

Stochastik

VON PETER PFAFFELHUBER

Version: 12. Januar 2021

Inhaltsverzeichnis

0	Grundlegendes	3
0.1	Vorbemerkung	3
0.2	Wahrscheinlichkeitsräume und Zufallsvariablen	3
0.3	Häufige Modelle: Münzwurf, Würfeln, Urnen	6
1	Kombinatorik und uniform verteilte Zufallsvariablen	8
1.1	Laplace-Experimente	8
1.2	Zufällige Permutationen	9
1.3	Zufällige Teilmengen	10
1.4	Zufällige Besetzungen	11
1.5	Kontinuierlich uniform verteilte Zufallsvariable	11
2	Verteilungen und deren Eigenschaften	12
2.1	Die Verteilungsfunktion	12
2.2	Bilder von Zufallsvariablen	14
2.3	Die Ein-Ausschlussformel	15
2.4	Die gemeinsame Verteilung und Unabhängigkeit	16
3	Weitere wichtige Verteilungen	18
3.1	Die Binomial- und die Multinomialverteilung	19
3.2	Die hypergeometrische Verteilung	19
3.3	Die Poisson-Verteilung und die Poisson-Approximation	20
3.4	Die geometrische und die Exponentialverteilung	21
3.5	Die Normalverteilung	23
4	Kenngößen von Zufallsvariablen	24
4.1	Der Erwartungswert	24
4.2	Die Varianz und die Kovarianz	30
4.3	Die erzeugende Funktion und die Laplace-Transformierte	32
4.4	Kenngößen wichtiger Verteilungen	34
5	Approximationssätze	37
5.1	Summen unabhängiger Zufallsvariablen	37
5.2	Das schwache Gesetz der großen Zahlen	39
5.3	Der zentrale Grenzwertsatz	40

6	Abhängige Zufallsvariable	44
6.1	Bedingte Wahrscheinlichkeiten	44
6.2	Bedingte Erwartungen	48
6.3	Mehrstufige Experimente	50
7	Markov-Ketten	51
7.1	Grundlegendes	51
7.2	Stationäre Verteilungen	56
7.3	Markov-Ketten-Konvergenzsatz	58

0 Grundlegendes

0.1 Vorbemerkung

Die *Stochastik* (oder Wahrscheinlichkeitstheorie) hat zwei Gesichter: Einerseits fasziniert es schon seit Jahrhunderten, intuitiv und analytisch Gewinnchancen etwa bei Glücksspielen einzuschätzen und zu berechnen. Dies verleiht der Stochastik einen angewandten und intuitiven Zugang. Andererseits geht es in der Stochastik als Teilgebiet der Mathematik auch um Mengen, Funktionen und deren Zusammenhänge. Letzteres bietet – wie in jedem Teilgebiet der Mathematik – das formale Grundgerüst.

In dieser Vorlesung werden wir von beiden Aspekten Gebrauch machen, wobei die intuitiven Ansätze zumindest in dieser grundlegenden Vorlesung überwiegen werden. In Nachfolgeveranstaltungen wie etwa der Vorlesung *Wahrscheinlichkeitstheorie*, kann dies anders aussehen. Der Grund hierfür ist, dass Wahrscheinlichkeiten durch Maße beschrieben werden, deren Theorie in der Vorlesung *Analysis III* behandelt wird.

Als zu diesem Skript ergänzende Literatur gibt es einige Bücher zu nennen, etwa:

- K. Bosch. Elementare Einführung in die Wahrscheinlichkeitsrechnung. Vieweg+Teubner Verlag, 2011
- H. O. Georgii. Stochastik: Einführung In Die Wahrscheinlichkeitstheorie Und Statistik. de Gruyter, 2015
- N. Henze. Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls. Springer, 2018
- G. Kersting und A. Wakolbinger. Elementare Stochastik. Birkhäuser, 2010

0.2 Wahrscheinlichkeitsräume und Zufallsvariablen

Wir fangen rein formal an und definieren zunächst, worum es in dieser Vorlesung ausschließlich geht.

Definition 0.1 (Wahrscheinlichkeitsräume). 1. Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{F}, \mathbf{P})$, bestehend aus einer Menge Ω , einer σ -Algebra¹ \mathcal{F} und einem Wahrscheinlichkeitsmaß (oder Wahrscheinlichkeitsverteilung) $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$, für das

$$\begin{aligned} \mathbf{P}(\Omega) &= 1, \\ \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} \mathbf{P}(A_i) \text{ falls } A_i \cap A_j = \emptyset \text{ für alle } i \neq j. \end{aligned} \tag{0.1}$$

gilt.

2. Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ heißt diskret, falls Ω höchstens abzählbar ist und $\mathcal{F} = \mathcal{P}(\Omega)$. In diesem Fall sagen wir auch, dass (Ω, \mathbf{P}) ein diskreter Wahrscheinlichkeitsraum ist. Die Funktion

$$p : \begin{cases} \Omega & \rightarrow [0, 1] \\ \omega & \mapsto \mathbf{P}(\{\omega\}) \end{cases}$$

¹Mehr dazu in Analysis III. Eine σ -Algebra – $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ – die Potenzmenge von Ω – ist ein Mengensystem mit (i) $\emptyset \in \mathcal{F}$, (ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$, (iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

heißt auch Zähldichte.

3. Das Wahrscheinlichkeitsmaß \mathbf{P} hat eine Dichte, falls $\Omega = \mathbb{R}^d$ für ein $d = 1, 2, \dots$ und es ein $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ gibt mit

$$\mathbf{P}(B_1 \times \cdots \times B_d) = \int_{B_1} \cdots \int_{B_d} f(x_1, \dots, x_d) dx_d \cdots dx_1$$

für alle Intervalle $B_k = [a_k, b_k]$, oder $B_k = (a_k, b_k)$, oder $B_k = (a_k, b_k]$, oder $B_k = [a_k, b_k)$, $k = 1, \dots, d$. Wir nennen $B_1 \times \cdots \times B_d$ auch einen Quader. Wir sagen dann auch, dass \mathbf{P} kontinuierlich ist.

4. Die Menge Ω heißt Grundraum oder Ergebnisraum, \mathcal{F} heißt auch Ereignisraum.
 5. Jede Abbildung $X : \Omega \rightarrow S$ heißt Zufallsvariable mit Zielbereich S oder S -wertige Zufallsvariable.²

Wir benennen zunächst elementare Eigenschaften von Wahrscheinlichkeitsmaßen.

Lemma 0.2 (Grundlegende Eigenschaften von Wahrscheinlichkeitsmaßen). Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum. Weiter sei $A, A_1, A_2, \dots \in \mathcal{F}$.

1. Es ist $\mathbf{P}(\emptyset) = 0, \mathbf{P}(\Omega) = 1$.
2. Es ist $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.
3. Ist $A_1 \subseteq A_2$, so ist $\mathbf{P}(A_1) \leq \mathbf{P}(A_2)$.
4. Es gilt

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

Beweis. 1. Da $\emptyset = \emptyset \uplus \emptyset$ und die rechte Seite disjunkte Mengen sind, gilt

$$\mathbf{P}(\emptyset) = \mathbf{P}(\emptyset \uplus \emptyset) = 2\mathbf{P}(\emptyset).$$

Da \mathbf{P} nur Werte in $[0, 1]$ annimmt, muss $\mathbf{P}(\emptyset) = 0$ gelten.

2. Es ist $\mathbf{P}(A^c) + \mathbf{P}(A) = \mathbf{P}(\Omega) = 1$, da $\Omega = A \uplus A^c$.
 3. Es gilt $A_2 = A_1 \uplus (A_2 \setminus A_1)$. Also ist

$$\mathbf{P}(A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2 \setminus A_1) \geq \mathbf{P}(A_1).$$

4. Wir schreiben

$$\begin{aligned} \mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbf{P}\left(\biguplus_{n=1}^{\infty} A_n \setminus (A_1 \cup \cdots \cup A_{n-1})\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n \setminus (A_1 \cup \cdots \cup A_{n-1})) \\ &\leq \sum_{n=1}^{\infty} \mathbf{P}(A_n) \end{aligned}$$

nach 2. □

²Hier sind wir etwas schlampig: eigentlich sind nur *messbare* Abbildungen Zufallsvariable. (Das bedeutet: Ist $\mathcal{F}' \subseteq \mathcal{P}(S)$ eine weitere σ -Algebra, so ist $X^{-1}(B) \in \mathcal{F}$ für alle $B \in \mathcal{F}'$.) Da es aber sehr schwierig ist, sich nicht-messbare Abbildungen auszudenken, verzichten wir auf diese Feinheit. Insbesondere ist für höchstens abzählbares Ω jede Abbildung messbar. Mehr dazu erfahren Sie in Analysis III und in der Wahrscheinlichkeitstheorie.

Bemerkung 0.3 (Monotonie von Wahrscheinlichkeitsmaßen). Es gibt weitere Eigenschaften von Wahrscheinlichkeitsmaßen, auf die wir hier ohne maßtheoretische Grundlagen nicht weiter eingehen können. Etwa gilt für $A_1 \subseteq A_2 \subseteq \dots$

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right)$$

und für $A_1 \supseteq A_2 \supseteq \dots$ mit $\bigcap_{n=1}^{\infty} A_n = \emptyset$ ist

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = 0.$$

In der Vorlesung Stochastik werden wir uns ausschließlich mit diskreten Wahrscheinlichkeitsräumen und mit Wahrscheinlichkeitsmaßen mit Dichten beschäftigen. In der Wahrscheinlichkeitstheorie werden dann Wahrscheinlichkeitsräume in ihrer vollen Allgemeinheit behandelt.

Wir beginnen mit einem grundlegenden Fall diskreter Wahrscheinlichkeitsräume. Das nächste Resultat besagt, dass es bei diskreten Wahrscheinlichkeitsräumen genügt, das Maß auf *Elementarereignissen* $\{\omega\}$ für alle $\omega \in \Omega$ zu kennen.

Lemma 0.4 (Zähldichte). *Sei Ω höchstens abzählbar. Eine Funktion $p : \Omega \rightarrow [0, 1]$ ist genau dann eine Zähldichte, wenn*

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

In diesem Fall ist

$$\mathbf{P} : A \mapsto \sum_{\omega \in A} p(\omega)$$

das einzige Wahrscheinlichkeitsmaß mit Zähldichte p .

Beweis. Jede Teilmenge $A \subseteq \Omega$ ist höchstens abzählbar und $A = \bigcup_{\omega \in A} \{\omega\}$ eine Zerlegung von A in disjunkte Teilmengen. Das Resultat folgt nun aus den Eigenschaften für Wahrscheinlichkeitsmaße. \square

Wir führen nun noch den wichtigen Begriff der Verteilung einer Zufallsvariablen ein.

