \\
&= \underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-k+1}{n}}_{\xrightarrow{n\to\infty} 1} \cdot \underbrace{\frac{1}{(1-p_n)^k}}_{\xrightarrow{n\to\infty} 1} \cdot \underbrace{\frac{(np_n)^k}{k!}}_{\xrightarrow{n\to\infty} \frac{\lambda^k}{k!}} \cdot \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\xrightarrow{n\to\infty} e^{-\lambda}}
\end{aligned}$$

Nach Voraussetzung gilt $np_n \xrightarrow{n\to\infty} \lambda$, und damit $p_n \xrightarrow{n\to\infty} 0$. Wegen

$$\lim_{n\to\infty} \left(1 + \frac{x_n}{n}\right)^n = e^x$$

für $x_n \xrightarrow{n\to\infty} x$ folgt

$$\mathbb{P}(X_n = k) \xrightarrow{n\to\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Damit ist die erste Behauptung gezeigt. Wegen $p_\lambda(k) \geq 0$ und

$$\sum_{k=0}^{\infty} p_\lambda(k) = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{x^k} = e^{-\lambda} \cdot e^{\lambda} = e^0 = 1$$

ist p_λ tatsächlich Zähldichte eines diskreten Wahrscheinlichkeitsmaßes auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$. □

3.2 Erwartungswert und Varianz

Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Die Menge Ω sei nachfolgend o.B.d.A. abzählbar mit $\Omega_0 \subset \Omega$.

Wir betrachten ein zufälliges Glücksspiel. Wie hoch ist der "erwartete" Gewinn?

Definition 3.5. *Sei X eine diskrete Zufallsvariable mit Zähldichte p, d.h. $\mathbb{P}^X(\{x\}) = \mathbb{P}(X^{-1}(\{x\})) = p(x)$. Wir sagen, dass der Erwartungswert von X existiert, falls*

$$\sum_{x \in \Omega^X} |x| \cdot p(x) < \infty. \tag{3.1}$$

In diesem Fall definieren wir den Erwartungswert von X als

$$\mathbb{E}X := \sum_{x \in \Omega^X} x \cdot p(x). \tag{3.2}$$

Bemerkung. *Die Forderung der absoluten Summierbarkeit* (3.1) *stellt sicher, dass* (3.2) *wohldefiniert ist, d.h. der Wert der (potentiell unendlichen) Reihe nicht von der Reihenfolge der Summation abhängt. Bedingung (3.1) wird typischerweise als* $\mathbb{E}|X| < \infty$ *abgekürzt.*

Beispiel 3.6. (i) *Sei* X *Laplace-verteilt auf* $\{1, \dots, N\}$*, d.h.* $\forall\, k \in \{1, \dots, N\}$ *ist* $\mathbb{P}(X = k) = \frac{1}{N}$ *. Dann gilt*

$$\mathbb{E}X = \sum_{j=1}^{N} j \cdot \frac{1}{N} = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}.$$

(ii) *Sei* X *Bernoulli-verteilt mit Parameter* $p \in (0,1)$*, d.h. es gelten* $\mathbb{P}(X = 1) = p$ *und* $\mathbb{P}(x = 0) = 1 - p$*. Dann folgt*

$$\mathbb{E}X = 1 \cdot p + 0 \cdot (1 - p) = p.$$

(iii) *Sei* X *binomialverteilt mit Parametern* n, p $(X \sim Bin(n,p))$*, d.h. es gilt*

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

für $k \in \{0, \dots, n\}$*. Wir berechnen:*

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{n} \underbrace{k \cdot \binom{n}{k}}_{= k\cdot \frac{n!}{k!(n-k)!} \;=\; n\cdot \frac{(n-1)!}{(k-1)!\cdot((n-1)-(k-1))!}} \cdot p^k \cdot (1-p)^{n-k} \\
&= np \cdot \sum_{k=1}^{n} \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)} \\
&= np \cdot \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^{k} \cdot (1-p)^{(n-1)-k}}_{\substack{\textit{Summe über die Zähldichte} \\ \textit{der } Bin(n-1,p)\textit{-Verteilung}}} \\
&= np.
\end{aligned}$$

Satz 3.7 (Transformationsformel für den Erwartungswert)**.** *Sei* X *eine diskrete Zufallsvariable mit Zähldichte* p_X*, und* $u : \Omega^X \to \mathbb{R}$ *eine Abbildung mit*

$$\sum_{x \in \Omega^X} |u(x)| \cdot p_X(x) < \infty$$

Dann gilt

$$\mathbb{E}[u(X)] = \sum_{x \in \Omega^X} u(x) \cdot p_X(x).$$

Beweis. Die Zähldichte von $Y := u(X)$ ist gegeben durch

$$p_Y(y) = \mathbb{P}(u(X) = y) = \sum_{\substack{x\in\Omega^X:\\ u(x)=y}} p_X(x).$$

Setzen wir dies in die Definition von $\mathbb{E}Y$ ein, so folgt

$$\mathbb{E}Y = \sum_{y\in\Omega^Y} y p_Y(y) = \sum_{y\in\Omega^Y} y \sum_{\substack{x\in\Omega^X:\\ u(x)=y}} p_X(x) = \sum_{y} \sum_{\substack{x\in\Omega^X:\\ u(x)=y}} u(x)p_X(x) = \sum_{x\in\Omega^X} u(x)p_X(x).$$

□

Satz 3.8 (Eigenschaften des Erwartungswertes)**.** *Es seien X, Y zwei diskrete Zufallsvariablen auf $(\Omega, \mathbb{P}(\Omega), \mathbb{P})$, deren Erwartungswerte existieren. Dann gilt für alle $a, b \in \mathbb{R}$:*

(i) $\mathbb{E}[aX] = a \cdot \mathbb{E}X$

(ii) $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$

(iii) $\mathbb{E}b = b$

(iv) $|\mathbb{E}X| \leq \mathbb{E}|X|$

Beweis. *(i)* Mit $u(x) := a \cdot x$ ergibt die Transformationsformel (Satz 3.7)

$$\mathbb{E}[aX] = \sum_{x\in\Omega^X} (ax) \cdot p_X(x) = a \cdot \sum_{x\in\Omega^X} x \cdot p_X(x) = a \cdot \mathbb{E}X$$

(ii) Mit $\Omega^X = \{x_1, x_2, \dots\}$, $\Omega^Y = \{y_1, y_2, \dots\}$, $A_i = X^{-1}(\{x_i\})$ und $B_j = Y^{-1}(\{y_j\})$ nimmt $X + Y$ Werte in $U := \{x_i + y_j \mid x_i \in \Omega^X y_j \in \Omega^Y\}$ an. Wir erhalten

$$\begin{aligned}
\mathbb{E}X + \mathbb{E}Y &= \sum_i x_i \mathbb{P}(A_i) + \sum_j y_j \mathbb{P}(B_j) \\
&= \sum_i x_i \mathbb{P}\Big(A_i \cap \sum_j B_j\Big) + \sum_j y_j \mathbb{P}\Big(\Big(\sum_i A_i\Big) \cap B_j\Big) \\
&= \sum_{i,j} x_i \mathbb{P}(A_i \cap B_j) + \sum_{i,j} y_j \mathbb{P}(A_i \cap B_j) \\
&= \sum_{i,j} (x_i + y_j)\mathbb{P}(A_i \cap B_j) \\
&= \sum_u \sum_{\substack{i,j\in\mathbb{N}:\\ x_i+y_j=u}} (x_i + y_j)\mathbb{P}(A_i \cap B_j) \\
&= \sum_u u \cdot \underbrace{\sum_{\substack{i,j\in\mathbb{N}:\\ x_i+y_j=u}} \mathbb{P}(A_i \cap B_j)}_{\mathbb{P}(X+Y=u)} = \mathbb{E}(X + Y).
\end{aligned}$$

(*iii*) Klar.

(*iv*) $\mathbb{E}|X| = \sum_{x\in\Omega^X} |x| p_X(x) = \sum_{x\in\Omega^X} \big|xp_X(x)\big| \geq \Big|\sum_{x\in\Omega^X} xp_X(x)\Big| = |\mathbb{E}X|$.

$\square$

Bemerkung. (*i*) *Es gilt im Allgemeinen NICHT:* $\mathbb{E}[g(X)] = g(\mathbb{E}X)$.

(*ii*) *Aus 3.8 (ii) folgt per Induktion*

$$\mathbb{E}\Big[\sum_{i=1}^n X_i\Big] = \sum_{i=1}^n \mathbb{E}X_i.$$

Beispiel 3.9. (*i*) *Sei* $X \sim Bin(n,p)$, *d.h.* X *ist die Anzahl von Erfolgen in* n *unabhängigen Bernoulli-Experimenten mit Parameter* p *– bspw. ist* X *die Anzahl der 6-er beim* n*-maligen Würfelwurf (in diesem Falle ist* $p = 1/6$*). Wir definieren für* $i \in \{1,\dots,n\}$ *das Ereignis*

$A_i :=$ *„das i-te Bernoulli-Experiment liefert Erfolg"* $= \{\omega \in \Omega_1 \times \ldots \times \Omega_n \mid \omega_i = 1\}$

(wobei $\Omega_j = \{0,1\}\ \forall\, j \in \{1,\dots,n\}$ *ist) sowie*

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \omega \in A_i \\ 0 & \omega \notin A_i \end{cases}$$

Dann gelten $X = \sum_{i=1}^n \mathbb{1}_{A_i}$ *und damit* $\mathbb{E}X = \sum_{i=1}^n \mathbb{E}\mathbb{1}_{A_i} = np$.

(*ii*) *Eine hypergeometrische Verteilung* $\mathcal{H}(N,R,n)$ *beschreibt ein* n*-faches Ziehen ohne Zurücklegen aus einer Urne mit* R *roten und* $N-R$ *weißen Kugeln. Es gilt*

$$\mathbb{P}(\underbrace{X = r}_{\text{genau } r \text{ rote Kugeln gezogen}}) = \frac{\binom{R}{r}\cdot\binom{N-R}{n-r}}{\binom{N}{n}},$$

wobei $\binom{n}{k} := 0$ *für* $k < 0$ *oder* $k > n$ *ist. Nun gibt es zwei Möglichkeiten zur Berechnung des Erwartungswertes.*

(*a*) *Wir rechnen mit demselben Trick wie bei der Binomialverteilung:*

$$\begin{aligned}\mathbb{E}X &= \binom{N}{n}^{-1} \cdot \sum_{r=1}^n r \cdot \binom{R}{r} \cdot \binom{N-R}{n-r} \\ &= \binom{N}{n}^{-1} \cdot \sum_{r=1}^n R \cdot \binom{R-1}{r-1} \cdot \binom{N-R}{n-r} \\ &= R \cdot \binom{N}{n}^{-1} \cdot \sum_{r=0}^{n-1} \binom{R-1}{r} \cdot \binom{N-R}{(n-1)-r}\end{aligned}$$

$$
\begin{aligned}
&= R \cdot \binom{N}{n}^{-1} \cdot \binom{N-1}{n-1} \qquad \text{nach Blatt 2, Aufgabe 1 (d)} \\
&= R \cdot \frac{n!(N-n)!}{N!} \cdot \frac{(N-1)!}{(n-1)! \cdot (N-n)!} \\
&= R \cdot \frac{n}{N}
\end{aligned}
$$

(*b*) *Wir definieren für $i \in \{1, \dots, n\}$ das Ereignis*

$$B_i := \text{„die } i\text{-te gezogene Kugel ist rot“.}$$

Damit ist $X = \sum_{i=1}^{n} \mathbb{1}_{B_i}$ und

$$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{E}\mathbb{1}_{B_i} = \sum_{i=1}^{n} \mathbb{P}(B_i) \overset{\text{Symmetrie}}{=} n\frac{R}{N}.$$

Definition 3.10. *Sei X eine diskrete Zufallsvariable mit $\mathbb{E}\big(X^2\big) < \infty$. Dann wird ihre Varianz definiert durch*

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

Bemerkung. *Man bezeichnet die Varianz von X auch häufig mit σ^2 oder σ_X^2. Die Größe $\sqrt{\operatorname{Var}(X)}$ nennt man Standardabweichung von X.*

Satz 3.11 (Eigenschaften der Varianz)**.** *Seien $a, b \in \mathbb{R}$ sowie X eine diskrete Zufallsvariable mit $\mathbb{E}\big(X^2\big) < \infty$. Dann gilt:*

(*i*) $\operatorname{Var}(aX + b) = a^2 \cdot \operatorname{Var}(X)\ \forall\, a, b \in \mathbb{R}$

(*ii*) $\operatorname{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$

(*iii*) $\mathbb{E}[(X - a)^2] = \operatorname{Var}(X) + (\mathbb{E}X - a)^2 \geq \operatorname{Var}(X)$ *mit Gleichheit genau dann, wenn $a = \mathbb{E}X$ ist.*

Beweis. (i) Seien $a, b \in \mathbb{R}$. Nach der Linearität des Erwartungswertes gilt

$$
\begin{aligned}
\operatorname{Var}(aX + b) &= \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] \\
&= \mathbb{E}[(aX + b - a\mathbb{E}X - b)^2] \\
&= \mathbb{E}[a^2(X - \mathbb{E}X)^2] \\
&= a^2\mathbb{E}[(X - \mathbb{E}X)^2] \\
&= a^2 \cdot \operatorname{Var}(X).
\end{aligned}
$$

(ii)

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X-\mathbb{E}X)^2] \\
&= \mathbb{E}[X^2 - 2X(\mathbb{E}X) + (\mathbb{E}X)^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X\cdot \mathbb{E}X] + \mathbb{E}[(\mathbb{E}X)^2] \quad \text{nach Linearität} \\
&= \mathbb{E}[X^2] - 2(\mathbb{E}X)(\mathbb{E}X) + (\mathbb{E}X)^2 \quad \text{nach Satz 3.8 (i) und (iii)} \\
&= \mathbb{E}[X^2] - (\mathbb{E}X)^2
\end{aligned}$$

(iii) Wegen $\mathbb{E}[X - \mathbb{E}X] = \mathbb{E}X - \mathbb{E}X = 0$ gilt

$$\begin{aligned}
\mathbb{E}[(X-a)^2] &= \mathbb{E}[(X-\mathbb{E}X+\mathbb{E}X-a)^2] \\
&= \mathbb{E}[(X-\mathbb{E}X)^2 + 2\,\underbrace{\mathbb{E}[(X-\mathbb{E}X)(\mathbb{E}X-a)]}_{=2(\mathbb{E}X-a)\mathbb{E}[X-\mathbb{E}X]=0} + \mathbb{E}[(\mathbb{E}X-a)^2]] \\
&= \mathrm{Var}(X) + (\mathbb{E}X-a)^2 \quad \text{nach 3.8 (iii).}
\end{aligned}$$

