Elementare Stochastik

Sommersemester 2018

4. Juni 2019

Inhaltsverzeichnis

Vorbemerkung

Wir werfen 3 Würfel gleichzeitig und ermitteln die Gesamtaugenzahl. Ist 11 ebenso wahrscheinlich wie 12? Mögliche Würfelkonstellationen:

$$\left.\begin{array}{ll} \text{„11“:} & 641, 632, 551, 542, 533, 443 \\ \text{„12“:} & 651, 642, 633, 552, 543, 444 \end{array}\right\} \text{jeweils 6 Möglichkeiten}$$

Glücksspieler des 17. Jahrhunderts „wussten“ schon, dass 11 häufiger ist als 12, was empirische Daten auch belegen.

Stochastische allgemeine Betrachtungsweise:

(i) Modellbildung: Präzisiere, welche Versuchsausgänge betrachtet werden sollen.

(ii) Man nimmt an, dass zu jedem Ereignis A eine Wahrscheinlichkeit (W-keit) $\mathbb{P}(A) \in [0, 1]$ gehört, die man für „einfache“ Ereignisse festlegt.

(iii) Man versucht auf der Grundlage konsistenter Rechenregeln aus Wahrscheinlichkeiten für einfache Ereignisse die Wahrscheinlichkeiten komplizierter Ereignisse zu bestimmen oder zu approximieren.

Für den Umgang mit Wahrscheinlichkeiten hat sich eine Axiomatik auf mengentheoretischer Grundlage bewährt, die 1933 von KOLMOGOROV entwickelt wurde. Wie beim Begriff des Vektorraums oder Hilberts Grundlagen der Geometrie wird der Umgang mit den Objekten formal vorgegeben und allein mit diesen Regeln argumentiert. Die Formalisierung eines angewandten Problems (etwa eines Würfelproblems) in den Formalismus der Stochastik (oben (i) und (ii)) ist deshalb der erste Teil der Problemlösung. Sodann können Aussagen der Wahrscheinlichkeitstheorie eingebracht werden.

1 Diskrete Wahrscheinlichkeitsräume

1 Wahrscheinlichkeitsräume und Kombinatorik

Definition 1.1. Ein *diskreter Wahrscheinlichkeitsraum* (W-Raum) ist ein Tripel $(\Omega, \mathfrak{A}, \mathbb{P})$ bestehend aus einer nichtleeren, höchstens abzählbaren Menge Ω, der Potenzmenge $\mathfrak{A} = \mathcal{P}(\Omega)$ von Ω und einer Abbildung $\mathbb{P} : \mathfrak{A} \to [0, 1]$ mit

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

(ii)
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
für jede Folge $(A_i)_{i\in\mathbb{N}}$ paarweise disjunkter Mengen $A_i \in \mathfrak{A}$. (*σ-Additivität*)

Lemma 1.2. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum. Dann gelten

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

(b) $\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$, falls $A_1, ..., A_n$ paarweise disjunkt. (*endl. Additvität*)

(c) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ für alle $A \in \mathfrak{A}$. $(A^c := \Omega \setminus A)$

(d) $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ für $A, B \in \mathfrak{A}$.

(e) $\mathbb{P}(A) \leq \mathbb{P}(B)$, falls $A \subseteq B$ für $A, B \in \mathfrak{A}$. (*Monotonie*)

(f) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(g) $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ für jede Folge $(A_i)_{i\geq 1}$ in $\mathfrak{A}$. (*Sub-σ-Additivität*)

Beweis.

(a) Wähle $A_i = \emptyset$ für alle $i \geq 1$. Da die A_i paarweise disjunkt sind, liefert die σ-Additivität
$$\mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$$
Es folgt also $\mathbb{P}(\emptyset) = 0$ (als Konsequenz eines Widerspruchs bei Annahme $\mathbb{P}(\emptyset) > 0$).

(b) Wähle $B_i = A_i$ für $i = 1, \ldots, n$ und $B_i = \emptyset$ für $i > n$. Dann gilt
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) \overset{(ii)}{=} \sum_{i=1}^{\infty} \mathbb{P}(B_i) \overset{(a)}{=} \sum_{i=1}^{n} \mathbb{P}(A_i).$$

(c) A, A^c sind disjunkt und $\Omega = A \cup A^c$. Damit gilt
$$1 \overset{(i)}{=} \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) \overset{(b)}{=} \mathbb{P}(A) + \mathbb{P}(A^c) \quad \Rightarrow \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

(d) Wir haben die disjunkte Zerlegung $B = (A \cap B) \cup (B \setminus A)$. Damit gilt

$$\mathbb{P}(B) \overset{(b)}{=} \mathbb{P}(B \cap A) + \mathbb{P}(B \setminus A) \quad \Rightarrow \quad \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(e) Da $A \subset B$ ist, gilt die disjunkte Zerlegung $B = A \cup (B \setminus A)$. Somit ist

$$\mathbb{P}(B) \overset{(b)}{=} \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A), \quad \text{da } \mathbb{P}(B \setminus A) \geq 0.$$

(f) Es gilt die disjunkte Zerlegung $A \cup B = A \cup (B \setminus A)$ und somit

$$\mathbb{P}(A \cup B) \overset{(b)}{=} \mathbb{P}(A) + \mathbb{P}(B \setminus A) \overset{(d)}{=} \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(B \cap A).$$

(g) Sei $B_1 = \emptyset$, $B_i := \bigcup_{j=1}^{i-1} A_j$ für $i \geq 1$. Es gilt somit die disjunkte Zerlegung

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \ldots = \bigcup_{i=1}^{\infty} (A_i \setminus B_i).$$

Es folgt schließlich

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \overset{(ii)}{=} \sum_{i=1}^{\infty} \mathbb{P}(A_i \setminus B_i) \overset{(e)}{\leq} \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

∎

Lemma 1.3. Die σ-Additivität aus Definition 1.1(ii) ist äquivalent zur gleichzeitigen Gültigkeit von

(i') *endliche Additivität* (vergleiche Lemma 1.2(b))

(ii') *Stetigkeit von unten*, d.h. für jede Folge $(A_i)_{i\geq 1}$ in $\mathfrak{A}$ mit $A_1 \subset A_2 \subset A_3 \subset \cdots$ gilt:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i).$$

Beweis.
„⇒“: Dass (i') gilt, haben wir bereits gezeigt. Zu (ii'): Sei $(A_i)_{i\geq 1}$ eine aufsteigende Folge in $\mathfrak{A}$, d.h. $A_1 \subset A_2 \subset A_3 \subset \ldots$ Dann gilt $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ mit $B_1 = A_1$ und $B_i = A_i \setminus A_{i-1}$ für $i \geq 2$. Dann sind die B_i nach Konstruktion paarweise disjunkt. Damit gilt $\mathbb{P}(B_i) = \mathbb{P}(A_i) - \mathbb{P}(A_{i-1})$ und

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) \overset{(ii)}{=} \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{i\to\infty} \sum_{j=1}^{i} \mathbb{P}(B_j) \\ &= \lim_{i\to\infty} (\mathbb{P}(A_1) + (\mathbb{P}(A_2) - \mathbb{P}(A_1))) + (\mathbb{P}(A_3) - \mathbb{P}(A_2)) + \ldots + (\mathbb{P}(A_i) - \mathbb{P}(A_{i-1})) \\ &= \lim_{i\to\infty} \mathbb{P}(A_i). \end{aligned}$$

„$\Leftarrow$“: Gelte nun (i’), (ii’). Sei $(A_i)_{i\geq 1}$ eine Folge paarweise disjunkter Mengen in $\mathfrak{A}$. Dann gilt mit $B_i = \bigcup_{j=1}^{i} A_j$ die Identität $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$. Die so konstruierte Folge $(B_i)_i$ ist ferner aufsteigend, d.h. $B_1 \subset B_2 \subset B_3 \subset \cdots$ und $\mathbb{P}(B_i) = \sum_{j=1}^{i} \mathbb{P}(A_j)$ nach (ii’). Ferner gilt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) \overset{(ii')}{=} \lim_{i\to\infty} \mathbb{P}(B_i) = \lim_{i\to\infty} \sum_{j=1}^{i} \mathbb{P}(A_j) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

∎

Bemerkung 1. Falls für alle $A_1 \supset A_2 \supset A_3 \supset \ldots$ gilt, dass $\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i)$, so spricht man von „Stetigkeit von oben“. Im vorigen Lemma kann „Stetigkeit von unten“ durch „Stetigkeit von oben“ ersetzt werden.

Bemerkung 2. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum und $\Omega = \{\omega_1, \omega_2, \ldots\}$. Dann ist $\mathbb{P} : \mathfrak{A} \to [0,1]$ bereits durch die Werte $p_i = \mathbb{P}(\{\omega_i\})$ für $i \geq 1$ vollständig festgelegt. Für jedes $A \in \mathfrak{A}$ existiert die Darstellung $A = \bigcup_{\omega_i \in A} \{\omega_i\}$ mit paarweise disjunkten Mengen. Damit gilt

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{\omega_i \in A} p_i = \sum_{i=1}^{\infty} \mathbb{1}_A(\omega_i) p_i.$$

Hierbei bezeichnet $\mathbb{1}_A$ die Indikatorfunktion von A, gegeben durch

$$\mathbb{1}_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{falls } \omega \notin A. \end{cases}$$

Bezeichnung 1. Wir haben bisher einige technische Begriffe der Stochastik verwendet, die wir im Folgenden mit einer Bedeutung versehen wollen.

(i) Ω heißt *Grundraum, Ergebnismenge, Stichprobenraum* oder *Ergebnisraum.*

(ii) Elemente von $\mathfrak{A}$ heißen *Ereignisse*, $A = \{\omega\}$ heißt *Elementarereignis.*

(iii) $\mathbb{P}$ heißt *W-Maß* oder *W-Verteilung*, $\mathbb{P}(A)$ bezeichnet die *W-keit des Ereignisses* A.

Beispiel 1.4 (Diskrete W-Räume)**.** (1) *Laplace-Modelle* (Gleichverteilung).
Allgemeines Prinzip: $\Omega = \{1, \ldots, n\}$ mit $\mathbb{P}(\{i\}) = \frac{1}{n} =: p_i$, also $p_1 = \cdots = p_n$.

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{|A|}{|\Omega|}.$$

Dabei bezeichnet $|A|$ die Kardinalität von A. Man spricht bei der Berechnung von Wahrscheinlichkeiten in Laplace-Modellen von „Günstige durch Mögliche“.

„Würfeln mit fairem Würfel“: $\Omega = \{1, \ldots, 6\}$, wobei $\mathbb{P}(\{i\}) =: p_i$ und $p_1 = p_2 = \cdots = p_6$. Wegen $1 = \mathbb{P}(\Omega) = \sum_{i=1}^{6} p_i = 6p_1$ folgt $p_i = 1/6$.

„Geburtstagsproblem": Es befinden sich N Personen in einem Raum, wir interessieren uns für das Ereignis A ="Mindestens zwei Anwesende haben am selben Tag Geburtstag".

$$\Omega = \{\omega = (\omega_1, \dots, \omega_N) \,|\, \omega_i \in \{1, \dots, 365\} \text{ für } i = 1, \dots, N\}.$$

Damit ist $A = \{\omega \in \Omega \,|\, \exists\, 1 \leq i < j \leq N : \omega_i = \omega_j\}$. Wir machen die (idealisierte) Modellannahme, dass die Geburtstage gleichverteilt seien. Dann gilt

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = 1 - \frac{|A^c|}{|\Omega|}, \quad |\Omega| = 365^N, \quad A^c = \{\omega \in \Omega \,|\, \forall\, 1 \leq i < j \leq N : \omega_i \neq \omega_j\}.$$

Das Ereignis A^c entspricht „es gibt keine zwei mit demselben Geburtstag". Die Anzahl der Möglichkeiten in A^c nimmt mit wachsendem N schnell ab

$$|A^c| = \begin{cases} \prod_{i=1}^{N} (366 - i), & N \leq 365 \\ 0, & N > 365. \end{cases}$$

N	20	23	40	150
$\mathbb{P}(A)$	0.411	0.507	0.891	$1 - 10^{-15}$

(2) *Poisson-Verteilung* Π_λ mit Parameter $\lambda > 0$.
Es sei $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ und $\Pi_\lambda(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}$ für $n \geq 0$. Damit ist Π_λ ein W-Maß, denn es gilt:

$$\Pi_\lambda(\Omega) = \sum_{n=0}^{\infty} \Pi_\lambda(\{n\}) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1.$$

Poisson-Verteilungen werden zur Modellierung der Anzahl seltener Phänomene pro Zeiteinheit verwendet. Beispiele sind die Anzahl fehlerhafter Teile in einer großen Produktion, Emission von α-Teilchen beim radioaktiven Zerfall oder die Anzahl der Druckfehler in einem Buch. Weshalb dabei jeweils zur Modellierung die Poisson-Verteilungen geeignet ist, werden wir später in Abschnitt 13 sehen.

(3) Einpunktverteilung (*Dirac-Maß*).
Sei Ω eine beliebige, höchstens abzählbare Menge und $\omega_0 \in \Omega$. Wir definieren

$$\delta_{\omega_0}(A) := \mathbb{1}_A(\omega_0) = \begin{cases} 1, & \text{falls } \omega_0 \in A, \\ 0, & \text{falls } \omega_0 \notin A, \end{cases}$$

für alle $A \in \mathfrak{A}$. Dann ist $(\Omega, \mathcal{P}(\Omega), \delta_{\omega_0})$ ein diskreter W-Raum.

2 Kombinatorik

Wir betrachten zunächst 4 Abzählprobleme.

I: Wie viele 10-stellige Dualzahlen gibt es?

II: Auf wie viele Arten können 3 verschiedene Autos auf 8 Parkplätzen parken?

III: Wie viele mögliche Ergebnisse gibt es beim Lotto „6 aus 49"?

IV: Auf wieviele Arten können 10 1-Euro-Münzen auf 3 Taschen verteilt werden?

Sei im Folgenden stets $M = \{1, \dots, n\}$.

Modell I: Stichprobe der Länge k aus M in Reihenfolge mit Zurücklegen.

$$\Omega_{\mathrm{I}} = M^k = M \times \cdots \times M = \{(\omega_1, \dots \omega_k) \mid \omega_i \in M \text{ für } i = 1, \dots, k\}.$$

Satz 2.1. Es gilt $|\Omega_{\mathrm{I}}| = n^k$.

Modell II: Stichprobe der Länge k aus M in Reihenfolge ohne Zurücklegen ($k \leq n$).

$$\Omega_{\mathrm{II}} = \left\{(\omega_1, \dots, \omega_k) \in M^k \,\middle|\, \omega_i \neq \omega_j \text{ für } i \neq j\right\}.$$

Satz 2.2. Es gilt $|\Omega_{\mathrm{II}}| = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$. Ist $k = n$, so befinden wir uns im Spezialfall $\Omega = \mathcal{S}_n$, der Menge aller Permutationen von M, auch symmetrische Gruppe von M genannt. Es folgt, dass $|\mathcal{S}_n| = n!$ gilt.

Modell III: Stichprobe der Länge k aus M ohne Reihenfolge ohne Zurücklegen.

$$\Omega_{\mathrm{III}} = \left\{\{\omega_1, \dots, \omega_k\} \,\middle|\, \omega_i \in M, \omega_i \neq \omega_j \text{ für alle } 1 \leq i < j \leq n\right\}.$$

Satz 2.3. Es gilt $|\Omega_{\mathrm{III}}| = \frac{n!}{k!(n-k)!} =: \binom{n}{k}$.

Beweis.
Betrachte zunächst $\Omega_{\mathrm{II}} = \{(\omega_1, \dots, \omega_k) \in M^k \mid \omega_i \neq \omega_j \text{ für } i \neq j\}$ und die Äquivalenzrelation $\sim$ auf $\Omega_{\mathrm{II}} : (\omega_1, \dots, \omega_k) \sim (\omega_1', \dots, \omega_k')$, falls eine Permutation π von $\{1, \dots, k\}$ existiert mit $\omega_i = \omega'_{\pi(i)}$ für $i = 1, \dots, k$. Offenbar gilt $\Omega_{\mathrm{III}} = \Omega_{\mathrm{II}}/\!\sim$. Jede Äquivalenzklasse hat $k!$ Elemente. Ein Repräsentant ist etwa jeweils $(\omega_1, \dots, \omega_k) \in \Omega_{\mathrm{II}}$ mit $\omega_1 < \omega_2 < \cdots < \omega_k$. Damit folgt $|\Omega_{\mathrm{III}}| = |\Omega_{\mathrm{II}}|/k! = n!/(k!(n-k)!)$. ∎

Aus dem Beweis folgt, dass man statt Ω_{III} alternativ auch

$$\Omega'_{\mathrm{III}} = \left\{(\omega_1, \dots, \omega_k) \in M^k \,\middle|\, \omega_1 < \omega_2 < \cdots < \omega_k\right\}$$

wählen kann.

Modell IV: Stichprobe der Länge k aus M ohne Reihenfolge mit Zurücklegen.

$$\Omega_{\mathrm{IV}} = \left\{(\omega_1, \dots, \omega_k) \in M^k \,\middle|\, \omega_1 \leq \omega_2 \leq \cdots \leq \omega_k\right\}.$$

Satz 2.4. Es gilt $|\Omega_{\mathrm{IV}}| = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$.

Beweis.

Wir betrachten $M^* = \{1, \ldots, n+k-1\}$ und

$$\Omega^*_{III} = \left\{ (\omega^*_1, \ldots, \omega^*_k) \in (M^*)^k \;\middle|\; \omega^*_1 < \omega^*_2 < \cdots < \omega^*_k \right\}$$

sowie die Abbildung $f : \Omega_{IV} \to \Omega^*_{III}$ mit

$$(\omega_1, \ldots, \omega_k) \mapsto (\omega_1, \omega_2 + 1, \ldots, \omega_k + k - 1).$$

Man sieht leicht, dass f bijektiv ist. Damit gilt $|\Omega_{IV}| = |\Omega^*_{III}|$. Nach Modell III gilt $|\Omega^*_{III}| = \binom{n+k-1}{k}$. ∎

Definition 2.5. Die Größen $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ für $k = 0, 1, \ldots, n$ heißen *Binomialkoeffizienten.* Wir setzen $\binom{n}{k} := 0$ für $k < 0$ oder $k > n$.

Interpretation der 4 Modelle:

a) k-maliges sukzessives Ziehen aus einer Urne mit n nummerierten Kugeln:
   - mit/ohne Zurücklegen,
   - mit/ohne Beachten der Reihenfolge des Ziehens.

b) Besetzung von n Zellen durch k Objekte
   - mit/ohne Mehrfachbesetzungen,
   - unterscheidbare/ununterscheidbare Objekte.

   Pauliprinzip: Mehrfachbesetzungen verboten.

Der Rest dieses Abschnitts besteht aus Anwendungen und Verallgemeinerungen der vier kombinatorischen Grundmodelle.

Korollar 2.6 (Binomischer Lehrsatz)**.** Für alle $x, y \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Beweis.

$$\begin{aligned}(x+y)^n &= (x+y)\cdots(x+y) = \sum_{A \subset \{1,\ldots,n\}} x^{|A|} y^{|A^c|} \\ &= \sum_{k=0}^{n} \sum_{A \subset \{1,\ldots,n\} : |A| = k} x^k y^{n-k} \overset{(2.3)}{=} \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.\end{aligned}$$

∎

Korollar 2.7. Für $n \in \mathbb{N}$ gelten

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \quad \sum_{k=0}^{n} \binom{n}{k} (-1)^k = 0.$$

Beweis.
Es gilt $\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k 1^{n-k} = (1+1)^n = 2^n$ und mit $x = 1, y = -1$ in Korollar 2.6 erhält man auch die zweite Summe. ∎

Die Binomialkoeffizienten geben an, auf wie viele Arten man n nummerierte Kugeln in zwei Gruppen teilen kann, so dass sich k Kugeln in Gruppe 1 befinden. Allgemeiner teilen wir nun in r nummerierte Gruppen der Größen $k_1, \ldots, k_r$ mit $\sum_{i=1}^{r} k_i = n$. Wieviele mögliche Arten gibt es?
Lösung: Für die erste Gruppe gibt es $\binom{n}{k_1}$ Möglichkeiten, zu jeder dieser Möglichkeiten gibt es für die zweite Gruppe $\binom{n-k_1}{k_2}$ Möglichkeiten, usw. Für die r-te Gruppe gibt es $\binom{n-k_1-\cdots-k_{r-1}}{k_r}$ Möglichkeiten. Die Gesamtanzahl ergibt sich durch Multiplikation als

$$\begin{aligned} &\binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{n-k_1-\ldots-k_{r-1}}{k_r} \\ =&\frac{n!(n-k_1)!\cdots(n-k_1-\ldots-k_{r-1})!}{k_1!(n-k_1)!k_2!(n-k_1-k_2)!\cdots k_r!(n-n)!} = \frac{n!}{k_1!k_2!\cdots k_r!}. \end{aligned}$$

Damit ist gezeigt:

Satz 2.8. Zu jeder Menge $M = \{1, \ldots, n\}$ und $k_1, \ldots, k_r \in \mathbb{N}_0$ mit $\sum_{i=1}^{r} k_i = n$ gibt es genau

$$\frac{n!}{k_1! \cdots k_r!} =: \binom{n}{k_1, \ldots, k_r} \tag{1}$$

viele geordnete Zerlegungen in Teilmengen $M_1, \ldots, M_r$ mit $|M_i| = k_i$. Die Zahlen in (1) heißen *Multinomialkoeffizienten.*

Korollar 2.9. Für $x_1, \ldots, x_r \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt

$$(x_1 + \cdots + x_r)^n = \sum_{\substack{k_1,\ldots,k_r \in \mathbb{N}_0 \\ \sum k_i = n}} \binom{n}{k_1, \ldots, k_r} x_1^{k_1} \cdots x_r^{k_r}.$$

Beweis.

$$\begin{aligned} (x_1 + \cdots + x_r)^n &= \sum_{\substack{A_1,\ldots,A_r \\ \text{Zerlegung von } \{1,\ldots,n\}}} \prod_{i=1}^{r} x_i^{|A_i|} = \sum_{\substack{k_1,\ldots,k_r \geq 0 \\ \sum k_i = n}} \sum_{\substack{A_1,\ldots,A_r \text{ Zerlegung} \\ |A_1| = k_1,\ldots,|A_r| = k_r}} \prod_{i=1}^{r} x_i^{k_i} \\ &\overset{(2.8)}{=} \sum_{\substack{k_1,\ldots,k_r \in \mathbb{N}_0 \\ \sum k_i = n}} \binom{n}{k_1, \ldots, k_r} x_1^{k_1} \cdots x_r^{k_r}. \end{aligned}$$

∎

Satz 2.10. Seien $p_1, \dots, p_r \geq 0$ mit $\sum_{i=1}^r p_i = 1$ und $n \in \mathbb{N}$. Dann ist auf $\Omega = \{(k_1, \dots, k_r) \in \mathbb{N}_0^r \mid \sum_{i=1}^r k_i = n\}$ durch

$$\mathbb{P}(\{(k_1, \dots, k_r)\}) = \binom{n}{k_1, \dots, k_r} p_1^{k_1} \cdots p_r^{k_r} \tag{2}$$

eine W-Verteilung gegeben. Sie heißt *Multinomialverteilung* zu den Parametern n und $p_1, \dots, p_r$.

Beweis.
Zu zeigen ist, dass für die Festlegung (2) gilt $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$. Dies folgt aber aus Korollar 2.9. ∎

Beispiel 2.11. Wie groß ist die W-keit, bei n Würfen mit einem fairen Würfel, k_1 mal 1, k_2 mal 2,...,k_6 mal 6 zu werfen?
Lösung: Wir wählen

$$\Omega = \{1, \dots, 6\}^n, \quad A = \{\omega \in \Omega : |\{1 \leq i \leq n : \omega_i = j\}| = k_j \ \text{ für } j = 1, \dots, 6\}.$$

Jedem $\omega \in A$ entspricht genau eine geordnete Zerlegung von $\{1, \dots, n\}$ in Gruppen mit Größen $k_1, \dots, k_6$. Nach Satz 2.8 also $|A| = \binom{n}{k_1, \dots, k_6}$. Da ein Laplace-Modell vorliegt, folgt

$$\mathbb{P}(A) = \frac{1}{6^n} \binom{n}{k_1, \dots, k_6}.$$

Beispiel 2.12. In einer Urne seien s schwarze und w weiße Kugeln, $n := s+w$. Es werden $k \leq n$ Kugeln ohne Zurücklegen gezogen. Was ist die W-keit, dass die Stichprobe genau ℓ schwarze und $k - \ell$ weiße Kugeln enthält.
Lösung: Seien

$$A = \{1, \dots, n\}, \quad A_s = \{1, \dots, s\}, \quad A_w = A \setminus A_s = \{s+1, \dots, n\}.$$

Ferner sei

$$\Omega = \left\{ (\omega_1, \dots, \omega_k) \in A^k \,\middle|\, \omega_1 < \omega_2 < \cdots < \omega_k \right\},$$

also $|\Omega| = \binom{n}{k}$ (Modell III). Es sei

$$\begin{aligned} B_\ell &= \text{„genau } \ell \text{ schwarze Kugeln unter } k \text{ gezogenen“} \\ &= \{\omega \in \Omega \mid \omega_i \in A_s \text{ für } i = 1, \dots, \ell, \text{ und } \omega_i \in A_w \text{ für } i = \ell+1, \dots, k\}. \end{aligned}$$

Es gilt $|B_\ell| = \binom{s}{\ell}\binom{w}{k-\ell}$. Das Laplace-Modell liefert

$$\mathbb{P}(B_\ell) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}} =: h(\ell; k, s+w, s).$$

Satz 2.13. Für die Parameter $s, w \in \mathbb{N}$ und $1 \leq k \leq s + w$ ist durch

$$p_\ell := h(\ell; k, s+w, s) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}}, \quad \ell = 0, \dots, k \tag{3}$$

eine W-Verteilung auf $\{0, \dots, k\}$ definiert. Sie heißt *hypergeometrische Verteilung.*

Beweis.
Seien Ω und B_ℓ wie oben. Die B_ℓ sind für $\ell = 0 \vee (k-w), \dots, k \wedge s$ paarweise disjunkt mit $\Omega = \bigcup B_\ell$. Es folgt

$$\sum_{\ell=0\vee(k-w)}^{k\wedge s} h(\ell; k, s+w, s) = 1.$$

Für $\ell < 0 \vee (k-w)$ oder $\ell > k \wedge s$ gilt $h(\ell; k, s+w, s) = 0$, da entsprechende Binomialkoeffizienten in (3) nach der Definition 2.5 Null sind. Wir haben also $\sum_{\ell=0}^{k} p_\ell = 1$. ∎

Beispiel 2.14 (Beispiele zur hypergeometrischen Verteilung)**.** (1) Die W-keit, genau ℓ Richtige im Lotto „6 aus 49" zu haben: $s = 6$, $w = 43$, $k = 6$, also $h(\ell; 6, 49, 6)$.

(2) Qualitätskontrolle: n Produktionsstücke, davon s defekt, $w = n - s$ nicht defekt. Stichprobe der Größe k. Die W-keit, dass genau ℓ Defekte unter der Stichprobe sind, ist $h(\ell, k, s+w, s)$.

(3) W-keit, dass Spieler A beim Skat 3 Asse erhält? (Er erhält 10 von 32 Karten, in denen sich insgesamt 4 Asse befinden.)
Lösung: $\frac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} = \frac{66}{899}$.

3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Beispiel 3.1.

1. Ein fairer Würfel werde geworfen. Modell: $\Omega = \{1, \dots, 6\}$, $\mathbb{P}(\{i\}) = 1/6$ für $i = 1, \dots, 6$. Ein Beobachter verrate, dass eine gerade Zahl geworfen wurde. Für die neue Situation gilt intuitiv

$$\widetilde{\mathbb{P}}(\{i\}) = \begin{cases} 0, & \text{falls } i \text{ ungerade} \\ \frac{1}{3}, & \text{falls } i \text{ gerade.} \end{cases}$$

2. Versicherungsproblem: Ein männlicher Bürger werde genau k Jahre alt mit W-keit p_k, $k \in \mathbb{N}$, mit $\sum_{k\geq 1} p_k = 1$. Gesucht: W-keit q_ℓ, dass er im ℓ-ten Lebensjahr stirbt, gegeben, dass er bereits das k-te Jahr erreicht hat. Dazu:

$$\begin{aligned} s_k &= \mathbb{P}\,(\text{wird mindestens } k \text{ Jahre alt}) \\ &= \mathbb{P}\left(\bigcup_{i=k}^{\infty}\{\text{wird genau } i \text{ Jahre alt}\}\right) = \sum_{i=k}^{\infty} p_i. \end{aligned}$$

Nun ist intuitiv (heuristisch über relative Häufigkeiten einsichtig)

$$q_\ell = \begin{cases} 0, & \text{für } \ell < k \\ p_\ell/s_k, & \text{für } \ell \geq k. \end{cases}$$

Allgemeines Konzept

Definition 3.2. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum. Sei $B \in \mathfrak{A}$ mit $\mathbb{P}(B) > 0$. Dann heißt

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

die *bedingte W-keit von* A unter (Bedingung) B.

Im Folgenden sei $(\Omega, \mathfrak{A}, \mathbb{P})$ stets ein diskreter W-Raum.

Lemma 3.3. Für $B \in \mathfrak{A}$ mit $\mathbb{P}(B) > 0$ ist $\mathbb{P}(\cdot \,|\, B) : \mathfrak{A} \to [0,1]$ eine auf B konzentrierte W-Verteilung.

Beweis.
Es gilt gemäß Monotonie $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ und somit $0 \leq \mathbb{P}(A \,|\, B) \leq 1$. Es gilt $\mathbb{P}(\Omega \,|\, B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$. Sei ferner $(A_i)_{i \geq 1}$ eine Folge paarweise disjunkter Ereignisse in $\mathfrak{A}$. Dann folgt

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B \right) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i \right) \cap B \right) = \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B) \right) \\ &\overset{\sigma\text{-add.}}{=} \frac{1}{\mathbb{P}(B)} \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \,|\, B). \end{aligned}$$

Damit sind alle Eigenschaften aus Definition 1.1 erfüllt und $\mathbb{P}(\cdot \,|\, B)$ ist eine Wahrscheinlichkeitsverteilung.
Auf B konzentriert: Für $A \subset B^c$ gilt $\mathbb{P}(A \,|\, B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(\emptyset)/\mathbb{P}(B) = 0$. ∎

Satz 3.4. Seien $A_1, \ldots, A_n$ Ereignisse mit $\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. Dann gilt

$$\begin{aligned} \mathbb{P}(A_1 \cap \ldots \cap A_n) &= \mathbb{P}(A_1)\mathbb{P}(A_2 \,|\, A_1)\mathbb{P}(A_3 \,|\, A_1 \cap A_2) \cdots \mathbb{P}(A_n \,|\, A_1 \cap \ldots \cap A_{n-1}) \\ &= \prod_{k=1}^{n} \mathbb{P}\left(A_k \,\middle|\, \bigcap_{j<k} A_j \right). \end{aligned}$$

Beweis.
Induktion nach n:
$n = 1$: $\mathbb{P}(A_1) = \mathbb{P}(A_1 \,|\, \Omega)$ nach Definition.

$n-1 \to n$:

$$\begin{aligned}\mathbb{P}(A_1 \cap \cdots \cap A_n) &= \mathbb{P}\left(A_n \,\middle|\, \bigcap_{j<n} A_j\right) \cdot \mathbb{P}\left(\bigcap_{j\leq n-1} A_j\right)\\ &\overset{\text{IV}}{=} \mathbb{P}\left(A_n \,\middle|\, \bigcap_{j<n} A_j\right) \prod_{k=1}^{n-1} \mathbb{P}\left(A_k \,\middle|\, \bigcap_{j<k} A_j\right).\end{aligned}$$

∎

Satz 3.5 (Satz von der totalen Wahrscheinlichkeit)**.** Sei $B_1, B_2, \ldots \in \mathfrak{A}$ eine endliche oder abzählbar unendliche Zerlegung von Ω, d.h. B_i sind paarweise disjunkt und $\bigcup_i B_i = \Omega$. Dann gilt für $A \in \mathfrak{A}$:

$$\mathbb{P}(A) = \sum_{\{i \,|\, \mathbb{P}(B_i)>0\}} \mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i).$$

Beweis.
Es ist $A = A \cap \Omega = A \cap \bigcup_i B_i = \bigcup_i (B_i \cap A)$, wobei die Mengen $B_i \cap A$ paarweise disjunkt sind für $i \in \mathbb{N}$. Es folgt

$$\mathbb{P}(A) = \sum_{\{i \,|\, \mathbb{P}(B_i)>0\}} \mathbb{P}(A \cap B_i) = \sum_{\{i \,|\, \mathbb{P}(B_i)>0\}} \mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i).$$

∎

Satz 3.6 (Bayes'sche Regel)**.** Sei $A \in \mathfrak{A}$ mit $\mathbb{P}(A) > 0$ und $B_1, B_2, \ldots \in \mathfrak{A}$ eine endliche oder abzählbar unendliche Zerlegung von Ω mit $\mathbb{P}(B_i) > 0$ für $i = 1, 2, \ldots$. Dann gilt

$$\mathbb{P}(B_i \,|\, A) = \frac{\mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i)}{\sum_{j\geq 1} \mathbb{P}(A \,|\, B_j)\mathbb{P}(B_j)}.$$

Beweis.

$$\mathbb{P}(B_i \,|\, A) = \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} \overset{(3.5)}{=} \frac{\mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i)}{\sum_{j\geq 1} \mathbb{P}(A \,|\, B_j)\mathbb{P}(B_j)}.$$

∎

Beispiel 3.7 (Test für seltene Krankheit)**.** Eine Krankheit trete insgesamt bei 0,5% der Bevölkerung auf.

$$\text{Ein Test führe bei } \begin{cases} 99\% \text{ der Kranken zur Reaktion,} \\ 2\% \text{ der Gesunden zur Reaktion.} \end{cases}$$

Gesucht ist die W-keit, dass eine Person, bei der der Test zur Reaktion führt, tatsächlich die Krankheit hat. Formalisierung: Sei X eine zufällig ausgewählte Person und $B =$ "X hat die Krankheit", d.h. $\mathbb{P}(B) = 0{,}005$. Ferner sei $A =$ "bei X führt Test zur Reaktion", d.h. $\mathbb{P}(A \mid B) = 0{,}99$ und $\mathbb{P}(A \mid B^c) = 0{,}02$. Die gesuchte Wahrscheinlichkeit ist $\mathbb{P}(B \mid A)$. Nach der Bayes'schen Regel und mit der disjunkten Zerlegung $\Omega = B \cup B^c$ gilt

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)} = \frac{495}{2485} \approx 0{,}2.$$

Von allen Personen, bei denen der Test zur Reaktion führt, sind also etwa 20% tatsächlich krank.

Definition 3.8. (a) Zwei Ereignisse $A, B \in \mathfrak{A}$ heißen (stochastisch) *unabhängig*, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ gilt.

(b) Sei $I \neq \emptyset$ eine Indexmenge und $A_i \in \mathfrak{A}$ für $i \in I$. Die Familie $(A_i)_{i \in I}$ heißt unabhängig, falls für jede endliche Teilfamilie $J \subset I$ gilt

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j) \qquad \text{„Produktformel“}.$$

(c) Die Familie $(A_i)_{i \in I}$ heißt *paarweise unabhängig*, falls für alle $i, j \in I$ mit $i \neq j$ gilt: A_i und A_j sind unabhängig.