Definition 0.5. *Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und X eine S -wertige Zufallsvariable (mit σ -Algebra \mathcal{F}'). Dann heißt die Abbildung*

$$X_* \mathbf{P} : \begin{cases} \mathcal{F}' & \rightarrow [0, 1] \\ A & \mapsto \mathbf{P}(\{\omega : X(\omega) \in A\}) \end{cases}$$

Bildmaß von X unter \mathbf{P} . Äquivalent schreiben wir

$$\mathbf{P}(X \in A) := \mathbf{P}(\{X \in A\}) := \mathbf{P}(X^{-1}(A)) := \mathbf{P}(\{\omega : X(\omega) \in A\}) = X_* \mathbf{P}(A).$$

Ist S höchstens abzählbar, so ist auch

$$\mathbf{P}(X \in A) = \sum_{s \in A} \mathbf{P}(X = s).$$

Ist $S \subseteq \mathbb{R}^d$ und besitzt $X_*\mathbf{P}$ eine Dichte f_X , so gilt

$$\mathbf{P}(X \in A) = \int_A f_X(x) dx$$

für jeden Quader A .

Beispiel 0.6 (Einfache Beispiele). 1. Die einfachste Zufallsvariable ist $X = \text{id}$. Hier ist $X_*\mathbf{P} = \mathbf{P}$, denn

$$\mathbf{P}(X \in A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(A).$$

2. Auch sehr einfach sind konstante Zufallsvariable. Sei also $X = c$ für ein $c \in S$. Dann ist

$$\mathbf{P}(X \in A) = \mathbf{P}(c \in A) = 1_{c \in A},$$

also $X_*\mathbf{P} = 1_c(\cdot)$.

0.3 Häufige Modelle: Münzwurf, Würfeln, Urnen

Wir behandeln zunächst einige häufige Beispiele.

Beispiel 0.7 (Münzwurf). Wenn man eine Münze wirft, zeigt sie bekanntlich *Kopf* oder *Zahl*. Bei einem p -Münzwurf ist die Wahrscheinlichkeit für *Kopf* gerade p . Damit ist die Wahrscheinlichkeit für *Zahl* gerade $q := 1 - p$. Bei üblichen Münzen denkt man oft an den Fall $p = 1/2$, weil Kopf und Zahl mit derselben Wahrscheinlichkeit oben zu liegen kommen. Wirft man jedoch eine Reißzwecke (was eine sehr spezielle Münze ist), die entweder mit der spitzen Seite oben oder unten zu liegen kommt, ist klar, dass auch $p \neq 1/2$ sein kann. Wirft man zweimal mit derselben Münze, kann man nach der Wahrscheinlichkeit fragen, zweimal *Kopf* zu werfen. Dies ist p^2 .

Um das Beispiel eines n -fach ausgeführten p -Münzwurfes in den Rahmen eines Wahrscheinlichkeitsraumes zu bringen, sei $\Omega = \{0, 1\}^n$, wobei $0 \equiv \text{Zahl}$ und $1 \equiv \text{Kopf}$. Für das Wahrscheinlichkeitsmaß geben wir die Zähldichte an, nämlich

$$\mathbf{P}((\omega_1, \dots, \omega_n)) = \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i} = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}. \quad (0.2)$$

Als Beispiel einer Zufallsvariable nehmen wir

$$X_i(\omega_1, \dots, \omega_n) = \omega_i, \quad X = X_1 + \dots + X_n.$$

Somit ist X_i das Ergebnis des i -ten Münzwurfes und X ist die Gesamtzahl an Münzwürfen mit Ausgang *Kopf*. Für die Verteilung von X_i und X gilt

$$\mathbf{P}(X_i = 1) = 1 - \mathbf{P}(X_i = 0) = p,$$

$$\mathbf{P}(X = k) = \mathbf{P}((\omega_1, \dots, \omega_n) : \omega_1 + \dots + \omega_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Die letzte Gleichheit erklärt man dabei wie folgt: Es gibt insgesamt $\binom{n}{k}$ Möglichkeiten, k Ausgänge *Kopf* auf die n Versuche zu verteilen – später mehr zu dieser Anzahl. Jede der Möglichkeiten, bei denen k Köpfe fallen, hat Wahrscheinlichkeit $p^k (1-p)^{n-k}$.

Wir werden ab Abschnitt 3.1 sagen, X habe eine Binomialverteilung mit Parametern n und p .

Beispiel 0.8 (Würfeln). Eine bekannte Herausforderung ist es (etwa bei *Mensch ärgere Dich nicht*), beim Würfeln mit einem Würfel möglichst schnell eine 6 zu werfen. Dies ist ganz ähnlich wie beim $(1/6)$ -Münzwurf, schließlich gibt es nur zwei Möglichkeiten *6* oder *nicht 6*. Damit ist die Wahrscheinlichkeit, k mal keine 6 zu werfen, gerade $(5/6)^k$. Wir berechnen noch die Wahrscheinlichkeit, dass die erste 6 genau im n -ten Wurf kommt. Da wir hierzu unbeschränkt oft den Würfel werfen müssen, sei $\Omega = \{1, \dots, 6\}^{\mathbb{N}}$. Weiter sei $X_i \in \{1, \dots, 6\}$ das Ergebnis des i -ten Wurfs und

$$T = \min\{n : X_n = 6\}$$

die Nummer des Versuches, bei dem zum ersten mal eine 6 fällt. Die Verteilung von T ist

$$\mathbf{P}(T = n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6},$$

denn die einzige Möglichkeit, wie $T = n$ zustande kommt, ist es, zunächst $(n - 1)$ -mal keine Sechs zu werfen und anschließend eine 6.

Beispiel 0.9 (Urne). In einer Urne befinden sich N Kugeln, und zwar K_1 Kugeln der Farbe 1, ..., K_n Kugeln der Farbe n . (Also ist $K_1 + \dots + K_n = N$.) Zieht man (mit verschlossenen Augen) aus der Urne, zieht man eine Kugel der Farbe i mit Wahrscheinlichkeit K_i/N . Zieht man anschließend nochmal aus der Urne (wobei die erste Kugel nicht zurückgelegt wurde), ist die Wahrscheinlichkeit nochmal eine Kugel der Farbe i zu ziehen $(K_i - 1)/(N - 1)$. Damit ist die Wahrscheinlichkeit, zwei Kugeln derselben Farbe zu ziehen, gerade

$$\sum_{i=1}^n \frac{K_i}{N} \frac{K_i - 1}{N - 1}.$$

Um dies zu formalisieren, sei $\Omega = \{1, \dots, n\}^2$ und $X_i \in \{1, \dots, n\}$ sei die Farbe des i -ten Zuges, $i = 1, 2$. Wir schreiben damit

$$\mathbf{P}(X_1 = X_2) = \sum_{i=1}^n \mathbf{P}(X_1 = X_2 = i) = \sum_{i=1}^n \frac{K_i}{N} \frac{K_i - 1}{N - 1}.$$

Im ersten Gleichheitszeichen haben wir verwendet, dass die Ereignisse $\{X_1 = X_2 = i\}, i = 1, \dots, n$ disjunkt sind.

Beispiel 0.10 (Geburtstagsproblem). In einem Raum befinden sich 23 Personen. Wie groß ist die Wahrscheinlichkeit, dass es zwei Personen gibt die am selben Tag Geburtstag haben?

Um diese Wahrscheinlichkeit zu berechnen, ist es hilfreich, das Gegenteil *Alle Personen haben an unterschiedlichen Tagen Geburtstag* zu betrachten. Wir stellen die Personen (in Gedanken) in einer Reihe auf. Die Wahrscheinlichkeit dass die zweite Person an einem anderen Tag als die erste Person Geburtstag hat ist $364/365$. (Von Schaltjahren und dem 29.2. sehen wir einmal ab.) Weiter ist die Wahrscheinlichkeit, dass die dritte Person an einem anderen Tag als die Personen eins und zwei Geburtstag hat, dann $363/365$. Überlegen wir das weiter, so ist die Wahrscheinlichkeit, dass alle Personen an unterschiedlichen Tagen Geburtstag haben, gerade

$$\frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - 22}{365} \approx 0.493.$$

Damit ist die Wahrscheinlichkeit, dass es zwei Personen gibt, die am gleichen Tag Geburtstag haben, etwa $1 - 0.493 = 0.507$.

1 Kombinatorik und uniform verteilte Zufallsvariablen

Der Grund, warum in der Stochastik zunächst das geschickte Abzählen (also Kombinatorik) so wichtig ist, sind uniforme Verteilungen oder Laplace-Experimente.

1.1 Laplace-Experimente

Unter einem Laplace-Experiment versteht man die Situation, wenn alle möglichen Ausgänge dieselbe Wahrscheinlichkeit besitzen. Beispiele sind der $\frac{1}{2}$ -Münzwurf oder das Würfeln mit einem fairen Würfel. Die zugehörigen Wahrscheinlichkeitsmaße heißen uniforme Verteilungen.

Definition 1.1 (Uniforme Verteilung). *Sei X eine S -wertige Zufallsvariable.*

1. Ist S endlich und gilt $\mathbf{P}(X = a) = \frac{1}{|S|}$ für alle $a \in S$, so nennen wir X (diskret) uniform verteilt auf S . Wir schreiben dann auch $X \sim U(S)$.
2. Ist $S \subseteq S' = \mathbb{R}^d$ und hat X die Dichte $f_X(x) = 1_S(x) \frac{1}{|S|}$ (wobei $|S|$ das Volumen von S im \mathbb{R}^d ist), so nennen wir X (stetig) uniform verteilt auf S . Wir schreiben dann ebenfalls $X \sim U(S)$.

Bemerkung 1.2. In beiden Fällen ist $\mathbf{P}(X \in A) = \frac{|A|}{|S|}$.

Beispiel 1.3 (Kollision von Kennzeichen). In Verallgemeinerung zu Beispiel 0.10 werde eine Menge von n Individuen mit $r \geq n$ verschiedenen Kennzeichen gekennzeichnet. Wir denken dabei an Kennzeichen wie r verschiedene Namen, Geburtstage, PINs etc. Wir gehen davon aus, dass alle Individuen jedes Kennzeichen mit gleicher Wahrscheinlichkeit erhalten. Wie groß ist die Wahrscheinlichkeit, dass keine zwei Individuen gleich gekennzeichnet sind?