□

Beispiel 3.12 (Varianz der Binomialverteilung)**.** *Sei $X \sim \mathrm{Bin}(n,p)$. Wir kennen bereits $\mathbb{E}X$. Anstatt $\mathbb{E}\big(X^2\big)$ zu berechnen, bestimmen wir $\mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}(X)$, damit wir aus dem Binomialkoeffizienten kürzen können.*

$$\begin{aligned}
\mathbb{E}(X(X-1)) &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \cdot \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} \\
&= n(n-1)p^2 \cdot \underbrace{\overbrace{\sum_{l=0}^{n-2} \binom{n-2}{l} p^l (1-p)^{(n-2)-l}}^{\text{Zähldichte von }\mathrm{Bin}(n-2,p)}}_{=1} \\
&= n(n-1)p^2
\end{aligned}$$

Es folgt

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 &= \mathbb{E}(X^2) - \mathbb{E}X + \mathbb{E}X - (\mathbb{E}X)^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= np(1-p).
\end{aligned}$$

Satz 3.13 (Markov- und Chebychev-Ungleichung)**.** *Sei X eine diskrete Zufallsvariable.*

(i) *Markov-Ungleichung: Falls $\mathbb{E}|X| < \infty$, so gilt $\forall\, \eta \in \mathbb{R}_+$:*

$$\mathbb{P}(|X| \geq \eta) \leq \frac{\mathbb{E}|X|}{\eta}$$

(ii) *Chebychev-Ungleichung: Ist $\mathbb{E}(X^2) < \infty$, so gilt $\forall\, \eta \in \mathbb{R}_+, a \in \mathbb{R}$:*

$$\mathbb{P}(|X - a| \geq \eta) \leq \frac{\mathbb{E}\left[(X-a)^2\right]}{\eta^2}$$

Für $a = \mathbb{E}X$ ist insbesondere

$$\mathbb{P}(|X - \mathbb{E}X| \geq \eta) \leq \frac{\mathbb{E}\left[(X-\mathbb{E}X)^2\right]}{\eta^2} = \frac{\mathrm{Var}(X)}{\eta^2}.$$

Beweis. (i)

$$\mathbb{P}(|X| \geq \eta) = \sum_{x \in \Omega^X : |x| \geq \eta} p(x) \leq \sum_{x \in \Omega^X : |x| \geq \eta} \frac{|x|}{\eta} p(x) \leq \frac{1}{\eta} \sum_{x \in \Omega^X} |x| p(x) = \frac{1}{\eta} \mathbb{E}X$$

(ii)

$$\mathbb{P}(|X - a| \geq \eta) = \mathbb{P}\left((X-a)^2 \geq \eta^2\right) \overset{(i)}{\leq} \frac{\mathbb{E}\left((X-a)^2\right)}{\eta^2}.$$

□

3.3 Mehrdimensionale Verteilungen

Angenommen, X und Y seien zwei Zufallsvariablen, bspw. X das Alter eines Fisches und Y seine Größe. Kann man aus den Werten von X auf die von Y schließen?

Definition 3.14. *Seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen auf einem gemeinsamen diskreten Wahrscheinlichkeitsraum. Das (diskrete) Maß $\mathbb{P}^{X_1,\ldots,X_n}$ auf $(\mathbb{R}^n, \mathcal{P}(\mathbb{R}^n))$, gegeben durch*

$$\mathbb{P}^{X_1,\ldots,X_n}(C) := \mathbb{P}((X_1, \ldots, X_n)^{-1}(C)) \quad \text{für } C \subset \mathbb{R}^n,$$

heißt gemeinsame Verteilung von $X_1, \ldots, X_n$. Die Funktion $F_{X_1,\ldots X_n} : \mathbb{R}^n \to [0,1]$ mit

$$F_{X_1,\ldots X_n}(x_1, \ldots, x_n) := \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}\left(\bigcap_{i=1}^n X_i^{-1}((-\infty, x_i])\right)$$

heißt gemeinsame Verteilungsfunktion von $X_1, \ldots, X_n$. $p_{X_1,\ldots X_n} : \mathbb{R}^n \to [0,1]$ mit

$$p_{X_1,\ldots X_n}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

nennen wir <u>*gemeinsame Zähldichte*</u> *von* $X_1, \dots, X_n$. *Für* $1 \leq i_1 < i_2 < \dots < i_k \leq n$ *heißt die gemeinsame Verteilung von* $X_{i_1}, \dots, X_{i_k}$ <u>*k-dimensionale Marginalverteilung*</u> *bzw.* <u>*k-dimensionale Randverteilung*</u>.

Bemerkung 3.15. *(i) Da der Träger von* $\mathbb{P}^{X_1,\dots,X_n}$ *bereits in* $\Omega^{X_1} \times \dots \times \Omega^{X_n}$ *enthalten ist, kann man die gemeinsame Verteilung* $\mathbb{P}^{X_1,\dots,X_n}$ *gleichermaßen auf*

$$\big(\Omega^{X_1} \times \dots \times \Omega^{X_n}, \mathcal{P}(\Omega^{X_1} \times \dots \times \Omega^{X_n})\big)$$

bzw. $\big(\Omega_0^{X_1} \times \dots \times \Omega_0^{X_n}, \mathcal{P}(\Omega_0^{X_1} \times \dots \times \Omega_0^{X_n})\big)$ *defnieren, wobei* Ω_0 *den Träger von* $\mathbb{P}$ *im strengen Sinne bezeichnet (siehe Bemerkung 1.4).*

(ii) Für $I \subset \{1, \dots n\}$ *gilt*

$$p_{X_i : i \in I}((x_i)_{i \in I}) = \mathbb{P}(X_i = x_i \ \forall\, i \in I) = \sum_{j \in I^C} \sum_{x_j \in \Omega_j^X} p(x_1, \dots, x_n).$$

Beispiel 3.16. *Betrachte den dreimaligen Würfelwurf, wobei „0“ Kopf und „1“ Zahl entspricht. Wir nutzen den Grundraum* $\Omega = \{0,1\}^3$, *und definieren*

$$X : \Omega \to \mathbb{R}, (\omega_1, \omega_2, \omega_3) \mapsto \omega_1 \text{ und } Y : \Omega \to \mathbb{R}, (\omega_1, \omega_2, \omega_3) \mapsto \sum_{i=1}^{3} \omega_i,$$

d.h. X *prüft, ob im ersten Wurf Zahl erhalten wurde, und* Y *entspricht der Anzahl an Zahlen in den 3 Würfen. Damit ist* $\Omega^X = \{0,1\}$ *und* $\Omega^Y = \{0,1,2,3\}$. *Wir wissen, dass* $\mathbb{P}^{(X,Y)}$ *ein diskretes Wahrscheinlichkeitsmaß auf* $(\Omega^X \times \Omega^Y, \mathcal{P}(\Omega^X \times \Omega^Y))$ *ist, wobei*

$$(X,Y) : \Omega \to \Omega^X \times \Omega^Y$$

und

$$\mathbb{P}^{(X,Y)}(C) = \mathbb{P}((X,Y)^{-1}(C)).$$

Ferner gilt:

$$\Omega^X \times \Omega^Y = \{(x,y) \mid x \in \Omega^X, y \in \Omega^Y\} = \{(0,0), \dots, (0,3), (1,0), \dots, (1,3)\}$$

Betrachte bspw. das Ereignis $C = \{(0,1), (1,2)\}$ *(dabei enspricht* $(0,1)$ *„1. Wurf Kopf, insgesamt 1x Zahl“, und* $(1,2)$ *„1. Wurf Zahl, insgesamt 2x Zahl“). Es gilt:*

$$(X,Y)^{-1}(C) = \{(0,1,0), (0,0,1), (1,1,0), (1,0,1)\}$$

$$\implies \mathbb{P}^{(X,Y)}(C) = \frac{|(X,Y)^{-1}(C)|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}.$$

Wir wollen nun noch die gemeinsame Verteilung von X und Y bestimmen:

$X \backslash Y$	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Damit gilt

$$p_X(0) = \mathbb{P}(X = 0, Y \text{ beliebig}) = \sum_{j=0}^{3} \mathbb{P}(X = 0, Y = j) = \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + 0 = \frac{1}{2}.$$

Definition 3.17. *Seien $X_1, \dots, X_n$ diskrete Zufallsvariablen. $X_1, \dots, X_n$ heißen stochastisch unabhängig, falls gilt:*

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \ \forall\, (x_1, \dots, x_n) \in \mathbb{R}^n$$

Bemerkung 3.18. *(i) Äquivalent zu obiger Definition ist:*

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \ \forall\, (x_1, \dots, x_n) \in \mathbb{R}^n$$

(ii) Seien $X_1, \dots, X_n$ stochastisch unabhängige, diskrete Zufallsvariablen. Dann gilt:

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) \ \forall\, B_1, \dots, B_n \subseteq \mathbb{R}$$

Notation. *Man verwendet bei Zufallsvariablen die Abkürzung "iid" für "unabhängig identisch verteilt" (independent identically distributed).*

Beispiel 3.19 (Summe von unabhängigen Zufallsvariablen)**.** *Seien $X \sim \mathrm{Bin}(n, p)$ und $Y \sim \mathrm{Bin}(m, p)$ stochastisch unabhängige Variablen. Wir wollen nun die Verteilung von $X + Y$ bestimmen.*

$$\begin{aligned}
\mathbb{P}(X + Y = k) &= \mathbb{P}(X + Y = k, X \text{ beliebig}) \\
&= \sum_{l=0}^{k} \mathbb{P}(X + Y = k, X = l) \\
&= \sum_{l=0}^{k} \mathbb{P}(X = l, Y = k - l)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=0}^{k} \mathbb{P}(X = l) \cdot \mathbb{P}(Y = k-l) && (X, Y \text{ stochastisch unabhängig})\\
&= \sum_{l=0}^{k} \binom{n}{l} p^l (1-p)^{n-l} \binom{m}{k-l} p^{k-l} (1-p)^{m-(k-l)}\\
&= \sum_{l=0}^{k} \binom{n}{l} \binom{m}{k-l} p^k (1-p)^{n+m-k}\\
&= p^k (1-p)^{n+m-k} \cdot \sum_{l=0}^{k} \binom{n}{l} \binom{m}{k-l}\\
&= \binom{n+m}{k} p^k (1-p)^{n+m-k} && (\text{Blatt 2, Aufgabe 1 (d)}),
\end{aligned}
$$

womit $X + Y \sim \mathrm{Bin}(n+m, p)$. *Induktiv folgt damit insbesondere für Zufallsvariablen* $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(p) = \mathrm{Bin}(1, p)$*:*

$$
\sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, p).
$$

Satz 3.20. *Seien* X *und* Y *zwei unabhängige diskrete Zufallsvariablen mit* $\mathbb{E}(X^2) < \infty$ *und* $\mathbb{E}(Y^2) < \infty$*. Dann gilt:*

(i) $\mathbb{E}(XY) = (\mathbb{E}X) \cdot (\mathbb{E}Y)$

(ii) $\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y$

Beweis. (i)

$$
\begin{aligned}
(\mathbb{E}X) \cdot (\mathbb{E}Y) &= \left(\sum_{x \in \Omega^X} x \cdot p_X(x) \right) \cdot \left(\sum_{y \in \Omega^Y} y \cdot p_Y(y) \right)\\
&= \sum_{x \in \Omega^X, y \in \Omega^Y} x \cdot p_X(x) \cdot y \cdot p_Y(y)\\
&= \sum_{u \in \Omega^{XY}} \sum_{\substack{x \in \Omega^X, y \in \Omega^Y \\ x \cdot y = u}} x \cdot p_X(x) \cdot y \cdot p_Y(y)\\
&= \sum_{u \in \Omega^{XY}} \sum_{\substack{x \in \Omega^X, y \in \Omega^Y \\ x \cdot y = u}} xy \cdot \underbrace{\mathbb{P}(X = x, Y = y)}_{\substack{= \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y) \\ \text{nach Unabhängigkeit}}}\\
&= \sum_{u \in \Omega^{XY}} u \cdot \sum_{\substack{x \in \Omega^X, y \in \Omega^Y \\ x \cdot y = u}} \mathbb{P}(X = x, Y = y)\\
&= \sum_{u \in \Omega^{XY}} u \cdot \underbrace{p_{XY}(u)}_{= \mathbb{P}(XY = u)} = \mathbb{E}(XY).
\end{aligned}
$$

(*ii*) Nach (*i*) ist

$$\begin{aligned}
\mathbb{E}\big((X-\mathbb{E}X)\cdot(Y-\mathbb{E}Y)\big) &= \mathbb{E}\big(XY - X\mathbb{E}Y - (\mathbb{E}X)Y + (\mathbb{E}X)(\mathbb{E}Y)\big) \\
&= \mathbb{E}(XY) - \mathbb{E}(X\mathbb{E}Y) - \mathbb{E}((\mathbb{E}X)Y) + \mathbb{E}((\mathbb{E}X)(\mathbb{E}Y)) \\
&= \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) - (\mathbb{E}X)(\mathbb{E}Y) + (\mathbb{E}X)(\mathbb{E}Y) \\
&= \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) \\
&= 0.
\end{aligned}$$

Damit gilt

$$\begin{aligned}
\operatorname{Var}(X+Y) &= \mathbb{E}\left((X+Y-\mathbb{E}(X+Y))^2\right) \\
&= \mathbb{E}\left(((X-\mathbb{E}X)+(Y-\mathbb{E}Y))^2\right) \\
&= \mathbb{E}\left((X-\mathbb{E}X)^2 + 2(X-\mathbb{E}X)(Y-\mathbb{E}Y) + (Y-\mathbb{E}Y)^2\right) \\
&= \mathbb{E}\left((X-\mathbb{E}X)^2\right) + \underbrace{\mathbb{E}(2(X-\mathbb{E}X)(Y-\mathbb{E}Y))}_{=0} + \mathbb{E}\left((Y-\mathbb{E}Y)^2\right) \\
&= \operatorname{Var} X + \operatorname{Var} Y.
\end{aligned}$$