Lemma 3.9. Sei $B \in \mathfrak{A}$ mit $\mathbb{P}(B) > 0$. Dann gilt:

$$A, B \text{ sind unabhängig } \Leftrightarrow \mathbb{P}(A \mid B) = \mathbb{P}(A).$$

Beweis.
„$\Rightarrow$“: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = (\mathbb{P}(A)\mathbb{P}(B))/\mathbb{P}(B) = \mathbb{P}(A)$.
„$\Leftarrow$“: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)$. ∎

Beispiel 3.10 (Zweimaliges Würfeln mit fairem Würfel)**.** Es ist

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\} = \{(\omega_1, \omega_2) \mid \omega_i \in \{1, \ldots, 6\} \text{ für } i = 1, 2\}.$$

Im Laplace-Modell gehen wir davon aus, dass die Wahrscheinlichkeit eines zweifachen Wurfes gegeben ist durch $\mathbb{P}(\{(\omega_1, \omega_2)\}) = 1/36$ für alle $(\omega_1, \omega_2) \in \Omega$. Seien $A =$“beim ersten Wurf < 4“ und $B =$“beim zweiten Wurf ≥ 3“.

$$\begin{aligned} A &= \{(\omega_1, \omega_2) \in \Omega : \omega_1 < 4\}, \quad |A| = 18, \quad \mathbb{P}(A) = \frac{1}{2} \\ B &= \{(\omega_1, \omega_2) \in \Omega : \omega_2 \geq 3\}, \quad |B| = 24, \quad \mathbb{P}(B) = \frac{2}{3}. \\ A \cap B &= \{(\omega_1, \omega_2) \in \Omega : \omega_1 < 4, \omega_2 \geq 3\}, \quad |A \cap B| = 12, \quad \mathbb{P}(A \cap B) = \frac{1}{3}. \end{aligned}$$

Es folgt $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, also sind A und B unabhängig.

Lemma 3.11. Sei $\{A_1, \dots, A_r\}$ eine unabhängige Familie von Ereignissen. Seien $B_1, \dots, B_r$ Ereignisse mit $B_i = A_i$ oder $B_i = A_i^c$ für $i = 1, \dots, r$. Dann ist die Familie $\{B_1, \dots, B_r\}$ unabhängig.

Beweis.
Zu zeigen ist, dass für jede Auswahl $(B_i)_{i\in J}$ mit $J \subset \{1, \dots, r\}$ die Produktformel gilt, d.h. $\mathbb{P}(\bigcap_{i\in J} B_i) = \prod_{i\in J} \mathbb{P}(B_i)$. Sei $|J| = s$ und o.B.d.A. $J = \{1, \dots, s\}$.
1. Fall: $B_i = A_i$ für $i = 1, \dots, s$. Dann folgt die Produktformel aus der Definition der Unabhängigkeit.
2. Fall: Gelte für genau ein $1 \le i \le s$: $B_i = A_i^c$ und $B_j = A_j$ für alle anderen $j \in J \setminus \{i\}$. O.B.d.A. sei $i = 1$, also $B_1 = A_1^c$, $B_j = A_j$ für $j = 2, \dots, s$. Dann ist eine disjunkte Vereinigung gegeben durch

$$\left(A_1^c \cap \bigcap_{\ell=2}^{s} A_\ell\right) \cup \left(A_1 \cap \bigcap_{\ell=2}^{s} A_\ell\right) = \bigcap_{\ell=2}^{s} A_\ell.$$

Also erhalten wir

$$\begin{aligned}\mathbb{P}\left(A_1^c \cap \bigcap_{\ell=2}^{s} A_\ell\right) &= \prod_{\ell=2}^{s} \mathbb{P}(A_\ell) - \mathbb{P}(A_1) \prod_{\ell=2}^{s} \mathbb{P}(A_\ell)\\ &= (1 - \mathbb{P}(A_1)) \prod_{\ell=2}^{s} \mathbb{P}(A_\ell) = \mathbb{P}(A_1^c) \prod_{\ell=2}^{s} \mathbb{P}(A_\ell).\end{aligned}$$

Der allgemeine Fall ergibt sich nun durch Induktion über die Anzahl der Indizes $1 \le i \le s$, für die $B_i = A_i^c$ gilt. Der Induktionsanfang ist gerade der 2. Fall, der Induktionsschritt kann mit einem ähnlichen Zerlegungsargument gezeigt werden. ∎

4 Produkträume

Wir wollen Modelle entwickeln, um n Zufallsexperimente unabhängig voneinander hintereinander ausführen zu können. Seien $(\Omega_1, \mathfrak{A}_1, \mathbb{P}_1), \dots, (\Omega_n, \mathfrak{A}_n, \mathbb{P}_n)$ diskrete W-Räume, die die Zufallsexperimente beschreiben. Betrachte

$$\begin{aligned}\Omega :=& \bigtimes_{i=1}^{n} \Omega_i = \Omega_1 \times \dots \times \Omega_n\\ =& \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \Omega_i \text{ für } i = 1, \dots, n\}.\end{aligned}$$

Sei $\mathfrak{A} := \mathcal{P}(\Omega)$ und seien π_i die Projektionen

$$\pi_i : \Omega \to \Omega_i, \quad \omega = (\omega_1, \dots, \omega_n) \mapsto \omega_i.$$

Für $A_i \in \mathfrak{A}_i$ sind Urbilder gegeben als

$$\begin{aligned}\pi_i^{-1}(A_i) &= \Omega_1 \times \Omega_2 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n\\ &= \text{„das } i\text{-te Teilexperiment hat Ausgang in } A_i\text{“}.\end{aligned}$$

Beispiel 4.1. Wir kommen zu Beispiel 3.10 zurück: $\Omega_i = \{1, \ldots, 6\}$ für $i = 1, 2$.
$\Omega = \Omega_1 \times \Omega_2 = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$,
A = "beim ersten Wurf < 4" $= \pi_1^{-1}(\{1, 2, 3\}) = \{1, 2, 3\} \times \Omega_2$,
B = "beim zweiten Wurf ≥ 3" $= \pi_2^{-1}(\{3, 4, 5, 6\}) = \Omega_1 \times \{3, 4, 5, 6\}$.

Die passende Wahl des W-Maßes in diesem Beispiel führt auf ein allgemeines Problem: Suche eine W-Verteilung $\mathbb{P}$ auf $(\Omega, \mathfrak{A})$, sodass:

(1) $\mathbb{P}\left(\pi_i^{-1}(A_i)\right) = \mathbb{P}_i(A_i)$ für alle $A_i \in \mathfrak{A}_i$ und $i = 1, \ldots, n$,

(2) $\left\{\pi_i^{-1}(A_i) : i = 1, \ldots, n\right\}$ soll eine unabhängige Familie in $(\Omega, \mathfrak{A}, \mathbb{P})$ sein für alle $A_i \in \mathfrak{A}_i$, $i = 1, \ldots, n$.

Satz 4.2. Seien $\Omega = \bigtimes_{i=1}^n \Omega_i$ sowie $\mathfrak{A}$ und π_i wie oben. Dann existiert genau eine W-Verteilung $\mathbb{P}$ auf $(\Omega, \mathfrak{A})$, sodass (1) und (2) gelten. Dabei ist $\mathbb{P}$ gegeben durch

$$\mathbb{P}(\{(\omega_1, \ldots, \omega_n)\}) = \prod_{i=1}^n \mathbb{P}_i(\{\omega_i\}), \quad (\omega_1, \ldots, \omega_n) \in \Omega. \tag{4}$$

Beweis.
Beweis hier für $n = 2$, der allgemeine Fall kann analog bewiesen werden.
Eindeutigkeit: Angenommen, $\mathbb{P}$ existiere mit (1) und (2). Wähle $A_1 = \{\omega_1\}$, $A_2 = \{\omega_2\}$ mit $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$. Dann ist $\pi_1^{-1}(A_1) = \{\omega_1\} \times \Omega_2$, $\pi_2^{-1}(A_2) = \Omega_1 \times \{\omega_2\}$ und $\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) = \{(\omega_1, \omega_2)\}$. Wir erhalten

$$\begin{aligned}\mathbb{P}\left(\{(\omega_1, \omega_2)\}\right) =& \mathbb{P}\left(\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2)\right) \overset{(2)}{=} \mathbb{P}\left(\pi_1^{-1}(A_1)\right) \mathbb{P}\left(\pi_2^{-1}(A_2)\right) \\ \overset{(1)}{=}& \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\})\end{aligned}$$

Also kann $\mathbb{P}$ höchstens die Form (4) haben.
Existenz: Z.z. $\mathbb{P}$ wie in (4) ist eine W-Verteilung mit (1) und (2).

- W-Maß: $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) = 1$, da $\mathbb{P}_1, \mathbb{P}_2$ W-Maße sind.

- Zu (1): Sei etwa $i = 1$:
$$\begin{aligned}\mathbb{P}\left(\pi_1^{-1}(A_1)\right) &= \mathbb{P}\left(A_1 \times \Omega_2\right) \overset{\sigma\text{-Add}}{=} \sum_{(\omega_1, \omega_2) \in A_1 \times \Omega_2} \mathbb{P}\left(\{(\omega_1, \omega_2)\}\right) \\ &\overset{(3)}{=} \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in \Omega_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) = \sum_{\omega_1 \in A_1} \mathbb{P}_1(\{\omega_1\}) = \mathbb{P}_1(A_1).\end{aligned}$$

- Zu (2): Seien $A_1 \in \mathfrak{A}_1$ und $A_2 \in \mathfrak{A}_2$. Dann $\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) = A_1 \times A_2$. Also
$$\begin{aligned}\mathbb{P}\left(\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2)\right) &= \mathbb{P}(A_1 \times A_2) = \sum_{(\omega_1, \omega_2) \in A_1 \times A_2} \mathbb{P}(\{(\omega_1, \omega_2)\}) \\ = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) &= \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \overset{(1)}{=} \mathbb{P}\left(\pi_1^{-1}(A_1)\right) \mathbb{P}\left(\pi_2^{-1}(A_2)\right),\end{aligned}$$
woraus die Unabhängigkeit folgt.

Es folgt die Behauptung. ∎

Definition 4.3. Der in Satz 4.2 definierte W-Raum $(\Omega, \mathfrak{A}, \mathbb{P})$ heißt das *Produkt der W-Räume* $(\Omega_i, \mathfrak{A}_i, \mathbb{P}_i)$. $\mathbb{P}$ wird *Produktmaß* genannt.

Anwendung: *Bernoulli-Experimente.*
Ein Zufallsexperiment mit zwei möglichen Ausgängen heißt Bernoulli-Experiment. Wir wollen die n-fache unabhängige Wiederholung modellieren. Sei $\Omega_i = \{0,1\}$ und $\mathbb{P}_i(\{1\}) = p = 1 - \mathbb{P}(\{0\})$. Man bezeichnet den Ausgang „1“ als Erfolg und nennt $p \in [0,1]$ die *Erfolgswahrscheinlichkeit.* Ein Modell für die n-fache unabhängige Wiederholung ist:

$$\Omega = \{0,1\}^n,\ \mathbb{P}(\{\omega\}) = p^k(1-p)^{n-k} \text{ für } \omega = (\omega_1, \dots, \omega_n) \text{ mit } \sum_{i=1}^{n} \omega_i = k.$$

Mit k wird also die Anzahl der Erfolge gezählt.
Zwei grundlegende Fragen dazu sind die Folgenden: Wie lassen sich W-keiten für die Anzahl der Erfolge und die Wartezeit auf den ersten Erfolg beschreiben? Wir betrachten analog zu Modell III aus Abschnitt 2

$$E_k = \left\{\omega \in \Omega : \sum_{i=1}^{n} \omega_i = k\right\} = \text{„genau } k \text{ Erfolge in } n \text{ Experimenten“}, \quad |E_k| = \binom{n}{k}.$$

Folglich ist $\mathbb{P}(E_k) = \sum_{\omega \in E_k} \mathbb{P}(\{\omega\}) = \sum_{\omega \in E_k} p^k(1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$.

Satz 4.4. Für die Parameter $n \in \mathbb{N}$ und $p \in [0,1]$ ist durch

$$b_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n$$

eine W-Verteilung auf $\{0, \dots, n\}$ definiert. Sie heißt *Binomialverteilung* mit Parameter $n \in \mathbb{N}$ und $p \in [0,1]$. Statt $b_{n,p}$ wird auch $B(n,p)$ geschrieben.

Korollar 4.5. Die Wahrscheinlichkeiten für die Anzahlen der Erfolge in n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in [0,1]$ sind durch die Binomialverteilung $b_{n,p}$ beschrieben.

Wir betrachten nun die Wahrscheinlichkeit, im k-ten Teilexperiment erstmals einen Erfolg zu haben. Dies ist das Ereignis

$$\begin{aligned} F_k &= \{\omega \in \Omega : \omega_1 = \dots = \omega_{k-1} = 0, \omega_k = 1\} \\ &= \{0\} \times \dots \times \{0\} \times \{1\} \times \Omega_{k+1} \times \dots \times \Omega_n, \end{aligned} \tag{5}$$

also ergibt sich

$$\mathbb{P}(F_k) = p(1-p)^{k-1}. \tag{6}$$

Man beachte, dass auch n Misserfolge möglich sind, weshalb die Zahlen in (6) für $k = 1, \dots, n$ keine W-Verteilung definieren können. Für $k \in \mathbb{N}$ liefert (6) allerdings eine W-Verteilung.

Satz 4.6. Zum Parameter $p \in (0, 1]$ ist durch

$$g_p(\{k\}) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

eine W-Verteilung auf $\mathbb{N}$ erklärt. Sie heißt *geometrische Verteilung* zum Parameter p.

Korollar 4.7. In einer Folge von unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in (0, 1]$ sind die Wahrscheinlichkeiten für den Index (Zeitpunkt), bei dem erstmals ein Erfolg eintritt, durch die geometrische Verteilung zum Parameter p beschrieben.

Verallgemeinerung: Zeitpunkt des r-ten Erfolgs, $r \in \mathbb{N}$. Wir betrachten die W-keit, im $(r+k)$-ten Teilexperiment den r-ten Erfolg zu beobachten ($k \in \mathbb{N}_0, r+k \leq n$). Dazu sei

$$G_k = \left\{ \omega \in \Omega : \sum_{i=1}^{r+k} \omega_i = r, \omega_{r+k} = 1 \right\}.$$

Für jedes $\omega \in G_k$ gilt $\mathbb{P}(\{\omega\}) = p^r(1-p)^{r+k-r} = p^r(1-p)^k$. Andererseits ist $|G_k| = \binom{r+k-1}{r-1}$, denn $r-1$ Indizes werden von $r+k-1$ gezogen ohne Zurücklegen ohne Reihenfolge. Damit gilt

$$\mathbb{P}(G_k) = \binom{r+k-1}{r-1} p^r (1-p)^k.$$

Satz 4.8. Zu den Parametern $p \in (0, 1]$ und $r \in \mathbb{N}$ ist durch

$$nb_{r,p}(\{k\}) = \binom{r+k-1}{r-1} p^r (1-p)^k, \quad k \in \mathbb{N}_0$$

eine W-Verteilung auf $\mathbb{N}_0$ gegeben. Sie heißt *negative Binomialverteilung* mit Parametern r und p (oder auch *Pascal*- oder *Pólya-Verteilung*).

Korollar 4.9. In einer Folge von unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in (0, 1]$ sind die W-keiten für die Anzahl der Misserfolge bis zum r-ten Erfolg durch die negative Binomialverteilung zu den Parametern r und p beschrieben.

Bemerkung 3. In der Literatur werden auch andere Vereinbarungen zum Gebrauch der Parameter der negativen Binomialverteilung gemacht. Beim Vergleich verschiedener Quellen sollte stets die individuelle Definition der negativen Binomialverteilung beachtet werden.

5 Diskrete Zufallsvariablen

Definition 5.1. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum, Ω' eine beliebige Menge. Jede Abbbildung

$$X : \Omega \to \Omega'$$

heißt Ω'-wertige *Zufallsvariable* (ZVe). Falls $\Omega' = \mathbb{R}$, so heißt X *reellwertige ZVe* (oder einfach nur ZVe), falls $\Omega' = \mathbb{R}^d$, so heißt X *Zufallsvektor*.

Satz 5.2. Sei $X : \Omega \to \Omega'$ eine Ω'-wertige diskrete ZVe und $\mathfrak{A}' = \mathcal{P}(\Omega')$. Dann ist durch

$$\mathbb{P}_X : \mathfrak{A}' \to [0,1], \quad A' \mapsto \mathbb{P}\left(X^{-1}(A')\right)$$

ein W-Maß auf $(\Omega', \mathfrak{A}')$ definiert. $\mathbb{P}_X$ heißt *Verteilung der ZVe X* (oder auch Bildmaß unter X).

Beweis.
$\mathbb{P}_X$ bildet offenbar in $[0,1]$ ab. Wir haben

$$\mathbb{P}_X(\Omega') = \mathbb{P}\left(X^{-1}(\Omega')\right) = \mathbb{P}(\Omega) = 1.$$

Sei $(A_i)_{i \geq 1}$ eine Folge paarweise disjunkter Mengen in $\mathfrak{A}'$. Dann folgt

$$\mathbb{P}_X\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{i\in\mathbb{N}} A_i\right)\right) = \mathbb{P}\left(\bigcup_{i\in\mathbb{N}} X^{-1}(A_i)\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(X^{-1}(A_i)\right) = \sum_{i=1}^{\infty} \mathbb{P}_X(A_i).$$

Es folgt die σ-Additivität. ∎

Bemerkung 4. Man beachte, dass Ω' i.A. nicht abzählbar ist (z.B. $\Omega' = \mathbb{R}$), jedoch bildet X nur auf eine höchstens abzählbare Menge $X(\Omega) = \{\omega' \in \Omega' \,|\, \exists\, \omega \in \Omega : X(\omega) = \omega'\}$ ab. $\mathbb{P}_X$ ist also auf $X(\Omega)$ ein diskretes W-Maß.

Notationen 1. Die folgenden Kurzschreibweisen sind gebräuchlich:
$X^{-1}(A) = \{\omega \in \Omega \,|\, X(\omega) \in A\} =: \{X \in A\}$.
Für reelle ZVe X: Sei $A = (-\infty, x]$, dann $X^{-1}(A) =: \{X \leq x\}$.

$$\mathbb{P}_X(A) = \mathbb{P}(\{X \in A\}) =: \mathbb{P}(X \in A),$$
$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X \in \{k\}) =: \mathbb{P}(X = k).$$

Beispiel 5.3 (n-maliger Münzwurf)**.** Wir betrachten n unabhängige Bernoulli Experimente mit Erfolgswahrscheinlichkeit $p \in [0,1]$. $\Omega = \{0,1\}^n$, $\mathbb{P}(\{(\omega_1, \ldots, \omega_n\}) = p^k(1-p)^{n-k}$ mit $k = \sum_{i=1}^n \omega_i$. Die ZVe

$$X : \Omega \to \mathbb{N}, \quad \omega \mapsto \sum_{i=1}^{n} \omega_i$$

beschreibt die Anzahl der „Erfolge" in Ω. Es ist $X(\Omega) = \{0, \ldots, n\}$. Wie lautet die Verteilung von X? Sei $A_k := \{(\omega_1, \ldots, \omega_n) \in \Omega : \sum_{i=1}^n \omega_i = k\}$. Dann

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(\{X = k\}) = \mathbb{P}\left(X^{-1}(\{k\})\right) = \mathbb{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Die Anzahl der Erfolge bei n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in [0,1]$ ist binomial $b_{n,p}$ verteilt (d.h. die Verteilung $\mathbb{P}_X$ von X ist die Binomialverteilung mit Parametern n und p, vgl. Korollar 4.5.)

Beispiel 5.4 (k-maliges Ziehen ohne Rücklegen aus einer Urne mit s schwarzen und w weißen Kugeln)**.** Betrachte eine ZVe $X =$"Anzahl gezogener schwarzer Kugeln". Seien $A = \{1, \dots, n\}$, $A_s = \{1, \dots, s\}$, $A_w = A \setminus A_s$, $\Omega = \left\{(\omega_1, \dots, \omega_k) \in A^k \,|\, \omega_1 < \dots < \omega_k\right\}$ (vgl. Beispiel 2.12). Damit haben wir

$$X : \Omega \to \{0, \dots, k\}, \quad (\omega_1, \dots, \omega_k) \mapsto \sum_{i=1}^{k} \mathbb{1}_{A_s}(\omega_i).$$

Wie in Abschnitt 2 gezeigt, liefert das Laplace-Modell auf Ω:

$$\mathbb{P}_X(\{\ell\}) = \mathbb{P}\left(X^{-1}(\ell)\right) = \mathbb{P}(B_\ell) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}}.$$

Die Anzahl X der schwarzen Kugeln ist folglich hypergeometrisch verteilt (zu entsprechenden Parametern).

Ebenso: Bei einer Folge von unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ ist die Wartezeit bis zum ersten Erfolg eine Zufallsvariable, die geometrisch g_p verteilt ist.

Definition 5.5. Seien $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum und $X_i : \Omega \to \Omega_i$ ZVe für $i \in I, I \neq \emptyset$. Die Familie $\{X_i \,|\, i \in I\}$ von ZVen heißt *(stochastisch) unabhängig*, falls für jede Wahl $A_i \in \Omega_i$ die Familie von Ereignissen $\{\{X_i \in A_i\} \,|\, i \in I\}$ unabhängig ist.

Satz 5.6. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum und $X_1, \dots, X_n$ ZVe auf Ω, $X_i : \Omega \to \Omega_i$. Dann sind äquivalent:

a) $X_1, \dots, X_n$ sind unabhängig.

b) Für alle $(x_1, \dots, x_n) \in \Omega_1 \times \dots \times \Omega_n$ gilt

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i = x_i\}\right) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

c) Für beliebige $A_i \in \Omega_i$ gilt

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

Beweis.
a) $\Rightarrow$ c): folgt aus den Definitionen 5.5 und 3.8 b).
c) $\Rightarrow$ b): Wähle $A_i = \{x_i\}$ für $i = 1, \dots, n$.
b) $\Rightarrow$ a): Seien $A_i \subset \Omega_i$ Mengen. Z.z. ist, dass $\{\{X_i \in A_i\} : i = 1, \dots, n\}$ Familie unabhängiger Mengen ist, d.h. für jede Teilfamilie die Produktformel gilt. Wir betrachten

o.E. die Teilfamilie $\{1,\dots,s\} \subset \{1,\dots,n\}$. Dann gilt

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{s}\{X_i \in A_i\}\right) &= \mathbb{P}\left(\bigcap_{i=1}^{s}\bigcup_{x_i\in A_i}\{X_i = x_i\}\right) = \mathbb{P}\left(\bigcup_{x_1\in A_1}\cdots\bigcup_{x_s\in A_s}\left(\bigcap_{i=1}^{s}\{X_i = x_i\}\right)\right)\\
&= \sum_{x_1\in A_1}\cdots\sum_{x_s\in A_s}\mathbb{P}\left(\bigcap_{i=1}^{s}\{X_i = x_i\}\right) = \sum_{x_1\in A_1}\cdots\sum_{x_s\in A_s}\prod_{i=1}^{s}\mathbb{P}(X_i = x_i)\\
&= \left(\sum_{x_1\in A_1}\mathbb{P}(X_1 = x_1)\right)\cdots\left(\sum_{x_s\in A_s}\mathbb{P}(X_s = x_s)\right)\\
&= \mathbb{P}(X_1 \in A_1)\cdots\mathbb{P}(X_s \in A_s).
\end{aligned}$$

Also gilt die Produktformel. ∎

Satz 5.7. Seien $X_1,\dots,X_n$ unabhängige ZVe, $X_i : \Omega \to \Omega_i$. Seien $f_i : \Omega_i \to \Gamma_i$ Funktionen für $i = 1,\dots,n$. Dann sind die ZVe $Y_1,\dots,Y_n$ mit $Y_i = f_i \circ X_i$ unabhängig.

Beweis.
Seien $A_i \subset \Gamma_i$ für $i = 1,\dots,n$ beliebig. Dann ist

$$\{Y_i \in A_i\} = \{f_i \circ X_i \in A_i\} = \left\{X_i \in f_i^{-1}(A_i)\right\}.$$

Da $X_1,\dots,X_n$ unabhängig sind, ist $\{\{X_i \in f_i^{-1}(A_i)\} : i = 1,\dots,n\}$ eine unabhängige Familie von Ereignissen. Folglich ist $\{\{Y_i \in A_i\} : i = 1,\dots,n\}$ eine unabhängige Familie. Es folgt die Unabhängigkeit von $Y_1,\dots,Y_n$. ∎

Als Beispiel zur stochastischen Unabhängigkeit betrachten wir die lokalen Ränge einer gleichverteilten, zufälligen Permutation:

Definition 5.8. Sei $\pi \in \mathcal{S}_n$ eine Permutation der Länge n. Dann heißt

$$R_i(\pi) = |\{1 \le j \le i : \pi_j \le \pi_i\}|$$

lokaler Rang von π_i *in* π. Falls $R_i = i$, so heißt π_i ein *(auf-)Rekord* in π, (falls $R_i = 1$, so heißt π_i ein *ab-Rekord.*)

Sei nun $\Omega = \mathcal{S}_n$ mit der Gleichverteilung $\mathbb{P}$ versehen.

$$\begin{aligned}
X_i &: \Omega \to \{1,\dots,i\}, \quad \omega \mapsto R_i(\omega),\\
Y_i &: \Omega \to \{0,1\}, \quad \omega \mapsto \mathbb{1}_{\{i\}}(X_i).
\end{aligned}$$

Satz 5.9. Der lokale Rang X_i einer zufälligen gleichverteilten Permutation ist gleichverteilt auf $\{1,\dots,i\}$ für alle $i = 1,\dots,n$. Die ZVe $X_1,\dots,X_n$ sind unabhängig. Die ZVe $Y_1,\dots,Y_n$ sind unabhängig.

Beweis.
Wir zeigen zunächst, dass X_i gleichverteilt auf $\{1, \dots, i\}$ ist. Sei dazu $k \in \{1, \dots, i\}$ beliebig gegeben und $\pi = (\pi_1, \dots, \pi_n)$. Nach dem Satz von der totalen W-keit 3.5 gilt

$$\mathbb{P}(X_i = k) = \sum_{\substack{A \subset \{1,\dots,n\} \\ |A| = i}} \mathbb{P}(X_i = k \,|\, \{\pi_1, \dots, \pi_i\} = A)\, \mathbb{P}(\{\pi_1, \dots, \pi_i\} = A).$$

Wir haben $\mathbb{P}(X_i = k \,|\, \{\pi_1, \dots, \pi_i\} = A) = \frac{1}{i}$, da gegeben, dass die ersten i Werte der Permutation die Elemente aus A sind, es gleichwahrscheinlich ist, welches dieser Elemente an Position i steht. Aus dem Laplace-Modell folgt ferner

$$\mathbb{P}(\{\pi_1, \dots, \pi_i\} = A) = \binom{n}{i}^{-1}.$$

Damit gilt

$$\mathbb{P}(X_i = k) = \sum_{\substack{A \subset \{1,\dots,n\} \\ |A| = i}} \frac{1}{i} \frac{1}{\binom{n}{i}} = \binom{n}{i} \frac{1}{i} \frac{1}{\binom{n}{i}} = \frac{1}{i}, \quad k = 1, \dots, i.$$

Zur Unabhängigkeit der $X_1, \dots, X_n$: Man beachte, dass zu jeder Wahl von Werten $1 \leq x_i \leq i$ für $i = 1, \dots, n$ genau eine Permutation $\pi \in \mathcal{S}_n$ existiert mit $R_i(\pi) = x_i$ für $i = 1, \dots, n$. Dies sieht man wie folgt ein: $R_n(\pi) = x_n$ legt den Wert $\pi_n = x_n$ fest. Nun legt aber $R_{n-1}(\pi) = x_{n-1}$ den Wert π_{n-1} fest, denn dies ist gerade die x_{n-1}-kleinste der Zahlen $\{1, \dots, n\} \setminus \{x_n\}$. Ebenso werden die weiteren Werte $\pi_{n-2}, \dots, \pi_1$ festgelegt. Diese Bijektion liefert

$$\begin{aligned}\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) &= \frac{1}{n!} = \frac{1}{1} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{n} \\ &= \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \cdots \mathbb{P}(X_n = x_n).\end{aligned}$$

Nach Satz 5.6 sind $X_1, \dots, X_n$ also unabhängig.
Schließlich haben die Y_i die Form $Y_i = f_i(X_i)$ mit $f_i(x) = \mathbb{1}_{\{i\}}(x)$. Nach Satz 5.7 sind damit auch $Y_1, \dots, Y_n$ unabhängig. ∎

6 Erwartungswert und Varianz

In diesem Abschnitt sei stets $(\Omega, \mathfrak{A}, \mathbb{P})$ ein diskreter W-Raum, auf dem alle auftretenden ZVe definiert sind.

Definition 6.1. Sei X reellwertige ZVe. Falls $\sum_{\omega \in \Omega} |X(\omega)|\, \mathbb{P}(\{\omega\}) < \infty$ ist, so existiert die *Erwartung* (Erwartungswert, EW) von X und ist gegeben durch

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

Lemma 6.2. Sei $\{x_1, x_2, \ldots\}$ eine Abzählung des Wertebereichs von X. Es existiere der EW von X. Dann gilt

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \, \mathbb{P}(X = x_i) = \sum_{i=1}^{\infty} x_i \, \mathbb{P}_X(\{x_i\}).$$

Beweis.
Es gilt

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \, \mathbb{P}(\{\omega\}) = \sum_{i=1}^{\infty} \sum_{\{\omega \,|\, X(\omega) = x_i\}} X(\omega) \, \mathbb{P}(\{\omega\}) = \sum_{i=1}^{\infty} x_i \, \mathbb{P}(X = x_i).$$

Die liefert die Behauptungen. ∎

Bemerkung 5. Der Erwartungswert von X hängt also nur von der Verteilung $\mathbb{P}_X$ der ZVen X ab. Derartige Größen heißen *Verteilungsgrößen*.

Satz 6.3. Seien X, Y reellwertige ZVe mit existierenden EWen. Dann gelten:

(i) Für $\lambda \in \mathbb{R}$ existiert der EW von λX, und es gilt $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$.

(ii) Der EW von $X + Y$ existiert, und es gilt $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

(iii) Sind X, Y unabhängig, so existiert der EW von XY und es gilt

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

(iv) Falls $X \geq 0$, so gilt $\mathbb{E}[X] \geq 0$.

(v) Falls $X \geq Y$ (punktweise), so gilt $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

(vi) Es ist $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ für alle $A \in \mathfrak{A}$.

Beweis.

(i) Existenz: $\sum_{\omega \in \Omega} |\lambda X(\omega)| \, \mathbb{P}(\{\omega\}) = |\lambda| \sum_{\omega \in \Omega} |X(\omega)| \, \mathbb{P}(\{\omega\}) < \infty$. Folglich ist

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

(ii) Mit der Dreiecksungleichung gilt

$$\begin{aligned} \sum_{\omega \in \Omega} |X(\omega) + Y(\omega)| \, \mathbb{P}(\{\omega\}) &\leq \sum_{\omega \in \Omega} (|X(\omega)| + |Y(\omega)|) \, \mathbb{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} |X(\omega)| \, \mathbb{P}(\{\omega\}) + \sum_{\omega \in \Omega} |Y(\omega)| \, \mathbb{P}(\{\omega\}) < \infty. \end{aligned}$$

Die gleiche Rechnung ohne Beträge (und Gleichheit statt der Dreiecksungleichung) liefert die Behauptung.

(iii) Seien $\{x_1, x_2, \dots\}$ und $\{y_1, y_2, \dots\}$ Abzählungen der Wertemengen von X und Y. Dann folgt wie im Beweis von Lemma 6.2, dass

$$\begin{aligned}\sum_{\omega\in\Omega} |X(\omega)Y(\omega)|\,\mathbb{P}(\{\omega\}) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |x_i y_j|\,\mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |x_i||y_j|\,\mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) \\ &= \left(\sum_{i=1}^{\infty} |x_i|\,\mathbb{P}(X = x_i)\right)\left(\sum_{j=1}^{\infty} |y_j|\,\mathbb{P}(Y = y_j)\right) < \infty.\end{aligned}$$

Die gleiche Rechnung ohne Beträge liefert $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(iv) Aus $X \geq 0$ folgt $\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\{\omega\}) \geq 0$.

(v) Sei $X(\omega) \geq Y(\omega)$ für jedes $\omega \in \Omega$, dann ist $Z(\omega) := X(\omega) - Y(\omega) \geq 0$. Die Behauptung folgt mit

$$0 \overset{(iv)}{\leq} \mathbb{E}[Z] \overset{(i),(ii)}{=} \mathbb{E}[X] - \mathbb{E}[Y].$$

(vi) Es ist $\mathbb{E}[\mathbb{1}_A] = \sum_{\omega\in\Omega} \mathbb{1}_A(\omega)\mathbb{P}(\{\omega\}) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}) = \mathbb{P}(A)$.

∎

Korollar 6.4. Die Menge aller auf $(\Omega, \mathfrak{A}, \mathbb{P})$ definierten reellwertigen ZVe mit existierendem Erwartungswert ist ein Vektorraum, der mit $\mathcal{L}_1(\Omega, \mathfrak{A}, \mathbb{P})$ bezeichnet wird. Der Erwartungswert ist ein lineares Funktional

$$\mathbb{E}[\,\cdot\,] : \mathcal{L}_1(\Omega, \mathfrak{A}, \mathbb{P}) \to \mathbb{R}.$$

Beweis.
Nach Satz 6.3 ist $\mathcal{L}_1(\Omega, \mathfrak{A}, \mathbb{P})$ abgeschlossen bzgl. Addition und Multiplikation mit Skalaren, also ein Untervektorraum der Menge aller reellwertigen ZVen auf $(\Omega, \mathfrak{A}, \mathbb{P})$. Die Abbildung $X \mapsto \mathbb{E}[X]$ ist linear nach (i) und (ii) in Satz 6.3. ∎

Beispiel 6.5. Sei X binomial $b_{n,p}$ verteilt mit $n \in \mathbb{N}$ und $p \in [0, 1]$. Dann gilt $\mathbb{E}[X] = np$.

Beweis.
Rechnerisch: Wegen $X \leq n$ existiert der EW von X. Es gilt

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

Man kann dies allerdings auch direkt einsehen: X ist verteilt wie die Anzahl der Erfolge bei n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p. Also gilt $X = \sum_{i=1}^n \mathbb{1}_{A_i}$, wobei $A_i = \{\text{Erfolg im } i\text{-ten Teilexperiment}\}$, also $\mathbb{P}(A_i) = p$ für $i = 1, \ldots, n$. Es folgt

$$\mathbb{E}[X] \overset{6.3(ii)}{=} \sum_{i=1}^n \mathbb{E}[\mathbb{1}_{A_i}] \overset{6.3(vi)}{=} \sum_{i=1}^n \mathbb{P}(A_i) = np.$$

∎

Wir betrachten nun Erwartungswerte passender Kompositionen: Sei $X : \Omega \to \Omega'$ eine ZVe und $f : \Omega' \to \mathbb{R}$ eine Funktion. Mit $\mathfrak{A}' = \mathcal{P}(\Omega')$ und $\mathbb{P}_X$ wird der Definitionsbereich von f zu einem Wahrscheinlichkeitsraum $(\Omega', \mathfrak{A}', \mathbb{P}_X)$. Deshalb kann man die Abbildung $f : \Omega' \to \mathbb{R}$ als reellwertige ZVe auffassen. Der Deutlichkeit halber bezeichne $\mathbb{E}_{\mathbb{P}}$ den EW von ZVen auf Ω, $\mathbb{E}_{\mathbb{P}_X}$ den EW von ZVen auf Ω'.