Hierzu sei $X = (X_1, \dots, X_n)$ mit Werten in $S = \{1, \dots, r\}^{\{1, \dots, n\}}$, so dass $1 \leq X_i \leq r$ das Kennzeichen des i -ten Individuums benennt. Wir suchen nun die Wahrscheinlichkeit von

$$\mathbf{P}(X_i \neq X_j \text{ für alle } i \neq j),$$

oder äquivalent

$$\mathbf{P}(X \in A) \text{ für } A = \{(x_1, \dots, x_n) : x_i \neq x_j \text{ für alle } i \neq j\}.$$

Um diese zu bestimmen, müssen wir $|A|$ berechnen. Da das erste Individuum r Kennzeichen zur Auswahl hat, das zweite Individuum (falls A zutrifft) dann noch $r - 1$ usw, ist

$$|A| = r(r - 1) \cdots (r - n + 1).$$

Damit ist

$$\mathbf{P}(X \in A) = \frac{r(r - 1) \cdots (r - n + 1)}{r^n} = \prod_{i=1}^{n-1} \left(1 - \frac{i}{r}\right)$$

Wir versuchen nun noch, diese Wahrscheinlichkeit für große n abzuschätzen. Verwenden wir $1 - t \leq e^{-t}$, so folgt

$$\mathbf{P}(X \in A) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{r}\right) \leq \exp\left(-\sum_{i=1}^{n-1} \frac{i}{r}\right) = \exp\left(-\frac{n(n-1)}{2r}\right)$$

Da

$$A^c = \bigcup_{i < j} B_{ij}, \quad B_{ij} := \{(a_1, \dots, a_n) : a_i = a_j\}$$

und

$$\mathbf{P}(X \in B_{ij}) = \frac{1}{r},$$

und da³ $\{(i, j) \in [1 : n]^2 : i < j\} = \frac{n(n-1)}{2}$, folgt, dass

$$\mathbf{P}(X \in A) = 1 - \mathbf{P}(X \in A^c) \geq 1 - \sum_{i < j} \mathbf{P}(X \in B_{ij}) = 1 - \frac{n(n-1)}{2r}$$

Es gilt also

$$1 - \frac{n(n-1)}{2r} \leq \mathbf{P}(X \in A) \leq \exp\left(-\frac{n(n-1)}{2r}\right).$$

Für große r muss also n mindestens so groß wie \sqrt{r} sein, damit es merklich zu Kollisionen kommt. (In Beispiel 0.10 war $r = 365$ und $n = 23$, also $n^2 = 529$.)

1.2 Zufällige Permutationen

Aus der linearen Algebra sind Permutationen bekannt. Wir beschäftigen uns nun mit uniform verteilten (also zufälligen) Permutationen. Wir erinnern zunächst an die Bedeutung und Schreibweise von Permutationen.

Bemerkung 1.4 (Permutation). Eine Permutation σ der Menge $[1 : n]$ ist eine bijektive Abbildung $\sigma : [1 : n] \rightarrow [1 : n]$. Um σ zu notieren, gibt es zwei verschiedene Möglichkeiten. Entweder man schreibt in zwei Zeilen jeweils i über $\sigma(i)$, was dann für $n = 7$ etwa so aussieht:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 2 & 7 & 3 & 1 & 4 & 6 \end{pmatrix}, \quad (1.1)$$

oder man verwendet Zykel von σ . Hierbei ist eine Zykel von σ der Länge j eine Folge k_1, \dots, k_j mit $\sigma(k_1) = k_2, \dots, \sigma(k_{j-1}) = k_j, \sigma(k_j) = k_1$. In obigem Beispiel ergibt sich

$$\sigma = (15)(2)(3764).$$

Hierbei werden in Klammern jeweils Zykel angegeben. Man beachte, dass die Zykel-Schreibweise nicht eindeutig ist, etwa gilt auch

$$\sigma = (51)(2)(6437).$$

Bekanntermaßen gibt es $n! = 1 \cdots n$ Permutationen von $[1 : n]$. Dies überlegt man sich am besten an der Schreibweise (1.1). Für $\sigma(1)$ gibt es nämlich genau n Möglichkeiten. Anschließend gibt es für $\sigma(2)$ noch $n-1$ Möglichkeiten usw. Zählt man also alle Permutationen durch, so kommt man auf $|[1 : n]| = n!$. Wir bezeichnen die Menge aller Permutationen von $[1 : n]$ mit $\mathbb{S}(n)$.

³Wir schreiben $[1 : n] := \{1, \dots, n\}$.

Beispiel 1.5 (Länge eines Zykel). Sei Σ eine zufällige Permutation von $[1 : n]$. (Das bedeutet: Auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ ist eine Zufallsvariable $\Sigma : \Omega \rightarrow \mathbb{S}(n)$ definiert, so dass die Verteilung von Σ uniform auf allen Permutationen ist.) Weiter sei

$$h(\sigma) := \text{Länge des Zykel, der } 1 \text{ enthält.}$$

Wir wollen nun zeigen, dass

$$\mathbf{P}(h(\Sigma) = b) = \frac{1}{n}.$$

Hierzu setzen wir

$$A := \{\sigma \in \mathbb{S}(n) : h(\sigma) = b\} = \{\sigma \in \mathbb{S}(n) : \sigma(1) \neq 1, \sigma^2(1) \neq 1, \dots, \sigma^{b-1}(1) \neq 1, \sigma^b(1) = 1\}.$$

Es gilt (mit derselben Überlegung, mit der wir $\mathbb{S}(n)$ durchgezählt haben)

$$|A| = (n-1)(n-2) \cdots (n-b+1) \cdot 1 \cdot (n-b) \cdots 1,$$

und damit

$$\mathbf{P}(h(\Sigma) = b) = \frac{|A|}{|\mathbb{S}|} = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

1.3 Zufällige Teilmengen

Während Permutationen Bijektionen beschreiben, beschäftigen wir uns nun mit zufälligen Teilmengen einer gegebenen Menge $[1 : n]$.

Bemerkung 1.6. Wir bezeichnen mit

$$S_k = \{A : A \subseteq [1 : n], |A| = k\}$$

die Menge der k -elementigen Teilmengen von $[1 : n]$. Sei X uniform auf S_k verteilt. Um nun die Verteilung von X zu bestimmen, benötigen wir $|S_k|$. Wir zeigen

$$|S_k| = \binom{n}{k} := \frac{n!}{k!(n-k)!} \quad \text{oder} \quad \mathbf{P}(X = \{1, \dots, k\}) = \frac{1}{\binom{n}{k}}.$$

Klar ist, dass aus der zweiten Aussage die erste folgt. Um $\{X = \{1, \dots, k\}\}$ zu erhalten, müssen wir als erstes Element eines in $[1 : k]$ wählen. Bei der zweiten Ziehung müssen wir eines der übrigen $k-1$ aus $[1 : k]$ ziehen usw. Daraus folgt

$$\mathbf{P}(X = \{1, \dots, k\}) = \frac{k}{n} \frac{k-1}{n-1} \cdots \frac{1}{n-k+1} = \frac{k!(n-k)!}{n!}.$$

Eine alternative Erklärung ist die Folgende: Um ein Element aus $|S_k|$ zu erhalten, wählen wir zunächst ein Element aus n aus, dann eines aus den restlichen $n-1$ usw. Die Einträge eines so erhaltenen k -Tupels bildet eine k -elementige Teilmenge von $[1 : n]$. Allerdings gibt es $k!$ verschiedene Reihenfolgen dieser Elemente, so dass $|S_k| = \binom{n}{k}$ folgt.

1.4 Zufällige Besetzungen

Bemerkung 1.7. Unter einer Besetzung verstehen wir eine Belegung von r Plätzen (Urnen) mit n Objekten (Kugeln). Formal setzen wir

$$S_{n,r} = \{k = (k_1, \dots, k_r) : k_j \in \mathbb{N}_0, k_1 + \dots + k_r = n\}$$

und betrachten nun wieder ein rein zufälliges Element X aus $S_{n,r}$. Wir zeigen nun

$$\mathbf{P}(X = (k_1, \dots, k_r)) = \frac{1}{\binom{n+r-1}{n}}$$

für jedes $(k_1, \dots, k_r) \in S_{n,r}$.

Um dies einzusehen, verwenden wir am besten eine Hilfskonstruktion: Sei

$$S' = \{01\text{-Folgen der Länge } n+r-1 \text{ mit genau } n \text{ Einsen}\}$$

$$h : \begin{cases} S_{n,r} & \rightarrow S' \\ (k_1, \dots, k_r) & \mapsto \underbrace{1 \dots 1}_{k_1\text{-mal}} 0 \underbrace{1 \dots 1}_{k_2\text{-mal}} 0 \dots \underbrace{1 \dots 1}_{k_r\text{-mal}} \end{cases}$$

also z.B. $h(2, 0, 3, 0) = 11001110$. Es ist klar, dass h eine Bijektion ist. Da S' alle n -elementigen Teilmengen von $[1 : (n+r-1)]$ beschreibt, ist

$$|S_{n,r}| = |S'| = \binom{n+r-1}{n}.$$

1.5 Kontinuierlich uniform verteilte Zufallsvariable

Wir erinnern zunächst an Definitionen 0.1.3 und 1.1. Danach ist eine Zufallsvariable X genau dann uniform auf $[a, b] \subseteq \mathbb{R}$ verteilt, wenn

$$\mathbf{P}(X \in [w, x]) = \frac{x-w}{b-a}$$

für alle $a \leq w \leq x \leq b$. Dies ist offenbar genau dann der Fall, wenn für $a \leq w \leq x \leq b$

$$\mathbf{P}(X \leq x) = \frac{x-a}{b-a}.$$

(Denn: Gilt dies, so ist $\mathbf{P}(X \in [w, x]) = \mathbf{P}(X \leq x) - \mathbf{P}(X \leq w)$.)

Für das nächste Resultat benutzen wir folgende Schreibweise für $x \in \mathbb{R}$: Der ganzzahlige Anteil von x ist $[x]$, und $\langle x \rangle = x - [x]$ ist der gebrochene Anteil von x .