□

Satz 3.21 (Schwaches Gesetz der großen Zahlen)**.** *Seien $X_1, \ldots, X_n$ iid diskrete Zufallsvariablen mit endlicher Varianz σ^2. Dann gilt $\forall\, \varepsilon > 0$:*

$$\mathbb{P}\left(\left|\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right) - \mathbb{E}X_1\right| \geq \varepsilon\right) \xrightarrow{n\to\infty} 0$$

Beweis. Da $X_1, \ldots, X_n$ iid sind, ist $\mathbb{E}X_i = \mathbb{E}X_1 \;\forall\, i \in \{1, \ldots, n\}$. Damit folgt mit der Chebychev-Ungleichung

$$\begin{aligned}
\mathbb{P}\left(\left|\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right) - \mathbb{E}X_1\right| \geq \varepsilon\right) &= \mathbb{P}\left(\left|\frac{1}{n}\cdot\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right| \geq \varepsilon\right) \\
&\leq \frac{\operatorname{Var}\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} \\
&\overset{3.11}{=\!=} \frac{1}{n^2}\cdot\frac{\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} \\
&\overset{3.20}{=\!=} \frac{1}{n^2\varepsilon^2}\cdot\sum_{i=1}^{n}\underbrace{\operatorname{Var}(X_i)}_{=\sigma^2} \\
&= \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0.
\end{aligned}$$

□

Bemerkung 3.22. *(i) Man nennt eine unendliche Folge* $(X_i)_{i\in\mathbb{N}}$ *von Zufallsvariablen „stochastisch unabhängig", falls je endlich viele von ihnen stochastisch unabhängig sind.*

(ii) Die Art der Konvergenz der Zufallsvariablen

$$Y_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

gegen $\mathbb{E}X_1$ *im schwachen Gesetz der großen Zahlen wird Konvergenz in Wahrscheinlichkeit oder stochastische Konvergenz genannt.*

Es existiert auch ein sogenanntes starkes Gesetz der großen Zahlen (später), dessen Konvergenzbegriff stärker ist, als der in Satz 3.21 (deswegen spricht man hier auch vom „schwachen Gesetz"). Insbesondere zeigt der Beweis, dass die Konvergenz im quadratischen Mittel

$$\mathbb{E}\left((Y_n - \mathbb{E}X_1)^2\right) \xrightarrow{n\to\infty} 0$$

die stochastische Konvergenz impliziert.

Definition 3.23. *Seien* X, Y *diskrete Zufallsvariablen mit* $\mathbb{E}(X^2) < \infty$ *und* $\mathbb{E}(Y^2) < \infty$. *Dann heißt*

$$\operatorname{cov}(X,Y) := \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$$

die Kovarianz von X *und* Y. *Ist* $\operatorname{cov}(X,Y) = 0$, *so heißen* X *und* Y *unkorreliert.*

Eine Verallgemeinerung von Satz 3.20 ist folgende.

Satz 3.24. *Seien* $X_1, \ldots, X_n$ *Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum mit* $\mathbb{E}(X_i^2) < \infty$ *für* $i \in \{1, \ldots, n\}$. *Dann existiert die Varianz* $\operatorname{Var}\left(\sum_{i=1}^n X_i\right)$, *und es gilt*

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + \sum_{i,j:i\neq j} \operatorname{cov}(X_i, X_j).$$

Beweis. Übungsaufgabe. □

Bemerkung. *Sind* X *und* Y *stochastisch unabhängige Zufallsvariablen mit* $\mathbb{E}(X^2) < \infty$ *und* $\mathbb{E}(Y^2) < \infty$, *so gilt nach Satz 3.20*

$$\operatorname{cov}(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = (\mathbb{E}X)(\mathbb{E}Y) - (\mathbb{E}X)(\mathbb{E}Y) = 0.$$

Im Allgemeinen impliziert Unkorreliertheit aber nicht die stochastische Unabhängigkeit.

Definition 3.25. *Seien X, Y diskrete Zufallsvariablen auf einem gemeinsamen, diskreten Wahrscheinlichkeitsraum mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$, $\operatorname{Var} X > 0$ und $\operatorname{Var} Y > 0$. Dann heißt*

$$\varrho(X,Y) := \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{Var} X} \cdot \sqrt{\operatorname{Var} Y}}$$

der Korrelationskoeffizient von X und Y.

Der Korrelationskoeffizient ist ein Maß für den linearen Zusammenhang zwischen zwei Zufallsvariablen.

Satz 3.26. *Seien X und Y diskrete Zufallsvariablen auf einem gemeinsamen, diskreten Wahrscheinlichkeitsraum mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$, $\operatorname{Var} X > 0$ und $\operatorname{Var} Y > 0$. Dann gelten $|\varrho(X,Y)| \leq 1$ sowie*

$$\varrho(X,Y) = \pm 1 \iff \exists\, a, b \in \mathbb{R} : \mathbb{P}(Y = a + bX) = 1.$$

Dabei ist $b > 0$, falls $\varrho(X,Y) = +1$, und $b < 0$, falls $\varrho(X,Y) = -1$.

Beweis. Die Aussage $|\varrho(X,Y)| \leq 1$ folgt aus der Cauchy-Schwarz-Ungleichung, denn die Kovarianz definiert ein Skalarprodukt auf der Menge der zentrierten Zufallsvariablen mit endlicher Varianz. Sei zunächst $\varrho(X,Y) = 1$. Dann gilt

$$\begin{aligned}\operatorname{Var}\left(\frac{X}{\sqrt{\operatorname{Var} X}} - \frac{Y}{\sqrt{\operatorname{Var} Y}}\right) &= \operatorname{Var}\left(\frac{X}{\sqrt{\operatorname{Var} X}}\right) + \operatorname{Var}\left(\frac{Y}{\sqrt{\operatorname{Var} Y}}\right) \\ &\quad - 2\operatorname{cov}\left(\frac{X}{\sqrt{\operatorname{Var} X}}, \frac{Y}{\sqrt{\operatorname{Var} Y}}\right) = 1 + 1 - 2 = 0\end{aligned}$$

$$\begin{aligned}&\iff \mathbb{P}\left(\frac{X}{\sqrt{\operatorname{Var} X}} - \frac{Y}{\sqrt{\operatorname{Var} Y}} = c\right) = 1 \text{ mit einem } c \in \mathbb{R} \\ &\implies \mathbb{P}(Y = a + bX) = 1 \text{ mit } b = \frac{\sqrt{\operatorname{Var} Y}}{\sqrt{\operatorname{Var} X}} > 0.\end{aligned}$$

Für $\varrho(X,Y) = -1$ betrachte analog:

$$\operatorname{Var}\left(\frac{X}{\sqrt{\operatorname{Var} X}} + \frac{Y}{\sqrt{\operatorname{Var} Y}}\right) = \ldots = 0 \implies \ldots \text{ mit } b = -\frac{\sqrt{\operatorname{Var} Y}}{\sqrt{\operatorname{Var} X}}.$$

□

Bemerkung und Definition 3.27. *Seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum mit $\mathbb{E}(X_i^2) < \infty\ \forall\, i \in \{1, \ldots, n\}$. Dann heißt die Matrix*

$$\Sigma := (\operatorname{cov}(X_i, X_j))_{i,j \in \{1,\ldots,n\}}$$

die Kovarianzmatrix von $X = (X_1, \dots, X_n)^\top$. Es gilt für $\gamma = (\gamma_1, \dots, \gamma_n)^\top \in \mathbb{R}^n$ (nach der Bilinearität der Kovarianz)

$$\underbrace{\mathrm{Var}(\gamma^\top X)}_{\geq 0} = \gamma^\top \Sigma \gamma.$$

Damit ist die Matrix Σ symmetrisch und positiv semidefinit.
Bspw. für $n = 2$ ist die Kovarianzmatrix gegeben durch

$$\Sigma = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{cov}(X_1, X_2) \\ \mathrm{cov}(X_1, X_2) & \mathrm{Var}(X_2) \end{pmatrix}.$$

3.4 Bedingte Verteilungen

Definition 3.28. *Seien X und Y Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Sei $x \in \mathbb{R}$ mit $\mathbb{P}(X = x) > 0$. Das Wahrscheinlichkeitsmaß auf $(\Omega^Y, \mathcal{P}(\Omega^Y))$, welches jeder Teilmenge $A \in \mathcal{P}(\Omega^Y)$ die Wahrscheinlichkeit $\mathbb{P}(y \in A \mid X = x)$ zuordnet, heißt bedingte Verteilung von Y gegeben $X = x$. Die bedingte Zähldichte von Y gegeben $X = x$ wird definiert durch*

$$p_{Y|X=x}(y) = \mathbb{P}(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Bemerkung 3.29. (*i*) *Die bedingte Verteilung von Y gegeben $X = x$ ist tatsächlich ein diskretes Wahrscheinlichkeitsmaß (im Sinne der Axiome von Kolmogorov) bzw. die bedingte Zähldichte tatsächlich die Zähldichte eines Wahrscheinlichkeitsmaßes ist. Denn aus $\mathbb{P}(X = x) = p_X(x) > 0$ folgt*

$$p_{Y|X=x}(y) = \frac{p_{X,Y}(x, y)}{p_X(x)} \geq 0$$

und

$$\sum_{y \in \Omega^Y} p_{Y|X=x}(y) = \sum_{y \in \Omega^Y} \frac{p_{X,Y}(x, y)}{p_X(x)} = \frac{1}{p_X(x)} \cdot \sum_{y \in \Omega^Y} p_{X,Y}(x, y) = \frac{1}{p_X(x)} \cdot p_X(x) = 1.$$

(*ii*) *Man schreibt auch $\mathbb{P}^{Y|X=x}(A)$ für $\mathbb{P}(Y \in A \mid X = x)$, bzw. $\mathbb{P}^{Y|X=x}$ für die bedingte Verteilung von Y gegeben $X = x$.*

Beispiel 3.30. *Seien X, Y stochastisch unabhängige Zufallsvariablen mit $X \sim \mathrm{Bin}(n, p)$ und $Y \sim \mathrm{Bin}(m, p)$. Dann gilt für die bedingte Zähldichte von X gegeben $X + Y = k$:*

$$p_{X|X+Y=k}(j) = \mathbb{P}(X = j \mid X + Y = k)$$

$$
\begin{aligned}
&= \frac{\mathbb{P}(X=j, X+Y=k)}{\mathbb{P}(X+Y=k)} \\
&= \frac{\mathbb{P}(X=j, Y=k-j)}{\mathbb{P}(X+Y=k)} \\
&= \frac{\mathbb{P}(X=j)\cdot\mathbb{P}(Y=k-j)}{\mathbb{P}(X+Y=k)} \\
&= \frac{\binom{n}{j}\cdot p^j\cdot(1-p)^{n-j}\cdot\binom{m}{k-j}\cdot p^{k-j}\cdot(1-p)^{m-(k-j)}}{\binom{n+m}{k}\cdot p^k\cdot(1-p)^{n+m-k}} \\
&= \frac{\binom{n}{j}\cdot\binom{m}{k-j}}{\binom{n+m}{k}},
\end{aligned}
$$

d.h. die bedingte Verteilung von X gegeben $X+Y=k$ ist die hypergeometrischen Verteilung $\mathcal{H}(n+m,n,k)$.

Definition 3.31. *Seien X und Y diskrete Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Für ein $x \in \mathbb{R}$ sei $\mathbb{P}(X=x)>0$.*

(i) Ist $\sum_y |y|\cdot p_{Y|X=x}(y) < \infty$, so wird der bedingte Erwartungswert von Y gegeben $X=x$ definiert durch

$$\mathbb{E}(Y|X=x) = \sum_y y\cdot p_{Y|X=x}(y).$$

(ii) Ist $\sum_y y^2\cdot p_{Y|X=x}(y) < \infty$, so wird durch

$$\operatorname{Var}(Y|X=x) := \mathbb{E}\left((Y-\mathbb{E}(Y\mid X=x))^2 \mid X=x\right)$$

die bedingte Varianz von Y gegeben $X=x$ definiert.

Damit ist der bedingte Erwartungswert von Y gegeben $X=x$ gleich dem Erwartungswert der bedingten Verteilung, und die bedingte Varianz ist gleich der Varianz der bedingten Verteilung. Alle Sätze und Identitäten, die wir bislang für Erwartungswert und Varianz bewiesen haben, stehen uns damit auch für den bedingten Erwartungswert und die bedingte Varianz zur Verfügung.

Definition 3.32. *Seien X, Y zwei diskrete Zufallsvariablen, sodass $\mathbb{E}(|Y| \mid X=x) < \infty$ $\forall\, x$ mit $\mathbb{P}(X=x)>0$ ist. Dann heißt die Zufallsvariable $g(X) = g\circ X$ mit*

$$g(x) := \begin{cases} \mathbb{E}(Y|X=x) & \textit{falls } p_X(x) > 0 \\ 0 & \textit{sonst} \end{cases}$$

die bedingte Erwartung von Y gegeben X und wird mit $\mathbb{E}(Y|X)$ bezeichnet. Entsprechend definieren wir die bedingte Varianz von Y gegeben X, $\operatorname{Var}(Y|X)$, indem die Funktion h, *gegeben durch*

$$h(x) := \begin{cases} \operatorname{Var}(Y|X=x) & \textit{falls } p_X(x) > 0 \\ 0 & \textit{sonst,} \end{cases}$$

mit der Zufallsvariable X verknüpft wird (vorausgesetzt, dass $\operatorname{Var}(Y|X = x)$ $\forall\, x$ *mit* $p_X(x) > 0$ *existiert).*

Damit sind bedingte Erwartung $\mathbb{E}(Y|X)$ und bedingte Varianz $\operatorname{Var}(Y|X)$ selbst Zufallsvariablen!

Satz 3.33 (Eigenschaften von $\mathbb{E}(Y|X)$)**.** *Seien X, Y, Z diskrete Zufallsvariablen mit $\mathbb{E}|Y| < \infty$ und $\mathbb{E}|Z| < \infty$. Dann gelten:*

(i) $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}Y$.

(ii) Sind X, Y stochastisch unabhängig, so ist $\mathbb{E}(Y|X) = \mathbb{E}Y$ mit Wahrscheinlichkeit 1 [Man sagt, ein Ereignis gilt „fast sicher“ (abgekürzt „f.s.“), falls es Wahrscheinlchkeit 1 hat. In diesem Sinne gilt also $\mathbb{E}(Y|X) = \mathbb{E}Y$ f.s.] Insbesondere gilt $\mathbb{E}(1|X) = 1$ fast sicher.