Satz 6.6 (Transformationssatz)**.** In der Situation der vorigen fünf Zeilen gilt: Es existiert der EW von $f \circ X$ (bez. $\mathbb{E}_{\mathbb{P}}$) genau dann, wenn der EW von f (bez. $\mathbb{P}_X$) existiert. In diesem Fall gilt

$$\mathbb{E}_{\mathbb{P}}[f \circ X] = \mathbb{E}_{\mathbb{P}_X}[f].$$

Beweis.
Sei $\{x_1, x_2, \ldots\}$ eine Abzählung der Werte von X und $A_i = \{X = x_i\} \subset \Omega$. Dann bilden $A_1, A_2, \ldots$ eine disjunkte Zerlegung von Ω und für $\omega \in A_i$ gilt $X(\omega) = x_i$. Damit folgt

$$\begin{aligned} \sum_{i=1}^\infty |f(x_i)| \, \mathbb{P}_X(\{x_i\}) &= \sum_{i=1}^\infty |f(x_i)| \, \mathbb{P}(A_i) = \sum_{i=1}^\infty \sum_{\omega \in A_i} |f(x_i)| \, \mathbb{P}(\{\omega\}) \\ &= \sum_{i=1}^\infty \sum_{\omega \in A_i} |f(X(\omega))| \, \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} |f(X(\omega))| \, \mathbb{P}(\{\omega\}). \end{aligned}$$

Damit hat f einen EW bez. $\mathbb{P}_X$ genau dann, wenn $f \circ X$ einen EW bzgl $\mathbb{P}$ hat. Die gleiche Rechnung ohne Beträge liefert die Behauptung. ∎

Bemerkung 6. Lemma 6.2 kann wie folgt umgeschrieben werden: Bezeichne $\mathrm{id} : \mathbb{R} \to \mathbb{R}$, $x \mapsto x$ die Identität auf $\mathbb{R}$. Dann gilt

$$\mathbb{E}[X] = \mathbb{E}_{\mathbb{P}}[X] = \mathbb{E}_{\mathbb{P}}[\mathrm{id} \circ X] = \mathbb{E}_{\mathbb{P}_X}[\mathrm{id}].$$

Der Erwartungswert kann als Maßzahl für den „Schwerpunkt" bzw. den „mittleren Wert" einer Verteilung aufgefasst werden. Wir betrachten zudem Maßzahlen für Streuung um den Erwartungswert.

Definition 6.7. Seien X, Y reelle ZVe, sodass X^2, Y^2 existierenden EW haben. Dann heißen:

- $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$ die *Varianz* von X,

- $\sigma_X := \sqrt{\mathrm{Var}(X)}$ die *Standardabweichung* von X,
- $\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$ die *Kovarianz* von X und Y,
- $\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$ der *Korrelationskoeffizient* von X und Y.

X und Y heißen *unkorreliert*, falls $\mathrm{Cov}(X,Y) = 0$ gilt.

Bemerkung 7. Alle EWe in Definition 6.7 sind definiert (existieren), da $|X| \leq 1 + X^2$, also existiert $\mathbb{E}[X]$ und $(X - \mathbb{E}[X])^2 \leq X^2 + 2|\mathbb{E}[X]||X| + (\mathbb{E}[X])^2$. Also existiert auch $\mathbb{E}[(X - \mathbb{E}[X])^2]$, ferner benutze man $|X \cdot Y| \leq X^2 + Y^2$. Falls X eine ZVe ist, so dass der EW von X^2 existiert, sagt man, dass X ein endliches *zweites Moment* habe.

Satz 6.8. Seien $X, Y, X_1, \ldots, X_n$ ZVe mit endlichem zweiten Moment. Dann gilt für alle $a, b, c, d \in \mathbb{R}$:

(i) $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$,

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

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

(iv) $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X,Y)$,

(v) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,

(vi) $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$,

(vii) X, Y unabhängig $\Rightarrow X, Y$ unkorreliert.

Beweis.
(i)-(v) folgen direkt aus der Definition. Z.B. für (i):

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

(vi) Wegen (ii) können wir o.E. $\mathbb{E}[X_i] = 0$ für $i = 1, \ldots, n$ annehmen. Dann folgt

$$\mathrm{Var}(X_1 + \cdots + X_n) \overset{(i)}{=} \mathbb{E}[(X_1 + \cdots + X_n)^2] = \sum_{i,j \leq n} \mathbb{E}[X_i X_j] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j].$$

Wegen $\mathbb{E}[X_i] = 0$ ist dies $\sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$.
(vii) Seien X, Y unabhängig. Nach Satz 5.7 sind dann $(X - \mathbb{E}[X]), (Y - \mathbb{E}[Y])$ unabhängig. Also

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X - \mathbb{E}[X]]\mathbb{E}[Y - \mathbb{E}[Y]] = 0 \cdot 0 = 0.$$

Somit sind X, Y unkorreliert. ∎

Satz 6.9 (Bienaymé)**.** Seien $X_1, \dots, X_n$ unabhängig mit endlichem zweiten Moment. Dann gilt

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Beweis.
O.E. gelte $\mathbb{E}[X_i] = 0$ für $i = 1, \dots, n$. Nach Satz 6.8 (vi) gilt $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathbb{E}[X_i X_j]$. Nach Satz 6.8 (vii) sind X_i, X_j unkorreliert für $i \neq j$. Damit ist $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = 0$. Es folgt die Behauptung. ∎

Beispiel 6.10. Sei X binomial $b_{n,p}$ verteilt mit $n \in \mathbb{N}$ und $p \in [0,1]$. Dann gilt

$$\mathrm{Var}(X) = np(1-p).$$

Beweis.
Wie im Beweis von Beispiel 6.5 haben wir $X = \sum_{i=1}^{n} \mathbb{1}_{A_i}$, wobei $\mathbb{P}(A_i) = p$ und $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ unabhängige ZVe sind. Es gilt $\mathrm{Var}(\mathbb{1}_{A_i}) = \mathbb{E}[\mathbb{1}_{A_i}^2] - \mathbb{E}[\mathbb{1}_{A_i}]^2 = p - p^2 = p(1-p)$. Nach Satz 6.9 liefert Unabhängigkeit der Indikatoren, dass $\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(\mathbb{1}_{A_i}) = np(1-p)$. ∎

7 Erzeugende Funktionen

Erzeugende Funktionen sind ein analytisches Hilfsmittel zum Studium von Wahrscheinlichkeitsverteilungen auf $\mathbb{N}_0$.

Definition 7.1. Die *erzeugende Funktion* einer W-Verteilung μ auf $\mathbb{N}_0$ mit $\mu(\{k\}) =: p_k$ ist gegeben durch

$$g(s) = g_\mu(s) = \sum_{k=0}^{\infty} p_k s^k. \tag{7}$$

Beobachtungen:

- Die Funktion in (7) ist innerhalb des Konvergenzradius (also im Konvergenzintervall) der Potenzreihe definiert. Für $|s| \leq 1$ gilt $\sum_{k=0}^{\infty} p_k |s^k| \leq \sum p_k = 1 < \infty$. Die Potenzreihe in (7) ist also mindestens für $|s| \leq 1$ definiert. (Für den Konvergenzradius r gilt $r \geq 1$.)
- Sei X eine ZVe in $\mathbb{N}_0$ mit Verteilung $\mathbb{P}_X$. Dann ist $g_{\mathbb{P}_X}(s) = \mathbb{E}[s^X]$.
- Ableitungen: Bezeichne $g^{(n)}$ die n-te Ableitung von g, $g^{(0)} := g$. Dann gilt

  $$g'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}, \quad g^{(n)}(s) = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} p_k s^{k-n}.$$

  Speziell für $s = 0$: $g^{(n)}(0) = n! p_n$ also, $p_n = g^{(n)}(0)/n!$, $n \in \mathbb{N}_0$.

Korollar 7.2. Eine Verteilung auf $\mathbb{N}_0$ ist eindeutig durch ihre erzeugende Funktion festgelegt, d.h. die Abbildung $\mu \mapsto g_\mu$ ist injektiv.

Beispiel 7.3. (i) Poisson-Verteilung Π_λ zum Parameter $\lambda > 0$: Es gilt $p_k = e^{-\lambda}\frac{\lambda^k}{k!}$. Damit folgt für alle $s \in \mathbb{R}$

$$g_{\Pi_\lambda}(s) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}s^k = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(s\lambda)^k}{k!} = e^{\lambda(s-1)}.$$

(ii) Sei X geometrisch g_p verteilt mit $p \in (0,1)$: Es ist dann $p_k = p(1-p)^{k-1}$ für $k \geq 1$. Damit gilt

$$g_{g_p}(s) = \sum_{k=1}^{\infty} p(1-p)^{k-1}s^k = \frac{p}{1-p}\sum_{k=1}^{\infty}((1-p)s)^k = \frac{ps}{1-s(1-p)}$$

für $|s| < 1/(1-p)$. Der Konvergenzradius ist hier also beschränkt.

(iii) Für die Binomialverteilung $b_{n,p}$ mit $n \in \mathbb{N}$ und $p \in [0,1]$ erhalten wir für $s \in \mathbb{R}$

$$g_{b_{n,p}}(s) = (ps + 1 - p)^n.$$

Ein allgemeines Problem stellt die Rückgewinnung von Information über μ aus g_μ dar. Wir diskutieren, wie Erwartungswert und Varianz einer Zufallsvariable X mit Verteilung $\mathbb{P}_X = \mu$, falls diese jeweils existieren, aus g_μ berechnet werden können.

Satz 7.4. Sei X eine ZVe in $\mathbb{N}_0$ mit Verteilung $\mathbb{P}_X$ und erzeugender Funktion $g = g_{\mathbb{P}_X}$. Der EW von X existiert genau dann, wenn der linksseitige Grenzwert

$$g'(1-) := \lim_{s\uparrow 1} g'(s)$$

existiert. Es gilt dann $\mathbb{E}[X] = g'(1-)$.

Beweis.
Für $|s| < 1$ ist $g'(s) = \sum_{k=1}^{\infty} kp_k s^{k-1}$ endlich. Falls $\mathbb{E}[X]$ existiert, ist also auch $\sum_{k=1}^{\infty} kp_k < \infty$. Nach Korollar 25.4 der Vorlesung Analysis 1 (was im Wesentlichen der Abelsche Stetigkeitssatz für Potenzreihen ist) folgt dann $\lim_{s\uparrow 1} g'(s) = \sum_{k=1}^{\infty} kp_k = \mathbb{E}[X]$. Falls der EW von X nicht existiert, gilt $\sum_{k\geq 1} kp_k = \infty$ und damit $\lim_{s\uparrow 1} g'(s) = \infty$. ∎

Bemerkung 8. Analog zeigt man, dass der EW von $X(X-1)\cdots(X-k+1)$ genau dann existiert, wenn $g^{(k)}(1-) := \lim_{s\uparrow 1} g^{(k)}(s)$ existiert. In diesem Falle gilt

$$\mathbb{E}[X(X-1)\cdots(X-k+1)] = g^{(k)}(1-).$$

Wegen $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = g''(1-) + g'(1-)$ folgt mit Satz 6.8 (i):

Korollar 7.5. Sei X eine ZVe in $\mathbb{N}_0$ mit Verteilung $\mathbb{P}_X$ und erzeugender Funktion $g = g_{\mathbb{P}_X}$. Falls $\lim_{s \uparrow 1} g''(s)$ existiert, so gilt

$$\mathrm{Var}(X) = g''(1-) + g'(1-) - (g'(1-))^2.$$

Beispiel 7.6. Sei X poissonverteilt zum Parameter $\lambda > 0$. Damit ist $g_{\mathbb{P}_X}(s) = e^{\lambda(s-1)}$ und wir erhalten $\mathbb{E}[X] = \lambda$ und $\mathrm{Var}(X) = \lambda$.

Satz 7.7. Sind X, Y unabhängige ZVe in $\mathbb{N}_0$ mit erzeugenden Funktionen $g_X := g_{\mathbb{P}_X}$ und $g_Y := g_{\mathbb{P}_Y}$. Dann gilt für die erzeugende Funktion $g_{X+Y} := g_{\mathbb{P}_{X+Y}}$

$$g_{X+Y}(s) = g_X(s) g_Y(s)$$

für alle s, für die sowohl g_X als auch g_Y definiert ist.

Beweis.
Nach Satz 5.7 sind die Zufallsvariablen s^X und s^Y unabhängig. Es folgt also

$$g_{X+Y}(s) = \mathbb{E}\left[s^{X+Y}\right] = \mathbb{E}\left[s^X \cdot s^Y\right] \overset{6.3(iii)}{=} \mathbb{E}\left[s^X\right] \mathbb{E}\left[s^Y\right] = g_X(s) g_Y(s).$$

∎

Beispiel 7.8. Seien X, Y unabhängige ZVe mit Poissonverteilungen $\mathbb{P}_X = \Pi_\lambda$ und $\mathbb{P}_Y = \Pi_\mu$ zu Parametern $\lambda, \mu > 0$. Dann ist $X + Y$ poissonverteilt zum Parameter $\lambda + \mu$, d.h. es gilt $\mathbb{P}_{X+Y} = \Pi_{\lambda+\mu}$.

Beweis.
Mit Satz 7.7 folgt $g_{X+Y}(s) = \exp(\lambda(s-1)) \exp(\mu(s-1)) = \exp((\lambda+\mu)(s-1))$. Dies ist die erzeugende Funktion der $\Pi_{\lambda+\mu}$-Verteilung. Nach Korollar 7.2 ist $X + Y$ damit $\Pi_{\lambda+\mu}$-verteilt. ∎

Analog zeigt man: Sind X und Y unabhängige ZVe mit Binomialverteilungen $\mathbb{P}_X = b_{n,p}$ und $\mathbb{P}_Y = b_{m,p}$ mit $n, m \in \mathbb{N}$ und gemeinsamem $p \in [0, 1]$. Dann ist $X+Y$ binomialverteilt zu den Parametern $n + m$ und p, d.h. es gilt $\mathbb{P}_{X+Y} = b_{n+m,p}$.

2 Allgemeine Modelle

In diesem Kapitel werden nun auch Wahrscheinlichkeitsräume definiert und untersucht, die nicht abzählbar zu sein brauchen.

8 Allgemeine Wahrscheinlichkeitsräume

Definition 8.1. Sei $\Omega \neq \emptyset$ eine beliebige Menge. Ein System (Familie) $\mathfrak{A}$ von Teilmengen von Ω heißt σ-Algebra (über Ω), falls gelten:

(a) $\Omega \in \mathfrak{A}$,

(b) $\forall A \in \mathfrak{A} : A^c \in \mathfrak{A}$,

(c) Für jede Folge $(A_i)_{i\geq 1}$ in $\mathfrak{A}$ gilt $\bigcup_{i\geq 1} A_i \in \mathfrak{A}$.

Lemma 8.2. Sei $\mathfrak{A}$ eine σ-Algebra. Dann gilt

(a) $\emptyset \in \mathfrak{A}$.

(b) Ist $(A_i)_{i\geq 1}$ eine Folge in $\mathfrak{A}$, so gilt $\bigcap_{i\geq 1} A_i \in \mathfrak{A}$.

(c) Seien $A_1, \dots, A_n \in \mathfrak{A}$, so gilt $A_1 \cap \dots \cap A_n \in \mathfrak{A}$ und $A_1 \cup \dots \cup A_n \in \mathfrak{A}$.

Beweis.

(a) $\emptyset = \Omega^c$,

(b) $\bigcap_{i\geq 1} A_i = \left(\bigcup_{i\geq 1} A_i^c\right)^c$,

(c) $A_1 \cap \dots \cap A_n = A_1 \cap \dots \cap A_n \cap \Omega \cap \Omega \cap \dots \in \mathfrak{A}$,
$A_1 \cup \dots \cup A_n = A_1 \cup \dots \cup A_n \cup \emptyset \cup \emptyset \cup \dots \in \mathfrak{A}$.

∎

Beispiel 8.3.

- Sei $\Omega \neq \emptyset$ eine beliebige Menge, so ist $\mathcal{P}(\Omega)$ eine σ-Algebra.
- Sei $A \subset \Omega$, so ist $\{\emptyset, A, A^c, \Omega\}$ eine σ-Algebra.
- Sei $\Omega' \subset \Omega$ und $\mathfrak{A}$ eine σ-Algebra über Ω. Dann ist $\mathfrak{A}' = \mathfrak{A}\cap\Omega' := \{A\cap\Omega' : A \in \mathfrak{A}\}$ eine σ-Algebra über Ω'. Sie heißt die Spur von $\mathfrak{A}$ in Ω' (Übung).
- Seien $\mathfrak{A}_i$ σ-Algebren über Ω für $i \in I$ mit beliebiger Indexmenge $I \neq \emptyset$. Dann ist $\bigcap_{i\in I} \mathfrak{A}_i$ σ-Algebra über Ω.

  Beweis.

  $\Omega \in \mathfrak{A}_i$ für alle $i \in I$, also $\Omega \in \bigcap_{i\in I} \mathfrak{A}_i$. Für $A \in \bigcap_{i\in I} \mathfrak{A}_i$ gilt $A \in \mathfrak{A}_i$ für alle $i \in I$. Damit ist auch $A^c \in \mathfrak{A}_i$ für alle $i \in I$, also $A^c \in \bigcap_{i\in I} \mathfrak{A}_i$. Sei $(A_j)_{j\geq 1}$ eine Folge in $\bigcap_{i\in I} \mathfrak{A}_i$. Dann ist $(A_j)_{j\geq 1}$eine Folge in $\mathfrak{A}_i$ für alle $i \in I$. Somit ist $\bigcup_{j\geq 1} A_j \in \mathfrak{A}_i$ für alle $i \in I$, also $\bigcup_{j\geq 1} A_j \in \bigcap_{i\in I} \mathfrak{A}_i$. ∎

Satz 8.4. Sei $\Omega \neq \emptyset$ und $\mathcal{F}$ eine beliebige Familie von Teilmengen von Ω. Dann existiert genau eine kleinste σ-Algebra $\mathfrak{A}(\mathcal{F})$, die $\mathcal{F}$ enthält (d.h. $\mathcal{F} \subset \mathfrak{A}(\mathcal{F})$). Dabei heißt $\mathfrak{A}(\mathcal{F})$ die *von $\mathcal{F}$ erzeugte σ-Algebra* und $\mathcal{F}$ *Erzeuger* von $\mathfrak{A}(\mathcal{F})$.

Beweis.
Es existiert mindestens eine σ-Algebra $\mathfrak{A}$ mit $\mathcal{F} \subset \mathfrak{A}$ (etwa $\mathfrak{A} = \mathcal{P}(\Omega)$). Sei $\{\mathfrak{A}_i : i \in I\}$ die Familie aller σ-Algebren $\mathfrak{A}_i$ mit $\mathcal{F} \subset \mathfrak{A}_i$. Wir setzen $\mathfrak{A}(\mathcal{F}) := \bigcap_{i\in I} \mathfrak{A}_i$. Offenbar gilt $\mathcal{F} \subset \mathfrak{A}(\mathcal{F})$ und $\mathfrak{A}(\mathcal{F})$ ist nach dem vorigen Beispiel eine σ-Algebra. Jede $\mathcal{F}$ enthaltende σ-Algebra ist beim Durchschnitt $\bigcap_{i\in I} \mathfrak{A}_i$ zugelassen. Damit ist $\mathfrak{A}(\mathcal{F})$ kleinstmöglich. ∎

Beispiel 8.5. Sei $\Omega = \mathbb{R}^d$. Für $a, b \in \mathbb{R}^d$ mit $a = (a_1, \ldots, a_d)$ und $b = (b_1, \ldots, b_d)$ bezeichne

$$[a, b) := [a_1, b_1) \times \cdots \times [a_d, b_d) = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d : a_i \leq x_i < b_i \text{ für } i = 1, \ldots, d\right\}.$$

Derartige Teilmengen des $\mathbb{R}^d$ heißen halboffen, $\mathcal{F} = \{[a, b) \subset \mathbb{R}^d : a, b \in \mathbb{R}^d\}$ ist die Menge der halboffenen Intervalle des $\mathbb{R}^d$.

Definition 8.6. Die σ-Algebra $\mathfrak{B}^d := \mathfrak{A}(\mathcal{F})$ über $\mathbb{R}^d$ mit $\mathcal{F} = \{[a, b) \subset \mathbb{R}^d : a, b \in \mathbb{R}^d\}$ heißt *Borelsche σ-Algebra* im $\mathbb{R}^d$. Die Mengen von $\mathfrak{B}^d$ heißen *Borelsche Mengen* oder Borelmengen. Es bezeichne $\mathfrak{B} := \mathfrak{B}^1$.

Bemerkung 9. Alle offenen und alle abgeschlossenen Mengen des $\mathbb{R}^d$ sind Borelmengen.

Definition 8.7. Ein *messbarer Raum* ist ein Paar $(\Omega, \mathfrak{A})$ bestehend aus einer nichtleeren Menge Ω und einer σ-Algebra über Ω. Ein *Wahrscheinlichkeitsmaß* $\mathbb{P}$ auf $\mathfrak{A}$ ist eine Abbildung $\mathbb{P} : \mathfrak{A} \to [0, 1]$ mit

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

(ii) $\mathbb{P}$ ist σ-additiv, d.h. für jede Folge $(A_i)_{i\geq 1}$ paarweise disjunkter Mengen in $\mathfrak{A}$ gilt

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

Das Tripel $(\Omega, \mathfrak{A}, \mathbb{P})$ heißt (allgemeiner) W-Raum, $\mathbb{P}$ auch W-Verteilung, die Elemente von $\mathfrak{A}$ heißen Ereignisse.

Satz 8.8. Die Aussagen und Begriffe aus Lemma 1.2, Lemma 1.3, Definition 3.2 sowie Lemma/Satz 3.4–3.11 für diskrete W-Räume gelten auch für allgemeine W-Räume.

Definition 8.9. Sei $\mathbb{P}$ eine W-Verteilung auf $(\mathbb{R}, \mathfrak{B})$. Dann heißt

$$F : \mathbb{R} \to [0, 1], \quad x \mapsto \mathbb{P}\left((-\infty, x)\right)$$

Verteilungsfunktion von $\mathbb{P}$.

Lemma 8.10. Sei F die Verteilungsfunktion einer W-Verteilung $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B})$. Dann gelten:

a) F ist monoton wachsend.

b) F ist linksseitig stetig.

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

d) $\mathbb{P}$ ist durch F eindeutig festgelegt.

Beweis.

a) Für $x \leq y$ gilt wegen der Monotonie des W-Maßes $F(x) = \mathbb{P}((-\infty, x)) \leq \mathbb{P}((-\infty, y)) = F(y)$.

b) Sei $(z_n)_{n \geq 1}$ eine Folge in $\mathbb{R}$ mit $z_n \uparrow z \in \mathbb{R}$. Dann definiert $A_n := (-\infty, z_n)$ eine aufsteigende Folge von Ereignissen in $\mathfrak{B}$ mit $\bigcup_{n \geq 1} A_n = (-\infty, z)$. Die Stetigkeit von unten (vgl. Lemma 1.3) liefert $\lim_{n \to \infty} F(z_n) = \lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A) = \mathbb{P}((-\infty, z)) = F(z)$, also ist F linksseitig stetig.

c) Mit $A_n := (-\infty, -n)$ ist eine absteigende Folge von Ereignissen mit $\bigcap_{n \geq 1} A_n = \emptyset$ definiert. Die Stetigkeit von oben liefert $\lim_{n \to \infty} F(-n) = \mathbb{P}(\emptyset) = 0$. Wegen der Monotonie von F gilt damit $\lim_{x \to -\infty} F(x) = 0$. Analog folgt mit $A_n := (-\infty, n)$, dass $\lim_{n \to \infty} F(n) = \mathbb{P}(\mathbb{R}) = 1$.

d) Jede Verteilung auf $(\mathbb{R}^d, \mathfrak{B}^d)$ wird durch ihre Werte auf $\mathcal{F} = \left\{[a, b) : a, b \in \mathbb{R}^d\right\}$ bereits vollständig festgelegt. (Dies ist ein maßtheoretischer Satz, der „Eindeutigkeitssatz", der hier ohne Beweis verwendet wird.) Speziell für $d = 1$ folgt:

$$\mathbb{P}([a, b)) = \mathbb{P}((-\infty, b) \setminus (-\infty, a)) = \mathbb{P}((-\infty, b)) - \mathbb{P}((-\infty, a)) = F(b) - F(a)$$

für alle $a \leq b$. Also legt F das W-Maß $\mathbb{P}$ fest.

∎

Bemerkung 10. Jede monoton wachsende, linksseitig stetige Funktion $G : \mathbb{R} \to [0, 1]$ mit $\lim_{x \to -\infty} G(x) = 0$ und $\lim_{x \to \infty} G(x) = 1$ heißt Verteilungsfunktion und definiert vermöge $\mathbb{P}([a, b)) = F(b) - F(a)$ eine eindeutige W-Verteilung auf $(\mathbb{R}, \mathfrak{B})$. (Die Existenz von $\mathbb{P}$ ist nichttrivial und wird hier nicht bewiesen.)

Ein wichtiger Spezialfall sind Verteilungen mit Dichten: Sei $f : \mathbb{R} \to \mathbb{R}_0^+$ eine nichtnegative Funktion mit

$$\int_{-\infty}^{\infty} f(x)\, dx = 1. \tag{8}$$

Dabei soll das Integral definiert sein, etwa $f|_{[a,b]}$ Regelfunktion sein für alle $a, b \in \mathbb{R}$ mit $a < b$. Für derartige Funktionen f definiert

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

eine Verteilungsfunktion. Für die zugehörige W-Verteilung $\mathbb{P}$ gilt

$$\mathbb{P}([a, b)) = \int_a^b f(x)\,dx. \tag{9}$$

Definition 8.11. Sei $\mathbb{P}$ eine W-Verteilung auf $(\mathbb{R}, \mathfrak{B})$ und $f : \mathbb{R} \to \mathbb{R}_0^+$, so dass (8) und (9) für alle $a, b \in \mathbb{R}$ mit $a < b$ gelten. Dann heißt f *Dichte* oder *W-Dichte* von $\mathbb{P}$.

Beispiel 8.12. 1) Gleichverteilung auf dem Intervall $[c, d]$: Für $c, d \in \mathbb{R}$ mit $c < d$ ist

$$f(x) = \frac{1}{d - c}\mathbb{1}_{[c,d]}(x)$$

eine Funktion mit (8). Die zugehörige Verteilung heißt *Gleichverteilung* auf $[c, d]$. Es ist f also die Dichte der Gleichverteilung auf $[c, d]$.

2) Für $\lambda > 0$ ist durch

$$f_\lambda(x) = \mathbb{1}_{[0,\infty)}(x)\lambda e^{-\lambda x}, \quad x \in \mathbb{R},$$

eine Funktion mit (8) definiert. Die zugehörige Verteilung heißt *Exponentialverteilung zum Parameter* $\lambda > 0$. Es wird später klar werden, dass die Exponentialverteilung ein stetiges Analogon zur geometrischen Verteilung ist. Sie wird etwa verwendet, um die Wartezeit auf den ersten radioaktiven Zerfall zu modellieren, oder allgemeiner für das erste Auftreten unter zufälligen Phänomenen mit „konstanter Rate pro Zeiteinheit“.

3) Für $\mu \in \mathbb{R}$ und $\sigma > 0$ ist durch

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

eine Funktion mit (8) gegeben. Die zugehörige Verteilung heißt *Normalverteilung mit Parameter* μ *und* σ^2. Diese bezeichnen wir auch mit $\mathcal{N}(\mu, \sigma^2)$. Speziell heißt $\mathcal{N}(0, 1)$ *Standardnormalverteilung.* Sie wird zur Modellierung von Messfehlern verwendet und zur Approximation von Verteilungen. Dies wird später durch den „Zentralen Grenzwertsatz“ begründet.

Verallgemeinerung auf $\mathbb{R}^d$: Sei $f : \mathbb{R}^d \to \mathbb{R}_0^+$ eine nichtnegative Funktion mit

$$\int_{\mathbb{R}^d} f(x)\,dx = 1, \tag{10}$$

und für $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ gelte

$$\mathbb{P}([a, b)) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(x_1, \dots, x_d)\,dx_d \cdots dx_1. \tag{11}$$

Definition 8.13. Sei $\mathbb{P}$ eine Verteilung auf $(\mathbb{R}^d, \mathfrak{B}^d)$ und $f : \mathbb{R}^d \to \mathbb{R}_0^+$, so dass (10) und (11) für alle $a, b \in \mathbb{R}^d$ gelten. Dann heißt f Dichte von $\mathbb{P}$.

9 Messbare Abbildungen und ZVe

Im Fall diskreter W-Räume $(\Omega, \mathfrak{A}, \mathbb{P})$ ist jede Abbildung $X : \Omega \to \Omega'$ eine ZVe. Wir wollen nun auch im allgemeinen Fall $(\Omega, \mathfrak{A}, \mathbb{P})$, wobei $\mathfrak{A}$ eine σ-Algebra, nicht notwendigerweise die Potenzmenge von Ω ist, (reelle) Zufallsvariable X einführen und insbesondere wieder Ereignissen der Form $\{X \leq 7\}$ oder $\{X \text{ gerade}\}$ Wahrscheinlichkeiten zuordnen. Dabei tritt das Problem auf, dass für eine allgemeine σ-Algebra $\mathfrak{A}$ über Ω nicht jedem $A \subset \Omega$ eine W-keit zugeordnet wird.

Definition 9.1. Seien $(\Omega, \mathfrak{A})$ und $(\Omega', \mathfrak{A}')$ messbare Räume. Eine Abbildung $f : \Omega \to \Omega'$ heißt *messbar*, falls für alle $B \in \mathfrak{A}'$ gilt:

$$f^{-1}(B) \in \mathfrak{A}.$$

Ist zudem $\mathbb{P}$ ein W-Maß auf $(\Omega, \mathfrak{A})$ (also $(\Omega, \mathfrak{A}, \mathbb{P})$ ein W-Raum), so heißt jede messbare Abbildung $X : \Omega \to \Omega'$ *Zufallsvariable*. Dann definiert $\mathbb{P}_X : \mathfrak{A}' \to [0, 1]$, $B \mapsto \mathbb{P}_X(B) := \mathbb{P}\left(X^{-1}(B)\right)$ eine W-Verteilung auf $(\Omega', \mathfrak{A}')$. Sie heißt W-Verteilung der ZVe X.

Lemma 9.2. Seien $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$ messbare Räume und $\mathcal{E} \subset \mathfrak{A}'$ ein Erzeuger von $\mathfrak{A}'$ (d.h. $\mathfrak{A}(\mathcal{E}) = \mathfrak{A}'$). Dann gilt:

$$f : \Omega \to \Omega' \text{ messbar } \Leftrightarrow f^{-1}(E) \in \mathfrak{A} \text{ für alle } E \in \mathcal{E}.$$

Beweis.
„$\Rightarrow$“: klar, da $\mathcal{E} \subset \mathfrak{A}' = \mathfrak{A}(\mathcal{E})$.
„$\Leftarrow$“: Seien $\mathcal{C} := \mathfrak{A}\left(\{f^{-1}(E) : E \in \mathcal{E}\}\right)$, $\mathcal{D} = \{B \in \mathfrak{A}' : f^{-1}(B) \in \mathcal{C}\}$. Nach Voraussetzung gilt $\mathcal{C} \subset \mathfrak{A}$. Wir zeigen unten, dass $\mathcal{D}$ eine σ-Algebra ist. Wegen $\mathcal{E} \subset \mathcal{D}$ folgt $\mathfrak{A}' = \mathfrak{A}(\mathcal{E}) \subset \mathcal{D}$. Für $B \in \mathfrak{A}'$ beliebig gilt damit $B \in \mathcal{D}$, also $f^{-1}(B) \in \mathcal{C}$ und wegen $\mathcal{C} \subset \mathfrak{A}$ dann $f^{-1}(B) \in \mathfrak{A}$. Es bleibt damit zu zeigen, dass $\mathcal{D}$ eine σ-Algebra über Ω' ist:

(i) $f^{-1}(\Omega') = \Omega \in \mathcal{C}$, also $\Omega' \in \mathcal{D}$.

(ii) Sei $B \in \mathcal{D}$. Dann gilt $f^{-1}(B^c) = \left(f^{-1}(B)\right)^c \in \mathcal{C}$, also $B^c \in \mathcal{D}$.

(iii) Sei $(B_i)_{i \geq 1}$ eine Folge in $\mathcal{D}$. Dann ist

$$f^{-1}\left(\bigcup_{i \geq 1} B_i\right) = \bigcup_{i \geq 1} f^{-1}(B_i) \in \mathcal{C},$$

also $\bigcup_{i \geq 1} B_i \in \mathcal{D}$.

∎

Lemma 9.3. Seien $(\Omega, \mathfrak{A})$ ein messbarer Raum und $f_i : \Omega \to \mathbb{R}$ messbare Funktionen (bez. $\mathfrak{B}$) für $i = 1, \ldots, d$. Dann ist $f : \Omega \to \mathbb{R}^d$, $f(x) = (f_1(x), \ldots, f_d(x))$ messbar (bez. $\mathfrak{B}^d$).

Beweis.
Für $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ gilt $f^{-1}([a,b)) = \bigcap_{i=1}^d f_i^{-1}([a_i, b_i)) \in \mathfrak{A}$, da $f_1, \dots, f_d$ messbar sind. Da die halboffenen Intervalle einen Erzeuger von $\mathfrak{B}^d$ bilden, folgt die Behauptung aus Lemma 9.2. ∎

Lemma 9.4. Seien $(\Omega, \mathfrak{A}), (\Omega', \mathfrak{A}'), (\Omega'', \mathfrak{A}'')$ messbare Räume, $f : \Omega \to \Omega', g : \Omega' \to \Omega''$ messbare Abbildungen. Dann ist $g \circ f$ messbar.

Beweis.
Für $B \in \mathfrak{A}''$ ist $(g \circ f)^{-1}(B) = f^{-1}(\underbrace{g^{-1}(B)}_{\in \mathfrak{A}'}) \in \mathfrak{A}$, da f, g messbar sind. ∎

Von besonderem Interesse sind messbare Funktionen mit Wertebereich $\mathbb{R}$ oder $\mathbb{R}^d$. Wenn nicht anders angegeben, seien $\mathbb{R}$ und $\mathbb{R}^d$ stets mit der Borelschen σ-Algebra $\mathfrak{B}$ bzw. $\mathfrak{B}^d$ versehen.

Bemerkung 11. Es ist leicht zu sehen (hier ohne Beweis), dass die Borelsche σ-Algebra neben den halboffenen Intervallen auch von den offenen Mengen des $\mathbb{R}^d$ erzeugt wird.

Lemma 9.5. Sei $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ stetig. Dann ist f messbar (jeweils bez. der Borelschen σ-Algebren $\mathfrak{B}^d$ und $\mathfrak{B}^{d'}$).

Beweis.
Urbilder offener Mengen unter stetigen Abbildungen sind offen. Das System offener Mengen des $\mathbb{R}^{d'}$ bildet einen Erzeuger von $\mathfrak{B}^{d'}$. Aus Lemma 9.2 folgt die Behauptung. ∎

Lemma 9.6. Seien $(\Omega, \mathfrak{A})$ ein messbarer Raum und $f_i : \Omega \to \mathbb{R}$ messbare Funktionen für $i \in \mathbb{N}$ sowie $(\alpha_i)_{i \geq 1}$ eine Folge in $\mathbb{R}$. Dann sind

$$\alpha_1 f_1 + \cdots + \alpha_n f_n, \quad f_1 \cdots f_n, \quad \sup_{i \in \mathbb{N}} f_i, \quad \inf_{i \in \mathbb{N}} f_i, \quad \limsup_{i \to \infty} f_i, \quad \liminf_{i \to \infty} f_i$$

messbare Funktionen mit Wertebereich $\overline{\mathbb{R}} := [-\infty, \infty]$.

Beweis.
Sei $f := (f_1, \dots, f_n) : \Omega \to \mathbb{R}^n$ und $g : \mathbb{R}^n \to \mathbb{R}$ die Abbildung $g(x_1, \dots, x_n) = \alpha_1 x_1 + \cdots + \alpha_n x_n$. Nach Lemma 9.3 ist f messbar, nach Lemma 9.5 ist g messbar, nach Lemma 9.4 ist $g \circ f = \alpha_1 f_1 + \cdots + \alpha_n f_n$ messbar. Ebenso folgt die Messbarkeit des Produkts. Das System $\{(-\infty, x) : x \in \mathbb{R}\}$ bildet einen Erzeuger von $\mathfrak{B}$, denn $[a, b) = (-\infty, b) \setminus (-\infty, a)$. Wir haben

$$\left\{\sup_{i \geq 1} f_i < x\right\} = \left\{\sup_{i \geq 1} f_i \in (-\infty, x)\right\} = \bigcap_{i=1}^{\infty} \{f_i \in (-\infty, x)\} \in \mathfrak{A}$$

für alle $x \in \mathbb{R}$. Nach Lemma 9.2 ist $\sup f_i$ messbar. Die restlichen Behauptungen folgen ähnlich. ∎

Definition 9.7. Seien $\mathbb{P}_1, \dots, \mathbb{P}_d$ W-Verteilungen auf $(\mathbb{R}, \mathfrak{B})$. Eine Verteilung $\mathbb{P}$ auf $(\mathbb{R}^d, \mathfrak{B}^d)$ heißt das *Produkt der Verteilungen* $\mathbb{P}_1, \dots, \mathbb{P}_d$, falls für alle $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ gilt:

$$\mathbb{P}([a, b)) = \prod_{i=1}^{d} \mathbb{P}_i([a_i, b_i)).$$

Bezeichnung: $\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_d$.