Lemma 1.8 (Verschiebungsinvarianz der uniformen Verteilung). *Sei X uniform auf $[0, r]$ verteilt für $r \in \mathbb{N}$ und $v \in \mathbb{R}$. Dann ist $\langle X + v \rangle \sim U([0, 1])$.*

Beweis. OBdA ist $r = 1$, der allgemeine Fall folgt wegen $\langle X \rangle \sim U([0, 1])$ und $\langle X + v \rangle = \langle \langle X \rangle + v \rangle$. OBdA ist $0 \leq v \leq 1$, ansonsten ersetzen wir v mit $\langle v \rangle$. Wir berechnen für $0 \leq x \leq 1$

$$\begin{aligned} \mathbf{P}(\langle X + v \rangle \leq x) &= \mathbf{P}(X + v \leq x) + \mathbf{P}(1 \leq X + v \leq 1 + x) \\ &= \mathbf{P}(X \leq x - v) + \mathbf{P}(1 - v \leq X \leq 1 + x - v) = \begin{cases} x - v + v, & x \geq v, \\ 0 + x, & x \leq v \end{cases} = x. \end{aligned}$$

Damit ist $\langle X + v \rangle \sim U([0, 1])$. □

Wir kommen nun zu einer interessanten Beobachtung: die führende Ziffer einer auf einem großen Intervall gezogenen uniformen Zufallsvariable ist mit größter Wahrscheinlichkeit eine 1:

Proposition 1.9 (Benford's Gesetz). *Sei X eine Zufallsvariable, so dass $\langle \log_{10} X \rangle \sim U([0, 1])$. Weiter sei*

$$h : \begin{cases} \mathbb{R}_+ & \rightarrow \{1, \dots, 9\} \\ x & \mapsto b \text{ falls } \log_{10} b \leq \langle \log_{10} x \rangle < \log_{10}(b+1). \end{cases}$$

(Mit anderen Worten ist $h(x)$ die führende Ziffer der Dezimaldarstellung von x .) Dann ist

$$\mathbf{P}(h(X) = b) = \log_{10} \left(1 + \frac{1}{b} \right).$$

Bemerkung 1.10. Erstaunlich viele Verteilungen für X sind so, dass $\langle \log_{10} X \rangle \sim U([0, 1])$ zumindest approximativ gilt. Um ein einfaches Beispiel zu nennen, sei $Y \sim U([0, r])$ für ein $r \in \mathbb{N}$ und $X = 10^Y$. Dann ist

$$\mathbf{P}(\langle \log_{10} X \rangle \leq x) = \mathbf{P}(\langle Y \rangle \leq x) = x$$

nach Lemma 1.8.

Beweis. Wir schreiben direkt

$$\mathbf{P}(h(X) = b) = \mathbf{P}(\langle \log_{10} X \rangle \in [\log_{10} b, \log_{10}(b+1))) = \log_{10}(b+1) - \log_{10}(b) = \log_{10} \left(1 + \frac{1}{b} \right).$$

□

2 Verteilungen und deren Eigenschaften

Der Begriff des Wahrscheinlichkeitsmaßes oder der Wahrscheinlichkeitsverteilung (siehe Definition 0.1) ist für unsere Zwecke zentral. In diesem Kapitel beschäftigen wir uns meist mit reellwertigen Zufallsvariablen und deren Bildmaßen. Wir gehen dabei nicht weiter auf das Wahrscheinlichkeitsmaß auf Ω ein, sondern interessieren uns ausschließlich für das Bildmaß.

2.1 Die Verteilungsfunktion

Wahrscheinlichkeitsverteilungen sind Abbildungen, die eine σ -Algebra als Definitionsmenge haben. Um solche Verteilungen veranschaulichen zu können, benötigen wir eine andere Darstellung, die wir nun mit den Verteilungsfunktionen kennen lernen werden.

Definition 2.1 (Verteilungsfunktion). *Sei X eine S -wertige Zufallsvariable und $S \subseteq \mathbb{R}$. Die Funktion*

$$F_X : \begin{cases} \mathbb{R} & \rightarrow [0, 1] \\ x & \mapsto \mathbf{P}(X \leq x) \end{cases}$$

heißt *Verteilungsfunktion (des Bildmaßes) von X* .

Lemma 2.2 (Verteilungsfunktion von diskreten und kontinuierlichen Zufallsvariablen). *Sei X eine S -wertige Zufallsvariable.*

1. Ist S diskret, so ist die Verteilungsfunktion

$$F_X(x) = \sum_{\substack{y \leq x \\ y \in S}} \mathbf{P}(X = y).$$

2. Besitzt X die Dichte f_X , so ist die Verteilungsfunktion

$$F_X(x) = \int_{-\infty}^x f_X(x) dx.$$

Ist f_X stückweise stetig, so ist $F'_X = f_X$ an allen Stetigkeitsstellen von f_X .

In beiden Fällen wird die Verteilung von X eindeutig durch F_X beschrieben.

Beweis. Die Aussagen 1. und 2. erhält man durch direktes Einsetzen von Zähldichte und Dichte in die Definition für die Verteilungsfunktion. Für die Eindeutigkeit der Verteilungsfunktion stellen wir im Falle einer diskreten Zufallsvariablen X und $x \in S$ fest, dass

$$\begin{aligned} \mathbf{P}(X = x) &= \lim_{\varepsilon \rightarrow 0} \mathbf{P}(x \leq X \leq x + \varepsilon) = \lim_{\varepsilon \rightarrow 0} \mathbf{P}(X \leq x + \varepsilon) - \mathbf{P}(X \leq x) \\ &= \lim_{\varepsilon \rightarrow 0} F_X(x + \varepsilon) - F_X(x). \end{aligned}$$

Für kontinuierliche Zufallsvariablen schreiben wir

$$\mathbf{P}(X \in (a, b]) = \mathbf{P}(X \leq b) - \mathbf{P}(X \leq a) = F_X(b) - F_X(a).$$

Also bestimmt F_X die Werte von $X_*\mathbf{P}$ auf allen Quadern und damit nach Definition 0.1 $X_*\mathbf{P}$ selbst. \square

Lemma 2.3 (Eigenschaften von Verteilungsfunktionen). *Sei F_X die Verteilungsfunktion einer Zufallsvariable X . Dann gilt:*

1. F_X ist monoton wachsend mit

$$F_X(x) \xrightarrow{x \rightarrow -\infty} 0, \quad F_X(x) \xrightarrow{x \rightarrow \infty} 1.$$

2. F_X ist rechtsstetig und hat linksseitige Grenzwerte. Hat X eine Dichte, dann ist F_X sogar stetig.

Beweis. 1. Wegen Lemma 0.2.3 gilt für $x \leq y$

$$F_X(x) = \mathbf{P}(X \leq x) \leq \mathbf{P}(X \leq y) = F_X(y)$$

und damit ist F_X monoton wachsend.

2. Der Fall mit Dichte ist klar wegen des Hauptsatzes der Differential- und Integralrechnung. Im Allgemeinen folgt mit Bemerkung 0.3

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} F_X(x + \varepsilon) &= \lim_{\varepsilon \downarrow 0} \mathbf{P}(X \leq x + \varepsilon) = \mathbf{P}(X \leq x) + \lim_{\varepsilon \downarrow 0} \mathbf{P}(x < X \leq x + \varepsilon) = \mathbf{P}(X \leq x), \\ \lim_{\varepsilon \downarrow 0} F_X(x - \varepsilon) &= \lim_{\varepsilon \downarrow 0} \mathbf{P}(X \leq x - \varepsilon) = \mathbf{P}(X < x). \end{aligned}$$

und 2. ist ebenfalls gezeigt. \square

Beispiel 2.4 (Verteilungsfunktion der kontinuierlich uniformen Verteilung). Sei $X \in U([a, b])$. Dann ist die Verteilungsfunktion von X

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x \geq b \end{cases}$$

Denn: Für $x \leq a$ und $x \geq b$ ist die Aussage klar, da $\mathbf{P}(X \leq a) = \mathbf{P}(X > b) = 0$. Die Zufallsvariable X hat nach Definition 1.1 die Dichte $f_X(x) = 1_{[a,b]}(x) \frac{1}{b-a}$. Für $a \leq x \leq b$ ist deshalb

$$\mathbf{P}(X \leq x) = \int_{-\infty}^x 1_{[a,b]}(y) \frac{1}{b-a} dy = \int_a^x \frac{1}{b-a} = \frac{x-a}{b-a}.$$

2.2 Bilder von Zufallsvariablen

Da eine S -wertige Zufallsvariable eine Abbildung $X : \Omega \rightarrow S$ ist, kann man ebenso Funktionen von Zufallsvariablen betrachten.

Definition 2.5 (Bild einer Zufallsvariablen). Sei X eine S -wertige Zufallsvariable und $h : S \rightarrow S'$ (messbar). Dann heißt $Y := h \circ X$ das Bild von X unter h und ist eine S' -wertige Zufallsvariable.

Die Verteilung von $Y := h \circ X$ heißt durch h transformierte Verteilung von X . Ist diese Verteilung identisch mit der einer Zufallsvariablen Z , so schreiben wir $Y \sim Z$.

Wir haben bereits Bilder von Zufallsvariablen betrachtet und wiederholen dies kurz.

Beispiel 2.6. 1. Sei Σ eine uniform auf $\mathbb{S}(n)$ verteilte Zufallsvariable (genau wie in Beispiel 1.5), sowie

$$h : \begin{cases} \mathbb{S}(n) & \rightarrow \mathbb{N} \\ \sigma & \mapsto \min\{k : \sigma^k(1) = 1\} \end{cases}$$

die Länge des Zyklus, der 1 enthält. Dann besagt Beispiel 1.5, dass die Zufallsvariable $Y := h(\Sigma)$ uniform auf $[1 : n]$ verteilt ist.

2. Genau wie in Lemma 1.8 sei $X \sim U([0, r])$ und $h : [0, r] \rightarrow [0, 1]$ durch $x \mapsto \langle x + v \rangle$ gegeben. Dann ist $Y := h \circ X$ nach $U([0, 1])$ verteilt, also $\langle X + v \rangle \sim U([0, 1])$.
3. Für eine \mathbb{R}_+ -wertige Zufallsvariable wird in Kapitel 4.3 das Bild von X unter der Abbildung $h_s : x \mapsto s^x$ für $s \in [0, 1]$, also die Zufallsvariable s^X , eine Rolle spielen.

Lemma 2.7 (Transformation von Verteilungen). Sei X eine $S \subseteq \mathbb{R}$ -wertige Zufallsvariable und $h : S \rightarrow \mathbb{R}$ *streng monoton wachsend*, sowie $Y := h \circ X$. Dann gilt für $x \in \mathbb{R}$

$$F_X(x) = F_Y(h(x)).$$

Ist insbesondere h streng monoton und differenzierbar mit differenzierbarer Umkehrabbildung, und hat X eine Dichte f_X , so hat Y die Dichte $y \mapsto f_Y(y) := f_X(h^{-1}(y)) |(h^{-1})'(y)|$.

Beweis. Wir schreiben direkt

$$F_X(x) = \mathbf{P}(X \leq x) = \mathbf{P}(h(X) \leq h(x)) = F_Y(h(x)).$$

Damit gilt wegen $F_X(h^{-1}(y)) = F_Y(y)$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(h^{-1}(y)) = f_X(h^{-1}(y))(h^{-1})'(y).$$

□

Beispiel 2.8. Sei $U \sim U([0, 1])$ und $h(u) = -\log u$. Dann hat $Y := h \circ X$ die Verteilung mit Dichte $f(x) = 1_{x \geq 0} e^{-x}$.

Denn: Es ist $h^{-1}(x) = e^{-x}$ und damit

$$f_Y(x) = 1_{e^{-x} \in [0, 1]} e^{-x} = 1_{x \geq 0} e^{-x}.$$

Lemma 2.9 (Simulationslemma). *Sei X eine Zufallsvariable mit Verteilungsfunktion F und F^{-1} die verallgemeinerte Inverse, gegeben durch*

$$F^{-1}(u) := \sup\{y : F(y) < u\}.$$

Ist $U \sim U([0, 1])$, so ist $F^{-1}(U) \sim X$.