(iii) <u>Linearität</u>:

$$\mathbb{E}(aY + bZ|X) = a \cdot \mathbb{E}(Y|X) + b \cdot \mathbb{E}(Z|X)$$

Beweis. (i)

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(Y|X)) &= \sum_{x:p_X(x)>0} \mathbb{E}(Y|X=x) \cdot p_X(x) \qquad \text{(nach der Transformationsformel)} \\
&= \sum_{x:p_X(x)>0} \left(\sum_y y \cdot p_{Y|X=x}(y) \right) \cdot p_X(x) \\
&= \sum_{x:p_X(x)>0} \sum_y y \cdot \underbrace{p_{Y|X=x}(y) \cdot p_X(x)}_{p_{X,Y}(x,y)} \\
&= \sum_x \sum_y y \cdot p_{X,Y}(x,y) \\
&= \sum_y y \cdot \underbrace{\left(\sum_x p_{X,Y}(x,y) \right)}_{=p_Y(y)} \\
&= \sum_y y \cdot p_Y(y) \\
&= \mathbb{E}Y
\end{aligned}$$

(*ii*) $\forall\, x$ mit $p_X(x) > 0$ gilt

$$\begin{aligned}
\mathbb{E}(Y|X=x) &= \sum_y y \cdot p_{Y|X=x}(y) \\
&= \sum_y y \cdot \frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)} \\
&= \sum_y y \cdot \frac{\mathbb{P}(Y=y) \cdot \mathbb{P}(X=x)}{\mathbb{P}(X=x)} \\
&= \sum_y y \cdot \mathbb{P}(Y=y) \\
&= \mathbb{E}Y.
\end{aligned}$$

(*iii*) Übungsaufgabe.

□

Satz 3.34. *Seien X und Y stochastisch unabhängige Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum, und sei $f : \mathbb{R}^2 \to \mathbb{R}$ eine Funktion mit $\mathbb{E}|f(X,Y)| < \infty$. Dann gilt für alle x_0 mit $\mathbb{P}(X = x_0) > 0$:*

$$\mathbb{E}(f(X,Y) \mid X = x_0) = \mathbb{E}(f(x_0, Y)).$$

Beweis. Wir definieren die Zufallsvariable $Z := f(X,Y)$. Damit ergibt sich

$$\begin{aligned}
\mathbb{E}(f(X,Y) \mid X = x_0) &= \mathbb{E}(Z \mid X = x_0) \\
&= \sum_z z \cdot p_{Z|X=x_0}(z) \\
&= \sum_z z \cdot \frac{\mathbb{P}(Z=z, X=x_0)}{\mathbb{P}(X=x_0)} \\
&= \sum_z z \cdot \sum_{y: f(x_0,y)=z} \frac{\mathbb{P}(Y=y, X=x_0)}{\mathbb{P}(X=x_0)} \\
&= \sum_z z \cdot \sum_{y: f(x_0,y)=z} \frac{\mathbb{P}(Y=y) \cdot \mathbb{P}(X=x_0)}{\mathbb{P}(X=x_0)} \\
&= \sum_z \sum_{y: f(x_0,y)=z} z \cdot \mathbb{P}(Y=y) \\
&= \sum_z \sum_{y: f(x_0,y)=z} f(x_0,y) \cdot \mathbb{P}(Y=y) \\
&= \sum_y f(x_0,y) \cdot p_Y(y).
\end{aligned}$$

□

Bemerkung 3.35. *(i) Ist f in Satz 3.34 von der Form $f(x,y) = y \cdot h(x)$, so folgt*

$$\mathbb{E}(Y \cdot h(X) \mid X = x_0) \stackrel{3.34}{=\!=} \mathbb{E}(Y \cdot h(x_0)) = h(x_0) \cdot \mathbb{E}Y =: g(x_0).$$

Damit gilt

$$\mathbb{E}(Y \cdot h(X) \mid X) = g(X) = h(X) \cdot \mathbb{E}Y \quad \mathbb{P}\text{-f.s.}$$

(ii) Insbesondere ist

$$\underbrace{\mathbb{E}(Y \mid X)}_{\text{Zufallsvariable}} = \underbrace{\mathbb{E}Y}_{\in \mathbb{R}} \quad \text{f.s.},$$

falls X und Y stochastisch unabhängig sind.

Durch Anwendung obiger Überlegungen kann man eine diskrete Variante des sogenannten Satzes von Fubini beweisen.

Satz 3.36 (Fubini diskret)**.** *Seien X und Y stochastisch unabhängige diskrete Zufallsvariablen und $f : \mathbb{R}^2 \to \mathbb{R}$ eine Funktion mit $\mathbb{E}|f(X,Y)| < \infty$. Sei $h(x) := \mathbb{E}(f(x,Y))$. Dann gilt:*

$$\mathbb{E}(f(X,Y)) = \mathbb{E}h(X)$$

Beweis. Nach Satz 3.34 gilt $\mathbb{E}(f(X,Y) \mid X = x) = \mathbb{E}f(x,Y) = h(x)$. Nach Satz 3.33 (i) folgt damit

$$\mathbb{E}h(x) = \mathbb{E}(\mathbb{E}(f(X,Y) \mid X = x)) = \mathbb{E}f(X,Y).$$

Alternativ:

$$\begin{aligned}
\mathbb{E}f(X,Y) &= \sum_{x,y} f(x,y) \cdot p_{X,Y}(x,y) \\
&= \sum_{x,y} f(x,y) \cdot p_X(x) \cdot p_Y(y) && \text{da } X, Y \text{ stochastisch unabhängig sind,} \\
&= \sum_x \underbrace{\left(\sum_y f(x,y) \cdot p_Y(y) \right)}_{=\mathbb{E}f(x,Y) = h(x)} \cdot p_X(x) \\
&= \sum_x h(x) \cdot p_X(x) \\
&= \mathbb{E}h(X) && \text{nach der Transformationsformel.}
\end{aligned}$$

□

Nach Satz 3.11 (iii) ist für eine diskrete Zufallsvariable X mit $\mathbb{E}(X^2) < \infty$ folgende Ungleichung erfüllt:

$$\mathbb{E}\left[(X-a)^2\right] \geq \mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}X)^2\right) \ \forall a \in \mathbb{R}.$$

Dabei gilt Gleichheit im Falle $a = \mathbb{E}X$. Insofern spricht man bei $\mathbb{E}X$ auch vom „besten konstanten Prädiktor“ von X bzgl. der mittleren quadratischen Abweichung. Eine entsprechende Optimalitätsaussage für bedingte Erwartungswerte ist Gegenstand des nächsten Satzes.

Satz 3.37. *Seien X, Y diskrete Zufallsvariablen mit $\mathbb{E}(Y^2) < \infty$ und sei $\varphi : \mathbb{R} \to \mathbb{R}$ eine Funktion, sodass $\mathbb{E}\left((\varphi(X))^2\right) < \infty$ ist. Dann gilt*

$$\mathbb{E}\left((Y - \varphi(X))^2\right) \geq \mathbb{E}\left((Y - \mathbb{E}(Y \mid X))^2\right).$$

Beweis. Wir definieren

$$g(x) := \begin{cases} \mathbb{E}\left[(Y - \varphi(X))^2 \mid X = x\right] & \text{falls } p_X(x) > 0 \\ 0 & \text{falls } p_X(x) = 0. \end{cases}$$

Gemäß Satz 3.33 (i) gilt damit

$$\mathbb{E}\left((Y - \varphi(X))^2\right) = \mathbb{E}\left(\mathbb{E}\left[(Y - \varphi(X))^2 \mid X\right]\right) = \mathbb{E}g(X).$$

Wir zeigen zunächst die Identität

$$\mathbb{E}\left((Y - \varphi(X))^2 \mid X = x\right) = \mathbb{E}\left((Y - \varphi(x))^2 \mid X = x\right). \tag{3.3}$$

Wie im Beweis von Satz 3.34 gilt mit $Z = (Y - \varphi(X))^2$ und $f(x, y) = (y - \varphi(x))^2$:

$$\begin{aligned} \mathbb{E}\left((Y - \varphi(X))^2 \mid X = x_0\right) &= \mathbb{E}(Z \mid X = x_0) \\ &= \sum_y f(x_0, y) \cdot p_{Y|X=x_0}(y) \\ &= \mathbb{E}(f(x_0, Y) \mid X = x_0) \\ &= \mathbb{E}\left((Y - \varphi(x_0))^2 \mid X = x_0\right). \end{aligned}$$

Insbesondere folgt $\mathbb{E}\left[(Y - \mathbb{E}(Y|X))^2 \mid X = x\right] = \mathbb{E}\left[(Y - \mathbb{E}(Y|X = x))^2 \mid X = x\right]$ für die spezielle Wahl $\varphi(X) = \mathbb{E}(Y|X)$, d.h. (3.3) ist verifiziert.

Z: $\mathbb{E}\left[\mathbb{E}\left[(Y - \mathbb{E}(Y|X))^2 \mid X\right]\right] \leq \mathbb{E}\left[\mathbb{E}\left((Y - \varphi(X))^2 \mid X\right)\right]$.

Für jedes x mit $p_X(x) > 0$ gilt

$$\begin{aligned} \mathbb{E}\left((Y - \mathbb{E}(Y|X))^2 \mid X = x\right) &\overset{(3.3)}{=} \mathbb{E}\left((Y - \mathbb{E}(Y|X = x))^2 \mid X = x\right) \\ &\overset{(*)}{\leq} \mathbb{E}\left((Y - \varphi(x))^2 \mid X = x\right) \overset{(3.3)}{=} \mathbb{E}\left((Y - \varphi(X))^2 \mid X = x\right). \end{aligned}$$

Dabei haben wir die Ungleichung $(*)$ bereits in Satz 3.11 (iii) bewiesen. Betrachten wir nun nicht mehr einen konkreten Wert x, sondern die gesamte Funktion, so erhalten wir

$$\mathbb{E}\left((Y-\mathbb{E}(Y|X))^2 \mid X\right) \le \mathbb{E}\left((Y-\varphi(X))^2 \mid X\right) \quad \text{f.s.}$$

Die Behauptung folgt nun aus der Monotonie des Erwartungswertes (Übungsblatt 6, Aufgabe 2 (a)). □

4 Stetige Zufallsvariablen

Diskrete Zufallsvariablen sind oftmals nicht gerechtfertigt. Beispielsweise die Größe von Schulkindern oder die Lebensdauer eine Glühbirne sind kontinuierliche Größen, denn z.B. die Glühbirne muss nicht zu einer exakten Sekunde oder Minute kaputtgehen.

Beispiel 4.1. *Student S wartet auf seine Freundin, die um 16 Uhr kommen möchte. Aus diesem Grund hält er alle Zeitpunkte zwischen 16 und 17 Uhr für gleichwahrscheinlich.*

Idee: (Approximation) Zerlege das Intervall $[a,b]=[16,17]$ in n Teilintervalle derselben Länge mit Mittelpunkten $x_1^n,\dots,x_n^n$ und ordne jedem die Wahrscheinlichkeit $1/n$ zu. Falls die Ankunftszeit X^n nur Werte in $\{x_1^n,\dots,x_n^n\}$ annehmen kann, ergibt sich

$$\mathbb{P}(c \le X^n \le d) = \sum_{c\le x_i^n \le d} \frac{1}{n} = \frac{b-a}{n} \sum_{c\le x_i^n\le d} \frac{1}{b-a} \underset{n\to\infty}{\longrightarrow} \int_c^d \frac{1}{b-a}dx.$$

Wir möchten also ein Wahrscheinlichkeitsmaß (das Bildmaß $\mathbb{P}^X$ der Ankunftszeit X) definieren, welches für alle $a \le c < d \le b$

$$\mathbb{P}^X([c,d]) = \frac{d-c}{b-a}$$

erfüllt. Dieses kann offenbar keinen abzählbaren Träger besitzen, denn $\mathbb{P}^X(\{c\})=0$ für alle $c\in\mathbb{R}$.

4.1 Allgemeine Wahrscheinlichkeitsräume

Will man ein Wahrscheinlichkeitsmaß auf $(\mathbb{R},\mathcal{P}(\mathbb{R}))$ wie in Beispiel 4.1 definieren, treten unüberwindbare mathematische Schwierigkeiten auf, die erzwingen, den Definitionsbereich eines solchen Maßes auf ein kleineres Mengensystem als die Potenzmenge zu beschränken. Wenn man unsere frühere Überlegung mit diskreten Wahrscheinlichkeitsmaßen $\mathbb{P}$ betrachtet, so sieht man allerdings, dass von der Potenzmenge $\mathcal{P}(\Omega)$ lediglich verwendet wurde,

- dass die interessierenden Mengen dazugehörten (insbesondere Ω und die Elemente $\omega\in\Omega$ mit $\mathbb{P}(\{\omega\})>0$),

- und dass man bestimmte Mengenoperationen (i.e. abzählbare Vereinigungen und Durchschnitte, Komplemente, ...) ausführen konnte, sodass $\mathbb{P}(A)$ für die so entstandenen Mengen A definiert war.

Definition 4.2 (σ-Algebra, messbarer Raum)**.** *Sei $\Omega \neq \varnothing$ eine Stichprobenmenge. Eine σ-Algebra über Ω ist ein Mengensystem $\mathscr{A}$ mit den folgenden Eigenschaften:*

(i) $\Omega \in \mathscr{A}$

(ii) $A \in \mathscr{A} \implies A^c \in \mathscr{A}$

(iii) $A_i \in A \;\forall\, i \in \mathbb{N} \implies \bigcup_{i \in \mathbb{N}} A_i \in \mathscr{A}$.

Das Paar $(\Omega, \mathscr{A})$ heißt messbarer Raum. Ein Element $A \in \mathscr{A}$ heißt messbare Menge.

Beispiele 4.3. *(i) Die Potenzmenge $\mathcal{P}(\Omega)$ ist die größte σ-Algebra über Ω, die sogenannte diskrete σ-Algebra.*

(ii) $\{\varnothing, \Omega\}$ ist die kleinste σ-Algebra über Ω, die sogenannte indiskrete σ-Algebra.

(iii) Sei $\mathcal{E}$ ein System von Mengen mit $\mathcal{E} \subset \mathcal{P}(\Omega)$. Dann ist

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathscr{A} \ \sigma\text{-Algebra über } \Omega \\ \mathcal{E} \subset A}} \mathscr{A}$$

ebenfalls eine σ-Algebra (Beweis: Blatt 7, Aufgabe 1). $\sigma(\mathcal{E})$ ist die kleinste σ-Algebra, die $\mathcal{E}$ enthält und wird auch die von $\mathcal{E}$ erzeugte σ-Algebra genannt.