Für reellwertige ZVe $X : \Omega \to \mathbb{R}$ oder Zufallsvektoren $X : \Omega \to \mathbb{R}^d$ wird im Folgenden stets $(\mathbb{R}, \mathfrak{B})$ bzw. $(\mathbb{R}^d, \mathfrak{B}^d)$ als messbarer Bildraum zugrunde gelegt.

Definition 9.8. Seien $X_1, \dots, X_d$ reellwertige ZVe auf $(\Omega, \mathfrak{A}, \mathbb{P})$ und $(X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$. Dann heißt $\mathbb{P}_X = \mathbb{P}_{(X_1,\dots,X_d)}$ die *gemeinsame Verteilung* von $X_1, \dots, X_d$.

Definition 9.9. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein W-Raum, $(\Omega_i, \mathfrak{A}_i)$ messbare Räume und $X_i : \Omega \to \Omega_i$ ZVe für $i \in I \neq \emptyset$. Die Familie von ZVen $\{X_i : i \in I\}$ heißt unabhängig, falls für jede Wahl $B_i \in \mathfrak{A}_i$ die Familie $\{\{X_i \in B_i\} : i \in I\}$ unabhängig ist (vgl. Definition 3.8).

Für reellwertige ZVe gilt folgende Charakterisierung der Unabhängigkeit.

Satz 9.10. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein W-Raum und $X_1, \dots, X_d$ reellwertige ZVe auf Ω. Dann gilt:

$$X_1, \dots, X_d \text{ unabhängig } \Leftrightarrow \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d} = \mathbb{P}_{(X_1,\dots,X_d)}.$$

Beweis.
"$\Rightarrow$": Für die Gleichheit von Verteilungen auf $(\mathbb{R}^d, \mathfrak{B}^d)$ genügt, die Gleichheit auf $\mathcal{F} = \{[a, b) : a, b \in \mathbb{R}^d\}$ zu zeigen, vgl. den Beweis von Lemma 8.10 (d). Für $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ gilt

$$\begin{aligned}
\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}([a, b)) &= \prod_{i=1}^{d} \mathbb{P}_{X_i}([a_i, b_i)) = \prod_{i=1}^{d} \mathbb{P}(\{X_i \in [a_i, b_i)\}) \\
&\overset{(*)}{=} \mathbb{P}\left(\bigcap_{i=1}^{d} \{X_i \in [a_i, b_i)\}\right) = \mathbb{P}((X_1, \dots, X_d) \in [a, b)) = \mathbb{P}_{(X_1,\dots,X_d)}([a, b)),
\end{aligned}$$

wobei für $(*)$ die Voraussetzung der Unabhängigkeit verwendet wird.
„$\Leftarrow$": Seien $B_1, \dots, B_d \in \mathfrak{B}$ beliebig. Dann gilt

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{d} \{X_i \in B_i\}\right) &= \mathbb{P}((X_1, \dots, X_d) \in B_1 \times \cdots \times B_d) \\
&\overset{(**)}{=} \prod_{i=1}^{d} \mathbb{P}_{X_i}(B_i) = \prod_{i=1}^{d} \mathbb{P}(\{X_i \in B_i\}),
\end{aligned}$$

wobei für $(**)$ die Voraussetzung verwendet wird. Also sind $X_1, \dots, X_d$ unabhängig. ∎

Lemma 9.11. Seien $X_1, \dots, X_d$ unabhängige reellwertige ZVe, $\mathbb{P}_{X_i}$ habe Dichte f_i für $i = 1, \dots, d$ und es sei $X = (X_1, \dots, X_d)$. Dann hat $\mathbb{P}_X$ die Dichte

$$f : (x_1, \dots, x_d) \mapsto \prod_{i=1}^{d} f_i(x_i).$$

Beweis.
Für $a = (a_1, \dots, a_d), b = (b_1, \dots, b_d) \in \mathbb{R}^d$ gilt

$$\begin{aligned}\mathbb{P}_X([a,b)) &\overset{(\spadesuit)}{=} \prod_{i=1}^{d} \mathbb{P}_{X_i}([a_i, b_i)) = \prod_{i=1}^{d} \int_{a_i}^{b_i} f(x_i)\, dx_i \\ &= \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f_1(x_1) \cdots f_d(x_d)\, dx_d \cdots dx_1,\end{aligned}$$

wobei für ($\spadesuit$) Satz 9.10 verwendet wurde. Nach Definition 8.13 ergibt diese Darstellung gerade die Behauptung. ∎

Satz 9.12. Seien X_1, X_2 unabhängige, reelle ZVe auf $(\Omega, \mathfrak{A}, \mathbb{P})$ mit Dichten f_1 und f_2. Dann hat $X_1 + X_2$ die Dichte

$$f_1 * f_2(u) = \int_{-\infty}^{\infty} f_1(u - v) f_2(v)\, dv, \quad u \in \mathbb{R}.$$

$f_1 * f_2$ heißt *Faltung* von f_1 und f_2.

Beweis.
Sei $s \in \mathbb{R}$ und $B = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq s\right\}$. Dann gilt mit der Substitution $u = x_1 + x_2$:

$$\begin{aligned}\mathbb{P}(X_1 + X_2 \leq s) &= \mathbb{P}_{(X_1,X_2)}(B) = \int_B f_1(x_1) f_2(x_2)\, d(x_1, x_2) \\ &= \int_{\{(x_1,x_2)\,|\,x_1+x_2 \leq s\}} f_1(x_1) f_2(x_2)\, d(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{s-x_2} f_2(x_2) f_1(x_1)\, dx_1\, dx_2 \\ &= \int_{-\infty}^{\infty} f_2(x_2) \int_{-\infty}^{s} f_1(u - x_2)\, du\, dx_2 = \int_{-\infty}^{s} \int_{-\infty}^{\infty} f_1(u - x_2) f_2(x_2)\, dx_2\, du.\end{aligned}$$

Damit folgt für alle $s < t$

$$\mathbb{P}(X_1 + X_2 \in [s, t)) = \int_s^t \int_{-\infty}^{\infty} f_1(u - v) f_2(v)\, dv\, du.$$

Nach Definition 8.11 ist damit der innere Integrand $u \mapsto \int_{-\infty}^{\infty} f_1(u - v) f_2(v)\, dv$ Dichte von $X_1 + X_2$. ∎

10 Erwartungswerte und höhere Momente

Sei X eine diskrete ZVe mit Werten $\{x_1, x_2, \ldots\}$ und existierendem Erwartungswert. Dann gilt $\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \mathbb{P}(\{X = x_i\})$. In allgemeinen W-Räumen kann der EW durch einen Grenzübergang aus dem EW für den diskreten Fall gewonnen werden: Sei $X : \Omega \to \mathbb{R}$ eine ZVe mit (allgemeinem) W-Raum $(\Omega, \mathfrak{A}, \mathbb{P})$ und Bildraum $(\mathbb{R}, \mathfrak{B})$. Dann setze man etwa für alle $n \geq 1$

$$A_{nk} := \left\{ \frac{k}{n} \leq X < \frac{k+1}{n} \right\} \text{ für } k \in \mathbb{Z}$$

und approximieren X durch die diskrete ZVe

$$X_n = \sum_{k=-\infty}^{\infty} \frac{k}{n} \mathbb{1}_{A_{nk}}, \quad n \geq 1.$$

Es gilt dann $X_n \leq X \leq X_n + \frac{1}{n}$. Falls die EWe der X_n existieren, so gilt $|\mathbb{E}[X_n] - \mathbb{E}[X_m]| \leq \frac{1}{n} + \frac{1}{m}$, d.h. $(\mathbb{E}[X_n])_{n \geq 1}$ ist eine Cauchy-Folge und damit konvergent in $\mathbb{R}$. In diesem Fall setzt man

$$\mathbb{E}[X] := \lim_{n \to \infty} \mathbb{E}[X_n]. \tag{12}$$

Andere Schreibweisen sind $\mathbb{E}_{\mathbb{P}}[X]$ oder $\int X \, d\mathbb{P}$.

Bemerkung 12. Man beachte, dass sich die Vorgehensweise von der Definition von Riemann-Integralen oder Integralen von Regelfunktionen dahingehend prinzipiell unterscheidet, dass nicht der Urbildraum, sondern der Bildraum (äquidistant) unterteilt wird. Dies entspricht dem Vorgehen zur Definition des Lebesgue-Integrals.

Da sich Grenzwerte der Form (12) nur selten explizit berechnen lassen, ist für praktische Zwecke folgender Zusammenhang zentral, den man auch als Definition des Erwartungswerts direkt verwenden könnte.

Satz 10.1. Sei X eine reellwertige ZVe, deren Verteilung $\mathbb{P}_X$ eine Dichte f besitze, die bis auf endlich viele Stellen stetig sei. Sei $g : \mathbb{R} \to \mathbb{R}$ stetig. Dann existiert der EW von $g(X)$ genau dann, wenn $\int_{-\infty}^{\infty} |g(x)| f(x) \, dx < \infty$, und in diesem Fall gilt

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx.$$

Beweisskizze.
Zu $\delta > 0$ existiert eine Folge $(x_n)_{n \in \mathbb{Z}}$ mit $x_n \to \infty$ für $n \to \infty$, $x_n \to -\infty$ für $n \to -\infty$ und $|g(x) - g(x_n)| \leq \delta$ für alle $x_n \leq x \leq x_{n+1}$. Sei $g_\delta(x) := g(x_n)$ für alle $x \in [x_n, x_{n+1})$. Damit ist eine Treppenfunktion g_δ definiert mit $|g_\delta(x) - g(x)| \leq \delta$ für alle $x \in \mathbb{R}$. Da $g_\delta(X)$ eine diskrete ZVe ist gilt

$$\begin{aligned}
\mathbb{E}[g_\delta(X)] &= \sum_{n=-\infty}^{\infty} g(x_n) \mathbb{P}(X \in [x_n, x_{n+1})) \\
&= \sum_{n=-\infty}^{\infty} g(x_n) \int_{x_n}^{x_{n+1}} f(x) \, dx \xrightarrow{\delta \downarrow 0} \int_{-\infty}^{\infty} g(x) f(x) \, dx,
\end{aligned}$$

falls dieses Integral existiert, d.h. falls $\int_{-\infty}^{\infty} |g(x)| f(x)\, dx < \infty$. ∎

Bemerkung 13. Es handelt sich hierbei nur um eine Beweisskizze, da in der Definition (12) die Unabhängigkeit des Grenzwerts von der verwendeten approximierenden Folge gezeigt werden müsste.

Korollar 10.2. Für eine reellwertige ZVe X, deren Verteilung eine Dichte f besitzt mit $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$, existiert der EW und es gilt

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

Beispiel 10.3. Sei X gleichverteilt auf $[c, d]$, d.h. $\mathbb{P}_X$ habe Dichte $x \mapsto \frac{1}{d-c} \mathbb{1}_{[c,d]}(x)\, dx$. Dann folgt mit Korollar 10.2

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{d-c} \mathbb{1}_{[c,d]}(x)\, dx = \frac{1}{d-c} \int_{c}^{d} x\, dx = \frac{1}{d-c} \frac{d^2 - c^2}{2} = \frac{d+c}{2}.$$

Definition 10.4. Sei X eine reellwertige ZVe mit Dichte f und $\int_{-\infty}^{\infty} |x|^m f(x)\, dx < \infty$ für ein $m \in \mathbb{N}$. Dann heißt

$$\mathbb{E}[X^m] = \int_{-\infty}^{\infty} x^m f(x)\, dx \qquad m\text{-tes Moment der ZVe } X,$$

$$\mathbb{E}[|X|^m] = \int_{-\infty}^{\infty} |x|^m f(x)\, dx \qquad m\text{-tes absolutes Moment der ZVe } X.$$

Hat X ein endliches zweites Moment (d.h. existiert der EW von X^2), so heißt

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)\, dx$$

Varianz von X.

Lemma 10.5. Die Rechenregeln für den Erwartungswert aus den Sätzen 6.3 und 6.6 und für die Varianz aus den Sätzen 6.8 und 6.9 gelten auch für allgemeine reellwertige ZVe.

Satz 10.6 (Jensensche Ungleichung)**.** Seien X eine reellwertige ZVe und $f : \mathbb{R} \to \mathbb{R}$ ein konvexe Funktion, so dass die EWe von X und $f \circ X$ existieren. Dann gilt

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

Beweis.
Sei f konvex, d.h. es gilt

$$\forall x, y \in \mathbb{R}, \lambda \in [0,1] : f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y).$$

An jeder Stelle $x_0 \in \mathbb{R}$ existiert eine Stützgerade $x \mapsto ax + b$, d.h. $a, b \in \mathbb{R}$ mit $f(x_0) = ax_0 + b$ und $f(x) \geq ax + b$ für alle $x \in \mathbb{R}$. Wir wählen $x_0 = \mathbb{E}[X]$. Dann folgt

$$f(E[X]) = a\mathbb{E}[X] + b = \mathbb{E}[aX + b] \leq \mathbb{E}[f(X)].$$

∎

Beispiel 10.7. • $x \mapsto |x|$ ist konvex, also gilt $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$, falls $\mathbb{E}[|X|]$ existiert.

• $x \mapsto |x|^p$ ist für $p \geq 1$ konvex, also gilt $|\mathbb{E}[X]|^p \leq \mathbb{E}[|X|^p]$.

Lemma 10.8 (Eigenschaften von Momenten)**.** Seien X, Y reellwertige ZVe.

a) $|X|^r$ habe EW für ein $r > 0$. Dann hat $|X|^s$ EW für alle $0 < s \leq r$.

b) Es gilt

$$\mathbb{E}[|X+Y|^r] \leq \begin{cases} 2^{r-1}\left(\mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r]\right) & \text{für } r \geq 1, \\ \mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r] & \text{für } 0 < r \leq 1. \end{cases}$$

Insbesondere ist

$$\mathcal{L}_r(\Omega, \mathfrak{A}, \mathbb{P}) := \{X : \Omega \to \mathbb{R} \text{ ZV } : |X|^r \text{ hat EW}\}$$

für $r > 0$ ein Vektorraum.

c) Es gilt $\left(\mathbb{E}[|X|^s]\right)^{1/s} \leq \left(\mathbb{E}[|X|^r]\right)^{1/r}$ für alle $0 < s \leq r$, falls $|X|^r$ einen EW hat.

Beweis.

a) Es gilt $|X|^s \leq 1 + |X|^r \Rightarrow \mathbb{E}[|X|^s] \leq 1 + \mathbb{E}[|X|^r] < \infty$.

b) Falls $r \geq 1$, so ist $x \mapsto |x|^r$ konvex. Damit ist $\left|\frac{x+y}{2}\right|^r \leq \frac{1}{2}\left(|x|^r + |y|^r\right)$ für alle $x, y \in \mathbb{R}$. Also $\mathbb{E}[|X+Y|^r] \leq 2^{r-1}\left(\mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r]\right)$. Falls $0 < r \leq 1$, so gilt $(a+b)^r \leq a^r + b^r$ für alle $a, b \geq 0$. Dies liefert die Behauptung im Fall $0 < r \leq 1$.

c) $x \mapsto |x|^{r/s}$ ist konvex. Die Jensensche Ungleichung angewandt auf $|X|^s$ liefert

$$\mathbb{E}\left[\left(|X|^s\right)^{r/s}\right] \geq \left(\mathbb{E}\left[|X|^s\right]\right)^{r/s}, \text{ also } \left(\mathbb{E}[|X|^r]\right)^{1/r} \geq \left(\mathbb{E}[|X|^s]\right)^{1/s}.$$

∎

Satz 10.9 (Cauchy-Schwarzsche Ungleichung)**.** Für beliebige $X, Y \in \mathcal{L}_2(\Omega, \mathfrak{A}, \mathbb{P})$ gilt

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right].$$

Beweis.
Für $\lambda \in \mathbb{R}$ ist $0 \leq \mathbb{E}[(\lambda X - Y)^2] = \lambda^2\mathbb{E}[X^2] - 2\lambda\mathbb{E}[XY] + \mathbb{E}[Y^2]$. Speziell für $\lambda = \mathbb{E}[XY]/\mathbb{E}[X^2]$ ergibt sich

$$0 \leq \frac{(\mathbb{E}[XY])^2}{\mathbb{E}[X^2]} - 2\frac{(\mathbb{E}[XY])^2}{\mathbb{E}[X^2]} + \mathbb{E}[Y^2] \Rightarrow (\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

∎

Bemerkung 14. a) Für $0 < s \leq r$ gilt $\mathcal{L}_r(\Omega, \mathfrak{A}, \mathbb{P}) \subset \mathcal{L}_s(\Omega, \mathfrak{A}, \mathbb{P})$.

b) Für $r \geq 1$ und $X \in \mathcal{L}_r(\Omega, \mathfrak{A}, \mathbb{P})$ definiert $\|X\|_r := \left(\mathbb{E}[|X|^r]\right)^{1/r}$ eine Semi-Norm auf $\mathcal{L}_r(\Omega, \mathfrak{A}, \mathbb{P})$. Es gilt $\|X\|_s \leq \|X\|_r$ für $1 \leq s \leq r$.

c) Für $X, Y \in \mathcal{L}_2(\Omega, \mathfrak{A}, \mathbb{P})$ und $\langle X, Y\rangle := \mathbb{E}[XY]$ liest sich die Cauchy-Schwarzsche Ungleichung $\langle X, Y\rangle \leq \|X\|_2\|Y\|_2$.

3 Summen unabhängiger Zufallsvariablen

In einem fairen Spiel zwischen zwei Spielern werde vielfach unabhängig eine Münze geworfen. Bei Kopf erhält jeweils Spieler A einen vorgegebenen Einsatz $E > 0$ von Spieler B, bei Zahl erhält Spieler B den Einsatz E von Spieler A. Was kann man über den Gewinn von Spieler A asymptotisch sagen, wenn das Spiel sehr lange dauert?
Wir modellieren dies wie folgt: Es sei $(X_i)_{i\geq 1}$ eine Folge unabhängiger ZVen mit $\mathbb{P}(X_i = 1) = \frac{1}{2} = \mathbb{P}(X_i = -1)$. Das Ereignis $\{X_i = 1\}$ bedeute, dass Spieler A im i-ten Spiel gewinnt. Dann ist der Gewinn von Spieler A gegeben durch

$$S_n = E \cdot \sum_{i=1}^{n} X_i,$$

falls n Spiele gespielt werden. Es ist S_n also eine Summe unabhängiger ZVe. Summen unabhängiger Zufallsvariablen treten bei der stochastischen Modellierung in zahlreichen Situationen auf. Deshalb untersuchen wir das asymptotische Verhalten solcher Summen. Es zeigt sich dabei, dass der Zufall nichts völlig Willkürliches ist, sondern Gesetzen folgt, die wir im Rahmen mathematischer Modellierungen beweisen können.

11 Die Gesetze großer Zahlen

In diesem Abschnitt untersuchen wir das Verhalten von S_n/n, wobei S_n eine Summe von n unabhängigen Zufallsvariablen ist.

Satz 11.1 (Markovsche Ungleichung)**.** Sei $\varphi : [0,\infty) \to [0,\infty)$ monoton wachsend und $\varepsilon > 0$ mit $\varphi(\varepsilon) > 0$. Dann gilt für jede reellwertige ZVe Z

$$\mathbb{P}\left(|Z| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left[\varphi(|Z|)\right]}{\varphi(\varepsilon)}.$$

Beweis.
Für $\omega \in \Omega$ sei

$$Y(\omega) := \begin{cases} \varphi(\varepsilon), & \text{falls } |Z(\omega)| \geq \varepsilon, \\ 0, & \text{falls } |Z(\omega)| < \varepsilon. \end{cases}$$

Dann gilt $Y \leq \varphi(|Z|)$ punktweise. Die Monotonie des EW liefert $\mathbb{E}\left[\varphi(|Z|)\right] \geq \mathbb{E}\left[Y\right] = \varphi(\varepsilon)\mathbb{P}(|Z| \geq \varepsilon)$. Dies ist die Behauptung. ∎

Bemerkung 15. Falls der EW von $\varphi(|Z|)$ nicht existiert, wird $\mathbb{E}\left[\varphi(|Z|)\right] = \infty$ gesetzt, die Behauptung gilt dann trivialerweise.

Korollar 11.2 (Chebyshevsche Ungleichung)**.** Sei X eine reellwertige ZVe mit $\mathrm{Var}(X) < \infty$. Dann gilt für alle $\varepsilon > 0$

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

Beweis.
Sei $Z := X - \mathbb{E}[X]$ und $\varphi(x) := x^2$. Dann liefert die Markovsche Ungleichung

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) = \mathbb{P}(|Z| \geq \varepsilon) \leq \frac{\mathbb{E}[\varphi(|Z|)]}{\varphi(\varepsilon)} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Dies ist die Behauptung. ∎

Satz 11.3 (Schwaches Gesetz großer Zahlen)**.** Seien $(X_i)_{i\geq 1}$ eine Folge unabhängiger ZVe mit $\mathbb{E}[X_i] = \mu$ und $\mathrm{Var}(X_i) \leq M$ für alle $i \in \mathbb{N}$ mit einer Schranke $M < \infty$. Bezeichne $S_n := \sum_{i=1}^n X_i$. Dann gilt für alle $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mu\right| \geq \varepsilon\right) \leq \frac{M}{\varepsilon^2 n} \to 0, \quad (n \to \infty).$$

Beweis.
Bezeichne $X := \frac{1}{n}S_n$. Dann gilt $\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$ und

$$\mathrm{Var}(X) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{M}{n},$$

wobei Satz 6.8 und der Satz von Bienaymé 6.9 verwendet werden. Die Chebyshevsche Ungleichung liefert

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mu\right| \geq \varepsilon\right) = \mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} \leq \frac{M}{\varepsilon^2 n} \to 0.$$

Dies ist die Behauptung. ∎

Bemerkung 16. Sei $(X_i)_{i\geq 1}$ eine Folge von Bernoulli-Experimenten, die unabhängig mit Erfolgswahrschinlichkeit $p \in [0,1]$ ausgeführt werden. Bezeichne $H_n(\omega) := \frac{1}{n}\sum_{i=1}^n X_i(\omega)$ für $\omega \in \Omega$ die relative Anzahl von Erfolgen, siehe Abbildung 1. Das schwache Gesetz großer Zahlen liefert

$$\mathbb{P}(|H_n - p| \geq \varepsilon) \leq \frac{1}{4\varepsilon^2 n}.$$

Für große n ist die W-keit, dass sich die relative Häufigkeit von der Erfolgswahrscheinlichkeit um mehr als ε unterscheidet, also klein.

Definition 11.4. Sei $(X_n)_{n\geq 1}$ eine Folge von ZVen und X eine ZVe auf $(\Omega, \mathfrak{A}, \mathbb{P})$. Die Folge $(X_n)_{n\geq 1}$ *konvergiert stochastisch* gegen X, falls gilt

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Stochastische Konvergenz bezeichnen wir mit $X_n \xrightarrow{\mathbb{P}} X$.

Bemerkung 17. Die Aussage des schwachen Gesetzes großer Zahlen ist also $\frac{1}{n}S_n \xrightarrow{\mathbb{P}} \mu$.

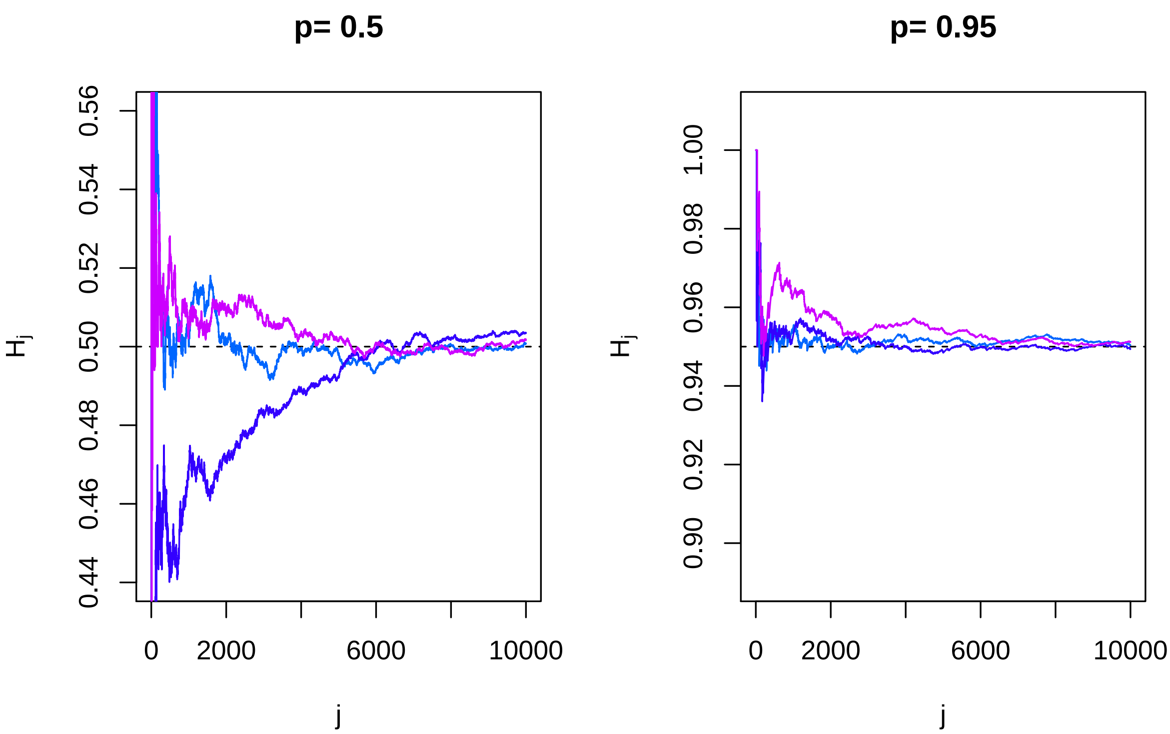


Abbildung 1: Gezeigt sind jeweils Simulationen $(H_j(\omega_i))_{j=1,\ldots,10000}$ für $i = 1, 2, 3$, vgl. Bemerkung 16. Im linken Bild ist $p = \frac{1}{2}$, rechts $p = 0.95$. Zur Simulation wurde folgender R-Code verwendet:

```
n <- 10000
p <- 0.5 # Hier entsprechend 0.95 ersetzen.
m <- 3
plot(0, 0, type="n", xlab="j", ylab=expression(paste(H[j])),
xlim=c(0,n), ylim=c(0,1))
abline(h=p, lty=2)
for (i in 1:m){
    X <- sample(c(1,0), prob=c(p,1-p), replace=TRUE, n)
    S <- cumsum(X)/(1:n)
    lines(S, col=rainbow(m, start=0.6, end=0.8)[i])
              }
```

Definition 11.5. Seien $(X_n)_{n\geq 1}$ eine Folge von ZVen und X eine ZVe auf $(\Omega, \mathfrak{A}, \mathbb{P})$. Die Folge $(X_n)_{n\geq 1}$ *konvergiert fast sicher* gegen X, falls

$$\mathbb{P}\left(\left\{\lim_{n\to\infty} X_n = X\right\}\right) = \mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Fast sichere Konvergenz bezeichnen wir mit $X_n \to X$ f.s.

Satz 11.6. Seien $(X_n)_{n\geq 1}$ eine Folge von ZVen und X eine ZVe auf $(\Omega, \mathfrak{A}, \mathbb{P})$. Dann gilt mit $n \to \infty$:

$$X_n \to X \text{ f.s. } \Longrightarrow X_n \xrightarrow{\mathbb{P}} X.$$

Beweis.
Sei $\varepsilon > 0$ beliebig und bezeichne

$$B_N := \{|X_n - X| < \varepsilon \text{ für alle } n \geq N\} = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon \text{ für alle } n \geq N\}.$$

$(B_N)_{N\geq 1}$ bildet eine aufsteigende Folge von Mengen mit

$$A := \{\lim_{n\to\infty} X_n = X\} \subset \bigcup_{N\geq 1} B_N.$$

Wegen $\mathbb{P}(A) = 1$ folgt, dass $\mathbb{P}(\bigcup_{N\geq 1} B_N) = 1$, mit der Stetigkeit von unten gilt also $\mathbb{P}(B_N) \to 1$ für $N \to \infty$. Damit gilt $\mathbb{P}(|X_N - X| \geq \varepsilon) \leq \mathbb{P}(B_N^c) \to 0$ für $N \to \infty$. ∎

Definition 11.7. Sei $(A_n)_{n\geq 1}$ eine Folge von Ereignissen. Dann heißt

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad \textit{Limessuperior} \text{ von } (A_n)_{n\geq 1},$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \quad \textit{Limesinferior} \text{ von } (A_n)_{n\geq 1}.$$

Bemerkung 18. Es gilt

$$\limsup_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\},$$
$$\liminf_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ für fast alle } n \in \mathbb{N}\}.$$

Satz 11.8 (Lemma von Borel-Cantelli)**.** Sei $(A_k)_{k\geq 1}$ eine Folge von Ereignissen.

a) Falls $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$, so gilt

$$\mathbb{P}\left(\limsup_{k\to\infty} A_k\right) = 0.$$

b) Sind $(A_k)_{k\geq 1}$ eine unabhängige Familie und $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$, so gilt

$$\mathbb{P}\left(\limsup_{k\to\infty} A_k\right) = 1.$$

Beweis.
Ad a): Es ist $\limsup_{k\to\infty} A_k = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k$, also für $n \geq 1$

$$\mathbb{P}\left(\limsup_{k\to\infty} A_k\right) \leq \mathbb{P}\left(\bigcup_{k=n}^\infty A_k\right) \leq \sum_{k=n}^\infty \mathbb{P}(A_k) \overset{n\to\infty}{\longrightarrow} 0,$$

da die Reihe $\sum_{k=1}^\infty \mathbb{P}(A_k)$ konvergiert.

Ad b): Die Unabhängigkeit der Ereignisse liefert für festes $n \in \mathbb{N}$ und $N > n$:

$$\mathbb{P}\left(\bigcap_{k=n}^N A_k^c\right) = \prod_{k=n}^N (1 - \mathbb{P}(A_k)) \leq \prod_{k=n}^N \exp\left(-\mathbb{P}(A_k)\right) = \exp\left(-\sum_{k=n}^N \mathbb{P}(A_k)\right) \overset{N\to\infty}{\longrightarrow} 0,$$

da die Reihe $\sum_{k=1}^\infty \mathbb{P}(A_k)$ divergiert. Hierbei wurde die für alle $x \in \mathbb{R}$ gültige Ungleichung $1 + x \leq e^x$ verwendet. Die Stetigkeit von oben liefert also $\mathbb{P}(\bigcap_{k\geq n} A_k^c) = 0$. Die Sub-$\sigma$-Additivität liefert nun

$$\mathbb{P}\left(\left(\limsup_{k\to\infty} A_k\right)^c\right) = \mathbb{P}\left(\bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k^c\right) \leq \sum_{n=1}^\infty \mathbb{P}\left(\bigcap_{k=n}^\infty A_k^c\right) = 0.$$

Es folgt die Behauptung. ∎

Satz 11.9 (Starkes Gesetz großer Zahlen)**.** Sei $(X_n)_{n\geq 1}$ eine Folge unabhängiger ZVe mit $\mathbb{E}\,[X_i^4] \leq M < \infty$ für alle $i \in \mathbb{N}$. Dann gilt

$$\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}\,[X_i]) \to 0 \text{ f.s. für } n \to \infty.$$

Bemerkung 19. Die Voraussetzung endlicher vierter Momente in Satz 11.9 kann zu endlichen ersten absoluten Momenten abgeschwächt werden. Dies erfordert einen aufwendigeren Beweis.