Beweis. Es gilt (wegen $F^{-1}(u) \leq x \iff u \leq F(x)$)

$$\mathbf{P}(F^{-1}(U) \leq x) = \mathbf{P}(U \leq F(x)) = F(x),$$

und die Behauptung ist gezeigt. □

2.3 Die Ein-Ausschlussformel

Für Wahrscheinlichkeitsmaße wissen wir, dass sie additiv sind, d.h. (0.1) gilt. Dies verallgemeinern wir nun auf nicht notwendigerweise disjunkte Mengen.

Proposition 2.10 (Ein- Ausschlussformel). *Für eine S -wertige Zufallsvariable X und (messbare) Mengen $A_1, \dots, A_n \subseteq S$ gilt*

$$\begin{aligned} & \mathbf{P}(X \in A_1 \cup \dots \cup A_n) \\ &= \sum_{i=1}^n \mathbf{P}(X \in A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(X \in A_i \cap A_j) + \dots \pm \mathbf{P}(X \in A_1 \cap \dots \cap A_n). \end{aligned}$$

Beweis. Wir zeigen die Behauptung mittels Induktion nach n . Für $n = 1$ ist die Behauptung klar, für $n = 2$ ist

$$\begin{aligned} \mathbf{P}(X \in A_1 \cup A_2) + \mathbf{P}(X \in A_1 \cap A_2) &= \mathbf{P}(X \in A_1) + \mathbf{P}(X \in A_2 \setminus A_1) + \mathbf{P}(X \in A_1 \cap A_2) \\ &= \mathbf{P}(X \in A_1) + \mathbf{P}(X \in A_2). \end{aligned}$$

Gilt die Behauptung für ein n , so schreiben wir

$$A_1 \cup \dots \cup A_{n+1} = A_1 \cup \dots \cup (A_n \cup A_{n+1}).$$

Damit ist (mit $A_i \cap (A_n \cup A_{n+1}) = (A_i \cap A_n) \cup (A_i \cap A_{n+1})$)

$$\begin{aligned}
\mathbf{P}(X \in A_1 \cup \dots \cup A_{n+1}) &= \sum_{i=1}^{n-1} \mathbf{P}(X \in A_i) + \mathbf{P}(X \in A_n \cup A_{n+1}) \\
&- \sum_{1 \leq i < j \leq n-1} \mathbf{P}(X \in A_i \cap A_j) - \sum_{1 \leq i \leq n-1} \mathbf{P}(X \in A_i \cap (A_n \cup A_{n+1})) \\
&\quad + \dots \pm \mathbf{P}(X \in A_1 \cap \dots \cap A_{n-1} \cap (A_n \cup A_{n+1})) \\
&= \sum_{i=1}^{n+1} \mathbf{P}(X \in A_i) - \mathbf{P}(X \in A_n \cap A_{n+1}) \\
&\quad - \sum_{1 \leq i < j \leq n-1} \mathbf{P}(X \in A_i \cap A_j) - \sum_{i=1}^{n-1} (\mathbf{P}(X \in A_i \cap A_n) + \mathbf{P}(X \in A_i \cap A_{n+1})) \\
&\quad \quad - \mathbf{P}(X \in A_i \cap A_n \cap A_{n+1}) \\
&\quad \pm \mathbf{P}(X \in A_1 \cap \dots \cap A_{n-1} \cap A_n) \pm \mathbf{P}(X \in A_1 \cap \dots \cap A_{n-1} \cap A_{n+1}) \\
&\quad \mp \mathbf{P}(X \in A_1 \cap \dots \cap A_{n-1} \cap A_n \cap A_{n+1}).
\end{aligned}$$

Daraus folgt die Behauptung. \square

Beispiel 2.11 (Fixpunkte in Permutationen). Sei Σ uniform verteilt auf $\mathbb{S}(n)$, d.h. Σ ist eine zufällige Permutation von $[1 : n]$. Wir wollen die Wahrscheinlichkeit bestimmen, mit der Σ mindestens einen Fixpunkt besitzt. Wir setzen $A_i := \{\sigma \in \mathbb{S}(n) : \sigma(i) = i\}$. Damit gilt für i_1, \dots, i_k verschieden

$$\mathbf{P}(\Sigma \in A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}.$$

Damit gilt

$$\begin{aligned}
\mathbf{P}(\Sigma \in A_1 \cup \dots \cup A_n) &= n \cdot \frac{1}{n} - \binom{n}{2} \frac{1}{n(n-1)} + \binom{n}{3} \frac{1}{n(n-1)(n-2)} \dots \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} \dots \pm \frac{1}{n!}.
\end{aligned}$$

Anders ausgedrückt haben genau

$$n! \left(1 - \frac{1}{2!} + \frac{1}{3!} \dots \pm \frac{1}{n!} \right)$$

Permutationen mindestens einen Fixpunkt.

2.4 Die gemeinsame Verteilung und Unabhängigkeit

Betrachtet man mehrere Zufallsvariablen gleichzeitig, so hat man es mit anderen Worten mit einem Vektor von Zufallsvariablen zu tun. Die Verteilung dieses Vektors wird auch gemeinsame Verteilung der Zufallsvariablen genannt. Eine besondere Situation tritt auf, wenn sich Wahrscheinlichkeiten für verschiedene Zufallsvariablen multiplizieren. Dann heißen diese (stochastisch) unabhängig.

Definition 2.12 (Gemeinsame Verteilung). Ist $X = (X_1, \dots, X_n)$ mit Zielbereich $S := S_1 \times \dots \times S_n$, so heißt $X_*\mathbf{P}$ auch die gemeinsame Verteilung von X_1, \dots, X_n . Für die Projektion $\pi_i : S \rightarrow S_i$ heißt das Bild von X unter π_i auch die i -te Marginal- oder Randverteilung von X .

Ist die gemeinsame Verteilung von X_1, \dots, X_n gerade das Produkt der Marginalverteilungen, d.h. gilt für alle (messbaren) A_1, \dots, A_n

$$\mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbf{P}(X_1 \in A_1) \cdots \mathbf{P}(X_n \in A_n),$$

so heißt die Familie (X_1, \dots, X_n) stochastisch unabhängig.

Sind X_1, \dots, X_n und haben die gleiche Marginalverteilung, so nennen wir X_1, \dots, X_n unabhängig und identisch verteilt, was wir mit *i.i.d.* abkürzen.

Lemma 2.13 (Berechnung der Marginalverteilungen). Ist $X = (X_1, \dots, X_n)$ mit Zielbereich $S := S_1 \times \dots \times S_n$. Ist X diskret, so ist die i -te Marginalverteilung gegeben durch

$$\mathbf{P}(X_i = a_i) = \sum_{a_1 \in S_1} \cdots \sum_{a_{i-1} \in S_{i-1}} \sum_{a_{i+1} \in S_{i+1}} \cdots \sum_{a_n \in S_n} \mathbf{P}(X_1 = a_1, \dots, X_n = a_n).$$

Ist X kontinuierlich mit Dichte f_X , so ist die i -te Marginalverteilung gegeben durch die Dichte

$$f_i(x_i) = \int \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Beweis. Beide Aussagen folgen direkt aus der Definition der Marginalverteilung. □

Beispiel 2.14 (p -Münzwurf). Wir betrachten den p -Münzwurf aus Beispiel 0.7. Es gilt: Ein Zufallsvektor $X = (X_1, \dots, X_n)$ ist genau dann ein p -Münzwurf, wenn X_1, \dots, X_n unabhängig sind und

$$\mathbf{P}(X_i = 1) = 1 - \mathbf{P}(X_i = 0) = p.$$

Denn dann können wir die Verteilung von X angeben mittels

$$\begin{aligned} \mathbf{P}(X = (\omega_1, \dots, \omega_n)) &= \mathbf{P}(X_1 = \omega_1, \dots, X_n = \omega_n) = \mathbf{P}(X_1 = \omega_1) \cdots \mathbf{P}(X_n = \omega_n) \\ &= \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i}, \end{aligned}$$

was (0.2) entspricht.

Beispiel 2.15 (Zufällige Permutation). Sei Σ uniform auf $\mathbb{S}(n)$ verteilt. Schreiben wir $\Sigma = (\Sigma(1), \dots, \Sigma(n))$, so erhalten wir einen Vektor von Zufallsvariablen. Die Marginalverteilungen sind

$$\mathbf{P}(\Sigma(i) = k) = \frac{1}{n}, \quad i, k = 1, \dots, n,$$

d.h. $\Sigma(i)$ ist $U([1 : n])$ -verteilt, $i = 1, \dots, n$. Anders ausgedrückt ist jede Marginalverteilung eine $U([1 : n])$ -verteilt. Allerdings ist die Familie $\Sigma(1), \dots, \Sigma(n)$ nicht unabhängig, etwa ist nämlich für $i \neq j$

$$\mathbf{P}(\Sigma(i) = 1, \Sigma(j) = 1) = 0 \neq \frac{1}{n^2} = \mathbf{P}(\Sigma(i) = 1) \cdot \mathbf{P}(\Sigma(j) = 1).$$

Lemma 2.16 (Eigenschaften unabhängiger Zufallsvariable). *Sei (X_1, \dots, X_n) eine unabhängige Familie von Zufallsvariablen.*

1. *Jede Teilfamilie X_{i_1}, \dots, X_{i_k} für i_1, \dots, i_k paarweise verschieden ist stochastisch unabhängig.*
2. *Seien h_1, \dots, h_n Funktionen auf den Zielbereichen S_1, \dots, S_n . Dann sind $h_1(X_1), \dots, h_n(X_n)$ unabhängig.*

Beweis. Für 1. wähle man $A_j = S_j$ für $j \notin \{i_1, \dots, i_k\}$ in der Produktformel. Für 2. schreiben wir

$$\begin{aligned} \mathbf{P}(h_1(X_1) \in A_1, \dots, h_n(X_n) \in A_n) &= \mathbf{P}(X_1 \in h_1^{-1}(A_1), \dots, X_n \in h_n^{-1}(A_n)) \\ &= \mathbf{P}(X_1 \in h_1^{-1}(A_1)) \cdots \mathbf{P}(X_n \in h_n^{-1}(A_n)) \\ &= \mathbf{P}(h_1(X_1) \in A_1) \cdots \mathbf{P}(h_n(X_n) \in A_n). \end{aligned}$$

□

Proposition 2.17 (Unabhängigkeit diskreter und stetiger Zufallsvariable). *Seien X_1, \dots, X_n Zufallsvariable mit Werten in S_1, \dots, S_n .*

1. *Im Falle diskreter Zufallsvariablen sind folgende Aussagen äquivalent:*

(i) X_1, \dots, X_n sind unabhängig.