Das nächste Lemma zeigt, dass σ-Algebren auch unter höchstens abzählbaren Durchschnitten und Mengendifferenzbildung abgeschlossen sind.

Lemma 4.4. *Für jede σ-Algebra $\mathscr{A}$ ûber Ω gelten folgende Aussagen:*

(i) $A, B \in \mathscr{A} \implies A \cap B \in \mathscr{A}$

(ii) $A, B \in \mathscr{A} \implies B \setminus A \in \mathscr{A}$

(iii) $A_i \in \mathscr{A} \;\forall\, i \in \mathbb{N} \implies \bigcap_{i \in \mathbb{N}} A_i \in \mathscr{A}$.

Beweis. *(i)* Seien $A, B \in \mathscr{A}$. Dann liegen auch A^c und B^c in $\mathscr{A}$, also auch $A^c \cup B^c$. $\Rightarrow A \cap B = (A^c \cup B^c)^c \in \mathscr{A}$.

(ii) Seien $A, B \in \mathscr{A}$. $\Rightarrow A^c, B^c \in \mathscr{A} \Rightarrow B \setminus A = A^c \cap B \in \mathscr{A}$.

(iii) Sei $A_i \in \mathscr{A} \;\forall\, i \in \mathbb{N}$. Damit ist auch $A_i^c \in \mathscr{A} \;\forall\, i \in \mathbb{N}$.

$$\xRightarrow{(iii)} \bigcup_{i \in \mathbb{N}} A_i^c \in \mathscr{A} \xRightarrow{(ii)} \bigcap_{i \in \mathbb{N}} A_i = \left(\bigcup_{i \in \mathbb{N}} A_i^c \right)^c \in \mathscr{A}$$

$\square$

Es folgen die Axiome von Kolmogorov für Wahrscheinlichkeitsmaße auf messbaren Räumen.

Definition 4.5 (Axiome von Kolmogorov (1933) im allgemeinen Fall)**.** *Sei $\mathscr{A}$ eine σ-Algebra über $\Omega \neq \varnothing$, d.h. das Paar $(\Omega, \mathscr{A})$ bildet einen messbaren Raum. Eine Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{A})$ ist eine Abbildung $\mathbb{P} : \mathscr{A} \to \mathbb{R}$ mit den folgenden Eigenschaften:*

(i) Nichtnegativität: $\mathbb{P}(A) \geq 0 \ \forall\, A \in \mathscr{A}$

(ii) Normiertheit: $\mathbb{P}(\Omega) = 1$

(iii) σ-Additivität: Es gilt für alle paarweise disjunkten Mengen $A_i \in \mathscr{A}$, $i \in \mathbb{N}$:

$$\mathbb{P}\left(\sum_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i)$$

Das Tripel $(\Omega, \mathscr{A}, \mathbb{P})$ heißt Wahrscheinlichkeitsraum. Eine Abbildung $\mu : \mathscr{A} \to \mathbb{R} \cup \{\infty\}$ mit den Eigenschaften (i) und (ii) und $\mu(\varnothing) = 0$ heißt Maß. Das Tripel $(\Omega, \mathscr{A}, \mu)$ heißt dann Maßraum.

Bemerkung 4.6. *Natürlich ist jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ insbesondere ein Maß, denn*

$$\mathbb{P}(\varnothing) = \mathbb{P}\left(\sum_{i \in \mathbb{N}} \varnothing\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(\varnothing) \in \mathbb{R}$$

impliziert $\mathbb{P}(\varnothing) = 0$.

4.2 Die Borel-σ-Algebra über $\mathbb{R}$

In Beispiel 4.1 sind wir natürlicherweise auf die Frage nach einem nicht-diskreten Wahrscheinlichkeitsmaß gestoßen, welches allen Intervallen eine Wahrscheinlichkeit proportional zur Intervalllänge zuordnet. Das legt nahe, als σ-Algebra über $\mathbb{R}$ diejenige zu studieren, welche von den Intervallen erzeugt wird.

Definition 4.7 (Borel-σ-Algebra)**.** *Die Borelsche σ-Algebra $\mathscr{B}(\mathbb{R})$ ist die kleinste σ-Algebra über $\mathbb{R}$, welche das Mengensystem $\mathcal{G}_1 := \{(a, b] \mid a, b \in \mathbb{R} \text{ mit } a \leq b\}$ enthält, d.h.*

$$\mathscr{B}(\mathbb{R}) = \bigcap_{\substack{\mathscr{A}\ \sigma\text{-Algebra über } \mathbb{R} \\ \mathcal{G}_1 \subset A}} \mathscr{A}$$

gemäß Beispiel 4.3.

Bemerkung. *Man schreibt auch $\mathscr{B}$ statt $\mathscr{B}(\mathbb{R})$, wenn kein Zweifel darüber besteht, dass $\mathbb{R}$ als Stichprobenraum gemeint ist.*

Das nächste Lemma zeigt, dass die Borel-σ-Algebra über $\mathbb{R}$ alle „praktisch relevanten" Teilmengen von $\mathbb{R}$ entält.

Lemma 4.8 (Eigenschaften von $\mathscr{B}(\mathbb{R})$)**.** *Die folgenden Mengen liegen in* $\mathscr{B}(\mathbb{R})$*:*

(*i*) *Jede ein-elementige Teilmenge von* $\mathbb{R}$*.*

(*ii*) *Jede endliche Teilmenge von* $\mathbb{R}$*.*

(*iii*) *Jede abzählbare Teilmenge von* $\mathbb{R}$*.*

(*iv*) *Jedes abgeschlossene, halb-offene oder offene Intervall von* $\mathbb{R}$*.*

(*v*) *Jede offene Teilmenge von* $\mathbb{R}$*.*

(*vi*) *Jede abgeschlossene Teilmenge von* $\mathbb{R}$*.*

Beweis. Übungsblatt 7, Aufgabe 2. □

Bemerkung 4.9. *Es gilt* $\mathscr{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$ *(hier ohne Beweis).*

4.3 Stetige Verteilungen und stetige Zufallsvariablen

Proposition und Definition 4.10. *Sei* $f : \mathbb{R} \to \mathbb{R}$ *Riemann-integrierbar mit* $f \geq 0$ *und* $\int_{-\infty}^{\infty} f(x)\, \mathrm{d}x = 1$*. Dann wird durch*

$$\mathbb{P}([a,b]) := \int_a^b f(x)\, \mathrm{d}x \quad \forall\, a \leq b$$

ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B})$ *definiert.* $\mathbb{P}$ *heißt* *stetige Verteilung.*

Beweis. Dass $\mathbb{P}$ eindeutig definiert und ein Wahrscheinlichkeitsmaß ist, wird in der Maßtheorie (Analysis III) bewiesen. □

Bemerkung 4.11. (*i*) *Es gilt* $\forall\, a \in \mathbb{R}$*:*

$$\mathbb{P}(\{a\}) = \int_a^a f(x)\, \mathrm{d}x = 0.$$

(*ii*) *Die Funktion* $F : \mathbb{R} \to \mathbb{R}$ *mit*

$$F(x) := \int_{-\infty}^{x} f(y)\, \mathrm{d}y$$

heißt *Verteilungsfunktion von* $\mathbb{P}$*. Ist* f *stetig, so gilt*

$$f(x) = F'(x)$$

nach dem Hauptsatz der Differential- und Integralrechnung.

(iii) *F ist monoton wachsend und (insbesondere) rechtsseitig stetig mit*

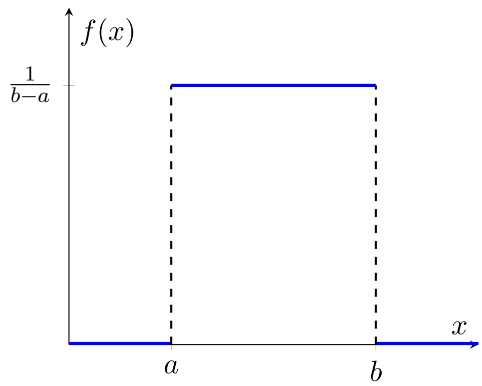
$$\lim_{x\to-\infty} F(x) = 0 \text{ und } \lim_{x\to\infty} F(x) = 1.$$

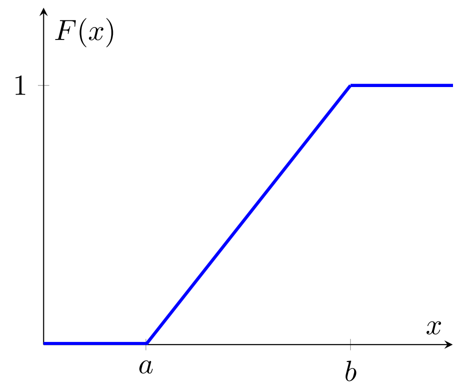
(iv) *f aus Definition 4.10 wird auch als <u>(Riemann-)Dichte von $\mathbb{P}$</u> bezeichnet.*

Beispiel 4.12. *(i) Die Gleichverteilung (auch „uniforme Verteilung“ genannt) auf $[a, b]$ (häufig mit $\mathcal{U}[a, b]$ bezeichnet):*

$$f(x) = \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x) \;\forall\; x \in \mathbb{R}$$

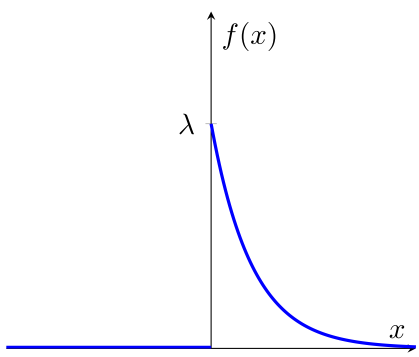
$$F(x) = \int_{-\infty}^{x} f(x)\ \mathrm{d}x = \begin{cases} 0 & \textit{für } x \leq a \\ \frac{x-a}{b-a} & \textit{für } a < x < b \\ 1 & \textit{für } x \geq b \end{cases}$$





(ii) Die <u>Exponentialverteilung</u> $\mathscr{E}(\lambda)$ ist gegeben durch

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \textit{für } x \geq 0 \\ 0 & \textit{für } x < 0. \end{cases}$$



Wir überprüfen, dass diese Funktion f tatsächlich eine Zähldichte ist:

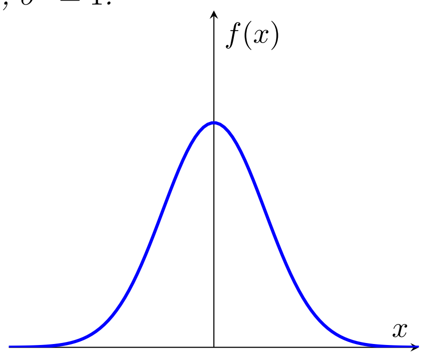
$$\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = \int_{0}^{\infty} \lambda \cdot e^{-\lambda x} = -e^{-\lambda x}|_0^\infty = 1.$$

Entsprechend erhält man auch die Verteilungsfunktion $F(x) = \mathbb{1}_{[0,\infty)}(x)(1 - e^{\lambda x})$.

(*iii*) *Die Normalverteilung $\mathscr{N}(\mu, \sigma^2)$ (auch Gaußverteilung genannt) ist gegeben durch die Dichte*

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

Beispielsweise für $\mu = 0$, $\sigma^2 = 1$:



Wir nutzen auch die folgenden Bezeichnungen:

$$\varphi(x) := f_{0,1}(x), \quad \Phi(x) := \int_{-\infty}^{x} \varphi(x)\ \mathrm{d}x.$$

Definition 4.13. *Für einen Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ heißt eine Abbildung $X : \Omega \to \mathbb{R}$ mit*

$$X^{-1}(B) \in \mathscr{A} \quad \forall\, B \in \mathscr{B}$$

(reellwertige) Zufallsvariable mit Bildmaß $\mathbb{P}^X$, gegeben durch

$$\mathbb{P}^X(B) := \mathbb{P}(X^{-1}(B)) \quad \forall\, B \in \mathscr{B}.$$

Besitzt $\mathbb{P}^X$ eine Riemann-Dichte f_X, d.h. es gilt

$$\mathbb{P}^X([a,b]) = \int_a^b f_X(x)\ \mathrm{d}x \quad \forall\, a, b \in \mathbb{R}, a < b,$$

so nennt man X stetig verteilte Zufallsvariable bzw. stetige Zufallsvariable.

$$\begin{tikzcd}
(\Omega, \mathscr{A}, \mathbb{P}) \arrow[rr, bend left, "X"] & & (\mathbb{R}, \mathscr{B}, \mathbb{P}^X)
\end{tikzcd}$$

Bemerkung 4.14 (Lineare Transformation von Zufallsvariablen)**.** *Sei X stetig verteilt mit stetiger Riemann-Dichte f_X und Verteilungsfunktion F_X (bspw. $X \sim \mathscr{N}(\mu, \sigma^2)$). Sei $Y = aX + b$ für ein $a > 0$ und ein $b \in \mathbb{R}$. Gesucht ist die Verteilung der Zufallsvariable Y. Es gilt:*

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

$$\implies f_Y(y) = F_Y'(y) = \frac{1}{a} \cdot F_X'\left(\frac{y-b}{a}\right) = \frac{1}{a} \cdot f_X\left(\frac{y-b}{a}\right)$$

Im Spezialfall $X \sim \mathscr{N}(0,1)$ ist damit $Y = \sigma X + \mu \sim \mathscr{N}(\mu, \sigma^2)$. Allgemeiner gilt im Fall $X \sim \mathscr{N}(\mu, \sigma^2)$, $Y = aX + b$:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \frac{1}{a\sqrt{\sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \frac{\left(\frac{y-b}{a} - \mu\right)^2}{\sigma^2}\right) \\ &= \frac{1}{2\pi} \frac{1}{\sqrt{a^2\sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{y - b - a\mu}{a\sigma}\right)^2\right), \end{aligned}$$

d.h. $Y \sim \mathscr{N}(a\mu + b, a^2\sigma^2)$.