Beweis von Satz 11.9.
Wir können ohne Einschränkung annehmen, dass $\mathbb{E}\,[X_i] = 0$ für alle $i \in \mathbb{N}$. Damit ist zu zeigen, dass $\frac{1}{n}\sum_{i=1}^n X_i \to 0$ f.s., d.h.

$$\mathbb{P}\left(\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = 0\right\}\right) = 1. \tag{13}$$

Wir haben für alle $\omega \in \Omega$, dass

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) = 0 \\
&\Leftrightarrow \forall\, \varepsilon > 0\, \exists\, N \in \mathbb{N}\, \forall\, n \geq N : \left|\frac{1}{n}\sum_{i=1}^n X_i(\omega)\right| \leq \varepsilon \\
&\Leftrightarrow \forall\, m \in \mathbb{N}\, \exists\, N \in \mathbb{N}\, \forall\, n \geq N : \left|\frac{1}{n}\sum_{i=1}^n X_i(\omega)\right| \leq \frac{1}{m}.
\end{aligned}$$

Damit lässt sich das Ereignis in (13) umformulieren in

$$\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = 0\right\} = \bigcap_{m=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i(\omega)\right| \leq \frac{1}{m}\right\}.$$

Für das Komplement dieses Ereignisses erhalten wir

$$\bigcup_{m=1}^{\infty}\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i(\omega)\right| > \frac{1}{m}\right\} = \bigcup_{m=1}^{\infty}\limsup_{n\to\infty}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i(\omega)\right| > \frac{1}{m}\right\}. \tag{14}$$

Wegen der Subadditivität des W-Maßes reicht es für alle $m \in \mathbb{N}$ zu zeigen, dass

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i(\omega)\right| > \frac{1}{m}\right\}\right) = 0.$$

Wegen der Unabhängigkeit der X_i und $\mathbb{E}[X_i] = 0$ für alle $i \in \mathbb{N}$ gilt $\mathbb{E}[X_iX_jX_kX_l] = 0$, außer $i, j, k, l \in \mathbb{N}$ sind paarweise gleich. Damit folgt

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{n}X_i\right)^4\right] &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\mathbb{E}[X_iX_jX_kX_\ell] \\ &\leq 3\sum_{i,j=1}^{n}\mathbb{E}[X_i^2X_j^2] \overset{(*)}{\leq} 3\sum_{i,j=1}^{n}\mathbb{E}[X_i^4]^{1/2}\mathbb{E}[X_j^4]^{1/2} \leq 3n^2M,\end{aligned}$$

wobei für $(*)$ die Cauchy-Schwarzsche Ungleichung, Satz 10.9, verwendet wurde. Dies liefert

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| \geq \varepsilon\right) &= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|^4 \geq \varepsilon^4\right) \\ &\overset{\text{Satz 11.1}}{\leq} \frac{1}{(\varepsilon n)^4}\mathbb{E}\left[(X_1 + \cdots + X_n)^4\right] \leq \frac{3M}{\varepsilon^4 n^2}.\end{aligned}$$

Es folgt

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| > \frac{1}{m}\right) \leq \sum_{n=1}^{\infty}\frac{3m^4M}{n^2} < \infty,$$

also nach dem Lemma von Borel-Cantelli 11.8 a)

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| > \frac{1}{m}\right\}\right) = 0.$$

Es folgt die Behauptung. ∎

12 Approximation der Binomialverteilung

Wir betrachten nochmals den Kontostand $S_n = \sum_{i=1}^n X_i$ von Spieler A (mit Einsatz $E = 1$) in Abschnitt 11, wobei $X_1, \dots, X_n$ unabhängig sind und wir hier $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = -1) = 1-p =: q$ annehmen mit $p \in (0,1)$. Mit $B_i := \frac{1}{2}(X_i+1)$ sind $B_1, \dots, B_n$ die Ausgänge von n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p. Wir erhalten

$$S_n = -n + 2\sum_{i=1}^n B_i.$$

Wir wissen, dass $\sum_{i \le n} B_i$ binomial $b_{n,p}$-verteilt ist. Damit gilt für $a < b$, dass

$$\mathbb{P}\left(S_n \in [a,b]\right) = b_{n,p}\left(\left[\frac{a+n}{2}, \frac{b+n}{2}\right]\right).$$

Die Verteilung $\mathbb{P}_{S_n}$ von S_n ist also vollständig beschrieben. Das schwache Gesetz großer Zahlen (auf $B_1, \dots, B_n$ angewandt) liefert dann

$$\forall\, \varepsilon > 0 : \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n B_i - p\right| \ge \varepsilon\right) \to 0, \quad (n \to \infty)$$

oder äquivalent, dass für alle $\varepsilon > 0$ gilt

$$\mathbb{P}\left(\sum_{i=1}^n B_i \in (n(p-\varepsilon), n(p+\varepsilon))\right) \to 1, \quad (n \to \infty).$$

Die Summe $\sum_{0 \le i \le n} B_i$ liegt also „mit hoher Wahrscheinlichkeit“ (genauer: mit gegen 1 konvergierender Wahrscheinlichkeit) im Intervall $(n(p-\varepsilon), n(p+\varepsilon))$. Für die Binomialverteilung bedeutet dies für alle $\varepsilon > 0$:

$$b_{n,p}\left((n(p-\varepsilon), n(p+\varepsilon))\right) \to 1, \quad (n \to \infty).$$

In diesem Abschnitt soll nun die Binomialverteilung $b_{n,p}$ in diesem Intervall genauer untersucht werden. Wir haben für $k \in \{0, \dots, n\}$

$$b_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Zur Approximation eignet sich die *Stirlingsche Formel*:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\vartheta(n)} \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

wobei $(12n+1)^{-1} \le \vartheta(n) \le (12n)^{-1}$. Für Folgen $(a_n)_{n \ge 1}, (b_n)_{n \ge 1}$ werden folgende Bezeichnungen verwendet:

$$\begin{aligned} a_n \sim b_n &:\Longleftrightarrow \frac{a_n}{b_n} \to 1 \text{ für } n \to \infty \quad \text{(asymptotische Äquivalenz)},\\ a_n = o(b_n) &:\Longleftrightarrow \frac{a_n}{b_n} \to 0 \text{ für } n \to \infty \quad \text{(klein-o-Notation)}. \end{aligned}$$

Wir betrachten ein von n abhängendes $k = k_n$ mit $\frac{k_n}{n} \to p$ für $n \to \infty$, d.h. insbesondere gilt $k \in (n(p-\varepsilon), n(p+\varepsilon))$ für alle n hinreichend groß. Die Stirlingsche Formel liefert dann

$$\begin{aligned} b_{n,p}(\{k\}) &= \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \sim \left(\frac{n}{2\pi k(n-k)}\right)^{1/2} \frac{n^n}{k^k(n-k)^{n-k}}p^k(1-p)^{n-k} \\ &= \frac{1}{\sqrt{2\pi}}\left(\frac{n}{k(n-k)}\right)^{1/2}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}, \end{aligned}$$

wobei $q = 1-p$ verwendet wird. Aus $k \sim np$ und $n-k \sim nq$ folgt

$$\left(\frac{n}{k(n-k)}\right)^{1/2} \sim \frac{1}{\sqrt{npq}} = \frac{1}{\sigma_n},$$

wobei σ_n^2 die Varianz einer $b_{n,p}$-verteilten ZVe bezeichne (vgl. Beispiel 6.10). Wir haben also

$$b_{n,p}(\{k\}) \sim \frac{1}{\sqrt{2\pi\sigma_n^2}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} =: \frac{1}{\sqrt{2\pi\sigma_n^2}}\chi(n,k). \tag{15}$$

Um $\chi(n,k)$ asymptotisch zu beschreiben, kürzen wir $t = t_n = \frac{k_n}{n} = \frac{k}{n}$ ab. Dann gilt

$$-\ln\chi(n,k) = n\left\{t\ln\frac{t}{p} + (1-t)\ln\frac{1-t}{q}\right\} =: ng(t). \tag{16}$$

Die Funktion g spielt auch in anderen Untersuchungen der Binomialverteilung eine wesentliche Rolle, sie heißt *Ratenfunktion* der Binomialverteilung $b_{n,p}$. Wir betrachten die Taylorentwicklung von g um p: Es ist $g(p) = 0$, $g'(p) = 0$, $g''(p) = \frac{1}{pq}$, also

$$g(t) = \frac{1}{2pq}(t-p)^2 + \psi(t-p),$$

wobei der Restterm ψ die Abschätzung $|\psi(t-p)| \le C|t-p|^3$ für eine passende Konstante $C > 0$ in einer Umgebung von p erfüllt. Nehmen wir nun für $t = t_n$ die stärkere Annahme

$$n(t-p)^3 \to 0 \text{ für } n \to \infty \tag{17}$$

an, so folgt $n\psi(t-p) \to 0$ und damit in (16):

$$\left|-\ln\chi(n,k) - n\frac{(t-p)^2}{2pq}\right| \to 0.$$

Mit der Abkürzung

$$x(n,k) := \frac{k-np}{\sigma_n} \tag{18}$$

folgt also

$$\left| -\ln \chi(n,k) - \frac{x(n,k)^2}{2} \right| \to 0 \text{ für } n \to \infty. \tag{19}$$

Die Bedingung (17) bedeutet für $x(n,k)$:

$$\frac{x(n,k)^3}{\sqrt{n}} \to 0. \tag{20}$$

Bezeichnen wir mit $\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ die Dichte der Standardnormalverteilung, so liefert Einsetzen von (19) in (15):

Satz 12.1 (Lokaler Grenzwertsatz für die Binomialverteilung)**.** Sei $0 < p < 1$ und $(k_n)_{n\geq 1}$ eine Folge mit (20), wobei $x(n,k)$ gegeben ist durch (18) mit $\sigma_n = \sqrt{npq}$. Dann gilt

$$b_{n,p}(\{k\}) \sim \frac{1}{\sigma_n} \varphi\left(x(n,k)\right). \tag{21}$$

Sind $(\alpha_n)_{n\geq 1}, (\beta_n)_{n\geq 1}$ Folgen mit (20), so ist die Konvergenz (21) gleichmäßig für alle $(k_n)_{n\geq 1}$ mit $\alpha_n \leq k_n \leq \beta_n$ für alle $n \geq 1$.

Eine typische Situation, in der Satz 12.1 angewandt wird, ist

$$k_n = np + x\sqrt{npq} + O(1)$$

mit $x \in \mathbb{R}$, wobei $O(1)$ (groß-O-Notation) einen von n anhängenden Term bezeichnet (also eine Folge), der beschränkt in n ist. Genauer: Eine Folge $(r_n)_{n\geq 1}$ ist $O(1)$, falls $\sup_{n\geq 1} |r_n| < \infty$.

Bemerkung 20 (Veranschaulichung von Satz 12.1)**.** Wir betrachten das Histogramm der Binomialverteilung, d.h. wir tragen über dem Intervall $[k - \frac{1}{2}, k + \frac{1}{2}]$ ein Rechteck der Fläche $b_{n,p}(\{k\})$ ein. Für große n wird das Histogramm sehr flach (da sich die Fläche aller Rechtecke stets zu 1 summiert). Man reskaliert deshalb wie folgt: Betrachte statt k nun $x(n,k) = \frac{k-np}{\sqrt{npq}}$ und trage auf $[x(n,k) - \frac{1}{2\sqrt{npq}}, x(n,k) + \frac{1}{2\sqrt{npq}}]$ ein Rechteck der Fläche $b_{n,p}(\{k\})$ ein. Für k wie in Satz 12.1, d.h. mit (20), konvergiert die Höhe des Rechtecks gegen $\varphi(x(n,k))$ nach Satz 12.1. Das reskalierte Histogramm konvergiert also in diesem Sinne gegen die Dichte der Standardnormalverteilung. Um Satz 12.1 für ZVen umzuschreiben, bezeichne S_n eine $b_{n,p}$-verteilte ZVe, also etwa die Summe der Erfolge bei n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in (0,1)$. Es ist dann

$$\mathbb{P}\left(S_n = k\right) = \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} = \frac{k-np}{\sqrt{npq}}\right) = \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} = x(n,k)\right) = \mathbb{P}\left(S_n^* = x(n,k)\right)$$

im Sinne der folgenden Definition.

Definition 12.2. Sei X eine ZVe in $\mathcal{L}_2(\Omega, \mathfrak{A}, \mathbb{P})$. Dann heißt

$$X^* = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}}$$

die *standardisierte Form* (oder ZVe) zu X.

Bemerkung 21. Es gilt stets $\mathbb{E}[X^*] = 0$ und $\mathrm{Var}(X^*) = 1$. Für die standardisierte ZVe S_n^* zu S_n gilt dann

$$\mathbb{P}(S_n^* = x(n,k)) \sim \frac{1}{\sigma_n} \varphi(x(n,k))$$

für $(k_n)_{n\geq 1}$ mit (20).
Im nächsten Schritt sollen nicht nur die Wahrscheinlichkeiten der „lokalen“ Ereignisse $\{S_n^* = x(n,k)\}$ approximiert werden, sondern auch Wahrscheinlichkeiten für Ereignisse der Form $\{a \leq S_n^* \leq b\}$ für $a < b$. Wir bezeichnen dazu mit

$$\Phi(x) := \int_{-\infty}^{x} \varphi(u)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

die *Verteilungsfunktion der Standardnormalverteilung.*

Satz 12.3 (Satz von de Moivre-Laplace)**.** Sei $0 < p < 1$ und S_n eine $b_{n,p}$-verteilte ZVe. Dann gilt für alle $a < b$:

$$\lim_{n\to\infty} \mathbb{P}(a \leq S_n^* \leq b) = \Phi(b) - \Phi(a).$$

Beweis.
Seien $a < b$ und, wie zuvor, $\sigma_n = \sqrt{npq}$. Es seien $\alpha_n := \lceil a\sigma_n + np \rceil$, $\beta_n := \lfloor b\sigma_n + np \rfloor$. Da S_n eine ganzzahlige ZVe ist, gilt dann

$$\{a \leq S_n^* \leq b\} = \{a\sigma_n + np \leq S_n \leq b\sigma_n + np\} = \{\alpha_n \leq S_n \leq \beta_n\}.$$

Ferner gilt nach Konstruktion

$$|x(n, \alpha_n) - a| \leq \frac{1}{\sigma_n}, \quad |x(n, \beta_n) - b| \leq \frac{1}{\sigma_n}.$$

Nach Satz 12.1 existieren eine Folge $(\varepsilon_n)_{n\geq 1}$ mit $\varepsilon_n \downarrow 0$ und

$$1 - \varepsilon_n \leq \frac{b_{n,p}(\{k\})}{\varphi(x(n,k))/\sigma_n} \leq 1 + \varepsilon_n \text{ für alle } \alpha_n \leq k \leq \beta_n.$$

Mit $R_n = \sum_{k=\alpha_n}^{\beta_n} \frac{1}{\sigma_n} \varphi(x(n,k))$ gilt also

$$(1 - \varepsilon_n) R_n \leq \mathbb{P}(a \leq S_n^* \leq b) \leq (1 + \varepsilon_n) R_n. \tag{22}$$

Andererseits ist aber R_n die Riemann-Summe zu

$$\int_{x\left(n,\alpha_n - \frac{1}{2}\right)}^{x\left(n,\beta_n + \frac{1}{2}\right)} \varphi(x)\, dx = \Phi\left(x\left(n, \beta_n + \frac{1}{2}\right)\right) - \Phi\left(x\left(n, \alpha_n - \frac{1}{2}\right)\right). \tag{23}$$

Mit $n \to \infty$ folgt aus (22) und $x(n, \beta_n + \frac{1}{2}) \to b$ sowie $x(n, \alpha_n - \frac{1}{2}) \to a$, dass

$$\lim_{n\to\infty} \mathbb{P}\,(a \le S_n^* \le b) = \Phi(b) - \Phi(a).$$

Dies ist die Behauptung. ∎

Beispiel 12.4. Es werden 600 faire Würfel geworfen. Die Wahrscheinlichkeit, mindestens 90 Sechsen und höchstens 100 Sechsen zu werfen, wird gesucht. Exakt erhält man

$$b_{n,p}(\{90,\dots,100\}) = 0{,}4024\dots$$

für $n = 600$ und $p = \frac{1}{6}$. Damit gilt $np = 100$, $\sigma_n = \sqrt{npq} = 9{,}13$. Die Approximation aus dem Satz von de Moivre-Laplace liefert

$$\mathbb{P}\,(90 \le S_n \le 100) = \mathbb{P}\left(\frac{90-100}{\sigma_n} \le S_n^* \le \frac{100-100}{\sigma_n}\right) \approx \Phi(0) - \Phi\left(-\frac{10}{9{,}13}\right) \approx 0{,}36,$$

wobei S_n die $b_{600,1/6}$-Verteilung hat.
Genauer kann mit den Korrekturtermen $\pm\frac{1}{2}$ in (23) im Beweis von Satz 12.3 approximiert werden: Statt des Integrals über $[(90-100)/\sigma_n, (100-100)/\sigma_n]$ nehme man das Integral über $[(90-\frac{1}{2}100)/\sigma_n, (100+\frac{1}{2}-100)/\sigma_n]$. Wegen $\sigma_n \to \infty$ für $n \to \infty$ macht dies asymptotisch im Grenzwert keinen Unterschied, für festes n liefert dies jedoch eine bessere Approximation:

$$\mathbb{P}\,(90 \le S_n \le 100) \approx \Phi\left(\frac{0{,}5}{9{,}13}\right) - \left(-\frac{10{,}5}{9{,}13}\right) \approx 0{,}397.$$

13 Poissonapproximation

In diesem Abschnitt werden Summen unabhängiger Indikatorvariablen (die also die Ausgänge unabhängiger Bernoulli Experimente beschreiben) approximiert, deren Erfolgswahrscheinlichkeiten nicht gleich zu sein brauchen. Die hier dargestellte Approximation durch die Poissonverteilung ist nützlich, falls die Erfolge vieler unabhängiger Bernoulli-Experimente mit kleinen Erfolgswahrscheinlichkeiten gezählt werden. Insbesondere kann die Binomialverteilung $b_{n,p}$ für große n und kleine p approximiert werden. Die Poissonverteilung Π_λ zum Parameter $\lambda > 0$ war definiert durch

$$\Pi_\lambda(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!} \text{ für } k \in \mathbb{N}_0.$$

In Beispiel 7.8 zu Satz 7.7 über erzeugende Funktionen erhielten wir

$$X, Y \text{ unabhängig mit } \mathbb{P}_X = \Pi_\lambda \text{ und } \mathbb{P}_Y = \Pi_\mu \quad \Rightarrow \quad \mathbb{P}_{X+Y} = \Pi_{\lambda+\mu} \text{ für } \lambda, \mu > 0. \tag{24}$$

Zur Approximation von Verteilungen auf $\mathbb{Z}$ verwenden wir folgenden Abstandsbegriff.

Definition 13.1. Seien Q_1 und Q_2 W-Verteilungen auf $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$. Dann heißt

$$d_{TV}(Q_1, Q_2) := \sum_{k\in\mathbb{Z}} |Q_1(\{k\}) - Q_2(\{k\})|$$

der *Totalvariationsabstand* von Q_1 und Q_2.

Bemerkung 22. Wir haben folgende offensichtliche Eigenschaften:

- Es gilt stets $d_{TV}(Q_1, Q_2) \leq 2$, denn

$$\sum_{k\in\mathbb{Z}} |Q_1(\{k\}) - Q_2(\{k\})| \leq \sum_{k\in\mathbb{Z}} (Q_1(\{k\}) + Q_2(\{k\})) \leq 2.$$

- Seien Q_n, Q W-Verteilungen auf $\mathbb{Z}$ für $n \geq 1$. Falls $d_{TV}(Q_n, Q) \to 0$, so $Q_n(\{k\}) \to Q(\{k\})$ für $n \to \infty$ für alle $k \in \mathbb{Z}$. Die Konvergenz gilt gleichmäßig in $k \in \mathbb{Z}$.

Satz 13.2 (Koppelungslemma)**.** Seien Q_1, Q_2 W-Verteilungen auf $\mathbb{Z}$ und X, Y ZVe auf einem W-Raum $(\Omega, \mathfrak{A}, \mathbb{P})$ mit $\mathbb{P}_X = Q_1$ und $\mathbb{P}_Y = Q_2$. Dann gilt

$$d_{TV}(Q_1, Q_2) \leq 2\mathbb{P}(X \neq Y).$$

Beweis.
Für $k \in \mathbb{Z}$ ist

$$\begin{aligned}
&|\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \\
&= |\mathbb{P}(X = k, Y = k) + \mathbb{P}(X = k, Y \neq k) - [\mathbb{P}(Y = k, X = k) + \mathbb{P}(Y = k, X \neq k)]| \\
&\leq \mathbb{P}(X = k, Y \neq k) + \mathbb{P}(Y = k, X \neq k).
\end{aligned}$$

Andererseits ist

$$\{X \neq Y\} = \bigcup_{k\in\mathbb{Z}} (\{X \neq Y\} \cap \{X = k\}) = \bigcup_{k\in\mathbb{Z}} (\{X = k\} \cap \{Y \neq k\})$$

eine paarweise disjunkte Vereinigung. Es folgt also

$$\sum_{k\in\mathbb{Z}} \mathbb{P}(X = k, Y \neq k) = \mathbb{P}(X \neq Y).$$

Insgesamt folgt

$$\begin{aligned}
d_{TV}(Q_1, Q_2) &= \sum_{k\in\mathbb{Z}} |Q_1(\{k\}) - Q_2(\{k\})| = \sum_{k\in\mathbb{Z}} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \\
&\leq \sum_{k\in\mathbb{Z}} (\mathbb{P}(X = k, Y \neq k) + \mathbb{P}(Y = k, X \neq k)) = 2\mathbb{P}(X \neq Y).
\end{aligned}$$

Dies ist die Behauptung. ∎

Satz 13.3. Seien $X_1, \dots, X_n$ unabhängige ZVe mit $\mathbb{P}(X_i = 1) = p_i$, $\mathbb{P}(X_i = 0) = 1 - p_i$ und $p_1, \dots, p_n \in [0, 1]$. Seien $S_n = \sum_{i=1}^n X_i$ und $\lambda_n = p_1 + \cdots + p_n$. Dann gilt

$$d_{TV}\left(\mathbb{P}_{S_n}, \Pi_{\lambda_n}\right) \leq 2 \sum_{i=1}^n p_i^2.$$

Beweis.
Um Satz 13.2 anwenden zu können, müssen S_n und ZVe T_n mit $\mathbb{P}_{T_n} = \Pi_{\lambda_n}$ auf einem W-Raum konstruiert werden, sodass S_n und T_n mit möglichst großer W-keit gleiche Werte annehmen. Dazu wählen wir $\Omega_i = \{-1, 0, 1, 2, \dots\}$ und je eine W-Verteilung auf Ω_i wie folgt für alle $i = 1, \dots, n$:

$$\begin{aligned} \mathbb{P}_i(\{0\}) &= 1 - p_i, \\ \mathbb{P}_i(\{k\}) &= e^{-p_i} \frac{p_i^k}{k!} \qquad \text{für } k \geq 1, \\ \mathbb{P}_i(\{-1\}) &= e^{-p_i} - (1 - p_i). \end{aligned}$$

(Wegen $1 + x \leq e^x$ für $x \in \mathbb{R}$ ist $\mathbb{P}_i(\{-1\}) \geq 0$ und damit durch die Werte oben tatächlich eine W-Verteilung $\mathbb{P}_i$ definiert.) Ferner seien $\Omega := \Omega_1 \times \cdots \times \Omega_n$ und $\mathbb{P}$ das Produktmaß der $\mathbb{P}_1, \dots, \mathbb{P}_n$ auf Ω (vgl. Definition 4.3), d.h. für $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ gilt $\mathbb{P}(\{\omega\}) = \prod_{1 \leq i \leq n} \mathbb{P}_i(\{\omega_i\})$. Wir definieren ZVe auf Ω durch

$$X_i(\omega) = \begin{cases} 0, & \text{falls } \omega_i = 0, \\ 1, & \text{sonst,} \end{cases} \qquad Y_i(\omega) = \begin{cases} k, & \text{falls } \omega_i = k \geq 1 \\ 0, & \text{sonst,} \end{cases}$$

für $i = 1, \dots, n$. Nach Satz 4.2 bilden $\{X_1, \dots, X_n\}$ und $\{Y_1, \dots, Y_n\}$ jeweils eine unabhängige Familie von ZVen. (Die Projektionen auf die einzelnen Komponenten sind unter einem Produktmaß unabhängig, die X_i sind Funktionen der Projektionen, es greift deshalb Satz 5.7. Für die Y_i ebenso.) Nach Konstruktion sind $X_1, \dots, X_n$ verteilt wie im Satz vorgegeben, $Y_1, \dots, Y_n$ sind poissonverteilt, $\mathbb{P}_{Y_i} = \Pi_{p_i}$. Nach (24) ist also $T_n := Y_1 + \cdots + Y_n$ poissonverteilt zum Parameter $\lambda_n > 0$, d.h. $\mathbb{P}_{T_n} = \Pi_{\lambda_n}$. Das Koppelungslemma (Satz 13.2) impliziert

$$d_{TV}\left(\mathbb{P}_{S_n}, \Pi_{\lambda_n}\right) \leq 2\mathbb{P}\left(S_n \neq T_n\right). \tag{25}$$

Wegen $\{S_n \neq T_n\} \subset \bigcup_{1 \leq i \leq n} \{X_i \neq Y_i\}$ schätzen wir $\mathbb{P}(X_i \neq Y_i)$ ab: Es ist

$$\mathbb{P}(X_i = Y_i) = \mathbb{P}_i(\{0\}) + \mathbb{P}_i(\{1\}) = 1 - p_i + e^{-p_i} p_i,$$

also $\mathbb{P}(X_i \neq Y_i) = p_i(1 - e^{-p_i}) \leq p_i^2$, wobei wieder die Ungleichung $e^x \geq 1 + x$ für $x \in \mathbb{R}$ verwendet wird. Mit (25) folgt deshalb

$$d_{TV}\left(\mathbb{P}_{S_n}, \Pi_{\lambda_n}\right) \leq 2\mathbb{P}(S_n \neq T_n) \leq 2 \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) \leq 2 \sum_{i=1}^n p_i^2.$$

Dies ist die Behauptung. ∎

Korollar 13.4. Sei $p = p(n)$ eine Folge in $[0,1]$ mit $n \cdot p \to \lambda > 0$ für $n \to \infty$. Dann gilt für die Binomialverteilung $b_{n,p}$:

$$b_{n,p}(\{k\}) \overset{n\to\infty}{\longrightarrow} \Pi_\lambda(\{k\}) \text{ für alle } k \in \mathbb{N}_0.$$

Beweis.
Für festes $n \in \mathbb{N}$ setzen wir $p_1 := \cdots := p_n := p(n)$. Es seien $X_1, \dots, X_n$ unabhängige ZVe mit $\mathbb{P}(X_i = 1) = p_i = \mathbb{P}(X_i = 0)$. Dann ist $S_n = \sum_{i=1}^n X_i$ binomial $b_{n,p(n)}$ verteilt. Nach Satz 13.3 gilt also

$$d_{TV}\left(b_{n,p(n)}, \Pi_{np(n)}\right) \le 2np^2(n) \sim \frac{2\lambda^2}{n} \to 0 \text{ für } n \to \infty.$$

Andererseits gilt für die Poissonverteilung

$$\Pi_{np(n)}(\{k\}) \overset{n\to\infty}{\longrightarrow} \Pi_\lambda(\{k\}),$$

da $np(n) \to \lambda$. Zusammen folgt die Behauptung. ∎

Man verwendet Korollar 13.4 häufig, um $b_{n,p}$ für große n und kleine p durch die Poissonverteilung Π_{pn} zu approximieren.

Beispiel 13.5. Es gebe 30 Selbstmorde pro 100000 Einwohner pro Jahr. Wie ist die Anzahl der Selbstmorde pro Jahr in einer Stadt mit 120000 Einwohnern approximativ verteilt? Nehmen wir an, jeder der Einwohner begehe unabhängig von den anderen Selbstmord mit W-keit $p = 3/10000 = 0{,}0003$. Bei 120000 Einwohnern wäre die Verteilung der Anzahl der Selbstmorde also binomial

$$b_{120000,0.0003} \approx \Pi_{36}.$$

Nach Satz 13.3 kann man die gefragte zufällige Anzahl also durch die Poissonverteilung Π_{36} approximieren.

14 Der Zentrale Grenzwertsatz

Sind $B_1, \dots, B_n$ unabhängige ZVe mit $\mathbb{P}(B_i = 1) = p = 1 - \mathbb{P}(B_i = 0)$ und $p \in (0,1)$, so lieferte (in Abschnitt 11) der Satz von de Moivre-Laplace 12.3, dass für die Summe $S_n := \sum_{1\le i\le n} B_i$ für alle $a < b$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left(a \le S_n^* \le b\right) = \Phi(b) - \Phi(a), \tag{26}$$

wobei S_n^* die standardisierte ZVe zu S_n bezeichnet (vgl. Definition 12.2). Die Verallgemeinerung dieses Resultats auf Summen unabhängiger ZVe, die alle dieselbe Verteilung haben (mit endlicher Varianz), bezeichnet man als „Zentralen Grenzwertsatz“ (wie auch weiterreichende Verallgemeinerungen, die hier nicht besprochen werden).

Definition 14.1. Zufallsvariablen $X_1, \dots, X_n$ heißen *identisch verteilt*, falls gilt:

$$\mathbb{P}_{X_1} = \mathbb{P}_{X_2} = \cdots = \mathbb{P}_{X_n}.$$

Definition 14.2. Sei $(X_n)_{n\geq 1}$ eine Folge reeller Zufallsvariable und X eine reelle Zufallsvariable mit Verteilungsfunktionen F_n bzw. F. Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert in Verteilung* gegen X, falls für alle $x \in \mathbb{R}$, in denen F stetig ist, gilt

$$F_n(x) \to F(x), \quad (n \to \infty).$$

Bezeichnungen: $X_n \xrightarrow{\mathcal{L}} X$ oder $X_n \xrightarrow{d} X$ für $n \to \infty$. (Dabei stehen $\mathcal{L}$ und d für engl. *law* bzw. *distribution*, was beides englische Begiffe für die Verteilung einer ZVe sind.

Bemerkung 23. Ist gezeigt, dass $(S_n^*)_{n\geq 1}$ in Verteilung gegen eine standardnormalverteilte Zufallsvariable konvergiert, so folgt insbesondere (26), da

$$\mathbb{P}\left(a \leq S_n^* < b\right) = F_{S_n^*}(b) - F_{S_n^*}(a) \overset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a),$$

da Φ stetig auf ganz $\mathbb{R}$ ist.

Satz 14.3 (Zentraler Grenzwertsatz)**.** Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter reeller ZVe mit $\mathrm{Var}(X_1) \in (0,\infty)$. Bezeichne $S_n = \sum_{1\leq i\leq n} X_i$, sowie

$$S_n^* = \frac{S_n - \mathbb{E}\left[S_n\right]}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n\mathbb{E}\left[X_1\right]}{\sigma\sqrt{n}}$$

mit $\sigma^2 := \mathrm{Var}(X_1)$. Sei Z eine standardnormalverteilte ZVe. Dann konvergiert $(S_n^*)_{n\in\mathbb{N}}$ in Verteilung gegen Z.

Der hier gegebene Beweis stützt sich auf folgendes Lemma.

Lemma 14.4. Sei $h : \mathbb{R} \to \mathbb{R}$ eine dreimal stetig differenzierbare Funktion mit beschränkter erster, zweiter und dritter Ableitung. Seien X_n, Z wie in Satz 14.3. Dann gilt

$$\mathbb{E}\left[h(S_n^*)\right] \to \mathbb{E}\left[h(Z)\right].$$

Bevor wir Lemma 14.4 beweisen, zeigen wir, wie daraus der Zentrale Grenzwertsatz 14.3 folgt.

Beweis von 14.3.
Sei $x \in \mathbb{R}$ beliebig gegeben. Dann existieren zwei Funktionen h_1, h_2 wie in Lemma 14.4 mit

$$\mathbb{1}_{(-\infty,x-\varepsilon)} \leq h_1 \leq \mathbb{1}_{(-\infty,x)} \leq h_2 \leq \mathbb{1}_{(-\infty,x+\varepsilon)}.$$

Wir können etwa

$$h_2(t) = \begin{cases} 1, & \text{falls } t \leq x, \\ \left(1 - \left(\frac{t-x}{\varepsilon}\right)^4\right)^4, & \text{falls } x \leq t \leq x+\varepsilon, \\ 0, & \text{falls } t \geq x+\varepsilon \end{cases}$$

wählen, h_1 entsprechend. Die Monotonie des Erwartungswertes liefert

$$\mathbb{E}\left[h_1(S_n^*)\right] \leq \mathbb{P}(S_n^* < x) \leq \mathbb{E}\left[h_2(S_n^*)\right]$$

sowie

$$\mathbb{P}(Z < x - \varepsilon) \le \mathbb{E}\left[h_1(Z)\right], \ \mathbb{E}\left[h_2(Z)\right] \le \mathbb{P}(Z < x + \varepsilon).$$

Mit Lemma 14.4 und $n \to \infty$ erhalten wir

$$\begin{aligned} \mathbb{P}(Z < x - \varepsilon) &\le \liminf_{n\to\infty} \mathbb{P}(S_n^* < x) \\ &\le \limsup_{n\to\infty} \mathbb{P}(S_n^* < x) \le \mathbb{P}(Z < x + \varepsilon). \end{aligned} \tag{27}$$

Es ist für $\varepsilon \downarrow 0$ jeweils

$$\begin{aligned} &\mathbb{P}(x \le Z \le x + \varepsilon) = \int_x^{x+\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy \le \frac{\varepsilon}{\sqrt{2\pi}} \to 0, \\ &\mathbb{P}(x - \varepsilon \le Z < x) \to 0. \end{aligned}$$

Mit $\varepsilon \downarrow 0$ in (27) folgt

$$\lim_{n\to\infty} \mathbb{P}(S_n^* < x) = \mathbb{P}(Z < x) = \Phi(x).$$

Da $x \in \mathbb{R}$ beliebig war, ist dies gerade die Konvergenz in Verteilung von S_n^* gegen Z. ∎

Beweis von Lemma 14.4.
Wir nehmen an, dass $\mathbb{E}\left[X_1\right] = 0$ und $\mathrm{Var}(X_1) = 1$ ist. (Andernfalls kann X_i durch seine standardisierte Version ersetzt werden.) Wir ergänzen $X_1, \dots, X_n$ durch standardnormalverteilte ZVe $Z_1, \dots, Z_n$, sodass die $2n$ ZVen $X_1, \dots, X_n, Z_1, \dots, Z_n$ unabhängig sind. Wir erzeugen künstlich eine Teleskopsumme durch schrittweises Ersetzen der X_i durch Z_i:

$$h\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right) - h\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i\right) = \sum_{i=1}^n \left(h\left(U_i + \frac{X_i}{\sqrt{n}}\right) - h\left(U_i + \frac{Z_i}{\sqrt{n}}\right)\right),$$

wobei $U_i := (X_1 + \dots + X_{i-1} + Z_{i+1} + \dots + Z_n)/\sqrt{n}$. Man beachte, dass (nach einer Übungsaufgabe) $(1/\sqrt{n}) \sum_{1 \le i \le n} Z_i$ wieder standardnormalverteilt ist. Taylorentwicklung von h um U_i liefert

$$\begin{aligned} &h\left(U_i + \frac{X_i}{\sqrt{n}}\right) - h\left(U_i + \frac{Z_i}{\sqrt{n}}\right) \\ &= h'(U_i)\frac{X_i - Z_i}{\sqrt{n}} + h''\left(U_i + \alpha\frac{X_i}{\sqrt{n}}\right)\frac{X_i^2}{2n} - h''\left(U_i + \alpha'\frac{Z_i}{\sqrt{n}}\right)\frac{Z_i^2}{2n} \\ &= h'(U_i)\frac{X_i - Z_i}{\sqrt{n}} + h''(U_i)\frac{X_i^2 - Z_i^2}{2n} + R_{in}, \end{aligned}$$

wobei

$$R_{in} = \left(h''\left(U_i + \alpha\frac{X_i}{\sqrt{n}}\right) - h''(U_i)\right)\frac{X_i^2}{2n} - \left(h''\left(U_i + \alpha'\frac{Z_i}{\sqrt{n}}\right) - h''(U_i)\right)\frac{Z_i^2}{2n}$$

und α, α' zufällig in $[0,1]$ sind. Bezeichne $c''' := \|h'''\|_\infty = \sup_{x\in\mathbb{R}} |h'''(x)| < \infty$. Nach dem Mittelwertsatz gilt stets $|h''(x) - h''(y)| \le c'''|x-y|$. Damit folgt

$$|R_{in}| \le \mathbb{1}_{\{|X_i|\le k\}} c''' \frac{k^3}{n^{3/2}} + \mathbb{1}_{\{|X_i|>k\}} 2c'' \frac{X_i^2}{n} + c''' \frac{|Z_i|^3}{n^{3/2}},$$

wobei $c'' := \|h''\|_\infty$ und $k > 0$ beliebig ist. Da U_i, X_i, Z_i unabhängig sind, folgt

$$\begin{aligned}\mathbb{E}\,[h'(U_i)(X_i - Z_i)] &= \mathbb{E}\,[h'(U_i)]\mathbb{E}\,[X_i - Z_i] = 0,\\ \mathbb{E}\,[h''(U_i)(X_i^2 - Z_i^2)] &= \mathbb{E}\,[h''(U_i)]\mathbb{E}\,[X_i^2 - Z_i^2] = 0.\end{aligned}$$

Also folgt

$$\begin{aligned}\left|\mathbb{E}\left[h\left(U_i + \frac{X_i}{\sqrt{n}}\right) - h\left(U_i + \frac{Z_i}{\sqrt{n}}\right)\right]\right| &= |\mathbb{E}\,[R_{in}]| \le \mathbb{E}\,[|R_{in}|]\\ &\le c''' \frac{k^3 + \mathbb{E}\,[|Z_1|^3]}{n^{3/2}} + \frac{2c''}{n}\mathbb{E}\left[X_1^2 \mathbb{1}_{\{|X_1|>k\}}\right]\end{aligned}$$

und mit Summation über $i = 1, \dots, n$

$$|\mathbb{E}\,[h(S_n^*)] - \mathbb{E}\,[h(Z)]| \le \frac{C}{\sqrt{n}} + 2c''\mathbb{E}\left[X_1^2 \mathbb{1}_{\{|X_1|>k\}}\right]$$

mit einer von k abhängenden Konstanten $C > 0$, also

$$\limsup_{n\to\infty} |\mathbb{E}\,[h(S_n^*)] - \mathbb{E}\,[h(Z)]| \le 2c''\mathbb{E}\left[X_1^2 \mathbb{1}_{\{|X_1|>k\}}\right].$$

Hieraus folgt die Behauptung, da $\mathbb{E}\,[X_1^2 \mathbb{1}_{\{|X_1|>k\}}] \to 0$ für $k \to \infty$, was aus $\mathbb{E}\,[X_1^2] < \infty$ folgt. Z.B. gilt im diskreten Fall ist

$$\mathbb{E}\left[X_1^2 \mathbb{1}_{\{|X_1|>k\}}\right] = \sum_{a:|a|>k} a^2 \mathbb{P}(X_1 = a)$$

und die Konvergenz dieser Reihen liefert die gewünschte Konvergenz. ∎

Bemerkung 24. Man kann (recht leicht) zeigen, dass die Konvergenz der Verteilungsfunktion in Satz 14.3 nicht nur punktweise, sondern sogar gleichmäßig gilt.