(ii) Es gilt

$$\mathbf{P}(X_1 = a_1, \dots, X_n = a_n) = \mathbf{P}(X_1 = a_1) \cdots \mathbf{P}(X_n = a_n).$$

2. *Im Falle kontinuierlicher Zufallsvariablen sei f_X die Dichte von X und f_1, \dots, f_n die Dichten von X_1, \dots, X_n . Folgende Aussagen sind äquivalent:*

(i) X_1, \dots, X_n sind unabhängig.

(ii) Es gilt

$$f_X(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Beweis. 1. '⇒' folgt direkt aus der Definition von Unabhängigkeit, wenn man $A_1 = \{a_1\}, \dots, A_n = \{a_n\}$ setzt. '⇐' Es gilt

$$\begin{aligned} \mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) &= \sum_{(a_1, \dots, a_n) \in A_1 \times \dots \times A_n} \mathbf{P}(X_1 = a_1, \dots, X_n = a_n) \\ &= \sum_{(a_1, \dots, a_n) \in A_1 \times \dots \times A_n} \mathbf{P}(X_1 = a_1) \cdots \mathbf{P}(X_n = a_n) \\ &= \sum_{a_1 \in A_1} \mathbf{P}(X_1 = a_1) \cdots \sum_{a_n \in A_n} \mathbf{P}(X_n = a_n) = \mathbf{P}(X_1 \in A_1) \cdots \mathbf{P}(X_n \in A_n). \end{aligned}$$

Eine analoge Rechnung liefert 2. □

3 Weitere wichtige Verteilungen

In diesem Kapitel lernen wir die wichtigsten Verteilungen kennen. Wir gehen wieder davon aus, dass Zufallsvariablen definiert sind, und geben diese wichtigen Verteilungen als Bildmaße an.

3.1 Die Binomial- und die Multinomialverteilung

Definition 3.1. Sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Eine Zufallsvariable X mit Zielbereich $\{0, \dots, n\}$ heißt binomialverteilt zu den Parametern n und p , falls (mit $q := 1 - p$)

$$\mathbf{P}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, \dots, n.$$

Wir schreiben dann $X \sim B(n, p)$.

Bemerkung 3.2. 1. Wir haben in Beispiel 0.7 gesehen: Ist $X = (X_1, \dots, X_n)$ ein p -Münzwurf, so ist $X_1 + \dots + X_n$ binomialverteilt zu den Parametern n und p .

Generell erhält man die Binomialverteilung immer dann, wenn man ein Experiment hintereinander unabhängig ausführt, und es in jedem Experiment mit derselben Wahrscheinlichkeit p zu einem *Erfolg* und mit Wahrscheinlichkeit $q := 1 - p$ zu einem *Misserfolg* kommen kann. Zählt man die Anzahl der *Erfolge*, so ist diese $B(n, p)$ -verteilt. Der Grund ist, dass es $\binom{n}{k}$ verschiedene Möglichkeiten gibt, die k Erfolge auf die n Versuche zu verteilen und jede einzelne Abfolge von Erfolgen und Misserfolgen dieselbe Wahrscheinlichkeit (nämlich $p^k q^{n-k}$) hat.

2. Der bekannte binomische Lehrsatz ergibt, dass

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = 1,$$

so dass es sich bei der Binomialverteilung in der Tat um ein Wahrscheinlichkeitsmaß handelt.

Unterscheidet man bei einem Experiment nicht nur zwischen Erfolgen und Misserfolgen, sondern zwischen ℓ verschiedenen Ausgängen, so erhält man die Multinomialverteilung.

Definition 3.3 (Multinomialverteilung). Sei $n \in \mathbb{N}$ und $p_1, \dots, p_\ell \in [0, 1]$ mit $p_1 + \dots + p_\ell = 1$. Eine $\Delta_\ell := \{(k_1, \dots, k_\ell) : k_1 + \dots + k_\ell = n\}$ -wertige Zufallsvariable $X = (X_1, \dots, X_\ell)$ heißt multinomialverteilt zu den Parametern n und p_1, \dots, p_ℓ , falls

$$\mathbf{P}(X = (k_1, \dots, k_\ell)) = \binom{n}{k_1 \dots k_\ell} p_1^{k_1} \dots p_\ell^{k_\ell}.$$

Hierbei ist

$$\binom{n}{k_1 \dots k_\ell} := \frac{n!}{k_1! \dots k_\ell!}$$

ein Multinomialkoeffizient.

3.2 Die hypergeometrische Verteilung

Gegeben seien N unterscheidbare Objekte, von denen K markiert sind. Wie viele Möglichkeiten gibt es, n Objekte (ohne Beachtung der Reihenfolge) so auszuwählen, dass genau k Markierte dabei sind? Die Antwort erhält man wie folgt: von den K Markierten muss man k auswählen, das sind schon mal $\binom{K}{k}$ Möglichkeiten. Bei den restlichen $n - k$ auszuwählenden

Objekten darf man dann nur noch nicht-Markierte nehmen. Dies sind dann $\binom{N-K}{n-k}$ Möglichkeiten. Insgesamt ergeben sich also

$$\binom{K}{k} \binom{N-K}{n-k}$$

Möglichkeiten. Diese Überlegung führt schon auf ein erstes Resultat.

Lemma 3.4 (Eine Identität mit Binomialkoeffizienten). *Es gilt für $N \geq n, K \geq 0$*

$$\sum_{k=0}^n \binom{K}{k} \binom{N-K}{n-k} = \binom{N}{n}.$$

Beweis. Einen anschaulichen Beweis erhält man, wenn man sich die Bedeutung der linken Seite ansieht. Nach obiger Erklärung wird hier über die Anzahl der Markierten Objekte summiert, die man beim Ziehen von n der N Objekte erhält. Insgesamt steht dort also die Anzahl der Möglichkeiten, (ohne Beachtung der Reihenfolge) aus den N Objekten gerade n auszuwählen, also $\binom{N}{n}$. Einen anderen Beweis erhält man durch Induktion. \square

Definition 3.5 (Hypergeometrische Verteilung). *Sei $n, N \in \mathbb{N}$ und $K \in \mathbb{N}_0$ mit $n, K \leq N$. Eine $[0 : n \wedge K]$ -wertige Zufallsvariable X heißt hypergeometrisch verteilt mit n, N, K , wenn*

$$\mathbf{P}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, n \wedge K.$$

Wir schreiben dann auch $X \sim \text{Hyp}(n, N, K)$.

Bemerkung 3.6 (Urnen in der Stochastik). *Betrachte eine Urne mit insgesamt N Kugeln, wovon K weiß sind. Wir ziehen n Kugeln heraus und betrachten*

$$\begin{aligned} X_i &:= 1_{i\text{-te Kugel ist weiß}}, & i = 1, \dots, n, \\ X &:= X_1 + \dots + X_n. \end{aligned}$$

Ziehen wir mit Zurücklegen, so haben wir in jedem Zug die gleiche Chance von $p = \frac{K}{N}$, eine weiße Kugel zu ziehen. In diesem Fall ist $X \sim B(n, p)$.

Ziehen wir ohne Zurücklegen, so sind die Zufallsvariablen X_1, \dots, X_n nicht unabhängig, da jeder Zug die Wahrscheinlichkeiten, im nächsten Zug eine weiße Kugel zu ziehen, ändert. In diesem Fall ist $X \sim \text{Hyp}(n, N, K)$.

3.3 Die Poisson-Verteilung und die Poisson-Approximation

Wie wir in Theorem 3.8 sehen werden, entsteht aus der Binomialverteilung bei sehr kleinen Erfolgswahrscheinlichkeiten, jedoch häufigen Versuchen, die Poisson-Verteilung.

Definition 3.7. *Sei $\lambda \in \mathbb{R}_+$. Eine \mathbb{N}_0 -wertige Zufallsvariable X heißt Poisson-verteilt mit Parameter λ , falls*

$$\mathbf{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

In diesem Fall schreiben wir $X \sim \text{Poi}(\lambda)$.

Theorem 3.8 (Gesetz der kleinen Zahlen). Sei $X_n \sim B(n, p_n)$ für $n = 1, 2, \dots$, so dass

$$n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda > 0.$$

Sei weiter $Z \sim \text{Poi}(\lambda)$. Dann gilt

$$\mathbf{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbf{P}(Z = k).$$

Beweis. Wir schreiben direkt

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

□

3.4 Die geometrische und die Exponentialverteilung

In Beispiel 0.8 haben wir die Verteilung der Wartezeit bis zur ersten 6 beim Würfelwurf berechnet. Solche Wartezeiten sind oftmals geometrisch verteilt.

Definition 3.9 (Geometrische Verteilung). Sei $p \in (0, 1]$. Eine \mathbb{N} -wertige Zufallsvariable X heißt geometrisch verteilt mit Parameter p , falls für $q = 1 - p$

$$\mathbf{P}(X > i) = q^i, \quad i = 1, 2, \dots$$

Wir schreiben dann auch $X \sim \text{Geo}(p)$. Es gilt dann also

$$\mathbf{P}(X = i) = \mathbf{P}(X > i - 1) - \mathbf{P}(X > i) = q^{i-1} - q^i = q^{i-1}p.$$

Bemerkung 3.10 (Interpretation beim p -Münzwurf). 1. Sei $X = (X_1, X_2, \dots)$ ein unendlicher (!) p -Münzwurf und

$$T := \min\{i : T_i = 1\}.$$

Dann ist $T \sim \text{Geo}(p)$.

2. In manchen Büchern nimmt eine geometrisch verteilte Zufallsvariable Y Werte in \mathbb{N}_0 an. In diesem Fall ist bei einem unendlichen p -Münzwurf Y nicht der Versuch des ersten Erfolges, sondern der letzte in der Serie von Misserfolgen, d.h. $Y = T - 1$ mit T wie in 1.
3. Für den unendlichen p -Münzwurf $X = (X_1, X_2, \dots)$ bemerken wir kurz, dass der Wahrscheinlichkeitsraum, auf dem X definiert ist, nicht diskret sein kann, da es überabzählbar viele $\{0, 1\}$ -wertige Folgen gibt. Immerhin können wir den unendlichen Münzwurf im Rahmen unserer Möglichkeiten für $p = 1/2$ dennoch definieren. Hierfür sei $\Omega = [0, 1]$ und \mathbf{P} eine $U([0, 1])$ -Verteilung. Für $x \in \Omega$ sei weiter $x = \sum_{i=1}^{\infty} x_i 2^{-i}$ die Binärdarstellung von x , also $x_i = [2^i x - \lfloor 2^i x \rfloor]$, $i = 1, 2, \dots$, wobei $[.]$ die Rundung und $\lfloor . \rfloor$ die Abrundung bedeutet. Dann ist mit $X_i(x) = x_i$ der Vektor $X = (X_1, X_2, \dots)$ der unendliche $1/2$ -Münzwurf.