Satz 4.15. *Sei X eine stetige Zufallsvariable mit Dichte f_X und $Y = g(X)$, wobei g eine streng monotone, differenzierbare Funktion sei. Dann ist auch Y stetig verteilt mit Dichte*

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right|$$

Beweis. Sei g streng monoton wachsend. Dann gilt:

$$\begin{aligned} \mathbb{P}^Y([a,b]) &= \mathbb{P}(Y \in [a,b]) \\ &= \mathbb{P}(g(X) \in [a,b]) \\ &= \mathbb{P}(X \in [g^{-1}(a), g^{-1}(b)]) = \int_{g^{-1}(a)}^{g^{-1}(b)} f_X(x)\ \mathrm{d}x = \int_a^b f_X(g^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\ \mathrm{d}y \end{aligned}$$

$$\implies \mathbb{P}(Y \in [a,b]) = \int_a^b \underbrace{f_X(g^{-1}(y)) \cdot \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)}_{\text{Riemann-Dichte von } Y}\ \mathrm{d}y \quad \forall\, a \leq b$$

□

Beispiel 4.16. *Betrachte $X \sim \mathscr{N}(0,1)$, $Y = X^2$. Frage: wie sieht f_Y aus? Das Problem hier ist, dass $g(x) = x^2$ nicht streng monoton ist. Wir wissen allerdings, dass $Y \geq 0$ ist,* *womit*

$$
\begin{aligned}
\mathbb{P}(Y \in [a,b]) &= \mathbb{P}(X \in [\sqrt{a}, \sqrt{b}]) + \mathbb{P}(X \in [-\sqrt{b}, -\sqrt{a}]) \\
&= 2 \cdot \mathbb{P}(X \in [\sqrt{a}, \sqrt{b}]) \qquad \text{da } \mathscr{N}(0,1) \text{ symmetrisch um den Punkt 0 ist} \\
&= 2 \cdot \int_{\sqrt{a}}^{\sqrt{b}} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right) \\
&= 2 \cdot \int_a^b \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-y}{2}\right) \cdot \frac{1}{2} \cdot y^{-1/2} \, \mathrm{d}y \qquad \text{durch die Substitution } y = \sqrt{x}
\end{aligned}
$$

$$
\implies f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \cdot e^{-y/2} \sqrt{y}^{-1} & \textit{für } y \geq 0 \\ 0 & \textit{für } y < 0. \end{cases}
$$

Y heißt dann χ^2-verteilt mit einem Freiheitsgrad $Y \sim \chi_1^2$.

4.3.1 Erwartungswert und Varianz

Definition 4.17. *Sei X eine stetige Zufallsvariable mit Riemann-Dichte f_X. Dann heißt*

$$
\mathbb{E}X := \int_{-\infty}^{\infty} x \cdot f_X(x) \, \mathrm{d}x
$$

der Erwartungswert von X, sofern

$$
\int_{-\infty}^{\infty} |x| \cdot f_X(x) \, \mathrm{d}x < \infty
$$

ist. Die Varianz wird wie gewohnt definiert als

$$
\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].
$$

Beispiel 4.18. *Für $X \sim \mathscr{N}(\mu, \sigma^2)$ ist*

$$
\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{\infty} x \cdot f_X(x) \, \mathrm{d}x \\
&= \underbrace{\int_{-\infty}^{\infty} \overset{\text{ungerade}}{\overset{\downarrow}{y}} \cdot \overbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot y^2\right)}^{\text{Symmetrisch um 0}} \, \mathrm{d}y}_{=0} + \mu \cdot \underbrace{\int_{-\infty}^{\infty} \overbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot y^2\right)}^{\text{Dichte der } \mathscr{N}(0,\sigma^2)\text{-Verteilung}} \, \mathrm{d}y}_{=1} \\
&\quad \text{durch die Substitution } y = x - \mu \\
&= \mu.
\end{aligned}
$$

Sinngemäß ist also der Erwartungswert bei stetigen Zufallsvariablen genauso definiert wie bei diskreten, nur wird eben die Summe durch ein Integral ersetzt. Analog zu Satz 3.7 gilt auch die Transformationsformel: Sei X stetig verteilt mit Riemann-Dichte f_X und g derart, dass die Integrale

$$\int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, \mathrm{d}x \text{ und } \int_{-\infty}^{\infty} |g(x)| \cdot f_X(x) \, \mathrm{d}x$$

existieren (und endlich sind). Dann gilt:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, \mathrm{d}x$$

4.3.2 Mehrdimensionale stetige Verteilungen

Definition 4.19. *Seien $X_1, \ldots, X_n$ Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum, und*

$$F(x_1, \ldots, x_n) := \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

die gemeinsame Verteilungsfunktion dieser Variablen. Ist $f : \mathbb{R}^n \to \mathbb{R}$ eine Funktion, sodass

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(y_1, \ldots, y_n) \, \mathrm{d}y_n \ldots \mathrm{d}y_1$$

für alle $x_1, \ldots, x_n \in \mathbb{R}$ gilt, so heißt f gemeinsame Dichte von $X_1, \ldots, X_n$. Für $I \subset \{1, \ldots, n\}$ ist

$$F(x_i : i \in I) = \int \ldots \int f(x_1, \ldots, x_n) \cdot \prod_{j \in I^C} \mathrm{d}x_j$$

die Marginaldichte von $(X_i)_{i \in I}$.

Beispiel 4.20. *Seien X_1, X_2 gemeinsam stetig verteilt mit Dichte f. Dann gilt für alle $a, b, c, d \in \mathbb{R}$ mit $a \leq b$ und $c \leq d$:*

$$\begin{aligned}
&\mathbb{P}(X_1 \in (a, b], X_2 \in (c, d]) \\
&\quad = \mathbb{P}(X_1 \leq b, X_2 \leq d) - \mathbb{P}(X_1 \leq a, X_2 \leq d) \\
&\qquad - \mathbb{P}(X_1 \leq b, X_2 \leq c) + \mathbb{P}(X_1 \leq a, X_2 \leq c) \\
&\quad = F(b, a) - F(a, d) - F(b, c) + F(a, c) \\
&\quad = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 - \ldots = \ldots = \int_{a}^{b} \int_{c}^{d} f(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2.
\end{aligned}$$

Allgemeiner: Sind $X_1, \ldots, X_n$ gemeinsam stetig verteilt mit Dichte $f_{X_1, \ldots, X_n}$, so ist

$$\mathbb{P}(X_1 \in (a_1, b_1], \ldots, X_n \in (a_n, b_n]) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \, \mathrm{d}x_n \ldots \, \mathrm{d}x_1.$$

Das nächste Beispiel illustriert, dass verschiedene gemeinsame Verteilungen ein- und dieselben Marginalverteilungen besitzen können.

Beispiel 4.21. *(i) Seien X, Y gemeinsam stetig verteilt mit Dichte*

$$f_{X,Y}(x,y) = \begin{cases} 2x + 2y - 4xy & \forall\,(x,y) \in [0,1]^2 \\ 0 & \text{sonst.} \end{cases}$$

Es gilt $f_{X,Y}(0,0) = 0$, und wir erhalten die Marginaldichten

$$\begin{aligned} f_X(x) &= \int_0^1 (2x + 2y - 4xy)\,\mathrm{d}y = 2x + 1 - 2x = 1\ \forall\, x \in [0,1], \\ f_Y(y) &= \int_0^1 (2x + 2y - 4xy)\,\mathrm{d}x = \cdots = 1\ \forall\, y \in [0,1]. \end{aligned}$$

Damit ist $X \sim \mathcal{U}[0,1]$ und $Y \sim \mathcal{U}[0,1]$.

(ii) Seien X, Y gemeinsam stetig verteilt mit Dichte

$$f_{X,Y}(x,y) = \begin{cases} 2 - 2x - 2y + 4xy & \forall\,(x,y) \in [0,1]^2 \\ 0 & \text{sonst.} \end{cases}$$

Hier sind $f_{X,Y}(0,0) = 2$ und

$$\begin{aligned} f_X(x) &= \int_0^1 (2 - 2x - 2y + 4xy)\,\mathrm{d}y = 1\ \forall\, x \in [0,1], \\ f_Y(y) &= \int_0^1 (2 - 2x - 2y + 4xy)\,\mathrm{d}x = 1\ \forall\, y \in [0,1]. \end{aligned}$$

Also gilt erneut $X \sim \mathcal{U}[0,1]$ und $Y \sim \mathcal{U}[0,1]$.

(iii) Seien nun X, Y gemeinsam stetig verteilt mit Dichte

$$f_{X,Y}(x,y) = \begin{cases} 1 & \forall\,(x,y) \in [0,1]^2 \\ 0 & \text{sonst.} \end{cases}$$

Einmal mehr bestimmen wir

$$f_X(x) = \int_0^1 1\,\mathrm{d}y = 1\ \forall\, x \in [0,1] \quad \textit{und} \quad f_Y(y) = \int_0^1 1\,\mathrm{d}x = 1\ \forall\, y \in [0,1],$$

also erneut $X \sim \mathcal{U}[0,1]$ und $Y \sim \mathcal{U}[0,1]$.

Obwohl in allen drei Fällen die (eindimensionalen) Marginalverteilungen identisch sind, ist die gemeinsame Verteilung in allen 3 Fällen unterschiedlich.

4.3.3 Stochastische Unabhängigkeit und bedingte Dichten

Definition 4.22. *Seien $X_1, \dots, X_n$ gemeinsam stetig verteilt mit Dichte $f_{X_1,\dots,X_n}$. Dann heißen $X_1, \dots, X_n$ stochastisch unabhängig, falls*

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) \quad \forall\, x_1, \dots, x_n \in \mathbb{R} \text{ gilt.}$$

Wie bei diskreten Zufallsvariablen faktorisiert die Dichte endlich vieler gemeinsam stetig verteilter stochastisch unabhängiger Zufallsvariablen.

Bemerkung 4.23. *Zusammen mit Beispiel 4.20 gilt damit im Falle der stochastischen Unabhängigkeit von X und Y:*

$$\begin{aligned}
\mathbb{P}(X \in (a,b], Y \in (c,d]) &= \int_c^d \int_a^b f_{X,Y}(x,y) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_c^d \int_a^b f_X(x) \cdot f_Y(y) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_c^d f_Y(y) \cdot \left(\int_a^b f_X(x) \,\mathrm{d}x \right) \,\mathrm{d}y \\
&= \left(\int_a^b f_X(x) \,\mathrm{d}x \right) \cdot \left(\int_c^d f_Y(y) \,\mathrm{d}y \right) \\
&= \mathbb{P}(X \in (a,b]) \cdot \mathbb{P}(Y \in (c,d])
\end{aligned}$$

Proposition und Definition 4.24. *Seien $X, Y : \Omega \to \mathbb{R}$ gemeinsam stetig verteilt mit Dichte $f_{X,Y}$, sowie*

$$f_{X|Y=y}(x) := \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & \text{falls } f_Y(y) \neq 0 \\ 0 & \text{sonst.} \end{cases}$$

Dann ist $f_{X|Y=y}(x) \geq 0$. Ferner gilt:

$$\int_{-\infty}^{\infty} f_{X|Y=y}(x) \,\mathrm{d}x = 1 \;\forall\, y \in \mathbb{R} \text{ mit } f_Y(y) > 0$$

und

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y=y}(x) \cdot f_Y(y) \,\mathrm{d}y.$$

Die Funktion $f_{X|Y=y}$ heißt bedingte (Riemann-)Dichte von X gegeben $Y = y$.

Beweis. Übungsaufgabe. □

Wie im diskreten Fall kann man nun auch bedingte Erwartungen, bedingte Varianzen, etc. definieren. Wir führen die Analogien an dieser Stelle aber nicht weiter aus.

5 Konvergenz von Zufallsvariablen und zentraler Grenzwertsatz

Definition 5.1. *Sei $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, X und X_n für $n \in \mathbb{N}$ reellwertige Zufallsvariablen. Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ für $n \to \infty$ gegen X*

(i) fast sicher, falls
$$\mathbb{P}(\{\omega \in \Omega \mid X_n(\omega) \to X(\omega)\}) = 1$$
ist. Wir schreiben dann: „$X_n \xrightarrow{\text{f.s.}} X$".

(ii) in Wahrscheinlichkeit oder stochastisch, falls $\forall\, \varepsilon > 0$ gilt:
$$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n\to\infty} 0.$$
Wir schreiben dann „$X_n \xrightarrow{\mathbb{P}} X$".

(iii) im p-ten Mittel (für ein $p \in \mathbb{R}$, $p \geq 1$), falls $\mathbb{E}\,[|X|^p] < \infty$ und $\mathbb{E}\,[|X_n|^p] < \infty$ $\forall\, n \in \mathbb{N}$ ist, und gilt:
$$\mathbb{E}\,[|X_n - X|^p] \xrightarrow{n\to\infty} 0.$$

(iv) in Verteilung, falls
$$F_{X_n}(x) \xrightarrow{n\to\infty} F_X(x)$$
für alle Stetigkeitspunkte x von F_X gilt. Wir schreiben dann „$X_n \xrightarrow{\mathscr{D}} X$".

Offenbar gelten (i) $\Rightarrow$ (ii) und nach der Markov-Ungleichung (iii) $\Rightarrow$ (ii). Weitere Implikationen werden später untersucht.

Beispiel 5.2. *Für $n \in \mathbb{N}$ sei X_n Laplace-verteilt auf $\left\{\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, 1\right\}$. Sei außerdem $X \sim \mathcal{U}[0,1]$. X ist also stetig verteilt, wohingegen die X_n für alle $n \in \mathbb{N}$ diskret sind. Für $t < 0$ ist $F_{X_n}(t) = F_X(t) = 0$. Für $t > 1$ ist $F_{X_n}(t) = F_X(t) = 1$. Für $t \in [0,1]$ gilt*
$$F_{X_n}(t) = \frac{\lfloor n \cdot t \rfloor}{n} \xrightarrow{n\to\infty} t = F_X(t),$$
also insgesamt $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$ $\forall\, t \in \mathbb{R}$ und damit $X_n \xrightarrow{\mathscr{D}} X$.

5.1 Zusammenhänge zwischen schwacher und stochastischer Konvergenz

Proposition 5.3. *Seien $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ Folgen von Zufallsvariablen und X eine Zufallsvariable mit $X_n \xrightarrow{\mathscr{D}} X$ und $X_n - Y_n \xrightarrow{\mathbb{P}} 0$. Dann gilt auch $Y_n \xrightarrow{\mathscr{D}} X$.*

Beweis. Z.z.: $F_{Y_n}(x) \xrightarrow{n\to\infty} F_X(x)$ für alle Stetigkeitspunkte x von F_X.