Definition 14.5. Seien X, Y reelle ZVe mit Verteilungsfunktionen F und G. Dann ist durch

$$\rho(X, Y) := \rho(F, G) := \sup_{x\in\mathbb{R}} |F(x) - G(x)| = \|F - G\|_\infty$$

eine Metrik auf der Menge der Verteilungsfunktionen definiert. Sie heißt *Kolmogorov-Metrik* (oder auch uniforme Metrik).

Bemerkung 25. In der Situation von Satz 14.3 gilt also sogar

$$\rho(S_n^*, Z) \to 0 \text{ für } n \to \infty.$$

4 Mathematische Statistik

Die Stochastik teilt sich in zwei Teilgebiete ein:

$$\text{Stochastik} \begin{cases} \text{Wahrscheinlichkeitstheorie (Kap. 1-3,5,6)} \\ \text{Statistik} \end{cases}$$

In der W-Theorie nimmt man an, W-Maße, die einfache Zufallsexperimente „steuern", zu kennen, und möchte daraus Eigenschaften komplizierterer Größen herleiten, z.B. Gesetze großer Zahlen, Grenzwertsätze, Poissonapproximation. Die Statistik wiederum lässt sich in zwei Gebiete unterteilen:

$$\text{Statistik} \begin{cases} \text{deskriptive Statistik (Darstellung von Daten: Tabellen, Graphiken)} \\ \text{schließende Statistik (Inferenzstatistik, induktive Statistik) (Kap. 4)} \end{cases}$$

Man kennt das W-Maß $\mathbb{P}$, das ein Zufallsexperiment steuert, nicht und möchte aus Beobachtungen (Realisierungen) von Versuchsausgängen auf $\mathbb{P}$ oder zumindest Eigenschaften von $\mathbb{P}$ schließen. Drei typische Beispiele sind:

1. Münzwurf (Bernoulliexperiment mit Erfolgsw-keit $p \in [0,1]$): Der Parameter p sei unbekannt. Das Experiment werde n-mal unabhängig ausgeführt. Dies liefert Daten $x_1, \dots, x_n \in \{0,1\}$.
   Problem: Rückschlüsse auf p.
   a) Gebe Schätzwert für p an – *Schätzproblem*.
   b) Gebe Intervall für p an – *Konfidenzintervall*.
   c) Entscheide etwa, ob die Münze fair ist, d.h. ob $p = \frac{1}{2}$ oder $p \neq \frac{1}{2}$ – *Test*.
2. Karpfen im Teich: In einem Teich befindet sich eine unbekannte Anzahl N von Fischen. Es werden s Fische gefangen, markiert und wieder ausgesetzt. Nachdem sich die Fische gut durchmischt haben, werden in einem zweiten Fang n Fische gefangen und die darunter Markierten gezählt. Wie schließt man auf N?
   Betrachte Verthältnisse: Sei x die Anzahl der markierten Fische unter den neu Gefangenen. Naheliegend ist $x/n \approx S/N$, also $N \approx \frac{Sn}{x}$.
3. Physikalische Messung: Eine Messung setze sich aus dem zu messenden Wert (deterministisch) und einem zufälligen Messfehler, der sich als Überlagerung vieler kleiner Einflüsse zusammensetze, zusammen. Der Zentrale Grenzwertsatz legt nahe, die Messungen als Realisierungen unabhängiger $\mathcal{N}(\mu, \sigma^2)$-verteilter ZVen mit unbekanntem $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ zu modellieren. Das statistische Problem ist, aus den Daten auf (μ, σ^2) rückzuschließen.
   Annahme: Der Fehler sei im Mittel 0. Messungen werden als Realisierungen einer $\mathcal{N}(\mu, \sigma^2)$-verteilter ZVe $X = \mu + Z$ aufgefasst, wobei Z normalverteilt und zentriert ist.

15 Schätzen

Es wird folgender allgemeine Rahmen für Schätzprobleme verwendet.

Definition 15.1. Ein *Schätzproblem* besteht aus

a) *Stichprobenraum* $(S, \mathcal{C})$: messbarer Raum.
(S beschreibt die Menge der möglichen Beobachtungsergebnisse.)

b) *Familie* $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ von W-Maßen auf $(S, \mathcal{C})$, Θ eine beliebige Parametermenge.
(Menge der möglichen Verteilungen aufgrund theoretischer Vorüberlegungen.)

c) $g : \Theta \to \Gamma \subset \mathbb{R}^d$ *zu schätzende Funktion.*
(Häufig $\Gamma = \Theta \subset \mathbb{R}$ und $g(\vartheta) = \vartheta$.)

Beispiel 15.2. Münze mit unbekannter Erfolgw-keit $p \in [0, 1]$ werde n-mal geworfen:

$$(S, \mathcal{C}) = (\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n)), \quad \Theta = [0, 1]$$

und für $\vartheta \in \Theta$ und $x = (x_1, \ldots, x_n) \in S$ sei

$$\mathbb{P}_\vartheta(\{x\}) = \vartheta^{\sum x_i}(1 - \vartheta)^{n - \sum x_i}$$

sowie $g(\vartheta) = \vartheta$.

Beispiel 15.3 (Karpfen im Teich)**.** N sei die Anzahl der Fische im Teich (unbekannt), s Anzahl markierter Fische, n Anzahl Fische im zweiten Fang.

$$(S, \mathcal{C}) = (\{0, 1, \ldots, n\}, \mathcal{P}(S)) .$$

Vorüberlegung: Angenommen, es seien N Fische im See. Dann gilt für die Anzahl x der markierten Fische im zweiten Fang

$$\mathbb{P}_N(\{x\}) = \frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}}, \ x \in S.$$

Die Anzahl der markierten Fische im zweiten Fang ist hypergeometrisch verteilt zu Parametern n, N, s, d.h. $\mathbb{P}_N = h(\,\cdot\,; n, N, s)$ mit h wie in Satz 2.13. Wir wählen deshalb

$$\Theta = \{\vartheta \in \mathbb{N} : \vartheta \geq s \vee n\} \text{ und zu } \vartheta \in \Theta : \mathbb{P}_\vartheta = h(\,\cdot\,; n, \vartheta, s).$$

Die zu schätzende Funktion ist $g(\vartheta) = \vartheta$.

Beispiel 15.4. Messungen $X_1, \ldots, X_n$ seien unabhängige $\mathcal{N}(\mu, \sigma^2)$-verteilte ZVe mit unbekannten $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times [0, \infty) =: \Theta$. Man beobachtet also Daten im Raum $(S, \mathcal{C}) = (\mathbb{R}^n, \mathfrak{B}^n)$. Zu $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta = \mathbb{P}_{(X_1, \ldots, X_n)}$. Dann hat $\mathbb{P}_\vartheta$ nach Lemma 9.11 die Dichte

$$f_\vartheta(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

für $x = (x_1, \ldots, x_n) \in S$. Die zu schätzende Funktion ist $g(\vartheta) = \vartheta$.

Definition 15.5. Für ein Schätzproblem gegeben durch Stichprobenraum $(S, \mathcal{C})$, Verteilungen $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ und zu schätzender Funktion $g : \Theta \to \Gamma \subset \mathbb{R}^d$ heißt jede messbare Abbildung $T : S \to \Gamma$ *Schätzer* für g. (Γ ist dabei mit der Spur von $\mathfrak{B}^d$ in Γ versehen, vgl. Beispiele 8.3).

Bemerkung 26. Statt Schätzer wird auch Punktschätzer oder Schätzfunktion gesagt.

Wir betrachten im Folgenden Methoden, um Schätzer zu konstruieren, anschließend Gütekriterien für Schätzer.

(A) Wir betrachten zunächst höchstens abzählbares S, d.h. $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ ist eine Familie diskreter W-Maße, $\mathcal{C} = \mathcal{P}(S)$.

Definition 15.6. Sei S höchstens abzählbar und $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ eine Familie von Verteilungen auf S. Zu $x \in S$ heißt

$$L_x : \Theta \to [0,1],\ \vartheta \mapsto \mathbb{P}_\vartheta(\{x\}) \quad \textit{Likelihood-Funktion}.$$

Falls L_x ein globales Maximum $\hat{\vartheta}(x)$ annimmt, so heißt $\hat{\vartheta}(x)$ *Maximum-Likelihood-Schätzer* (ML-Schätzer) von ϑ, und $g(\hat{\vartheta}(x))$ heißt ML-Schätzer von $g(\vartheta)$.

Der ML-Schätzer schätzt also zu gegebener Beobachtung $x \in S$ denjenigen Parameter $\hat{\vartheta}(x)$, für den die beobachteten Daten die größte W-keit haben.

Beispiel 15.7 (ML-Schätzer für Beispiel 15.2)**.** Zu $x \in \{0,1\}^n$ ist $L_x(\vartheta) = \vartheta^{\sum x_i}(1-\vartheta)^{n-\sum x_i}$ und $\mathcal{L}_x(\vartheta) := \log L_x(\vartheta) = (\sum x_i)\log\vartheta + (n - \sum x_i)\log(1-\vartheta)$. Da der Logarithmus monoton wachsend ist, ist L_x maximal genau dann, wenn $\mathcal{L}_x$ maximal ist.

$$\frac{d\mathcal{L}_x}{d\vartheta}(\vartheta) = \frac{1}{\vartheta}\sum x_i - \frac{1}{1-\vartheta}\left(n - \sum x_i\right) \overset{!}{=} 0. \tag{28}$$

In der Nullstelle $\hat{\vartheta}(x) = \frac{1}{n}\sum x_i$ hat L_x ein globales Maximum. Es ist also

$$\hat{\vartheta}(x) = \frac{1}{n}\sum_{1\le i\le n} x_i$$

ML-Schätzer für ϑ.

Bemerkung 27. Die Gleichung (28),

$$\frac{d\mathcal{L}_x}{d\vartheta}(\vartheta) = 0$$

heißt *ML-Gleichung*.

Beispiel 15.8 (ML-Schätzer für Beispiel 15.3)**.** Im Setting von Beispiel 15.3 gilt

$$\frac{\mathbb{P}_\vartheta(\{x\})}{\mathbb{P}_{\vartheta-1}(\{x\})} = \frac{(\vartheta - s)(\vartheta - n)}{\vartheta(\vartheta - s - n + x)}$$

und damit

$$\mathbb{P}_\vartheta(\{x\}) > \mathbb{P}_{\vartheta-1}(\{x\}) \Leftrightarrow (\vartheta - s)(\vartheta - n) > \vartheta(\vartheta - s - n + x) \Leftrightarrow \vartheta < \frac{sn}{x}.$$

Es ist zu $x \in S$ also $\mathbb{P}_\vartheta(\{x\})$ maximal für

$$\begin{cases} \hat{\vartheta}(x) = \lfloor \frac{ns}{x} \rfloor, & \text{falls } \frac{ns}{x} \notin \mathbb{N}, \\ \hat{\vartheta}(x) = \frac{ns}{x} \text{ oder } \hat{\vartheta}(x) = \frac{ns}{x} - 1, & \text{falls } \frac{ns}{x} \in \mathbb{N}. \end{cases}$$

Bei beliebiger Wahl der beiden Werte im Fall $\frac{ns}{x} \in \mathbb{N}$ ist $\hat{\vartheta}$ ML-Schätzer für ϑ. Der ML-Schätzer braucht also im Allgemeinen nicht eindeutig zu sein.

(B) Wir betrachten nun $S \subset \mathbb{R}^n$ und nehmen an, dass alle Verteilungen $\mathbb{P}_\vartheta$ eine Dichte haben, die wir mit f_ϑ bezeichnen. Man beachte, dass dann für alle $x \in S$ und $\vartheta \in \Theta$ gilt: $\mathbb{P}(\{x\}) = 0$. Das Konzept der ML-Schätzer muss deshalb modifiziert werden.

Definition 15.9. Sei $S \subset \mathbb{R}^n$ und $\mathbb{P}_\vartheta$ habe die Dichte f_ϑ für alle $\vartheta \in \Theta$. Für $x \in S$ heißt dann $L_x : \Theta \to \mathbb{R}_0^+$, $\vartheta \mapsto f_\vartheta(x)$ *Likelihood-Funktion.* Falls L_x ein globales Maximum in $\hat{\vartheta}(x)$ annimmt, so heißt $\hat{\vartheta}(x)$ *ML-Schätzer* von ϑ, und $g(\hat{\vartheta}(x))$ heißt ML-Schätzer von $g(\vartheta)$.

Beispiel 15.10 (ML-Schätzer für Beispiel 15.4)**.** Es sei $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times [0, \infty) =: \Theta$ unbekannt. Wir haben

$$f_\vartheta(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right).$$

Damit

$$\mathcal{L}_x(\vartheta) := \log L_x(\vartheta) = -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

Wir lösen

$$\begin{aligned} &\left(\frac{\partial \mathcal{L}_x}{\partial \mu}, \frac{\partial \mathcal{L}_x}{\partial \sigma}\right)(\mu, \sigma) \stackrel{!}{=} (0, 0) \\ &\Leftrightarrow -\frac{1}{2\sigma^2}\sum_{i=1}^n 2(x_i - \mu) = 0 \text{ und } -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (x_i - \mu)^2 = 0 \\ &\Leftrightarrow \mu = \frac{1}{n}\sum_{i=1}^n x_i \text{ und } \sigma^2 = \frac{1}{n}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{i=1}^n x_i\right)^2. \end{aligned}$$

An dieser Stelle liegt auch tatsächlich ein Maximum vor. Damit ist

$$\hat{\vartheta}(x) = \left(\frac{1}{n}\sum_{i=1}^n x_i, \frac{1}{n}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{i=1}^n x_i\right)^2\right)$$

ML-Schätzer für $\vartheta = (\mu, \sigma^2)$.

Wir kommen nun zu Gütekriterien für Schätzer. Im Folgenden betrachten wir Schätzprobleme mit zu schätzender Funktion $g : \Theta \to \Gamma$ mit $\Gamma \subset \mathbb{R}$. Für Zufallsvariablen $X : S \to \mathbb{R}$ werden Erwartungswerte bezüglich $\mathbb{P}_\vartheta$ mit $\mathbb{E}_\vartheta[X]$ bezeichnet.

Definition 15.11. Sei $T : S \to \Gamma$ ein Schätzer für eine zu schätzende Funktion $g(\vartheta)$. Der Schätzer T heißt *erwartungstreu*, falls

$$\mathbb{E}_\vartheta[T] = g(\vartheta) \text{ für alle } \vartheta \in \Theta.$$

Erwartungstreue bedeutet also, dass der Schätzer im Mittel (d.h. im Erwartungswert) die zu schätzende Funktion korrekt schätzt; unabhängig davon, welches $\mathbb{P}_\vartheta$ das Wahre ist.

Bemerkung 28. Man bezeichnet auch $\mathrm{Bias}_\vartheta(T) := \mathbb{E}_\vartheta[T] - g(\vartheta)$ den *Bias des Schätzers* T. Es ist T also erwartungstreu („unbiased“), falls $\mathrm{Bias}_\vartheta(T) = 0$ für alle $\vartheta \in \Theta$ gilt.

Beispiel 15.12 (ML-Schätzer in Beispiel 15.2)**.** In Beispiel 15.2 waren $S = \{0,1\}^n$, $\Theta = [0,1]$, $\mathbb{P}_\vartheta(\{x\}) = \vartheta^{\sum x_i}(1-\vartheta)^{n-\sum x_i}$ für $\vartheta \in \Theta$, $x = (x_1, \dots, x_n) \in S$ und $g(\vartheta) = \vartheta$. Wir hatten den ML-Schätzer bestimmt zu

$$T(x) = \hat{\vartheta}(x) = \frac{1}{n}\sum_{i=1}^n x_i, \quad x \in S.$$

Damit gilt für alle $\vartheta \in \Theta$:

$$\mathbb{E}_\vartheta[T] = \sum_{k=0}^n \frac{k}{n}\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k} = \frac{1}{n}\mathbb{E}[X] = \frac{1}{n}(n\vartheta) = \vartheta,$$

wobei X eine $b_{n,\vartheta}$-verteilte ZVe bezeichne. Der ML-Schätzer für Beispiel 15.2 ist also erwartungstreu.

Beispiel 15.13 (ML-Schätzer in Beispiel 15.4)**.** Beispiel 15.4 waren $(S, \mathcal{C}) = (\mathbb{R}^n, \mathfrak{B}^n)$, $\Theta = \mathbb{R} \times [0,\infty)$, $\mathbb{P}_\vartheta = \mathbb{P}_{(X_1,\dots,X_n)}$ mit $X_1, \dots, X_n$ unabhängig und identisch $\mathcal{N}(\mu, \sigma^2)$ verteilt, wobei $\vartheta = (\mu, \sigma^2)$. Wir wollen nun $g(\vartheta) = \sigma^2$ schätzen und hatten bereits den ML-Schätzer bestimmt zu

$$T(x) = \frac{1}{n}\sum_{i=1}^n\left(x_i - \frac{1}{n}\sum_{i=1}^n x_i\right)^2, \quad x = (x_1, \dots, x_n) \in S.$$

Mit $\mathbb{P}_\vartheta = \mathbb{P}_{(X_1,\dots,X_n)}$ und dem Transformationssatz (Lemma 10.5 bzw. Satz 6.6) gilt

$$\mathbb{E}_\vartheta[T] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2\right]. \tag{29}$$

Eine einfache Rechnung liefert

$$\frac{1}{n}\sum_{i=1}^n\left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2. \tag{30}$$

Zwischenüberlegung: Für unabhängige X, Y mit $\mathbb{P}_X = \mathcal{N}(\mu_1, \sigma_1^2)$, $\mathbb{P}_Y = \mathcal{N}(\mu_2, \sigma_2^2)$ gilt $\mathbb{P}_{X+Y} = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (vgl. Übungsaufgaben). Damit hat $\sum_{1\le i\le n} X_i$ die Verteilung $\mathcal{N}(n\mu, n\sigma^2)$ und $\frac{1}{n}\sum_{1\le i\le n} X_i$ die Verteilung $\mathcal{N}(\mu, \frac{1}{n}\sigma^2)$. Ferner gilt nach einer Übungsaufgabe, dass für X mit $\mathbb{P}_X = \mathcal{N}(\mu_1, \sigma_1^2)$ gilt: $\mathbb{E}[X] = \mu_1$, $\mathrm{Var}(X) = \sigma_1^2$. Insbesondere gilt also

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right] = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right]^2 = \frac{\sigma^2}{n} + \mu^2.$$

Wir erhalten aus (29) und (30), dass

$$\mathbb{E}_\vartheta[T] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i^2] - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{n-1}{n}\sigma^2.$$

Der ML-Schätzer für $g(\vartheta) = \sigma^2$ ist also nicht erwartungstreu. Erwartungstreu ist der Schätzer

$$\widetilde{T}(x) := \frac{n}{n-1}T(x) = \frac{1}{n-1}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{i=1}^n x_i\right)^2,$$

denn es gilt

$$\mathbb{E}_\vartheta[\tilde{T}] = \frac{n}{n-1}\mathbb{E}_\vartheta[T] = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

Der ML-Schätzer T unterschätzt g also im Mittel.

Bemerkung 29. Erwartungstreue ist eine wünschenswerte Eigenschaft, allerdings existieren erwartungstreue Schätzer nicht immer und falls sie existieren, sind sie nicht unbedingt gute Schätzer (vgl. Übungsaufgaben). Ein anderes Maß für die Güte von Schätzern ist der mittlere quadratische Fehler.

Definition 15.14. Sei $T : S \to \Gamma \subset \mathbb{R}$ ein Schätzer für eine zu schätzende Funktion g. Dann heißt

$$\mathrm{MSE}(\vartheta) := \mathbb{E}_\vartheta[(T - g(\vartheta))^2]$$

mittlerer quadratischer Fehler (MSE=mean squared error).

Bemerkung 30. Es gilt

$$\begin{aligned}\mathrm{MSE}(\vartheta) &= \mathbb{E}_\vartheta[(T - \mathbb{E}_\vartheta[T] + \mathbb{E}_\vartheta[T] - g(\vartheta))^2]\\ &= \mathbb{E}_\vartheta[(T - \mathbb{E}_\vartheta[T])^2] + (E_\vartheta[T] - g(\vartheta))^2\\ &= \mathrm{Var}_\vartheta(T) + (\,\mathrm{Bias}_\vartheta(T))^2\,.\end{aligned}$$

Man möchte also $\mathrm{MSE}(\vartheta)$ klein halten, damit die W-keit dass Schätzwerte nahe an der zu schätzenden Größe liegen, groß ist. Allerdings existieren im Allgemeinen keine Schätzer T, der MSE gleichmäßig in ϑ minimiert.

Beispiel 15.15. Gegeben seien Realisierungen $x_1, \dots, x_n$ unabhängiger ZVe $X_1, \dots, X_n$, die identisch gleichverteilt auf $[0, \vartheta]$ seien mit unbekanntem $\vartheta > 0$. Wir haben also $S = [0, \infty)^n$ und $\mathbb{P}_\vartheta$ ist die Gleichverteilung auf $[0, \vartheta]^n$. Wir wollen $g(\vartheta) = \vartheta$ schätzen. Dazu betrachten wir $M_n(x_1, \dots, x_n) = \max\{x_1, \dots, x_n\}$ sowie die Schätzer

$$\hat{\vartheta}_1(x) = \frac{n+1}{n} M_n(x), \quad \hat{\vartheta}_2(x) = \frac{n+2}{n+1} M_n(x), \qquad x = (x_1, \dots, x_n) \in S. \tag{31}$$

Wir benötigen zunächst ein technisches Hilfsmittel:

Lemma 15.16. Unter $\mathbb{P}_\vartheta$ hat M_n die Dichte

$$f_\vartheta(x) = \mathbb{1}_{[0,\vartheta]}(x) \frac{n}{\vartheta^n} x^{n-1}, \quad x \in \mathbb{R},$$

und es gilt

$$\mathbb{E}_\vartheta[M_n] = \frac{n}{n+1} \vartheta, \quad \mathbb{E}_\vartheta[M_n^2] = \frac{n}{n+2} \vartheta^2.$$

Beweis.
Bezeichne F_{X_i} die Verteilungsfunktion der ZVe X_i. Dann gilt

$$F_{X_i}(x) = \mathbb{P}(X_i \leq x) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{\vartheta}, & 0 \leq x \leq \vartheta \\ 1, & x \geq \vartheta. \end{cases}$$

Damit folgt für die Verteilungsfunktion F_{M_n} von M_n:

$$F_{M_n}(x) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \leq x\} \right) = \begin{cases} 0, & x \leq 0 \\ \left(\frac{x}{\vartheta}\right)^n, & 0 \leq x \leq \vartheta \\ 1, & x \geq \vartheta. \end{cases}$$

Für $0 \leq a \leq b \leq \vartheta$ gilt also

$$\mathbb{P}_{M_n}([a, b)) = F_{M_n}(b) - F_{M_n}(a) = \int_a^b F'_{M_n}(x)\, dx = \int_a^b \frac{n}{\vartheta^n} x^{n-1}\, dx = \int_a^b f_\vartheta(x)\, dx.$$

Damit folgt

$$\begin{aligned} \mathbb{E}_\vartheta[M_n] &= \int_0^\vartheta x \frac{n}{\vartheta^n} x^{n-1}\, dx = \frac{n}{\vartheta^n} \left[\frac{1}{n+1} x^{n+1} \right]_0^\vartheta = \frac{n}{n+1} \vartheta, \\ \mathbb{E}_\vartheta[M_n^2] &= \int_0^\vartheta x^2 \frac{n}{\vartheta^n} x^{n-1}\, dx = \frac{n}{\vartheta^n} \left[\frac{1}{n+2} x^{n+2} \right]_0^\vartheta = \frac{n}{n+2} \vartheta^2. \end{aligned}$$

∎

In Beispiel 15.15 gilt damit:

Satz 15.17. Für die Schätzer in (31) gilt: Der Schätzer $\hat{\vartheta}_1$ für ϑ ist erwartungstreu, $\hat{\vartheta}_2$ ist nicht erwartungstreu. Der Schätzer $\hat{\vartheta}_2$ hat gleichmäßig kleineren mittlerene quadratischen Fehler als $\hat{\vartheta}_1$.

Beweis.
Es gilt mit Lemma 15.16:

$$\mathbb{E}_\vartheta[\hat{\vartheta}_1] = \mathbb{E}_\vartheta\left[\frac{n+1}{n}M_n\right] = \vartheta = g(\vartheta),$$

d.h. $\hat{\vartheta}_1$ ist erwartungstreu. Ebenso folgt

$$\mathbb{E}_\vartheta[\hat{\vartheta}_2] = \frac{n(n+2)}{(n+1)^2}\vartheta = \left(1 - \frac{1}{(n+1)^2}\right)\vartheta,$$

d.h. der Schätzer $\hat{\vartheta}_2$ unterschätzt $g(\vartheta)$ im Mittel.
Für den mittleren quadratischen Fehler MSE_1 von $\hat{\vartheta}_1$ gilt

$$\begin{aligned}\mathrm{MSE}_1(\vartheta) &= \mathbb{E}_\vartheta\left[\left(\hat{\vartheta}_1 - \vartheta\right)^2\right]\\ &= \mathbb{E}_\vartheta\left[\left(\frac{n+1}{n}M_n\right)^2\right] - 2\vartheta\mathbb{E}_\vartheta\left[\frac{n+1}{n}M_n\right] + \vartheta^2\\ &= \frac{\vartheta^2}{(n+1)^2-1}.\end{aligned}$$

Analog folgt für den mittleren quadratischen Fehler MSE_2 von $\hat{\vartheta}_2$:

$$\mathrm{MSE}_2(\vartheta) = \frac{\vartheta^2}{(n+1)^2},$$

also $\mathrm{MSE}_2(\vartheta) < \mathrm{MSE}_1(\vartheta)$ für alle $\vartheta > 0$. ∎

16 Testen

Wie bei Schätzproblemen haben wir einen Stichprobenraum $(S, \mathcal{C})$ der möglichen Beobachtungsergebnisse sowie eine Familie von Verteilungen $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$, die aufgrund theoretischer Vorüberlegungen die für das Experiment in Frage kommt. Wir wollen nun nicht mehr ϑ oder eine Funktion $g(\vartheta)$ schätzen, sondern nur feststellen, ob der wahre Parameter ϑ zu einer Mengen $\Theta_0 \subset \Theta$ oder zu $\Theta_1 = \Theta \setminus \Theta_0$ gehört. In den drei Beispielen aus Abschnitt 15 wollen wir also etwa testen, ob die Münze fair ist oder nicht, ob im Teich eine gewisse Anzahl von Fischen überschritten wird oder nicht, oder, ob der Messwert aufgrund der Messungen zu einem Intervall gehört oder nicht.
Im Folgenden sei stets $(S, \mathcal{C})$ ein Stichprobenraum und $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ eine Familie von Verteilungen auf $(S, \mathcal{C})$. Dies wird auch als *parametrisches Modell* bezeichnet. Folgende Bezeichnungen sind üblich.

Bezeichnung 2. • *Nullhypothese* (H_0): „Der wahre Wert ϑ gehört zu Θ_0“, wobei $\Theta_0 \subset \Theta$.

- *Alternative* (H_1): „Der wahre Wert ϑ gehört zu Θ_1“, wobei $\Theta_1 = \Theta \setminus \Theta_0$.

Definition 16.1. Ein *Test* ist eine messbare Abbildung $T : S \to \{0, 1\}$, wobei $T(x) = 1$ die Annahme von H_1 und $T(x) = 0$ die Annahme von H_0 bedeutet. Die Menge $K = \{x \in S : T(x) = 1\}$ heißt *kritischer Bereich* von T.

Bemerkung 31. Man spricht von einem *Fehler 1. Art*, falls man die Hypothese fälschlich ablehnt, von einem *Fehler 2. Art*, falls man die Hypothese fälschlich annimmt.

Definition 16.2. Ein Test hat *(Signifikanz-)Niveau* $\alpha \in [0, 1]$, falls

$$\forall \vartheta \in \Theta_0 : \mathbb{P}_\vartheta(K) \leq \alpha.$$

Bemerkung 32. Bei einem Test mit Niveau α ist die W-keit für einen Fehler 1. Art also durch α beschränkt. Häufig liegt eine Asymmetrie bezüglich der Fehler 1. und 2. Art vor, z.B. soll getestet werden, ob eine gewisse Krankheit vorliegt (H_0) oder nicht (H_1), um gegebenenfalls eine Behandlung durchzuführen. Führt nun die Nichtbehandlung eines Kranken zu irreparablem Schaden, die Behandlung eines Gesunden nur zu materiellem Schaden, so muss der Fehler 1. Art kontrolliert werden. Typische Vorgehensweise: Man fixiert ein $\alpha \in [0, 1]$ und sucht unter den Tests zum Niveau α, d.h. mit $\mathbb{P}_\vartheta(K) \leq \alpha$ für $\vartheta \in \Theta_0$ denjenigen Test, der die W-keit für Fehler 2. Art minimiert.

Definition 16.3. Die Funktion $\beta : \Theta \to [0, 1]$,

$$\beta(\vartheta) = \mathbb{P}_\vartheta(K)$$

heißt *Gütefunktion* des Tests. Für $\vartheta \in \Theta_1$ heißt $\beta(\vartheta)$ die *Macht des Tests* an der Stelle $\vartheta \in \Theta_1$.

Bemerkung 33. Für $\vartheta \in \Theta_0$ ist $\beta(\vartheta)$ die W-keit für einen Fehler 1. Art, für $\vartheta \in \Theta_1$ ist $1 - \beta(\vartheta)$ die W-keit für den Fehler 2. Art.

Beispiel 16.4. Wir betrachten Beispiel 15.2: $(S, \mathcal{C}) = (\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n))$, $\Theta = [0, 1]$, $\mathbb{P}_\vartheta(\{x\}) = \vartheta^{\sum x_i}(1 - \vartheta)^{n - \sum x_i}$. Wir wählen als Hypothese H_0, dass die Münze fair ist, und testen gegen die Alternative H_1, dass $\vartheta \neq 1/2$ ist, d.h. $\Theta_0 = \{1/2\}$ und $\Theta_1 = [0, 1] \setminus \{1/2\}$. Wir legen das Niveau zu $\alpha = 0{,}05$ fest.
Es ist plausibel, die Hypothese abzulehnen, falls $|\sum_{1 \leq i \leq n} x_i - n/2| > c$ für einen kritischen Wert $c > 0$. Wir wählen also $c > 0$ minimal mit

$$\mathbb{P}_{\frac{1}{2}}\left(\left|\sum_{i=1}^{n} X_i - \frac{n}{2}\right| > c\right) = \sum_{k : |k - n/2| > c} \binom{n}{k} \left(\frac{1}{2}\right)^n \leq \alpha = 0{,}05,$$

wobei $X_1, \ldots, X_n$ unabhängig sind mit $\mathbb{P}(X_i = 1) = 1/2 = \mathbb{P}(X_i = 0)$. Z.B. für $n = 100$ erhält man $c = 10$. Damit ist der kritische Bereich $K = \{x \in S : |\sum_{i \leq 100} x_i - 50| > 10\}$. Was passiert mit dem Fehler 2. Art? Falls etwa $\vartheta = 0{,}6$, so ist $\beta(\vartheta) = 0{,}462$, d.h. der Fehler 2. Art hat W-keit $0{,}538$ für $\vartheta = 0{,}6$. Die Daten reichen nicht aus für eine bessere Trennschärfe. Möchte man etwa $\mathbb{P}_{1/2}(K) \leq 0{,}05$ und $\mathbb{P}_{0,6}(K) \geq 0{,}9$, so muss n erhöht werden.

Bezeichnung 3. Falls $|\Theta_0| = 1$, so heißt die Hypothese *einfach*, falls $|\Theta_1| = 1$, so heißt die Alternative einfach. Im Falle $|\Theta_0| > 1$ bzw. $|\Theta_1| > 1$ heißen Hypothese bzw. Alternative *zusammengesetzt*.

Der Fall einer einfachen Alternative, d.h. $\Theta_1 = \{\vartheta_1\}$, führt auf ein Optimierungsproblem: Suche $K \subset S$, so dass $\mathbb{P}_{\vartheta_1}(K)$ maximal unter der Nebenbedingung $\mathbb{P}_\vartheta(K) \leq \alpha$ für $\vartheta \in \Theta_0$ wird. Der Fall einer zusammengesetzten Alternative, d.h. $|\Theta_1| > 1$: Falls eine Menge $K \subset S$ existiert, die für alle $\vartheta_1 \in \Theta_1$ optimal ist unter der Nebenbedingung $\mathbb{P}_\vartheta(K) \leq \alpha$ für $\vartheta \in \Theta_0$, so spricht man von einem gleichmäßig mächtigsten Test zum Niveau α. (UMP-Test, uniformly most powerful)

Konstruktion von Tests: Likelihood-Quotienten-Tests. Wir betrachten — wie bei ML-Schätzern — wieder zwei Fälle.

- S abzählbar, d.h. $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ ist eine Familie diskreter W-Verteilungen.
- $S \subset \mathbb{R}^n$ und alle $\mathbb{P}_\vartheta$ haben Dichten f_ϑ.

Im Fall einfacher Hypothese und Alternative sei $\Theta = \{\vartheta_0, \vartheta_1\}$, die Hypothese H_0 sei durch $\Theta_0 = \{\vartheta_0\}$ gegeben. Wir betrachten für jedes feste $x \in S$ wieder die Likelihood-Funktion

$$L_x(\vartheta) = \begin{cases} \mathbb{P}_\vartheta(\{x\}), & \text{falls } \mathbb{P}_{\vartheta_0}, \mathbb{P}_{\vartheta_1} \text{ diskret,} \\ f_\vartheta(x), & \text{falls } \mathbb{P}_{\vartheta_0}, \mathbb{P}_{\vartheta_1} \text{ mit Dichten,} \end{cases} \qquad \vartheta \in \Theta.$$

Definition 16.5. Der Quotient

$$\frac{L_x(\vartheta_1)}{L_x(\vartheta_0)}$$

heißt *Likelihood-Quotient*. Ein Likelihood-Quotienten-Test (LQT) von $\Theta_0 = \{\vartheta_0\}$ gegen $\Theta_1 = \{\vartheta_1\}$ ist ein Test $T : S \to \{0, 1\}$ der Form

$$T(x) = \begin{cases} 1, & \text{falls } L_x(\vartheta_1)/L_x(\vartheta_0) \geq c, \\ 0, & \text{sonst,} \end{cases} \tag{32}$$

mit $c > 0$.

Bemerkung 34. Die Idee eines LQT ist, dass hohe Werte des Likelihood-Quotienten für ϑ_1, d.h. für Ablehnung der Hypothese, sprechen. Der kritische Bereich des LQT ist gegeben durch

$$K = \left\{ x \in S : \frac{L_x(\vartheta_1)}{L_x(\vartheta_0)} \geq c \right\} \text{ und das Signifikanzniveau } \alpha \geq \mathbb{P}_{\vartheta_0}(K).$$

Man sagt auch, der LQT habe Signifikanzniveau $\alpha = \mathbb{P}_{\vartheta_0}(K)$.