Wartezeiten sind nicht immer diskrete Anzahlen von Versuchen, sondern können auch kontinuierliche Größen sein. Dies wird zumeist durch Exponentialverteilungen dargestellt.

Definition 3.11 (Exponentialverteilung). *Sei $\lambda > 0$. Eine \mathbb{R} -wertige Zufallsvariable X heißt exponentialverteilt mit Rate λ , falls sie die Dichte*

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0} \quad (3.1)$$

und die Verteilungsfunktion

$$F_X(x) = (1 - e^{-\lambda x}) \mathbf{1}_{x \geq 0} \quad (3.2)$$

besitzt. Dies ist genau dann der Fall, wenn

$$\mathbf{P}(X > x) = e^{-\lambda x}, \quad x \geq 0.$$

Wir schreiben dann $X \sim \text{Exp}(\lambda)$.

Proposition 3.12 (Eigenschaften der Exponentialverteilung). *Sei $\lambda, \mu > 0$.*

1. *Sei $X \sim \text{Exp}(\lambda)$. Dann ist $\mu X \sim \text{Exp}(\lambda/\mu)$*
2. *Sei $U \sim U([0, 1])$. Dann ist $-(\log U)/\lambda \sim \text{Exp}(\lambda)$.*

Beweis. Sei $x \geq 0$. 1. Für $Y := \mu X$ gilt für $x \geq 0$

$$\mathbf{P}(Y > x) = \mathbf{P}(X > x/\mu) = e^{-\lambda x/\mu}.$$

Daraus folgt bereits die Behauptung. 2. folgt einerseits mit Lemma 2.9. Alternativ schreiben wir für $Z := (-\log U)/\lambda$

$$\mathbf{P}(Z > x) = \mathbf{P}(-(\log U)/\lambda > x) = \mathbf{P}(U > e^{-\lambda x}) = e^{-\lambda x}.$$

Wieder folgt die Behauptung. □

Proposition 3.13 (Exponentialapproximation). *Sei $Z \sim \text{Exp}(1)$ und X_1, X_2, \dots eine Folge von geometrisch verteilten Zufallsvariablen zu den Parametern p_1, p_2, \dots mit $p_n \downarrow 0$. Dann gilt für $0 \leq c < d \leq \infty$*

$$\mathbf{P}(c \leq p_n X_n \leq d) \xrightarrow{n \rightarrow \infty} \mathbf{P}(c \leq Z \leq d).$$

Beweis. Sei

$$c_n := (\lceil c/p_n \rceil - 1)p_n, \quad d_n := (\lfloor d/p_n \rfloor - 1)p_n,$$

also $c_n \xrightarrow{n \rightarrow \infty} c, d_n \xrightarrow{n \rightarrow \infty} d$. Dann

$$\begin{aligned} \mathbf{P}(c \leq p_n X_n \leq d) &= \mathbf{P}(X_n \geq c/p_n) - \mathbf{P}(X_n > d/p_n) \\ &= (1 - p_n)^{\lceil c/p_n \rceil - 1} - (1 - p_n)^{\lfloor d/p_n \rfloor - 1} \\ &= \left((1 - p_n)^{1/p_n} \right)^{c_n} - \left((1 - p_n)^{1/p_n} \right)^{d_n} \\ &\xrightarrow{n \rightarrow \infty} e^{-c} - e^{-d} = \int_c^d e^{-x} dx. \end{aligned}$$

□

3.5 Die Normalverteilung

In der Statistik werden wir sehen, dass die Normalverteilung grundlegend für viele Anwendungen ist. Sie wird oft mit der Gauß'schen Glockenkurve in Verbindung gebracht. Dies ist der Graph der Funktion

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Aus der Analysis bekannt ist die Tatsache⁴

$$\int e^{-x^2} dx = \sqrt{\pi}.$$

Daraus folgt, dass $\int f(x)dx = 1$ und damit f eine Dichte ist.

Definition 3.14. Sei $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$. Eine \mathbb{R} -wertige Zufallsvariable heißt normalverteilt mit Erwartungswert μ und Varianz σ^2 , falls sie die Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3.3)$$

besitzt. Wir schreiben dann auch $X \sim N(\mu, \sigma^2)$. Im Fall $\mu = 0, \sigma^2 = 1$ sagen wir auch, X sei standardnormalverteilt.

Lemma 3.15 (Eigenschaften der Normalverteilung). Sei $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$.

1. Ist $Z \sim N(0, 1)$, so ist $\sigma Z + \mu \sim N(\mu, \sigma^2)$.
2. Ist $X \sim N(\mu, \sigma^2)$, so ist $\frac{X-\mu}{\sigma} \sim N(0, 1)$.

⁴Vermutlich wird dies erst am Ende der Vorlesung Analysis 3 bewiesen. Ein elementarer Beweis ist der folgende, basierend auf der Gamma-Funktion

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Wir schreiben

$$\begin{aligned} \Gamma\left(\frac{1}{2}\right)^2 &= \int_0^\infty \int_0^\infty s^{-1/2} t^{-1/2} e^{-(s+t)} ds dt \\ &\stackrel{r=t+s}{=} \int_0^\infty \int_0^r s^{-1/2} (r-s)^{-1/2} e^{-r} ds dr \\ &\stackrel{u=s/r}{=} \int_0^\infty \int_0^1 r u^{-1/2} r^{-1/2} r^{-1/2} (1-u)^{-1/2} e^{-r} du dr \\ &= \int_0^\infty e^{-r} dr \cdot \int_0^1 \frac{1}{\sqrt{u(1-u)}} du \\ &\stackrel{s=2u-1}{=} \int_{-1}^1 \frac{ds}{\sqrt{1-s^2}} = \arcsin(s)|_{-1}^1 = \pi. \end{aligned}$$

Daraus folgt

$$\int_{-\infty}^\infty e^{-x^2} dx = 2 \int_0^\infty e^{-x^2} dx \stackrel{y=x^2}{=} \int_0^\infty e^{-y} y^{-\frac{1}{2}} dy = \Gamma(1/2) = \sqrt{\pi}.$$

Beweis. 1. Für $x \in \mathbb{R}$ schreiben wir

$$\begin{aligned} \mathbf{P}(\sigma Z + \mu \leq x) &= \mathbf{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &\stackrel{y=\sigma z + \mu}{=} \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) dy. \end{aligned}$$

2. folgt analog. □

4 Kenngrößen von Zufallsvariablen

Um sich etwas unter einer Verteilung vorstellen zu können, ist es oftmals hilfreich, diese durch Kenngrößen zu beschreiben.

4.1 Der Erwartungswert

Wenn man eine Verteilung einer Zufallsvariable durch eine einzige Zahl beschreiben müsste, würde man wohl einen Lageparameter der Verteilung wählen. Diese beschreiben grob, wo Werte von X am ehesten liegen. Der häufigste Lageparameter ist der Erwartungswert.

Definition 4.1 (Erwartungswert). *Sei X eine $S \subseteq \mathbb{R}$ -wertige Zufallsvariable.*

1. *Ist X diskret, so ist der Erwartungswert von (der Verteilung von) X*

$$\mathbf{E}[X] := \sum_{x \in S} x \cdot \mathbf{P}(X = x),$$

falls die Summe wohldefiniert ist.

2. *Ist X kontinuierlich mit Dichte f , so ist Erwartungswert von (der Verteilung von) X*

$$\mathbf{E}[X] := \int x f(x) dx,$$

falls das Integral wohldefiniert ist.

Bemerkung 4.2 (k -tes Moment; weitere Lageparameter). 1. Allgemeiner als der Erwartungswert ist das k -te Moment einer Zufallsvariable X , gegeben als $\mathbf{E}[X^k]$, $k = 1, 2, \dots$

2. Neben dem Erwartungswert gibt es noch weitere Lageparameter einer Verteilung, nämlich Median und Modalwert. Der Median der Verteilung von X ist dabei jede Zahl m mit

$$\mathbf{P}(X \geq m) \geq \frac{1}{2}, \quad \mathbf{P}(X \leq m) \geq \frac{1}{2}.$$

Der Modalwert ist $\operatorname{argmax} \mathbf{P}(X = x)$ im Falle diskreter und $\operatorname{argmax} f_X(x)$ im Falle einer Zufallsvariablen mit Dichte.

Der Erwartungswert ist von allen Lageparametern oft am besten handhabbar, was vor allem an seiner Linearität liegt (Proposition 4.5).

Lemma 4.3 (Transformation von Erwartungswerten). *Sei X eine S -wertige Zufallsvariable und $h : S \rightarrow \mathbb{R}$.*

1. Ist X diskret, so gilt

$$\mathbf{E}[h(X)] = \sum_{x \in S} h(x) \mathbf{P}(X = x),$$

falls die Summe existiert.

2. Ist X kontinuierlich mit Dichte f , und ist h streng monoton und differenzierbar mit differenzierbarer Umkehrabbildung, so gilt

$$\mathbf{E}[h(X)] = \int h(x) f(x) dx.$$

Beweis. 1. Wir schreiben

$$\begin{aligned} \mathbf{E}[h(X)] &= \sum_{y \in h(S)} y \mathbf{P}(h(X) = y) = \sum_{y \in h(S)} y \sum_{x \in h^{-1}(y)} \mathbf{P}(X = x) \\ &= \sum_{y \in h(S)} \sum_{x \in h^{-1}(y)} h(x) \mathbf{P}(X = x) \\ &= \sum_{x \in S} h(x) \mathbf{P}(X = x) \end{aligned}$$

2. OBdA sei h streng monoton wachsend. (Ansonsten wechseln wir zu $-h$.) Nach Lemma 2.7 hat $h(X)$ die Dichte $y \mapsto f(h^{-1}(y)) |(h^{-1})'(y)|$. Deshalb gilt

$$\mathbf{E}[h(X)] = \int y f(h^{-1}(y)) |(h^{-1})'(y)| dy \stackrel{z=h^{-1}(y)}{=} \int h(z) f(z) dz.$$

□

Bemerkung 4.4. In 2. mussten wir starke Annahmen an h machen, um $\mathbf{E}[h(X)]$ berechnen zu können. Richtig ist allerdings, dass die Aussage auch für viel allgemeinere Funktionen h gilt, nämlich für alle, für die $\int f(x) h(x) dx$ existiert. Dies kann man allerdings erst mit Maßtheorie einsehen. Wir werden zur Vereinfachung diese Aussage auch für allgemeine Funktionen h verwenden.

Wir kommen nun zu Eigenschaften des Erwartungswertes.