Sei $Z_n := X_n - Y_n$, d.h. $Z_n \xrightarrow{\mathbb{P}} 0$. Seien weiter $x \in \mathbb{R}$ und eine Nullfolge $(\varepsilon_k)_{k\in\mathbb{N}}$ mit $\varepsilon_k \searrow 0$ so gewählt, dass x und $x \pm \varepsilon_k$ $\forall\, k \in \mathbb{N}$ Stetigkeitspunkte von F_X sind. Das ist möglich, da F_X höchstens abzählbar viele Sprungstellen besitzt. Dann gilt

$$\begin{aligned}
F_{Y_n}(x) &= \mathbb{P}(Y_n \leq x) = \mathbb{P}(X_n - Z_n \leq x) = \mathbb{P}(X_n \leq x + Z_n) \\
&= \mathbb{P}(X_n \leq x + Z_n, |Z_n| < \varepsilon_k) + \mathbb{P}(X_n \leq x + Z_n, |Z_n| \geq \varepsilon_k) \\
&\leq \underbrace{\mathbb{P}(X_n \leq x + \varepsilon_k)}_{\xrightarrow{n\to\infty} F_X(x+\varepsilon_k)} + \underbrace{\mathbb{P}(|Z_n| \geq \varepsilon_k)}_{\xrightarrow{n\to\infty} 0}
\end{aligned}$$

nach Voraussetzung, also

$$\limsup_{n\to\infty} F_{Y_n}(x) \leq F_X(x + \varepsilon_k) \quad \forall\, k \in \mathbb{N}.$$

Gleichermaßen folgt

$$\liminf_{n\to\infty} F_{Y_n}(x) \geq F_X(x - \varepsilon_k) \quad \forall\, k \in \mathbb{N},$$

womit $\lim_{n\to\infty} F_{Y_n}(x) = F_X(x)$, da x Stetigkeitspunkt fon F_X ist. □

Proposition 5.4. *Seien $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ Folgen von Zufallsvariablen und $c \in \mathbb{R}$ eine Konstante. Dann gilt:*

(*i*) $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathscr{D}} X$

(*ii*) $X_n \xrightarrow{\mathscr{D}} X, Y_n \xrightarrow{\mathbb{P}} 0 \implies X_n \cdot Y_n \xrightarrow{\mathbb{P}} 0$

(*iii*) $X_n \xrightarrow{\mathscr{D}} X, Y_n \xrightarrow{\mathbb{P}} c \implies X_n + Y_n \xrightarrow{\mathscr{D}} X + c, X_n \cdot Y_n \xrightarrow{\mathscr{D}} c \cdot X$ *und* $\frac{X_n}{Y_n} \xrightarrow{\mathscr{D}} \frac{X}{c}$ *falls* $c \neq 0$.

Beweis. (*i*) Folgt aus Proposition 5.3.

(*ii*) Sei $\varepsilon, k > 0$. Wir zerlegen $\mathbb{P}(|X_n \cdot Y_n| \geq \varepsilon)$ wie folgt:

$$\begin{aligned}
\mathbb{P}(|X_n \cdot Y_n| \geq \varepsilon) &= \mathbb{P}\left(|X_n \cdot Y_n| \geq \varepsilon, |Y_n| < \frac{\varepsilon}{k}\right) + \mathbb{P}\left(|X_n \cdot Y_n| \geq \varepsilon, |Y_n| \geq \frac{\varepsilon}{k}\right) \\
&\leq \underbrace{\mathbb{P}(|X_n| \geq k)}_{\substack{\xrightarrow{n\to\infty} \mathbb{P}(|X| \geq k) \\ \text{falls } -k \text{ und } k \\ \text{Stetigkeitspunkte} \\ \text{von } F_X \text{ sind}}} + \underbrace{\mathbb{P}\left(|Y_n| \geq \frac{\varepsilon}{k}\right)}_{\xrightarrow{n\to\infty} 0},
\end{aligned}$$

also

$$\limsup_{n\to\infty} \mathbb{P}(|X_n \cdot Y_n| \geq \varepsilon) \leq \mathbb{P}(|X| \geq k),$$

sofern $-k$ und k Stetigkeitspunkte von F_X sind. Außerdem gilt $\mathbb{P}(|X| > k) \xrightarrow{k\to\infty} 0$ nach Lemma 1.8 (v) (Stetigkeit von oben), womit

$$\limsup_{n\to\infty} \; \mathbb{P}(|X_n \cdot Y_n| \geq \varepsilon) = 0.$$

(iii) Aus $X_n \xrightarrow{\mathscr{D}} X$ folgt bereits $X_n + c \xrightarrow{\mathscr{D}} X + c$, denn

$$F_{X_n+c}(x) = \mathbb{P}(X_n \leq x - c) = F_{X_n}(x - c) \underset{n\to\infty}{\longrightarrow} F_X(x - c) = F_{X+c}(x),$$

falls $x - c$ Stetigkeitspunkt von F_X (d.h. x Stetigkeitspunkt von F_{X+c}) ist.
$\implies$ $(X_n + Y_n) - (X_n + c) = Y_n - c \xrightarrow{\mathbb{P}} 0 \overset{5.3}{\Longrightarrow} X_n + Y_n \xrightarrow{\mathscr{D}} X + c$.

Aus $X_n \xrightarrow{\mathscr{D}} X$ folgt entsprechend $c \cdot X_n \xrightarrow{\mathscr{D}} c \cdot X$, denn

$$F_{cX}(x) = \mathbb{P}(cX \leq x) = \begin{cases} F_X(x/c) & \text{falls } c > 0 \\ \mathbb{1}_{[0,\infty)} & \text{falls } c = 0 \\ 1 - F_X(x/c) & \substack{\text{falls } c<0 \text{ und } x/c \\ \text{Stetigkeitspunkt von } F_X \text{ ist.}} \end{cases}$$

Damit ergibt sich $X_n \cdot Y_n - X_n \cdot c = X_n \cdot (Y_n - c) \xrightarrow{\mathbb{P}} 0$ nach Proposition 5.4 (ii) und folglich $X_n \cdot Y_n \xrightarrow{\mathscr{D}} X \cdot c$ nach Proposition 5.3. Die letzte Aussage folgt analog.

□

Beispiel 5.5. *Seien $X_n \sim \mathscr{N}(0, \sigma_n^2)$, $n \in \mathbb{N}$, mit $\lim_{n\to\infty} \sigma_n^2 = 0$. Wegen*

$$\sigma_n^2 = \mathbb{E}(X_n^2) = \mathbb{E}((X_n - 0)^2) \xrightarrow{n\to\infty} 0$$

konvergiert (X_n) im quadratischen Mittel gegen 0. Damit folgt auch

$$\mathbb{P}(|X_n - 0| \geq \varepsilon) \leq \frac{\mathbb{E}(|X_n - 0|^2)}{\varepsilon^2} \xrightarrow{n\to\infty} 0 \quad \forall\, \varepsilon > 0$$

nach der Markov-Ungleichung, also $X_n \xrightarrow{\mathbb{P}} 0$. Proposition 5.4 (i) impliziert schließlich $X_n \xrightarrow{\mathscr{D}} 0$.

Beispiel 5.6. *Sei $X \sim \mathscr{N}(0,1)$ und $X_n := -X \;\forall\, n \in \mathbb{N}$. Die Standardnormalverteilung ist symmetrisch um den Punkt 0, also gilt auch $X_n \sim \mathscr{N}(0,1) \;\forall\, n \in \mathbb{N}$ und insbesondere $X_n \xrightarrow{\mathscr{D}} X$. Wir zeigen nun, dass (X_n) nicht stochastisch gegen X konvergiert:*

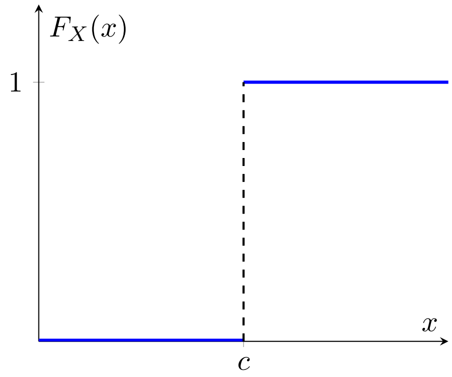
$$\begin{aligned}\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(2 \cdot |X| \geq \varepsilon) &= \mathbb{P}(2X \geq \varepsilon) + \mathbb{P}(2X \leq -\varepsilon) \\ &= \left(1 - \Phi\left(\frac{\varepsilon}{2}\right)\right) + \Phi\left(-\frac{\varepsilon}{2}\right) \neq 0.\end{aligned}$$

Die Umkehrung von Proposition 5.4 (i) gilt also im Allgemeinen nicht. Bemerkenswerterweise ist sie aber gültig, wenn X konstant ist.

Proposition 5.7. *Seien $X_1, X_2, \ldots$ Zufallsvariablen und $c \in \mathbb{R}$ eine Konstante mit $X_n \xrightarrow{\mathscr{D}} c$. Dann gilt auch $X_n \xrightarrow{\mathbb{P}} c$.*

Beweis. Sei $X_n \xrightarrow{\mathscr{D}} c$. Dann gilt $\forall t \neq c$:

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t) = \begin{cases} 1 & t > c \\ 0 & t < c \end{cases}$$



Hierbei lassen wir den Punkt c aus, da die Funktion F_X im Punkt c unstetig ist, d.h. die Verteilungskonvergenz garantiert nicht, dass obige Aussage auch für $t = c$ gilt.
Sei $\varepsilon > 0$ beliebig. Dann gilt

$$\begin{aligned} \mathbb{P}(|X_n - X| > \varepsilon) &= \mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon) \\ &\leq \mathbb{P}(X_n \leq c - \varepsilon) + 1 - \mathbb{P}(X_n \leq c + \varepsilon) = \underbrace{F_{X_n}(c - \varepsilon)}_{\to 0} + 1 - \underbrace{F_{X_n}(c + \varepsilon)}_{\to 1}, \end{aligned}$$

also $\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \varepsilon) = 0$, d.h. $X_n \xrightarrow{\mathbb{P}} X$. $\square$

5.2 Der zentrale Grenzwertsatz

Seien $X_1, \ldots, X_n$ iid mit $\mathbb{E}X_i =: \mu$ und $\mathrm{Var}(X_i) =: \sigma^2 < \infty$. Wir definieren

$$\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Es ist bereits bekannt, dass $\overline{X_n} \xrightarrow{\mathbb{P}} \mu$ gilt (Satz 3.13). Nach Bemerkung 4.14 gilt zudem speziell für $X_i \sim \mathscr{N}(\mu, \sigma^2)$:

$$\overline{X_n} \sim \mathscr{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ bzw. } \sqrt{n} \cdot \frac{\overline{X_n} - \mu}{\sigma} \sim \mathscr{N}(0,1).$$

Im allgemeinen Fall werden wir zeigen, dass die Verteilung von $\sqrt{n}\frac{\overline{X_n}-\mu}{\sigma}$ im Sinne der Konvergenz aus Definition 5.1 (iv) gegen die $\mathscr{N}(0,1)$-Verteilung konvergiert, d.h. dass gilt:

$$\mathbb{P}\left(\sqrt{n}\cdot\frac{\overline{X_n}-\mu}{\sigma}\le x\right)\xrightarrow{n\to\infty}\Phi(x)\quad\forall\, x\in\mathbb{R}$$

mit der Verteilungsfunktion Φ der Standardnormalverteilung $\mathcal{N}(0,1)$. Wegen

$$\frac{\overline{X_n}-\mu}{\sigma}=\frac{1}{n}\cdot\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}\quad\text{und}\quad\mathbb{E}\Big(\frac{X_i-\mu}{\sigma}\Big)=0,\ \mathrm{Var}\Big(\frac{X_i-\mu}{\sigma}\Big)=1$$

und können wir o.B.d.A. annehmen, dass $\mu=0$ und $\sigma^2=1$ ist.

Satz 5.8 (Lindeberg-Trick). *Seien $X_1,\dots,X_n$ iid mit $\mathbb{E}X_i=0$ und $\mathrm{Var}(X_i)=1$. Sei $f:\mathbb{R}\to\mathbb{R}$ zweimal stetig differenzierbar, sodass f, f' und f'' beschränkt sind und f'' gleichmäßig stetig ist. Wir definieren $\forall\, n\in\mathbb{N}$:*

$$Z_n:=\sqrt{n}\cdot\overline{X_n}=\frac{1}{\sqrt{n}}\cdot\sum_{i=1}^{n}X_i$$

Dann gilt $\mathbb{E}[f(Z_n)]\xrightarrow{n\to\infty}\mathbb{E}[f(Z)]$ mit $Z\sim\mathcal{N}(0,1)$.