Satz 16.6 (Lemma von Neyman-Pearson)**.** Jeder LQT T ist im folgenden Sinne optimal: Ist $\widetilde{T}$ ein weiterer Test mit $\mathbb{P}_{\vartheta_0}(\widetilde{K}) \leq \mathbb{P}_{\vartheta_0}(K)$, wobei $K, \widetilde{K}$ die kritischen Bereiche von T und $\tilde{T}$ bezeichnen, so hat $\widetilde{T}$ eine mindestens ebenso große Fehlerwahrscheinlichkeit 2. Art wie T:

$$\mathbb{P}_{\vartheta_0}(\widetilde{K}) \leq \mathbb{P}_{\vartheta_0}(K) \quad \Rightarrow \quad \mathbb{P}_{\vartheta_1}(\widetilde{K}) \leq \mathbb{P}_{\vartheta_1}(K).$$

Beweis.
Wir betrachten den Fall diskreter Räume S. Für den Fall mit Dichten kann man analog schließen. Wir haben $\mathbb{P}_{\vartheta_0}(\widetilde{K}) \leq \mathbb{P}_{\vartheta_0}(K)$. Sei $A := \{x \in S : T(x) > \widetilde{T}(x)\}$. Für $x \in A$ ist $T(x) = 1$, gemäß (32) also $\mathbb{P}_{\vartheta_1}(\{x\}) \geq c\mathbb{P}_{\vartheta_0}(\{x\})$. Auf $B := \{x \in S : T(x) < \widetilde{T}(x)\}$ ist $T(x) = 0$, also $\mathbb{P}_{\vartheta_1}(\{x\}) < c\mathbb{P}_{\vartheta_0}(\{x\})$. Damit folgt

$$\begin{aligned}
\mathbb{P}_{\vartheta_1}(K) - \mathbb{P}_{\vartheta_1}(\widetilde{K}) &= \mathbb{E}_{\vartheta_1}[\mathbb{1}_K] - \mathbb{E}_{\vartheta_1}[\mathbb{1}_{\widetilde{K}}] = \mathbb{E}_{\vartheta_1}[T] - \mathbb{E}_{\vartheta_1}[\widetilde{T}] \\
&= \sum_{x \in S} \left(T(x) - \widetilde{T}(x)\right) \mathbb{P}_{\vartheta_1}(\{x\}) \\
&= \sum_{x \in A} \left(T(x) - \widetilde{T}(x)\right) \mathbb{P}_{\vartheta_1}(\{x\}) + \sum_{x \in B} \left(T(x) - \widetilde{T}(x)\right) \mathbb{P}_{\vartheta_1}(\{x\}) \\
&\geq \sum_{x \in A} \left(T(x) - \widetilde{T}(x)\right) c\mathbb{P}_{\vartheta_0}(\{x\}) + \sum_{x \in B} \left(T(x) - \widetilde{T}(x)\right) c\mathbb{P}_{\vartheta_0}(\{x\}) \\
&= c \sum_{x \in S} \left(T(x) - \widetilde{T}(x)\right) \mathbb{P}_{\vartheta_0}(\{x\}) = c \left(\mathbb{E}_{\vartheta_0}[T] - \mathbb{E}_{\vartheta_0}[\widetilde{T}]\right) \\
&= c \left(\mathbb{P}_{\vartheta_0}(K) - \mathbb{P}_{\vartheta_0}(\widetilde{K})\right) \geq 0.
\end{aligned}$$

∎

Im Fall zusammengesetzter Hypothesen bzw. Alternativen, also $|\Theta| > 2$, kann man analog vorgehen: Für $x \in S$ sei

$$\lambda(x) = \frac{\sup_{\vartheta \in \Theta} L_x(\vartheta)}{\sup_{\vartheta \in \Theta_0} L_x(\vartheta)}.$$

Dann gilt $\lambda(x) \geq 1$ für alle $x \in S$. Große Werte von $\lambda(x)$ legen wieder nahe, dass $\vartheta \in \Theta_1$ gilt. Man wählt dann den Test mit kritischem Bereich $K = \{x \in S : \lambda(x) \geq c\}$ mit einem $c > 1$.

Bemerkung 35. Falls $\Theta = \{\vartheta_0, \vartheta_1\}$ und $c > 1$, so gilt

$$\lambda(x) = \frac{\max\{L_x(\vartheta_0), L_x(\vartheta_1)\}}{L_x(\vartheta_0)} \geq c \quad \Leftrightarrow \quad \frac{L_x(\vartheta_1)}{L_x(\vartheta_0)} \geq c.$$

Die Vorgehensweise für zusammengesetzte Hypothesen bzw. Alternativen ist also tatsächlich eine Verallgemeinerung des LQT.

5 Informationstheorie

In diesem Kapitel wird die Entropie definiert und der Quellenkodierungssatz erläutert.

17 Entropie

Definition 17.1. Eine (Informations-)*Quelle* ist ein Paar $(S, \mathbb{P})$, bestehend aus einer höchstens abzählbaren Menge $S \neq \emptyset$ von *Symbolen* (oder *Signalen*) und einem diskreten Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(S, \mathcal{P}(S))$.

Es soll der „Informationsgehalt" von Quelle gemessen werden. Die Quelle $(S, \mathbb{P})$ liefert Symbole, die zunächst unbekannt sind und von der Quelle generiert werden. Die Symbole sind zufällig und gemäß $\mathbb{P}$ verteilt. An den Informationsgehalt von Quellen stellen wir axiomatisch einige Anforderungen. Zunächst entwickeln wir eine Maßzahl I für den Informationsgehalt eines einzelnen generierten Symbols: Sei $s \in S$ und $p_s := \mathbb{P}(\{s\})$ die W-keit, dass die Quelle s generiert.

- $p_s = 1$: In diesem Fall liefert die Quelle (fast sicher) das Symbol s. Dies interpretieren wir als keine Unsicherheit/Überraschung bzw. keine Information: $I(p_s) = I(1) := 0$.
- $p_s > 0$ sehr klein: Generiert die Quelle trotz kleinem p_s dennoch das Symbol s, so ist die Unsicherheit/Überraschung groß, $I(p_s)$ soll deshalb groß sein.

Wir fordern:

(a) $I(1) = 0$,

(b) $p \mapsto I(p)$ ist monoton fallend,

(c) I ist stetig,

(d) $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$.

Eigenschaft (d) bedeutet: Werden zwei Symbole unabhängig voneinander empfangen, so addieren sich deren Informationen. Insbesondere folgt aus c) und d), dass $I(p^r) = rI(p)$ für alle $r > 0$ und $p \in [0, 1]$.
In der Analysis zeigt man, dass eine Funktion I mit (a)–(d) von der Form $I(p) = K \cdot \log_b p$ ist, mit $b > 0, K < 0$.

Definition 17.2. Für eine Quelle $(S, \mathbb{P})$ ist die *Information* (Unsicherheit) eines generierten Symbols $s \in S$ mit $p_s = \mathbb{P}(\{s\})$ definiert als:

$$I(p_s) := -\log_2 p_s, \quad I(0) := 0.$$

Mit der Information eines generierten Symbols kann nun die Information der Quelle definiert werden: Die Information einer Quelle soll nun die mittlere Information eines durch die Quelle generierten Symbols sein, d.h. der Erwartungswert der Information eines zufälligen von der Quelle gesendeten Signals.

Definition 17.3. Die *Entropie* einer Quelle $(S, \mathbb{P})$ mit $S = \{s_1, s_2, \ldots\}$ und $p_i := \mathbb{P}(\{s_i\})$ ist gegeben durch

$$H_2(\mathbb{P}) := \sum_{i=1}^{\infty} p_i I(p_i) = -\sum_{i=1}^{\infty} p_i \log_2 p_i.$$

Dabei wird $x \log_2 x := 0$ für $x = 0$ gesetzt. (Dies setzt die Funktion $x \mapsto x \log_2 x$ stetig nach 0 fort.)

Bemerkung 36. Für endliches $S = \{s_1, \ldots, s_n\}$ und $p_i = \mathbb{P}(\{s_i\})$ wird auch

$$H(p_1, \ldots, p_n) := H(\mathbb{P})$$

geschrieben. Für jedes $n \in \mathbb{N}$ ist H dann eine Funktion $H : \Delta_n \to [0, \infty)$ mit dem Simplex $\Delta_n = \{(p_1, \ldots, p_n) \in [0,1]^n : \sum_{i=1}^n p_i = 1\}$. Damit kann H auch als Funktion auf $\bigcup_{n=1}^{\infty} \Delta_n$ definiert werden, ist aber zudem für abzählbar unendliche Vektoren $(p_i)_{i \geq 1}$ in $[0,1]$ mit $\sum_{i=1}^{\infty} p_i = 1$ definiert.
Für ZVe X mit Werten in einer höchstens abzählbaren Menge S wird die Entropie von X durch

$$H(X) := H(\mathbb{P}_X)$$

definiert. Wir untersuchen nun zunächst den Wertebereich von $H(\mathbb{P})$ für W-Verteilungen $\mathbb{P}$ mit endlichem S.

Lemma 17.4. Für $(p_1, \ldots, p_n), (q_1, \ldots, q_n) \in [0,1]^n$, $1 = \sum_{i=1}^n p_i \geq \sum_{i=1}^n q_i$ gilt:

$$-\sum_{i=1}^{n} p_i \log_2 p_i \leq -\sum_{i=1}^{n} p_i \log_2 q_i.$$

Gleichheit gilt genau dann, wenn $(p_1, \ldots, p_n) = (q_1, \ldots, q_n)$.

Beweis.
Es gilt $\log_2 x \leq c(x-1)$ für alle $x > 0$ mit passendem $c > 0$, z.B. $c = 1/\ln 2$. Dabei gilt Gleichheit genau für $x = 1$. Seien $p_i, q_i > 0$. Dann folgt

$$\log_2 \frac{q_i}{p_i} \leq c\left(\frac{q_i}{p_i} - 1\right)$$

mit Gleichheit genau für $p_i = q_i$, also $p_i \log_2 \frac{q_i}{p_i} \leq c q_i - c p_i$ und damit

$$p_i \log_2 \frac{1}{p_i} \leq p_i \log_2 \frac{1}{q_i} + c q_i - c p_i \tag{33}$$

mit Gleichheit genau für $p_i = q_i$. Ungleichung (33) ist ebenfalls wahr für $p_i = 0$ (wegen $0 \leq q_i$) und für $p_i \neq 0$ und $q_i = 0$ (rechte Seite ist dann ∞). Die Ungleichung (33) gilt also für beliebige $p_i, q_i \geq 0$ mit Gleichheit genau für $p_i = q_i$. Summation in (33) über i liefert nun

$$\sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log_2 \frac{1}{q_i} + \underbrace{c \sum_{i=1}^{n} q_i - c \sum_{i=1}^{n} p_i}_{\leq 0} \leq \sum_{i=1}^{n} p_i \log_2 \frac{1}{q_i}.$$

Dies ist die Behauptung. Gleichheit gilt genau für $p_i = q_i$ für $i = 1, \ldots, n$. ∎

Satz 17.5. Sei $\mathbb{P}$ eine W-Verteilung auf $S = \{s_1, \dots, s_n\}$. Dann gilt

$$0 \leq H_2(\mathbb{P}) \leq \log_2 n.$$

Zudem gilt $H_2(\mathbb{P}) = \log_2 n$ genau für die Gleichverteilung $\mathbb{P}$ auf S und $H_2(\mathbb{P}) = 0$ genau, falls $\mathbb{P}(\{s_i\}) = 1$ für ein $i \in \{1, \dots, n\}$.

Beweis.
Seien $p_i = \mathbb{P}(\{s_i\})$ und $q_i := \frac{1}{n}$ für $i = 1, \dots, n$. Dann folgt mit Lemma 17.4

$$H_2(\mathbb{P}) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log_2 \frac{1}{1/n} = \log_2 n \sum_{i=1}^{n} p_i = \log_2 n.$$

Gleichheit gilt genau dann, wenn $p_i = q_i = 1/n$ für $i = 1, \dots, n$ gilt, d.h. wenn $\mathbb{P}$ die Gleichverteilung auf S ist. Falls $H_2(\mathbb{P}) = 0$ ist, so gilt $p_i \log \frac{1}{p_i} = 0$ für alle $i = 1, \dots, n$ also $p_i \in \{0, 1\}$ für $i = 1, \dots, n$. ∎

18 Codierung von Quellen

Ziel dieses Abschnitts ist es, Symbole einer Informationsquelle möglichst effizient zu codieren.

Definition 18.1. Sei $(S, \mathbb{P})$ eine Quelle. Ein *binärer Code* ist eine injektive Abbildung

$$k : S \to \bigcup_{n \geq 1} \{0, 1\}^n.$$

Für $s \in S$ wird $k(s)$ als *Codewort von* s bezeichnet. Ferner sei $\ell : S \to \mathbb{N}$, $s \mapsto$"Anzahl der Komponenten von $k(s)$". Die Zahl $\ell(s)$ heißt *Codewortlänge* von s.

Bemerkung 37. Um effizient zu codieren, sind kurze Codewortlängen von Vorteil, allerdings muss dabei gewährleistet sein, dass Codewörter eindeutig zu entziffern sind. Wir fordern deshalb für Codes die *Präfix-Eigenschaft*: Kein Codewort ist Präfix eines anderen Codeworts. Man sagt auch, der Code sei *präfixfrei.* Im Weiteren werden wir nur präfixfreie binäre Codes betrachten. Wir können diese veranschaulichen, indem die Codewörter mit Blättern eines (gewurzelten) Binärbaums identifiziert werden, vgl. Abbildung 2.

Satz 18.2 (Fano-Kraft Ungleichung)**.** Sei k ein präfixfreier binärer Code für S und $\ell : S \to \mathbb{N}$ die Funktion der Codewortlängen. Dann gilt

$$\sum_{s \in S} 2^{-\ell(s)} \leq 1.$$

Gleichheit gilt genau dann, wenn der Code einem vollständigen Binärbaum entspricht.

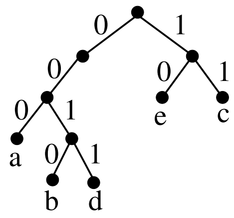


Abbildung 2: Beispiel eines binären Baumes zum Code k für $S = \{a, b, c, d, e\}$ mit $k(a) = (0, 0, 0)$, $k(b) = (0, 0, 1, 0)$, $k(c) = (1, 1)$, $k(d) = (0, 0, 1, 1)$, $k(a) = (1, 0)$. Der gezeigte Baum ist nicht vollständig, der Code kann etwa um das Codewort $(0, 1)$ erweitert werden.

Beweis.
Wir betrachten einen binären Baum, dessen Blätter die Codewörter von k enthalten und konstruieren in diesem Baum einen zufälligen Pfad. Dazu starten wir an der Wurzel und werfen eine faire Münze, um festzulegen, ob der zufällige Pfad der 0-Kante oder der 1-Kante folgt. Falls die entsprechende Kante im Baum nicht existiert, terminiert das Verfahren. Andernfalls führt die Kante zu einem weiteren Knoten. Ist dieser ein Blatt, so stoppen wir, andernfalls werfen wir unabhängig eine faire Münze und iterieren das Verfahren, um den Pfad fortzusetzen. Es bezeichne A_s das Ereignis, dass das Verfahren in einem Blatt mit Codewort $k(s)$ stoppt. Nach Konstruktion folgt, dass $\mathbb{P}(A_s) = (1/2)^{\ell(s)}$ gilt für alle $s \in S$. Zudem sind die Ereignisse A_s für $s \in S$ paarweise disjunkt. Es folgt also

$$1 \geq \mathbb{P}\left(\bigcup_{s \in S} A_s\right) = \sum_{s \in S} \mathbb{P}(A_s) = \sum_{s \in S} 2^{-\ell(s)}.$$

Dies ist die Fano-Kraft Ungleichung. Das Argument zeigt zudem, dass Gleichheit genau dann gilt, wenn jeder innere Knoten zwei Kinder besitzt, der binäre Baum also vollständig ist. ∎

Korollar 18.3. Sei $S = \{s_1, \ldots, s_m\}$ und $l_1, \ldots, l_m \in \mathbb{N}$ mit $\sum_{j=1}^m 2^{-l_j} \leq 1$. Dann existiert ein binärer präfixfreier Code für S mit Codewortlängen $\ell(s_j) = l_j$ für $j = 1, \ldots, m$.

Beweis.
Wir nehmen ohne Einschränkung $l_1 \leq \cdots \leq l_m$ an und führen Induktion über m. Für $m = 1$ ist die Behauptung trivial. Sei nun die Behauptung für je $m - 1$ Codewortlängen bewiesen. Für das Teilalphabet $\{s_1, \ldots, s_{m-1}\}$ existiert dann nach Induktionsvorraussetzung ein binärer präfixfreier Code. Für seine Codewortlängen gilt dann echte Ungleichung in Satz 18.2. Damit lässt sich dieser Code fortsetzen. Da für s_m ein Codewort mit maximaler Länge l_m benötigt wird, kann der Code für $\{s_1, \ldots, s_{m-1}\}$ passend zu einem Code für S fortgesetzt werden. ∎

Definition 18.4. Seien $(S, \mathbb{P})$ eine Quelle, $k : S \to \bigcup_n \{0,1\}^n$ ein präfixfreier Code sowie $\ell : S \to \mathbb{N}$ die Funktion der Codewortlängen. Sei ferner X eine ZVe in S mit Verteilung $\mathbb{P}_X = \mathbb{P}$ und für alle $s \in S$ sei $p_s = \mathbb{P}(X = s)$. Als *mittlere Codewortlänge* bezeichnen wir dann

$$\mathbb{E}\,[\ell(X)] = \sum_{s \in S} p_s \ell(s).$$

Satz 18.5 (Quellencodierungssatz)**.** Sei $(S, \mathbb{P})$ eine Quelle und X eine ZVe in S mit Verteilung $\mathbb{P}_X = \mathbb{P}$. Für jeden präfixfreien binären Code gilt

$$\mathbb{E}\,[\ell(X)] \geq H_2(X) = H_2(\mathbb{P}_X).$$

Beweis.
Es existert eine Konstante $c > 0$ mit $\log_2 x \leq c(x-1)$ für alle $x > 0$, z.B. $c = 1/\ln(2)$. Für jeden präfixfreien binären Code gilt

$$\begin{aligned} H_2(X) - \mathbb{E}\,[\ell(X)] &= \sum_{s \in S} p_s \left(\log_2 \frac{1}{p_s} - \ell(s) \right) = \sum_{s \in S} p_s \log_2 \frac{2^{-\ell(s)}}{p_s} \\ &\leq \sum_{s \in S} p_s c \left(\frac{2^{-\ell(s)}}{p_s} - 1 \right) = c \underbrace{\sum_{s \in S} 2^{-\ell(s)}}_{\leq 1} - \underbrace{c \sum_{s \in S} p_s}_{= c \cdot 1} \\ &\leq 0. \end{aligned}$$

Es folgt die Behauptung. ∎

Beispiel 18.6 (Shannon-Code)**.** Sei $(S, \mathbb{P})$ eine Quelle. Es bezeichne $p_s = \mathbb{P}(\{s\})$ und $\ell_s = \lceil -\log_2 p_s \rceil$ für $s \in S$. Es gilt also

$$-\log_2 p_s \leq \ell_s \leq -\log_2 p_s + 1.$$

Nach Korollar 18.3 existiert ein zugehöriger Code mit Codewortlängen $\ell(s) = \ell_s$, denn es gilt $\sum_{s \in S} 2^{-\ell_s} \leq \sum_{s \in S} p_s = 1$. Solch einen Code bezeichnet man als *Shannon-Code*. Er benötigt zur Codierung im Mittel höchstens ein Bit pro Symbol mehr als jeder andere präfixfreie binäre Code, denn

$$\mathbb{E}\,[\ell(X)] = \sum_{s \in S} p_s \ell_s \leq -\sum_{s \in S} p_s \log_2 p_s + \sum_{s \in S} p_s = H(\mathbb{P}) + 1,$$

und $H(\mathbb{P})$ ist nach dem Quellencodierungssatz eine untere Schranke für die mittlere Codewortlänge für jeden präfixfreien binären Code.

Beispiel 18.7 (Huffman-Code)**.** Sei $(S, \mathbb{P})$ eine Quelle mit endlichem S und $p_s = \mathbb{P}(\{s\})$. Ein (bezüglich der mittleren Codewortlänge) optimaler präfixfreier binärer Code für $(S, \mathbb{P})$ muss die folgenden Eigenschaften haben, denn andernfalls könnte man ihn jeweils direkt verbessern:

(i) Ein Knoten, der kein Blatt ist, hat genau zwei Kinder.

(ii) Falls $p_{s_1} < p_{s_2}$ für $s_1, s_2 \in S$, so gilt $\ell(s_1) \geq \ell(s_2)$. Andernfalls tausche man die Codewörter für s_1 und s_2 und vermindert damit die mittlere Codewortlänge.

(iii) Haben $s_1, s_2 \in S$ die kleinsten Wahrscheinlichkeiten, d.h. $p_{s_1} \leq p_{s_2} \leq p_s$ für alle $s \in S \setminus \{s_1, s_2\}$, dann gilt: $\ell(s_1) = \ell(s_2) \geq \ell(s)$ für alle $s \in S \setminus \{s_1, s_2\}$.

(iv) O.E. sind $s_1, s_2 \in S$ mit minimalen W-keiten also $p_{s_1} \leq p_{s_2} \leq p_s$ für alle $s \in S \setminus \{s_1, s_2\}$ Geschwister (haben einen gemeinsamen direkten Vorfahren im Baum). Mit S muss dann auch

$$\tilde{S} = (S \setminus \{s_1, s_2\}) \cup \{\langle s_1 s_2 \rangle\}$$

betrachtet werden, wobei s_1, s_2 zu einem neuen Buchstaben $\langle s_1 s_2 \rangle$ verschmelzen. Die neue W-Verteilung $\tilde{\mathbb{P}}$ auf $\tilde{S}$ ist dann gegeben durch

$$\tilde{\mathbb{P}}(\{s\}) = \mathbb{P}(\{s\}), \text{ falls } s \in S \setminus \{s_1, s_2\} \quad \text{und} \quad \tilde{\mathbb{P}}(\{\langle s_1 s_2 \rangle\}) = \mathbb{P}(\{s_1\}) + \mathbb{P}(\{s_2\}). \tag{34}$$

Dies liefert umgekehrt die Konstruktion der *Huffman-Codes*: Man verschmilzt gemäß (iv) zwei Symbole kleinster Wahrscheinlichkeit zu einem neuen Symbol mit entsprechend aktualisierten Wahrscheinlichkeiten gemäß (34) und wiederholt die Prozedur iterativ bis ein Symbol mit W-keit 1 verbleibt. Dann liest man den so entstandenen Baum von der Wurzel (dem Symbol mit Gewicht 1) zu den Blättern, um den Code zu erhalten, vgl. Abbildung 2.

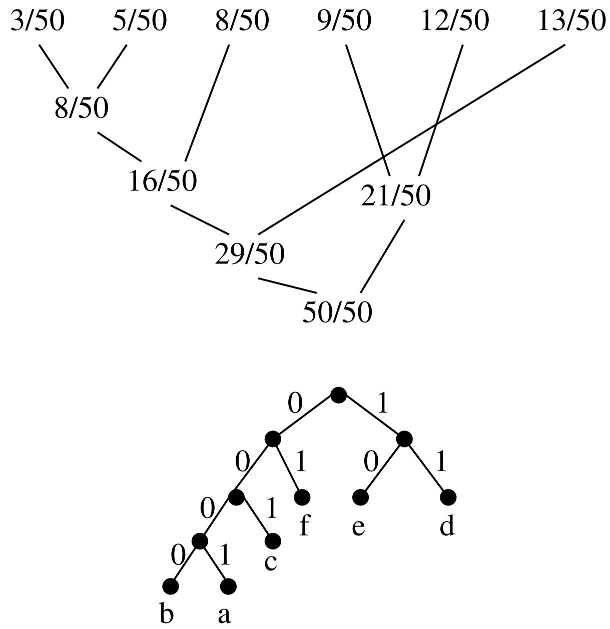


Abbildung 3: Beispiel zur Konstruktion eines Huffman-Codes für $(\mathsf{S}, \mathbb{P})$ mit $\mathsf{S} = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}\}$ mit $\mathbb{P}(\{\mathsf{a}\}) = \frac{3}{50}$, $\mathbb{P}(\{\mathsf{b}\}) = \frac{5}{50}$, $\mathbb{P}(\{\mathsf{c}\}) = \frac{8}{50}$, $\mathbb{P}(\{\mathsf{d}\}) = \frac{9}{50}$, $\mathbb{P}(\{\mathsf{e}\}) = \frac{12}{50}$ und $\mathbb{P}(\{\mathsf{f}\}) = \frac{13}{50}$. Die mittlere Codewortlänge des Huffman-Codes ergibt sich zu 2,48. Der Shannon-Code für diese Quelle hat eine mittlere Codewortlänge von 3,02. Die Entropie der Quelle ist $\mathsf{H}_2(\mathbb{P}) = 2{,}443$.

6 Markov-Ketten

Es werden nun am Beispiel der Markov-Ketten stochastische Modelle betrachtet, die zusätzlich eine zeitliche Dynamik enthalten.

19 Die Markovsche Eigenschaft

An einem Skilift starten zu den Zeitpunkten $n \in \mathbb{N}_0$ Tellerbügel, die je eine Person befördern können. Zwischen den Zeitpunkten n und $n+1$ kommen Y_n neue Skifahrer am Lift an. Es sei $(Y_n)_{n\geq 0}$ eine Folge unabhängiger Zufallsvariable. Sei X_n die Länge der Warteschlange unmittelbar vor der Abfahrt des Tellerbügels zur Zeit n. In diesem Warteschlangenmodell gilt offenbar für all $n \geq 1$, dass

$$X_n = \max\{0, X_{n-1} - 1\} + Y_{n-1}.$$

Es sei $X_0 = i_0$ eine bekannte Zahl zu Beobachtungsbeginn. Da Y_n unabhängig von $Y_0, \ldots, Y_{n-1}$ ist, ist Y_n auch unabhängig von $(X_0, \ldots, X_n)$, da $(X_0, \ldots, X_n)$ eine Funktion von $(Y_0, \ldots, Y_{n-1})$ ist, vgl. Satz 5.7. Damit gilt für $i_0, i_1, \ldots, i_{n+1} \in \mathbb{N}_0$,

$$\begin{aligned}
&\mathbb{P}\left(X_{n+1} = i_{n+1}, X_n = i_n, \ldots, X_0 = i_0\right) \\
&= \mathbb{P}\left(Y_n = i_{n+1} - \max\{i_n - 1, 0\}, X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) \\
&= \mathbb{P}\left(Y_n = i_{n+1} - \max\{i_n - 1, 0\}\right) \cdot \mathbb{P}\left(X_n = i_n, \ldots, X_0 = i_0\right),
\end{aligned}$$

und es folgt

$$\mathbb{P}\left(X_{n+1} = i_{n+1} \,|\, X_n = i_n, \ldots, X_0 = i_0\right) = \mathbb{P}\left(Y_n = i_{n+1} - i_n - 1\right).$$

Die bedingte W-keit hängt also gar nicht von den Größen der Warteschlange zu den Zeitpunkten $0, 1, \ldots, n-1$ ab. Dies ist eine wesentliche, häufig auftretende Eigenschaft eines sich „zeitlich entwickelnden zufälligen Systems“.

Definition 19.1. Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein W-Raum, $T \neq \emptyset$ beliebige Indexmenge und $(S, \mathcal{S})$ ein messbarer Raum. Eine Familie $\{X_t : t \in T\}$ von ZVen mit Werten in S heißt *stochastischer Prozess* mit *Parameterbereich* T und *Zustandsraum* S.

Bemerkung 38. Im Folgenden wird S stets höchstens abzählbar sein, $\mathcal{S} = \mathcal{P}(S)$ und $T = \mathbb{N}_0$. Statt $\{X_t : t \in T\}$ schreibt man auch $(X_t)_{t\in T}$.

Definition 19.2. Eine *Markov-Kette* ist ein stochastischer Prozess $(X_n)_{n\in\mathbb{N}_0}$ mit höchstens abzählbarem Zustandsraum S, der die *Markovsche Eigenschaft* besitzt: Für alle $n \in \mathbb{N}$ und $s_0, \ldots, s_{n+1} \in S$ mit $\mathbb{P}\left(X_0 = s_0, \ldots, X_n = s_n\right) > 0$ gilt

$$\mathbb{P}\left(X_{n+1} = s_{n+1} \,|\, X_0 = s_0, \ldots, X_n = s_n\right) = \mathbb{P}\left(X_{n+1} = s_{n+1} \,|\, X_n = s_n\right).$$

Bemerkung 39. Folgende Interpretation ist nützlich: X_n beschreibt den Zustand eines Systems zur Zeit n. Die Markovsche Eigenschaft bedeutet, dass die W-keit, zur Zeit $n+1$ in einen beliebigen Zustand s_{n+1} zu gelangen, nur vom Zustand s_n zur Zeit n (und n) abhängt, nicht aber von den Zuständen, in welchen sich das System früher befand.

Lemma 19.3. Sei $(X_n)_{n\geq 0}$ eine Markov-Kette. Für alle $n \in \mathbb{N}$ und $s_0, \dots, s_n \in S$ gilt:

$$\mathbb{P}(X_0 = s_0, \dots, X_n = s_n) = \mathbb{P}(X_0 = s_0)\,\mathbb{P}(X_1 = s_1 | X_0 = s_0) \cdots \mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1}).$$

Beweis.
Es ist

$$\begin{aligned}
&\mathbb{P}(X_0 = s_0, \dots, X_n = s_n) \\
&= \mathbb{P}(X_n = s_n | X_0 = s_0, \dots, X_{n-1} = s_{n-1}) \cdot \mathbb{P}(X_0 = s_0, \dots, X_{n-1} = s_{n-1}) \\
&\overset{\text{ME}}{=} \mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1}) \cdot \mathbb{P}(X_0 = s_0, \dots, X_{n-1} = s_{n-1}) \\
&\overset{\text{Ind}}{=} \mathbb{P}(X_n = s_n) \cdot \mathbb{P}(X_{n-1} = s_{n-1} | X_{n-2} = s_{n-2}) \cdots \mathbb{P}(X_1 = s_1 | X_0 = s_0).
\end{aligned}$$

Dies ist die Behauptung. ∎

Satz 19.4. Sei $(X_n)_{n\geq 0}$ eine Markov-Kette und $0 < n < N$. Dann gilt für alle $s_n \in S$ und $E \subset S^n$, $F \subset S^{N-n}$

$$\mathbb{P}((X_{n+1}, \dots, X_N) \in F \,|\, X_n = s_n, (X_0, \dots, X_{n-1}) \in E) = \mathbb{P}((X_{n+1}, \dots, X_N) \in F \,|\, X_n = s_n).$$

Zum Beweis von Satz 19.4 zeigen wir zunächst:

Lemma 19.5. Seien $A, B, C_1, C_2, \dots$ Ereignisse, wobei $C_1, C_2, \dots$ paarweise disjunkt sind mit $\mathbb{P}(B \cap C_i) > 0$ für alle $i \in \mathbb{N}$. Die W-keiten $\mathbb{P}(A \,|\, B \cap C_i)$ seien für $i \in \mathbb{N}$ alle gleich. Dann gilt

$$\mathbb{P}(A \,|\, B \cap C_1) = \mathbb{P}\Big(A \,\Big|\, B \cap \bigcup_{i\geq 1} C_i\Big).$$

Beweis.
Es ist mit $C := \bigcup_{i\geq 1} C_i$

$$\begin{aligned}
\mathbb{P}(A \,|\, B \cap C_1) \cdot \mathbb{P}(B \cap C) &= \sum_{i\geq 1} \mathbb{P}(A \,|\, B \cap C_i) \cdot \mathbb{P}(B \cap C_i) = \sum_{i\geq 1} \mathbb{P}(A \cap B \cap C_i) \\
&= \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \,|\, B \cap C) \cdot \mathbb{P}(B \cap C).
\end{aligned}$$

Kürzen von $\mathbb{P}(B \cap C)$ liefert die Behauptung. ∎

Beweis von Satz 19.4.
Wir betrachten zunächst den Spezialfall $F = \{(s_{n+1}, \dots, s_N)\}$ mit $(s_{n+1}, \dots, s_N) \in S^{N-n}$ und bezeichnen $p_k(j|i) := \mathbb{P}(X_{k+1} = s_j | X_k = s_i)$. Für $(s_0, \dots, s_{n-1}) \in S^n$ beliebig gilt nach Lemma 19.3

$$\begin{aligned}
&\mathbb{P}((X_{n+1}, \dots, X_N) \in F \,|\, X_n = s_n, (X_0, \dots, X_{n-1}) = (s_0, \dots, s_{n-1})) \\
&= \frac{\mathbb{P}(X_0 = s_0, \dots, X_N = s_N)}{\mathbb{P}(X_0 = s_0, \dots, X_n = s_n)} = \frac{\mathbb{P}(X_0 = s_0) p_0(1|0) \cdots p_{N-1}(N|N-1)}{\mathbb{P}(X_0 = s_0) p_0(1|0) \cdots p_{n-1}(n|n-1)} \\
&= p_n(n+1|n) p_{n+1}(n+2|n+1) \cdots p_{N-1}(N|N-1) =: p.
\end{aligned}$$

Damit ist p insbesondere unabhängig von $s_0, \dots, s_{n-1}$. Nach Lemma 19.5 gilt dann für beliebige disjunkte Vereinigungen C von Mengen der Form $\{(X_0, \dots, X_{n-1}) = (s_0, \dots, s_{n-1})\}$, dass

$$\mathbb{P}\left((X_{n+1}, \dots, X_N) \in F \,|\, \{X_n = s_n\} \cap C\right) = p.$$

Mit $C = \{(X_0, \dots, X_{n-1}) \in E\}$ und $C = \Omega$ erhält man die Behauptung für den Fall $F = \{(s_{n+1}, \dots, s_N)\}$. Nun kann man für allgemeines $F \subset S^{N-n}$ aber über $(s_{n+1}, \dots, s_N) \in F$ summieren und erhält mit der σ-Additivität die Behauptung. ∎

Satz 19.6 (Chapman-Kolmogorov-Gleichung)**.** Sei $(X_n)_{n\geq 0}$ eine Markov-Kette. Dann gilt für alle $0 \leq k < m < n$ und alle $u, v \in S$:

$$\mathbb{P}\left(X_n = v \,|\, X_k = u\right) = \sum_{s\in S} \mathbb{P}\left(X_m = s \,|\, X_k = u\right) \cdot \mathbb{P}(X_n = v \,|\, X_m = s).$$

Beweis.
Es ist

$$\begin{aligned}
\mathbb{P}\left(X_k = u, X_n = v\right) &= \sum_{s\in S} \mathbb{P}\left(X_k = u, X_n = v, X_m = s\right) \\
&= \sum_{s\in S} \mathbb{P}\left(X_k = u, X_m = s\right) \cdot \mathbb{P}\left(X_n = v \,|\, X_k = u, X_m = s\right) \\
&\overset{\text{Satz 19.4}}{=} \sum_{s\in S} \mathbb{P}\left(X_k = u, X_m = s\right) \cdot \mathbb{P}\left(X_n = v \,|\, X_m = s\right).
\end{aligned}$$

Dividieren durch $\mathbb{P}(X_k = u)$ auf beiden Seiten liefert die Behauptung. ∎

Bisher hingen die W-keiten $\mathbb{P}(X_{n+1} = v \,|\, X_n = u)$ formal auch von n ab. In vielen Anwendungen trifft man auf Markov-Ketten, bei denen diese Übergangswahrscheinlichkeit unabhängig von n sind.

Definition 19.7. Eine Markov-Kette heißt *homogen* (oder Kette mit stationären Übergangswahrscheinlichkeiten), falls für alle $u, v \in S$

$$\mathbb{P}\left(X_{n+1} = v \,|\, X_n = u\right) =: p_{uv}$$

unabhängig von n ist.

Bemerkung 40. Mit p_{uv} wie in der vorigen Definition ist $P := (p_{uv})_{u,v\in S}$ eine stochastische Matrix, d.h. es gilt für alle $u, v \in S$, dass $p_{uv} \geq 0$ und für alle $u \in S$ gilt $\sum_{v\in S} p_{uv} = 1$.

Definition 19.8. Sei $(X_n)_{n\geq 0}$ eine homogene Markov-Kette. Dann heißt $P = (p_{uv})_{u,v\in S}$ *Matrix der Übergangswahrscheinlichkeiten* und die Verteilung $\pi = \mathbb{P}_{X_0}$ *Startverteilung.* Ferner bezeichnet $\pi_u := \pi(\{u\})$ für $u \in S$.

Bemerkung 41. Durch Startverteilung und Matrix der Übergangswahrscheinlichkeiten sind die gemeinsamen Verteilungen der X_n festgelegt: Für alle $(s_0, \dots, s_n) \in S^{n+1}$ gilt

$$\mathbb{P}\left((X_0, \dots, X_n) = (s_0, \dots, s_n)\right) = \pi_{s_0} p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{n-1},s_n}. \tag{35}$$

Beispiel 19.9. Im vorigen Beispiel der Warteschlange am Tellerlift ist für unabhängige $(Y_n)_{n\geq 0}$ die Länge der Warteschlange zu den einzelnen Zeitpunkten gegeben durch

$$X_n = \max\{0, X_{n-1} - 1\} + Y_{n-1}.$$

Wir haben

$$\begin{aligned}\mathbb{P}\left(X_{n+1} = i_{n+1} \,|\, X_n = i_n, \dots, X_0 = i_0\right) &= \mathbb{P}\left(Y_n = i_{n+1} - \max\{i_n - 1, 0\}\right) \\ &= \mathbb{P}\left(X_{n+1} = i_{n+1} \,|\, X_n = i_n\right).\end{aligned}$$

$(X_n)_{n\geq 1}$ ist also eine Markov-Kette. Sie ist homogen, falls $Y_0, Y_1, \dots$ identisch verteilt sind. Es ist dann

$$p_{ij} = \mathbb{P}\left(Y_n = j - \max\{i - 1, 0\}\right)$$

unabhängig von n.

Häufig wird die Verteilung einer homogenen Markov-Kette direkt durch Angabe der Startverteilung und Übergangsmatrix (Matrix der Übergangswahrscheinlichkeiten) definiert:

Beispiel 19.10 (Einfache symmetrische Irrfahrt auf $\mathbb{Z}^d$)**.** Ein Teilchen bewege sich in jedem Zeitabschnitt $n \to n+1$ von seinem Ort $x \in \mathbb{Z}^d$ auf dem Gitter $\mathbb{Z}^d$ mit gleicher Wahrscheinlichkeit zu einem der $2d$ benachbarten Punkte. Der Zustandsraum ist $S = \mathbb{Z}^d$, und für $x, y \in \mathbb{Z}^d$ ist

$$p_{xy} = \begin{cases} \frac{1}{2d}, & \text{falls } \|x - y\| = 1, \\ 0, & \text{sonst.} \end{cases}$$

(Hier bezeichnet $\|\cdot\|$ die euklidische Norm.) Gibt man zudem eine Startverteilung π vor, so ist die Verteilung der Irrfahrt nach (35) festgelegt. Man nennt diese Markov-Kette die *einfache symmetrische Irrfahrt* auf $\mathbb{Z}^d$.

Definition 19.11. Für eine homogene Markov-Kette $(X_n)_{n\geq 0}$ heißen

$$p_{uv}^{(m)} := \mathbb{P}\left(X_{n+m} = v \,|\, X_n = u\right)$$

die m-Schritt-Übergangswahrscheinlichkeiten von u nach v.

Bemerkung 42. Die m-Schritt-Übergangswahrscheinlichkeiten hängen nicht von n ab. Für $m = 1$ ist dies gerade die Definition, für $m \geq 2$ folgt dies induktiv aus der Chapman-Kolmogorov-Gleichung. Dies lautet im homogenen Fall gerade

$$p_{uv}^{(\ell+m)} = \sum_{s\in S} p_{us}^{(\ell)} p_{sv}^{(m)}.$$

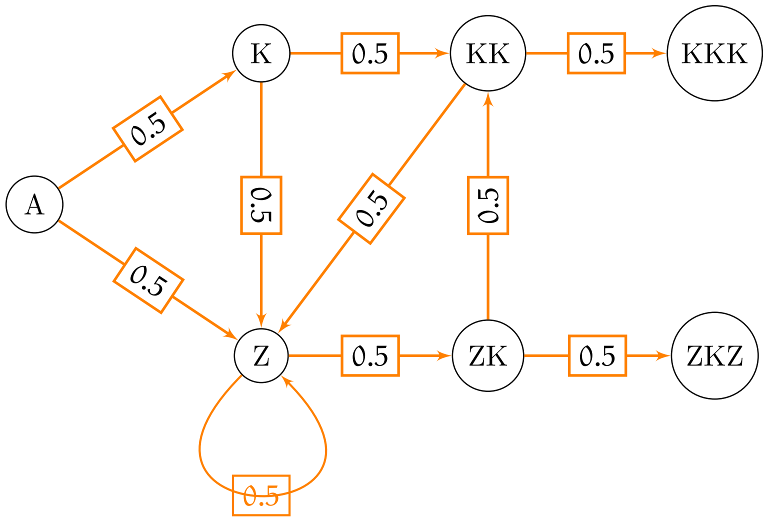
Bezeichnet P die Übergangsmatrix und $P^{(m)}$ die entsprechende Matrix der m-Schritt-Übergangswahrscheinlichkeiten, so folgt mit Induktion, dass

$$P^{(m)} = P^m,$$

wobei rechts die m-te Potenz der Matrix P steht (im Sinne des Matrizenprodukts).

20 Absorptionswahrscheinlichkeiten

Es werde eine faire Münze solange geworfen, bis entweder dreimal Kopf oder „Zahl, Kopf, Zahl" in Serie gefallen ist. Spieler A gewinnt im ersten Fall KKK, Spieler B im zweiten Fall ZKZ. Offenbar muss man sich, während das Spiel noch läuft, stets nur den letzten oder die beiden letzten Würfe merken, um zu entscheiden, wer das Spiel gewinnt. Es interessieren die Wahrscheinlichkeiten, dass Spieler A bzw. Spieler B gewinnt. Da für den Spielverlauf stets nur die letzten beiden Würfe (bzw. der letzte Wurf) relevant sind, kann man wie folgt modellieren:



Wir nehmen als Zustandsraum $S = \{A, K, Z, KK, ZK, KKK, ZKZ\}$ und Übergangswahrscheinlichkeit $1/2$ in Richtung der Pfeile und zudem Wahrscheinlichkeit 1 vom Zustand KKK und ZKZ jeweils zu sich selbst. Starten wir eine Markov-Kette $(X_n)_{n\geq 0}$ mit Zustandsraum S, diesen Übergangswahrscheinlichkeiten und $X_0 = A$ (d.h., Startverteilung $\pi_A = 1$ und $\pi_s = 0$ für $s \in S \setminus \{A\}$), so gewinnt offenbar Spieler A gerade, falls die Kette den Zustand KKK erreicht (und dort dann bleibt), Spieler B, falls die Kette im Zustand ZKZ „terminiert". Dieses Fragestellung führt auf die allgemeine Problematik der Trefferwahrscheinlichkeiten von Markov-Ketten.

Wir betrachten im Folgenden homogene Markov-Ketten $(X_n)_{n\geq 0}$ mit Übergangsmatrix $(p_{uv})_{u,v\in S}$. Die Kette kann in verschiedenen Zuständen gestartet werden, d.h. wir betrachten verschiedene Startverteilungen π für eine Kette mit derselben Übergangsmatrix $(p_{uv})_{u,v\in S}$. Man schreibt dann $\mathbb{P}_x(\,\cdot\,)$ und $\mathbb{E}_x[\,\cdot\,]$ falls die Kette im Zustand $x \in S$ gestartet wird und allgemeiner $\mathbb{P}_\pi(\cdot)$ und $\mathbb{E}_\pi[\cdot]$, falls die Kette mit Startverteilung π gestartet wird.

Definition 20.1. Sei $(X_n)_{n\geq 0}$ eine homogene Markov-Kette mit Übergangsmatrix $(p_{uv})_{u,v\in S}$. Sei $\emptyset \neq B \subset S$ eine Teilmenge des Zustandsraumes S. Dann heißt

$$M := \min\{n \geq 0 : X_n \in B\}$$

der *Zeitpunkt des ersten Eintritts* in B. Die Wahrscheinlichkeiten

$$w(x) = \mathbb{P}_x\left(M < \infty, X_M = z\right),$$

von Startpunkt x aus bei Eintritt in B den Zustand $z \in B$ zu erreichen, heißen *Trefferwahrscheinlichkeiten.*

Beispiel 20.2. Im vorigen Beispiel lassen sich die Wahrscheinlichkeiten, dass Spieler A bzw. Spieler B gewinnt, wie folgt in diesem Rahmen einordnen: Wir starten die Markov-Kette im Zustand $x = A$, wählen $B = \{KKK, ZKZ\}$ und betrachten dann $z = KKK$ bzw. $z = ZKZ$. Die Trefferwahrscheinlichkeiten sind dann gerade die gesuchten Wahrscheinlichkeiten.

Lemma 20.3. Sei $(X_n)_{n\geq 0}$ eine homogene Markov-Kette mit Übergangsmatrix $(p_{uv})_{u,v\in S}$. Seien $\emptyset \neq B \subset S$ und $z \in B$. Dann gilt für die Trefferwahrscheinlichkeiten

$$w(x) = \begin{cases} 1, & \text{falls } x = z \\ 0, & \text{falls } x \in B \setminus \{z\} \\ \sum_{y\in S} p_{xy} w(y), & \text{falls } x \in S \setminus B. \end{cases} \tag{36}$$

Beweis.
Für $x \in B$ ist die Behauptung klar. Für $x \notin B$ führen wir eine „Zerlegung nach dem ersten Schritt“ durch: Es gilt

$$w(x) = \mathbb{P}_x\left(M < \infty, X_M = z\right) = \sum_{y\in S} \mathbb{P}_x(X_1 = y)\mathbb{P}_x\left(M < \infty, X_M = z | X_1 = y\right) \tag{37}$$

nach dem Satz von der totalen W-keit, und es ist $\mathbb{P}_x(X_1 = y) = p_{xy}$. Wir unterscheiden nun die Fälle $y \in B$ und $y \in S \setminus B$. Im Falle $y \in B$ ist wegen $x \in S \setminus B$

$$\mathbb{P}_x(M < \infty, X_M = z \,|\, X_1 = y) = \mathbb{1}_{\{z\}}(y) = w(y) \quad \text{für alle } \; y \in B.$$

Sei nun $y \in S \setminus B$. Wegen $x, y \in S \setminus B$ gilt dann $M \geq 2$. Den zweiten Faktor in (37) schreiben wir mit der σ-Additivität, angewandt auf $\{M < \infty\} = \bigcup_{\ell \geq 0} \{M = \ell\}$, um:

$$\begin{aligned}
\mathbb{P}_x(M < \infty, X_M = z | X_1 = y) &= \sum_{\ell=2}^{\infty} \mathbb{P}_x(M = \ell, X_M = z | X_1 = y) \\
&= \sum_{\ell=2}^{\infty} \mathbb{P}_x\left(X_0, X_1, \ldots, X_{\ell-1} \notin B, X_\ell = z | X_1 = y\right) \\
&\overset{\text{ME}}{=} \sum_{\ell=2}^{\infty} \mathbb{P}_y\left(X_0, \ldots, X_{\ell-2} \notin B, X_{\ell-1} = z\right) \\
&= \sum_{\ell=1}^{\infty} \mathbb{P}_y\left(X_0, \ldots, X_{\ell-1} \notin B, X_\ell = z\right) \\
&= \sum_{\ell=0}^{\infty} \mathbb{P}_y\left(M = \ell, X_M = z\right) \qquad (38) \\
&= \mathbb{P}_y\left(M < \infty, X_M = z\right) = w(y),
\end{aligned}$$

wobei in (38) verwendet wurde, dass unter $X_0 = y \notin B$ dann $M \geq 1$ gilt. Zusammen folgt also $w(x) = \sum_{y \in S} p_{xy} w(y)$. ∎

Bemerkung 43. Gezeigt ist mit Lemma 20.3, dass die Trefferwahrscheinlichkeiten $w(x)$, die Bedingungen (36) erfüllen. Sie werden durch (36) i.A. allerdings nicht eindeutig bestimmt. Die Trefferwahrscheinlichkeiten sind charakterisiert dadurch, dass sie minimal unter allen nichtnegativen Lösungen von (36) sind.

Satz 20.4. Sei $f : S \to \mathbb{R}_0^+$ eine nichtnegative Funktion mit $f(z) = 1$, $f(x) = 0$ für $x \in B \setminus \{z\}$ und

$$\sum_{y \in S} p_{xy} f(y) \leq f(x)$$

für alle $x \notin B$. Dann gilt $w(x) \leq f(x)$ für alle $x \in S$.

Beweis.
Für $x \in B$ ist die Behauptung klar. Für $x \notin B$ zeigen wir zunächst $\mathbb{P}_x(X_M = z, M \leq \ell) \leq f(x)$ durch Induktion nach ℓ. Der Induktionsanfang $\ell = 0$ ist wahr, da $\mathbb{P}_x(X_M = z, M \leq 0) = 0$ gilt für $x \notin B$. Für den Induktionsschrit $\ell \to \ell + 1$ führen wir wieder eine Zerlegung nach dem ersten Schritt durch: Es ist

$$\begin{aligned}
\mathbb{P}_x\left(X_M = z, M \leq \ell + 1\right) &= \sum_{y \in S} \mathbb{P}_x(X_1 = y) \mathbb{P}_x\left(X_M = z, M \leq \ell + 1 | X_1 = y\right) \\
&= \sum_{y \in S} p_{xy} \mathbb{P}_y\left(X_M = z, M \leq \ell\right) \\
&\overset{\text{IV}}{\leq} \sum_{y \in S} p_{xy} f(y) \leq f(x).
\end{aligned}$$

Mit der Stetigkeit von unten folgt

$$\mathbb{P}_x(X_M = z, M \le \ell) \uparrow \mathbb{P}_x(X_M = z, M < \infty) = w(x),$$

also $w(x) \le f(x)$. ∎

Bemerkung 44. Die eingangs gestellte Frage kann nun gelöst werden. Es ist leicht zu finden, dass

$$\mathbb{P}(\text{Spieler A gewinnt}) = \frac{5}{12}, \quad \mathbb{P}(\text{Spieler B gewinnt}) = \frac{7}{12}.$$

Beispiel 20.5 (Gambler's ruin)**.** Eine Spielerin habe Kapital $K \in \mathbb{N}$ mit $1 \le K < a$ und $a \in \mathbb{N}$. Sie spiele ein Spiel in Runden. In jeder Runde gewinnt die Spielerin mit Wahrscheinlichkeit $p \in (0,1)$ den Einsatz 1, mit Wahrscheinlichkeit $q = 1 - p$ verliert sie den Einsatz. Sie hört auf, falls sie ihr Kapital verspielt hat oder Kapital a erreicht hat. Gesucht ist die Ruinwahrscheinlichkeit.
Wir modellieren dies mit einer Markov-Kette: Der Zustandsraum ist $S = \{0, 1, \dots, a\}$. Die Übergangswahrscheinlichkeiten sind für $x \in \{1, \dots, a-1\}$

$$p_{xy} = \begin{cases} p, & \text{falls } y = x+1, \\ q, & \text{falls } y = x-1, \\ 0, & \text{sonst.} \end{cases}$$

Ferner wird $p_{00} = p_{aa} = 1$ gesetzt. Die gesuchte Wahrscheinlichkeit ergibt sich als Trefferwahrscheinlichkeit wie folgt: Wir wählen $B = \{0, a\}$, $z = 0$. Dann ist die Trefferwahrscheinlichkeit $w(K)$ die gesuchte Ruinwahrscheinlichkeit. Nach Lemma 20.3 gilt für $1 \le x \le a-1$

$$w(x) = p \cdot w(x+1) + q \cdot w(x-1) \tag{39}$$

sowie $w(0) = 1$, $w(a) = 0$. Das Gleichungssystem (39) lässt sich lösen:

$$w(x) = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^a - 1} \text{ für } p \neq \frac{1}{2}, \quad w(x) = \frac{a-x}{a} \text{ für } p = \frac{1}{2}.$$

21 Rekurrenz und Transienz

Die Zustände einer Markov-Kette unterscheidet man danach, ob sie im Verlauf der Zeit immer wieder besucht werden, oder, ob es einen letzten Zeitpunkt gibt, nach dem die Kette nicht mehr zum Zustand zurückkehrt. Im Folgenden sei stets $(X_n)_{n \ge 0}$ eine homogene Markov-Kette mit Zustandsraum S. Bei Start in $x \in S$ bezeichne

$$T_x := \min\{n \ge 1 \,|\, X_n = x\}$$

den *Zeitpunkt der ersten Rückkehr* nach x. Wie üblich wird $\min \emptyset := \infty$ vereinbart.

Definition 21.1. Ein Zustand $x \in S$ heißt *rekurrent*, falls $\mathbb{P}_x(T_x < \infty) = 1$. Ein Zustand $x \in S$ heißt *transient*, falls $\mathbb{P}_x(T_x < \infty) < 1$. Eine Markov-Kette heißt rekurrent (bzw. transient), falls alle ihre Zustände rekurrent (bzw. transient) sind.

Satz 21.2. Für einen transienten Zustand $x \in S$ gilt bei beliebiger Startverteilung π, dass

$$\mathbb{P}_\pi(X_n = x \text{ für unendlich viele } n) = 0.$$

Für einen rekurrenten Zustand $x \in S$ gilt

$$\mathbb{P}_x(X_n = x \text{ für unendlich viele } n) = 1.$$

Beweis.
Sei $C_x = |\{n \geq 1 : X_n = x\}|$ die Anzahl der Besuche in x. Mit der Markov-Eigenschaft gilt für $m \geq 1$

$$\begin{aligned}
\mathbb{P}_\pi(C_x \geq m) &= \sum_{\ell=1}^{\infty} \mathbb{P}_\pi(X_1, \ldots, X_{\ell-1} \neq x, X_\ell = x, X_n = x \text{ noch } \geq m-1 \text{ mal für } n > \ell) \\
&= \sum_{\ell=1}^{\infty} \mathbb{P}_\pi(X_1, \ldots, X_{\ell-1} \neq x, X_\ell = x) \\
&\qquad \times \mathbb{P}_\pi(X_n = x \text{ noch } \geq m-1 \text{ mal für } n > \ell \,|\, X_1, \ldots, X_{\ell-1} \neq x, X_\ell = x) \\
&\overset{\text{ME}}{=} \mathbb{P}_\pi(C_x \geq 1)\, \mathbb{P}_x(C_x \geq m-1).
\end{aligned}$$

Iteration des Arguments liefert

$$\begin{aligned}
\mathbb{P}_\pi(C_x \geq m) &= \mathbb{P}_\pi(C_x \geq 1) \cdot (\mathbb{P}_x(C_x \geq 1))^{m-1} \\
&= \mathbb{P}_\pi(T_x < \infty) \cdot (\mathbb{P}_x(T_x < \infty))^{m-1}. \qquad (40)
\end{aligned}$$

Falls nun x transient ist, so folgt mit $m \to \infty$, dass $\mathbb{P}_\pi(C_x = \infty) = 0$. Falls andererseits x rekurrent ist, so folgt aus (40) mit $\pi = \delta_x$, dem Dirac-Maß in x, (d.h. es werde in x gestartet), dass

$$\mathbb{P}_x(C_x \geq m) = \mathbb{P}_x(T_x < \infty)^m = 1, \quad m \geq 1.$$

Es gilt $\{C_x \geq m\} \downarrow \{C_x = \infty\}$. Die Stetigkeit von oben liefert dann

$$\mathbb{P}_x(X_n = x \text{ für unendlich viele } n) = \mathbb{P}_x(C_x = \infty) = 1.$$

Dies ist die Behauptung. ∎

Satz 21.3. Ein Zustand $x \in S$ ist transient genau dann, wenn

$$\sum_{n \geq 1} \mathbb{P}_x(X_n = x) < \infty.$$

Beweis.
„$\Leftarrow$“: Es ist $\{X_n = x \text{ für unendlich viele } n\} = \limsup_{n\to\infty}\{X_n = x\}$. Die Endlichkeit der Reihe liefert mit dem Lemma von Borel-Cantelli 11.8 a), dass

$$\mathbb{P}_x(X_n = x \text{ für unendlich viele } n) = 0.$$

Nach Satz 21.2 ist x nicht rekurrent. Nach Definition 21.1 ist x somit transient.
„$\Rightarrow$“: Zunächst muss beachtet werden, dass Satz 11.8 b) nicht auf $\{X_n = x\}$ angewandt werden kann, da diese Ereignisse im Allgemeinen nicht unabhängig sind. Nach dem Beweis von Satz 21.2 gilt

$$\mathbb{P}_x(C_x = m) = \mathbb{P}_x(C_x \geq m) - \mathbb{P}_x(C_x \geq m+1) = q^m - q^{m+1} = q^m(1-q)$$

mit $q := \mathbb{P}_x(T_x < \infty) < 1$, da x transient ist. Bei Start in x ist C_x also geometrisch verteilt zum Parameter $1 - q > 0$. Andererseits ist $\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = x\}} = C_x$. Damit folgt nun

$$\sum_{n=1}^{\infty} \mathbb{P}_x(X_n = x) = \mathbb{E}_x[C_x] = \frac{1}{1-q} < \infty.$$

∎

Dies ist die Behauptung.

Als Anwendung betrachten wir die einfache symmetrische Irrfahrt aus Abschnitt 19.

Satz 21.4. Die einfache symmetrische Irrfahrt ist rekurrent für Dimensionen $d = 1, 2$ und transient für $d \geq 3$.

Beweis.
Wir bestimmen die W-keiten $\mathbb{P}_x(X_n = x)$. Offenbar kann man nur in einer geraden Anzahl von Schritten vom Zustand $x \in \mathbb{Z}^d$ zurück nach x kommen. Geht man n_i Schritte in positive Richtung der i-ten Koordinate, so muss man auch n_i Schritte in die entgegengesetzte Richtung gehen. Damit folgt

$$\begin{aligned} \mathbb{P}_x(X_{2n} = x) &= \sum_{n_1+\cdots+n_d=n} \binom{2n}{n_1, n_1, n_2, n_2, \ldots, n_d, n_d} (2d)^{-2n} \\ &= \binom{2n}{n} \sum_{n_1+\cdots+n_d=n} \binom{n}{n_1, \ldots, n_d}^2 (2d)^{-2n}. \end{aligned} \tag{41}$$

Für $d = 1$ liefert dies mit der Stirlingschen Formel asymptotisch für $n \to \infty$

$$\mathbb{P}_x(X_{2n} = x) = \binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}}.$$

Da die Reihe $\sum_{n\geq 1} 1/\sqrt{n}$ divergiert, folgt aus Satz 21.3 die Rekurrenz der Irrfahrt.
Für $d = 2$ folgt aus (41)

$$\mathbb{P}_x(X_{2n} = x) = \binom{2n}{n} \sum_{j=0}^{n} \binom{n}{j}\binom{n}{n-j} 4^{-2n} = \binom{2n}{n}^2 4^{-2n} \sim \frac{1}{\pi n}.$$

Die zugehörige Reihe in Satz 21.3 ist wieder divergent, die Irrfahrt also rekurrent. (Für $\sum_{j=0}^{n}\binom{n}{j}\binom{n}{n-j}=\binom{2n}{n}$ vgl. hypergeometrische Verteilung.)
Für $d \geq 3$ betrachten wir zunächst die Multinomialkoeffizienten $\binom{n}{n_1,\ldots,n_d}$. Es ist leicht zu sehen, dass für $m := \lceil n/d \rceil$ gilt

$$\binom{n}{n_1,\ldots,n_d} \leq \binom{dm}{m,\ldots,m}.$$

Damit folgt aus (41)

$$\begin{aligned}\mathbb{P}_x(X_{2n}=x) \leq& \binom{2n}{n}2^{-2n}\binom{dm}{m,\ldots,m}d^{-n}\underbrace{\sum_{n_1+\cdots+n_d=n}\binom{n}{n_1,\ldots,n_d}d^{-n}}_{=1}\\ =&\binom{2n}{n}2^{-2n}\binom{dm}{m,\ldots,m}d^{-n} \sim \left(\frac{2dm}{n}\right)^{1/2}(2\pi m)^{-d/2}\, d^{dm-n}.\end{aligned}$$

Wegen $dm \leq n+d$ liefert dies $\mathbb{P}_x(X_{2n}=x) = O(n^{-d/2})$ für $n \to \infty$. Damit konvergiert die Reihe in Satz 21.3 für jedes $d \geq 3$. Die Irrfahrt ist also transient für alle $d \geq 3$. ∎

22 Stationäre Verteilungen von Markov-Ketten

In diesem Abschnitt betrachten wir homogene Markov-Ketten $(X_n)_{n\geq 0}$ mit endlichem Zustandsraum S und Übergangsmatrix $P = (p_{uv})_{u,v\in S}$. Wir studieren die Verteilung $\mathbb{P}_{X_n}$ für große n.

Definition 22.1. Sei $(X_n)_{n\geq 0}$ eine homogene Markov-Kette mit Übergangsmatrix $P = (p_{uv})_{u,v\in S}$. Eine Verteilung π auf S heißt *stationäre Verteilung* (oder auch *Gleichgewichtsverteilung*) für $(X_n)_{n\geq 0}$, falls gilt

$$\pi_x = \sum_{y\in S}\pi_y p_{yx} \quad \text{für alle } x \in S.$$

Bemerkung 45. Wählt man als Startverteilung eine Gleichgewichtsverteilung π, so gilt $\mathbb{P}_{X_n} = \pi$ für alle $n \in \mathbb{N}_0$.

Beweis.
Für $x \in S$ folgt mit dem Satz von der totalen Wahrscheinlichkeit

$$\mathbb{P}_\pi(X_1=x) = \sum_{y\in S}\mathbb{P}_\pi(X_1=x\,|\,X_0=y)\mathbb{P}_\pi(X_0=y) = \sum_{y\in S}p_{yx}\pi_y = \pi_x,$$

also $\mathbb{P}_{X_1} = \pi$. Mit Induktion nach n folgt die Behauptung. ∎

Eine Gleichgewichtsverteilung braucht i.A. nicht zu existieren. Wir schränken uns deshalb später auf eine wichtige spezielle Klasse von Markov-Ketten ein.

Definition 22.2. Sei $(X_n)_{n\geq 0}$ eine homogene Markov-Kette.

a) Wir schreiben $x \rightsquigarrow y$, falls ein $m \in \mathbb{N}$ existiert mit $\mathbb{P}_x(X_m = y) > 0$. Der Zustand x *kommuniziert* mit y (Bez. $x \leftrightsquigarrow y$), falls $x \rightsquigarrow y$ und $y \rightsquigarrow x$.

b) $(X_n)_{n\geq 0}$ heißt *irreduzibel*, falls $x \leftrightsquigarrow y$ für alle $x, y \in S$, andernfalls reduzibel.

c) Für $x \in S$ heißt $d(x) = \text{ggT}\{n \geq 1 : \mathbb{P}_x(X_n = x) > 0\}$ *Periode* von x.

d) $(X_n)_{n\geq 0}$ heißt *aperiodisch*, falls $d(x) = 1$ für alle $x \in S$.

e) $(X_n)_{n\geq 0}$ heißt *ergodisch*, falls $(X_n)_{n\geq 0}$ irreduzibel und aperiodisch ist.

Satz 22.3. Sei $(X_n)_{n\geq 0}$ eine ergodische Markov-Kette mit endlichem Zustandsraum S. Dann existiert ein $M \in \mathbb{N}$, sodass für alle $x, y \in S$ und $n \geq M$ gilt:

$$\mathbb{P}_y(X_n = x) > 0.$$

Der Beweis benutzt das folgende zahlentheoretische Lemma.

Lemma 22.4. Sei $A = \{a_1, a_2, \ldots\} \subset \mathbb{N}$ mit $\text{ggT}(a_1, a_2, \ldots) = 1$ und abgeschlossen bez. Addition, d.h. für $a, a' \in A$ gilt $a + a' \in A$. Dann existiert ein $N < \infty$ mit $n \in A$ für alle $n \geq N$. (ohne Beweis)

Beweis von Satz 22.3.
Zu $x \in S$ sei $A_x = \{n \geq 1 \,|\, \mathbb{P}_x(X_n = x) > 0\}$. Die Menge A_x erfüllt die Voraussetzungen von Lemma 22.4: $\text{ggT}(A_x) = 1$, da $(X_n)_{n\geq 0}$ aperiodisch ist, und für $a, a' \in A_x$ gilt mit der Chapman-Kolmogorov Gleichung

$$\begin{aligned}
\mathbb{P}_x(X_{a+a'} = x) &= \sum_{y\in S} \mathbb{P}(X_a = y|X_0 = x)\mathbb{P}(X_{a+a'} = x|X_a = y)\\
&\geq \mathbb{P}(X_a = x|X_0 = x)\mathbb{P}(X_{a+a'} = x|X_a = x)\\
&= \mathbb{P}_x(X_a = x)\mathbb{P}_x(X_{a'} = x) > 0,
\end{aligned}$$

also $a + a' \in A_x$. Nach Lemma 22.4 existiert ein $N_x \in \mathbb{N}$ mit $\mathbb{P}_x(X_n = x) > 0$ für alle $n \geq N_x$. Seien nun $x, y \in S$ beliebig. Wegen der Irreduzibilität existiert ein $m_{xy} \in \mathbb{N}$ mit $\mathbb{P}_x(X_{m_{xy}} = y) > 0$. Für $m \geq M := \max\{N_x + m_{xy} \,|\, x, y \in S\}$ gilt dann

$$\begin{aligned}
\mathbb{P}_y(X_m = x) &= \sum_{s\in S} \mathbb{P}(X_{m-m_{xy}} = s \,|\, X_0 = y)\mathbb{P}(X_m = x \,|\, X_{m-m_{xy}} = s)\\
&\geq \underbrace{\mathbb{P}\left(X_{m-m_{xy}} = y \,|\, X_0 = y\right)}_{>0,\text{ da } m-m_{xy}\geq N_y} \cdot \underbrace{\mathbb{P}\left(X_m = x \,|\, X_{m-m_{xy}} = y\right)}_{=\mathbb{P}_y(X_{m_{xy}}=x)>0} > 0.
\end{aligned}$$

Es ist $M < \infty$, da S endlich ist. ∎

Satz 22.5. Jede ergodische Markov-Kette $(X_n)_{n\geq 0}$ mit endlichem Zustandsraum hat eine stationäre Verteilung.

Beweis.
Kann elementar geführt werden, wird hier aber ausgelassen. ∎

Satz 22.6 (Ergodensatz für Markov-Ketten)**.** Sei $(X_n)_{n\geq 0}$ eine ergodische Markov-Kette mit endlichem Zustandsraum S und Startverteilung μ. Sei π eine stationäre Verteilung für $(X_n)_{n\geq 0}$. Für $n \to \infty$ gilt dann

$$d_{TV}(\mathbb{P}_{X_n}, \pi) \to 0.$$

Beweis.
Wir wählen X_0 mit $\mathbb{P}_{X_0} = \mu$ und konstruieren die Markov-Kette $X_1, X_2, \ldots$ dann explizit wie in einer Übungsaufgabe: Sei $(U_n)_{n\geq 1}$ eine Folge unabhängiger, auf $[0,1]$ gleichverteilter ZVe. Dann kann $X_n = f(X_{n-1}, U_n)$ gewählt werden mit einer deterministischen Funktion f (vgl. Übungsaufgabe 40). Zudem konstruieren wir eine Markov-Kette $(X'_n)_{n\geq 0}$ mit Startverteilung $\mathbb{P}_{X'_0} = \pi$ und $X'_n = f(X'_{n-1}, U'_n)$ mit einer zweiten Folge $(U'_n)_{n\geq 1}$ unabhängiger, auf $[0,1]$ gleichverteilter ZVe ebenfalls unabhängig von $(U_n)_{n\geq 1}$. Zudem sei X'_0 unabhängig von X_0. Nach Bemerkung 45 gilt nun $\mathbb{P}_{X'_n} = \pi$ für alle $n \in \mathbb{N}_0$. Wir betrachten den Zeitpunkt, zu dem sich die Ketten erstmals treffen:

$$T = \min\{n \geq 0 \,|\, X_n = X'_n\}.$$

Nach Satz 22.3 existiert ein $M \in \mathbb{N}$, sodass für alle $x, y \in S$ gilt $\mathbb{P}_x(X_M = y) > 0$. Zu festem $y \in S$ sei

$$\alpha := \min_{x\in S} \mathbb{P}_x(X_M = y) > 0.$$

Damit folgt

$$\begin{aligned}
\mathbb{P}(T \leq M) &\geq \mathbb{P}(X_M = X'_M) \geq \mathbb{P}(X_M = y, X'_M = y) = \mathbb{P}(X_M = y) \cdot \mathbb{P}(X'_M = y) \\
&= \left(\sum_{x\in S} \mathbb{P}(X_M = y, X_0 = x)\right) \cdot \left(\sum_{x\in S} \mathbb{P}(X'_M = y, X'_0 = x)\right) \\
&= \left(\sum_{x\in S} \mathbb{P}(X_0 = x)\mathbb{P}_x(X_M = y)\right) \cdot \left(\sum_{x\in S} \mathbb{P}(X'_0 = x)\mathbb{P}_x(X'_M = y)\right) \\
&\geq \left(\sum_{x\in S} \mathbb{P}(X_0 = x) \cdot \alpha\right) \left(\sum_{x\in S} \mathbb{P}(X'_0 = x) \cdot \alpha\right) = \alpha^2.
\end{aligned}$$

Es folgt also $\mathbb{P}(T > M) \leq 1 - \alpha^2$. Mit demselben Argument folgt nun

$$\begin{aligned}
\mathbb{P}(T > 2M) &= \mathbb{P}(T > M)\mathbb{P}(T > 2M \,|\, T > M) \leq (1-\alpha^2)\mathbb{P}(X_{2M} \neq X'_{2M} \,|\, T > M) \\
&\leq (1-\alpha^2)^2.
\end{aligned}$$

Mit Induktion folgt

$$\mathbb{P}(T > \ell M) \leq (1-\alpha^2)^\ell \overset{\ell\to\infty}{\longrightarrow} 0, \quad \text{also} \quad \lim_{n\to\infty} \mathbb{P}(T > n) = 0. \tag{42}$$

Wir definieren nun eine dritte Markov-Kette durch $X''_0 = X_0$ und

$$X''_{n+1} := \begin{cases} f(X''_n, U_{n+1}), & \text{falls } X''_n \neq X'_n, \\ f(X''_n, U'_{n+1}), & \text{falls } X''_n = X'_n. \end{cases}$$

Die Folge $(X''_n)_{n\geq 0}$ ist also identisch mit $(X_n)_{n\geq 0}$ bis sich X_n und X'_n treffen. Danach ist $(X''_n)_{n\geq 0}$ identisch mit $(X'_n)_{n\geq 0}$. Nach Konstruktion ist $(X''_n)_{n\geq 0}$ ebenfalls eine Markov-Kette mit Übergangsmatrix wie $(X_n)_{n\geq 0}$ und $(X'_n)_{n\geq 0}$ und Startverteilung $\mathbb{P}_{X''_0} = \mathbb{P}_{X_0} = \mu$. (Begründung: Die Übergänge für X''_n werden mit einer Folge unabhängiger, uniform auf $[0,1]$ verteilter ZVen, die as (U_n) und (U'_n) ausgewählt werden, mittels der Funktion f konstruiert.) Insgesamt haben wir nun

- $\mathbb{P}_{X''_n} = \mathbb{P}_{X_n}$ für alle $n \in \mathbb{N}_0$,
- $\mathbb{P}_{X'_n} = \pi$ für alle $n \in \mathbb{N}_0$,
- $\{X''_n \neq X'_n\} \subset \{T > n\}$.

Lemma 13.2 liefert nun mit (42)

$$d_{TV}(\mathbb{P}_{X_n}, \pi) = d_{TV}(\mathbb{P}_{X''_n}, \mathbb{P}_{X'_n}) \leq 2\mathbb{P}(X''_n \neq X'_n) \leq 2\mathbb{P}(T > n) \xrightarrow{n\to\infty} 0.$$

Dies ist die Behauptung. ∎

Bemerkung 46. Die Idee des vorigen Beweises, die von Wolfgang Döblin stammt, besteht darin, X''_n mit Verteilung von X_n zu konstruieren und gleichzeitig an X'_n zu koppeln, d.h. Gleichheit mit hoher W-keit zu erreichen. Man spricht bei derartigen Argumenten von „Coupling“.