Beweis. Seien $Y_1,\dots,Y_n$ stochastisch unabhängig $\mathcal{N}(0,1)$-verteilt, sodass $X_1,\dots,X_n$, $Y_1,\dots,Y_n$ gemeinsam stochastisch unabhängig sind. Es gilt $\sqrt{n}\overline{Y_n}\sim\mathcal{N}(0,1)$, d.h. $\mathbb{E}[f(Z)]=\mathbb{E}[f(\sqrt{n}\cdot\overline{Y_n})]$. Wir ersetzen nun nacheinander X_i durch Y_i:

$$\begin{aligned}
f(Z_n)-f(\sqrt{n}\cdot\overline{Y_n})&=f\left(\frac{X_1+\ldots+X_n}{\sqrt{n}}\right)-f\left(\frac{Y_1+X_2+\ldots+X_n}{\sqrt{n}}\right)\\
&\quad+f\left(\frac{Y_1+\ldots+X_n}{\sqrt{n}}\right)-f\left(\frac{Y_1+Y_2+X_3+\ldots+X_n}{\sqrt{n}}\right)\\
&\quad+\ldots\\
&\quad+f\left(\frac{Y_1+\ldots+Y_{n-1}+X_n}{\sqrt{n}}\right)-f\left(\frac{Y_1+\ldots+Y_n}{\sqrt{n}}\right)\\
&=V_1+\ldots+V_n
\end{aligned}$$

mit

$$V_i:=f\left(U_i+\frac{X_i}{\sqrt{n}}\right)-f\left(U_i+\frac{Y_i}{\sqrt{n}}\right)\ \text{ und } U_i:=\frac{Y_1+\ldots+Y_{i-1}+X_{i+1}+\ldots+X_n}{\sqrt{n}}.$$

Mithilfe der Taylor-Entwicklung nach U_i folgt

$$\begin{aligned}
V_i&=f\left(U_i+\frac{X_i}{\sqrt{n}}\right)-f\left(U_i+\frac{Y_i}{\sqrt{n}}\right)\\
&=\frac{X_i-Y_i}{\sqrt{n}}f'(U_i)+\frac{1}{2n}\cdot X_i^2\cdot f''\left(U_i+\frac{\theta_1\cdot X_i}{\sqrt{n}}\right)-\frac{1}{2n}\cdot Y_i^2\cdot f''\left(U_i+\frac{\theta_2\cdot Y_i}{\sqrt{n}}\right),
\end{aligned}$$

wobei $\theta_1, \theta_2 \in [0,1]$ geeignete Zufallsvariablen sind. Wir betrachten nun den Stetigkeitsmodul

$$\delta(h) := \sup_{\substack{x,y\in\mathbb{R}\\ |x-y|\leq h}} |f''(x) - f''(y)|.$$

Nach Voraussetzung ist f'' gleichmäßig stetig, womit $\delta(h) \to 0$ für $h \to 0$. Damit ist

$$V_i = \frac{X_i - Y_i}{\sqrt{n}} f'(U_i) + \frac{1}{2n}(X_i^2 - Y_i^2) \cdot f''(U_i) + R_i,$$

wobei das Restglied R_i wie folgt abgeschätzt werden kann:

$$|R_i| \leq \frac{1}{2n} \cdot X_i^2 \cdot \delta\left(\frac{|X_i|}{\sqrt{n}}\right) + \frac{1}{2n} \cdot Y_i^2 \cdot \delta\left(\frac{|Y_i|}{\sqrt{n}}\right).$$

Nun sind aber X_i, Y_i, U_i gemeinsam stochastisch unabhängig, womit

$$\begin{aligned}
\mathbb{E}(V_i) &= \frac{1}{\sqrt{n}} \cdot \mathbb{E}\left[(X_i - Y_i) \cdot f'(U_i)\right] + \frac{1}{2n} \cdot \mathbb{E}\left[(X_i^2 - Y_i^2) \cdot f''(U_i)\right] + \mathbb{E}R_i \\
&= \frac{1}{\sqrt{n}} \cdot (\underbrace{\mathbb{E}X_i}_{=0} - \underbrace{\mathbb{E}Y_i}_{=0}) \cdot \mathbb{E}f'(U_i) + \frac{1}{2n} \cdot (\underbrace{\mathbb{E}(X_i^2)}_{=1} - \underbrace{\mathbb{E}(Y_i^2)}_{=1}) \cdot \mathbb{E}[f''(U_i)] + \mathbb{E}R_i \\
&= \mathbb{E}R_i.
\end{aligned}$$

Den Absolutbetrag $|\mathbb{E}R_i|$ beschränken wir mithilfe des Stetigkeitsmoduls durch

$$|\mathbb{E}R_i| \leq \mathbb{E}|R_i| \leq \frac{1}{2n} \cdot \mathbb{E}\left[X_i^2 \cdot \delta\left(\frac{|X_i|}{\sqrt{n}}\right)\right] + \frac{1}{2n} \cdot \mathbb{E}\left[Y_i^2 \cdot \delta\left(\frac{|Y_i|}{\sqrt{n}}\right)\right].$$

Für den linken Term ist

$$\begin{aligned}
\mathbb{E}\left[X_i^2 \cdot \delta\left(\frac{|X_i|}{\sqrt{n}}\right)\right] &= \mathbb{E}\left[X_i^2 \cdot \delta\left(\frac{|X_i|}{\sqrt{n}}\right) \cdot \left(\mathbb{1}_{\{|X_i|\leq\lambda\sqrt{n}\}} + \mathbb{1}_{\{|X_i|>\lambda\sqrt{n}\}}\right)\right] \\
&\leq \underbrace{\mathbb{E}\left[X_i^2 \cdot \delta(\lambda)\right]}_{<\varepsilon \text{ für } \lambda \text{ hinreichend klein}} + 2 \cdot ||f''||_{\sup} \cdot \underbrace{\mathbb{E}\left[X_i^2 \cdot \mathbb{1}_{\{|X_i|>\lambda\sqrt{n}\}}\right]}_{\xrightarrow{n\to\infty} 0}.
\end{aligned}$$

Für den rechten Term gehen wir analog vor. Damit ist $|\mathbb{E}R_i| \leq \frac{1}{n} \cdot a_n$ für eine Nullfolge $(a_n)_{n\in\mathbb{N}}$, also gilt die Behauptung. $\square$

Wir kommen zum Hauptresultat dieses Abschnittes.

Satz 5.9 (Der zentrale Grenzwertsatz)**.** *Seien* $X_1, X_2, \ldots$ *iid mit* $\mathbb{E}X = \mu$ *und* $\operatorname{Var}(X_i) = \sigma^2 \; \forall\, i \in \mathbb{N}$. *Dann gilt*

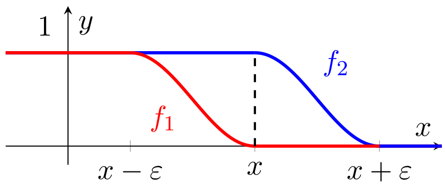
$$\sqrt{n}\,\frac{\overline{X_n} - \mu}{\sigma} \longrightarrow_{\mathcal{D}} Z \sim \mathcal{N}(0,1).$$

Beweis. Es ist zu zeigen, dass die Aussage von Satz 5.8 auch mit $f = \mathbb{1}_{(-\infty,x]}$ für alle

$x \in \mathbb{R}$ gilt (Φ ist stetig). Sei $\varepsilon > 0$ beliebig. Wir wählen nun Funktionen f_1 und f_2, die die Voraussetzungen von Satz 5.8 erfüllen, mit $f_1 \leq f \leq f_2$ und

$$\int \big(f_2(t) - f_1(t)\big) \; \mathrm{d}t \leq \varepsilon.$$

Diese Funktionen könnten beispielsweise wie folgt aussehen:



Satz 5.8 impliziert nun einerseits

$$\begin{array}{ccccc} \mathbb{E}f_1(Z_n) & \leq \mathbb{E}f(Z_n) \leq & \mathbb{E}f_2(Z_n) & & \text{(Monotonie des Erwartungswertes)} \\ \downarrow & & \downarrow & & \\ \mathbb{E}f_1(Z) & \leq \; \mathbb{E}f(Z) \; \leq & \mathbb{E}f_2(Z) & & \text{(Monotonie des Erwartungswertes).} \end{array}$$

Andererseits ist aber

$$\begin{aligned} \mathbb{E}f_2(Z) - \mathbb{E}f_1(Z) &= \int_{-\infty}^{\infty} f_2(z) \cdot \Phi(z) \; \mathrm{d}z - \int_{-\infty}^{\infty} f_1(z) \cdot \Phi(z) \; \mathrm{d}z \\ &= \int_{-\infty}^{\infty} (f_2(z) - f_1(z)) \cdot \underbrace{\Phi(z)}_{\leq \Phi(0)} \; \mathrm{d}z \\ &\leq \Phi(0) \cdot \int_{-\infty}^{\infty} (f_2(z) - f_1(z)) \; \mathrm{d}z \\ &\leq \Phi(0) \cdot \varepsilon \end{aligned}$$

Wegen $\varepsilon > 0$ beliebig folgt $\mathbb{E}f(Z_n) \to \mathbb{E}f(Z)$, also die Behauptung. □

Bemerkung 5.10. *(i) Der zentrale Grenzwertsatz gilt auch noch, wenn man die Voraussetzung „unabhängig“ bzw. „identisch“ verteilt etwas abschwächt. Wir verweisen hier auf die Litteratur.*

(ii) Allgemeiner gilt $\forall\, B \in \mathscr{B}(\mathbb{R})$:

$$\mathbb{P}\left(\sqrt{n} \cdot \frac{\overline{X_n} - \mu}{\sigma} \in B\right) \xrightarrow{n \to \infty} \mathbb{P}(Z \in B).$$

(iii) Von dem zentralen Grenzwertsatz existiert auch ein mehrdimensionales Analogon.

Anwendung 5.11 (Binomialapproximation)**.** *Wir wissen, dass für iid Zufallsvariablen $X_i \sim \mathrm{Ber}(p)$ gilt, dass $\sum_{i=1}^n X_i \sim \mathrm{Bin}(n,p)$ ist, mit Erwartungswert np und Varianz $np(1-p)$. Ist p „klein“ (d.h. $p = p_n$, sodass $np_n \xrightarrow{n\to\infty} \lambda > 0$), so wissen wir nach dem Poisson-Grenzwertsatz, dass*

$$\mathbb{P}\left(\sum_{i=1}^n X_i = k\right) \xrightarrow{n\to\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

Ist p hingegen nicht klein, so können wir gemäß dem zentralen Grenzwertsatz eine Normalapproximation verwenden:

$$\sqrt{n} \cdot \frac{\overline{X_n} - p}{\sqrt{p(1-p)}} \xrightarrow{\mathscr{D}} Z \sim \mathscr{N}(0,1)$$

$$\begin{aligned}
\Longrightarrow \mathbb{P}\left(\sum_{i=1}^k X_i = k\right) &\overset{\substack{\text{„Trick“}\\ \downarrow}}{=} \mathbb{P}\left(k - \frac{1}{2} < \sum_{i=1}^n X_i \leq k + \frac{1}{2}\right) \\
&= \mathbb{P}\left(\sqrt{n} \cdot \frac{\frac{1}{n}\cdot\left(k-\frac{1}{2}\right) - p}{\sqrt{p(1-p)}} < \underbrace{\sqrt{n} \cdot \frac{\frac{1}{n}\cdot\sum_{i=1}^n X_i - p}{\sqrt{p(1-p)}}}_{\text{hierfür gilt der zentrale Grenzwertsatz}} \leq \sqrt{n} \cdot \frac{\frac{1}{n}\cdot\left(k+\frac{1}{2}\right) - p}{\sqrt{p(1-p)}}\right) \\
&\approx \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

Beispiel 5.12. *Seien $X_1, X_2, \ldots$ iid mit $\mathbb{E}X_1 = \mu$, $\mathrm{Var}(X_1) = \sigma^2$ und $\mathbb{E}X_1^4 < \infty$.*

(i) *Wir definieren*

$$\widehat{\sigma_n}^2 := \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \mu)^2.$$

$\widehat{\sigma_n}^2$ ist ein Schätzer für die Varianz σ^2 (bei bekanntem μ), denn

$$\mathbb{E}\,\widehat{\sigma_n}^2 = \mathbb{E}\left[\frac{1}{n} \cdot \sum_{i=1}^n (X_i - \mu)^2\right] = \frac{1}{n} \cdot \sum_{i=1}^n \underbrace{\mathbb{E}\left[(X_i - \mu)^2\right]}_{=\mathrm{Var}(X_1)} = \sigma^2$$

und damit $\widehat{\sigma_n}^2 \xrightarrow{\mathbb{P}} \sigma^2$ nach Satz 3.21 (schwaches Gesetz). Nach dem zentralen Grenzwertsatz und Proposition 5.4 (iii) folgt

$$\sqrt{n} \cdot \frac{\overline{X_n} - \mu}{\widehat{\sigma_n}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1).$$

(ii) Wir definieren

$$\widehat{s_n}^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X_n})^2.$$

$\widehat{s_n}^2$ ist ein Schätzer für die Varianz (bei unbekanntem μ) mit $\mathbb{E}\ \widehat{s_n}^2 = \sigma^2$ (nachrechnen!). Wir wollen beweisen, dass

$$Z_n := \sqrt{n} \cdot (\widehat{s_n}^2 - \sigma^2) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \mu_4 - \sigma^4) \tag{5.1}$$

mit $\mu_4 := \mathbb{E}((X_1 - \mu)^4)$. Sei $Y_i := (X_i - \mu)^2$, $i \in \mathbb{N}$. Dann gelten $\mathbb{E}(Y_i) = \operatorname{Var}(X_i) = \sigma^2$ und $\operatorname{Var}(Y_i) = \mathbb{E}(Y_i^2) - (\mathbb{E}(Y_i))^2 = \mu_4 - \sigma^4$. Der zentrale Grenzwertsatz impliziert bereits

$$\tilde{Z}_n := \sqrt{n} \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \mu)^2 - \sigma^2 \right) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \mu_4 - \sigma^4).$$

Nach Proposition 5.3 folgt daraus (5.1), falls $Z_n - \tilde{Z}_n \xrightarrow{\mathbb{P}} 0$. Letzteres gilt wegen

$$\begin{aligned}
&\sqrt{n}\left(\widehat{s_n}^2 - \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right) \\
&= \sqrt{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}\Big((X_i-\mu)^2 - 2(X_i-\mu)(\overline{X_n}-\mu) + (\overline{X_n}-\mu)^2\Big)\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. - \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) \\
&= \sqrt{n}\left(\frac{1}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{n}{n-1}(\overline{X_n}-\mu)^2\right) \\
&= \underbrace{\underbrace{\frac{\sqrt{n}}{n-1}}_{\to 0}\ \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}_{\xrightarrow{\mathbb{P}}\sigma^2 \text{ nach } 3.21}}_{} - \underbrace{\underbrace{\sqrt{n}(\overline{X_n}-\mu)}_{\substack{\xrightarrow{\mathscr{D}}\mathscr{N}(0,\sigma^2) \\ \text{nach dem ZGWS}}} \cdot \underbrace{\frac{n}{n-1}}_{\to 1} \cdot \underbrace{(\overline{X_n}-\mu)}_{\substack{\xrightarrow{\mathbb{P}}0 \\ \text{nach } 3.21}}}_{\xrightarrow{\mathbb{P}}0 \text{ nach Prop. } 5.4\ (i)} \xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

wobei wir im letzten Schritt nachfolgendes Lemma verwendet haben.

Lemma 5.13. *Seien $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ Folgen von Zufallsvariablen und X, Y Zufallsvariablen mit $X_n \xrightarrow{\mathbb{P}} X$ und $Y_n \xrightarrow{\mathbb{P}} Y$. Dann gilt $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$.*

Beweis. Die Behauptung folgt unmittelbar aus

$$\mathbb{P}\Big(|X_n + Y_n - (X+Y)| \geq \varepsilon\Big) \leq \mathbb{P}\Big(|X_n - X| \geq \frac{\varepsilon}{2}\Big) + \mathbb{P}\Big(|Y_n - Y| \geq \frac{\varepsilon}{2}\Big) \text{ für alle } \varepsilon > 0.$$

$\square$