Proposition 4.5 (Linearität des Erwartungswertes). 1. Sei X eine S -wertige Zufallsvariable und $A \subseteq S$ (messbar). Dann gilt

$$\mathbf{E}[1_A(X)] = \mathbf{P}(X \in A).$$

2. Seien $c, d \in \mathbb{R}$ und X, Y Zufallsvariablen mit Zielbereichen $S, T \subseteq \mathbb{R}$ und wohldefinier-tem Erwartungswert. Dann gilt

$$\mathbf{E}[cX + dY] = c\mathbf{E}[X] + d\mathbf{E}[Y].$$

Beweis. 1. Es gilt

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

2. Sind X, Y diskret, so gilt mit Lemma 4.3

$$\begin{aligned}\mathbf{E}[cX + dY] &= \sum_{X \in S} \sum_{y \in T} (cx + dy) \mathbf{P}(X = x, Y = y) \\ &= c \sum_{x \in S} x \mathbf{P}(X = x) + d \sum_{y \in T} y \mathbf{P}(Y = y) \\ &= c\mathbf{E}[X] + d\mathbf{E}[Y].\end{aligned}$$

Haben X, Y die gemeinsame Dichte f , und haben die Marginalverteilungen f_X und f_Y , so ist

$$\begin{aligned}\mathbf{E}[cX + dY] &= \int (cx + dy) f(x, y) dx dy \\ &= c \int x \int f(x, y) dy dx + d \int y \int f(x, y) dx dy \\ &= c \int x f_X(x) dx + d \int y f_Y(y) dy \\ &= c\mathbf{E}[X] + d\mathbf{E}[Y].\end{aligned}$$

□

Beispiel 4.6 (Fixpunkte in Permutationen). Sei Σ eine zufällige Permutation von $[1 : n]$. Weiter sei

$$X = \sum_{i=1}^n X_i, \quad X_i = 1_{\Sigma(i)=i}$$

die Anzahl an Fixpunkten in Σ . Wir zeigen nun, dass

$$\mathbf{E}[X] = 1.$$

Denn mit Linearität des Erwartungswertes gilt

$$\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \sum_{i=1}^n \mathbf{P}(\Sigma(i) = i) = n \cdot \frac{1}{n} = 1.$$

Beispiel 4.7 (Runs in Münzwürfen). Sei $X = (X_1, \dots, X_n)$ ein p -Münzwurf. Ein *Run* ist ein maximaler Block aus 0ern oder 1ern, also enthält beispielsweise 1100010 genau vier Runs. Bezeichnen wir mit Y die Anzahl der Runs, so können wir das formalisieren mittels

$$Y_n = 1 + \sum_{i=2}^n 1_{E_i}, \quad E_i = \{X_i \neq X_{i-1}\}.$$

Mit dieser Formel lässt sich nun ganz einfach die erwartete Anzahl von Runs berechnen. Es gilt nämlich

$$\mathbf{E}[1_{E_i}] = \mathbf{P}(X_i \neq X_{i-1}) = 2pq$$

und damit

$$\mathbf{E}[Y] = 1 + (n-1)\mathbf{E}[1_{E_1}] = 1 + 2(n-1)pq.$$

Beispiel 4.8. In einer Urne befinden sich $1 \leq w \leq g$ markierte und $g - w$ unmarkierte Kugeln. Wir ziehen die Kugeln ohne Zurücklegen aus der Urne, und schreiben $X_i = 1$, falls die i -te Kugel markiert ist und $X_i = 0$ sonst. Wir wollen $\mathbf{E}[T]$ für $T := \min\{i : T_i = 1\}$, also den ersten Zug einer markierten Kugel, berechnen. Wir behaupten

$$\mathbf{E}[T] = \frac{g+1}{w+1}.$$

Eine erste Möglichkeit ist die folgende: Die Verteilung von T lässt sich angeben als

$$\mathbf{P}(T = k+1) = \frac{(g-w) \cdots (g-w-k+1)}{g \cdots (g-k+1)} \frac{w}{g-k} = \frac{\binom{g-w}{k} w}{\binom{g-1}{k} g}, \quad k = 0, \dots, g-w.$$

Wir zeigen die Behauptung nun mittels Induktion nach g . Für $g = 1$ muss $w = 1$ sein und die Behauptung ist klar. Gilt die Behauptung für $g-1$, so ist für g

$$\begin{aligned} \mathbf{E}[T] &= \sum_{k=0}^{g-w} (k+1) \mathbf{P}(T = k+1) = 1 + \sum_{k=0}^{g-w} k \mathbf{P}(T = k+1) \\ &= 1 + \frac{w}{g} \sum_{k=0}^{g-w} k \frac{\binom{g-w}{k}}{\binom{g-1}{k}} = 1 + \frac{w}{g} \sum_{k=0}^{g-w-1} (k+1) \frac{\binom{g-w}{k+1}}{\binom{g-1}{k+1}} \\ &= 1 + \frac{g-w}{g} \underbrace{\sum_{k=0}^{g-w-1} (k+1) \frac{\binom{g-1-w}{k}}{\binom{g-2}{k}} \frac{w}{g-1}}_{= \frac{g}{w+1} \text{ nach Induktionsannahme}} \\ &= 1 + \frac{g-w}{w+1} = \frac{g+1}{w+1}. \end{aligned}$$

Eine zweite Möglichkeit ist die folgende: Wir definieren induktiv

$$T_0 := 0, \quad T_{\ell+1} := \min\{i > T_\ell : X_i = 1\}, \quad \ell = 0, \dots, w-1$$

sowie $T_{w+1} = g+1$. Dann sind $T_{\ell+1} - T_\ell, \ell = 0, \dots, w$ identisch verteilt mit

$$g+1 = T_{w+1} - T_0 = \sum_{\ell=0}^w T_{\ell+1} - T_\ell.$$

Deshalb gilt

$$g+1 = \sum_{\ell=0}^w \mathbf{E}[T_{\ell+1} - T_\ell] = (w+1) \mathbf{E}[T_1] = \mathbf{E}[T]$$

und die Behauptung folgt.

Proposition 4.9 (Positivität und Monotonie des Erwartungswertes). *1. Sei $X \geq 0$ eine \mathbb{R} -wertige Zufallsvariable. Dann gilt*

$$\mathbf{E}[X] \geq 0$$

und

$$\mathbf{E}[X] = 0 \quad \text{genau dann, wenn} \quad \mathbf{P}(X = 0) = 1.$$

2. Für \mathbb{R} -wertige Zufallsvariable X_1, X_2 mit $X_1 \leq X_2$ mit wohldefinierten Erwartungswerten gilt $\mathbf{E}[X_1] \leq \mathbf{E}[X_2]$.

Beweis. 1. Ist X diskret, so ergeben sich die ersten beiden Aussagen aus der Definition

$$\mathbf{E}[X] = \sum_{x \in \mathbb{R}} x \mathbf{P}(X = x).$$

Hat X die Dichte f , so muss $f(x) = 0$ für $x \leq 0$ gelten, da sonst $\mathbf{P}(X < 0) > 0$ wäre. Dann ist $\mathbf{E}[X] = \int_0^\infty x f(x) dx > 0$ wegen der Positivität des Integranden. Insbesondere kann die zweite Aussage dann nicht zutreffen. Die letzte Aussage folgt aus der ersten, da

$$0 \leq \mathbf{E}[X_2 - X_1] = \mathbf{E}[X_2] - \mathbf{E}[X_1].$$

□

Proposition 4.10 (Jensen'sche Ungleichung). Sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ konvex und X eine \mathbb{R} -wertige Zufallsvariable mit endlichem Erwartungswert. Dann ist

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)].$$

Insbesondere gilt

$$|\mathbf{E}[X]| \leq \mathbf{E}[|X|], \quad \mathbf{E}[X]^2 \leq \mathbf{E}[X^2].$$

Beweis. Da φ konvex ist, gibt es für jedes $d \in \mathbb{R}$ ein $c \in \mathbb{R}$, so dass $\varphi(d) + c(x - d) \leq \varphi(x)$ für alle $x \in \mathbb{R}$. Für $d = \mathbf{E}[X]$ ist also

$$\varphi(\mathbf{E}[X]) = \mathbf{E}[\varphi(\mathbf{E}[X]) + c(X - \mathbf{E}[X])] \leq \mathbf{E}[\varphi(X)].$$

□

Beispiel 4.11 (Arithmetisches und geometrisches Mittel). Seien $x_1, \dots, x_n \geq 0$ reelle Zahlen. Bekanntlich ist $\bar{x}_{\text{ar}} := \frac{1}{n} \sum_{i=1}^n x_i$ das arithmetische und $\bar{x}_{\text{geo}} = \left(\prod_{i=1}^n x_i \right)^{1/n}$ das geometrische Mittel. Wir zeigen nun

$$\bar{x}_{\text{ar}} \geq \bar{x}_{\text{geo}}.$$

Denn: Sei X uniform auf x_1, \dots, x_n verteilt, d.h. $\mathbf{P}(X = x_i) = \frac{1}{n}$. Da $x \mapsto \log x$ konkav ist, verwenden wir die Jensen'sche Ungleichung und sehen

$$\log(\bar{x}_{\text{ar}}) = \log\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \log \mathbf{E}[X] \geq \mathbf{E}[\log X] = \frac{1}{n} \sum_{i=1}^n \log x_i = \frac{1}{n} \log \prod_{i=1}^n x_i = \log(\bar{x}_{\text{geo}})$$

Proposition 4.12 (Cauchy-Schwarz Ungleichung). Seien X, Y reellwertige Zufallsvariable mit $\mathbf{E}[X^2], \mathbf{E}[Y^2] < \infty$. Dann gilt

$$\mathbf{E}[XY]^2 \leq \mathbf{E}[X^2] \mathbf{E}[Y^2].$$

Beweis. Im Fall $\mathbf{P}(X = 0) = 1$ oder $\mathbf{P}(Y = 0) = 1$ ist die Ungleichung trivial. Andernfalls betrachte

$$U := \frac{X}{\sqrt{\mathbf{E}[X^2]}}, \quad V := \frac{Y}{\sqrt{\mathbf{E}[Y^2]}}.$$

Wegen $(U - V)^2, (U + V)^2 \geq 0$ ist klar, dass

$$\pm 2UV \leq U^2 + V^2, \text{ also } \pm \mathbf{E}[UV] \leq 1.$$

Multiplizieren der letzten Gleichung mit $\sqrt{\mathbf{E}[X^2]} \cdot \sqrt{\mathbf{E}[Y^2]}$ und Quadrieren liefert die Behauptung. \square

Proposition 4.13 (Multiplikation von Erwartungswerten bei Unabhängigkeit). *Seien X, Y reellwertige Zufallsvariable mit endlichen Erwartungswerten. Sind X, Y unabhängig, so gilt*

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

Beweis. Falls X, Y diskret sind, schreiben wir

$$\begin{aligned} \mathbf{E}[XY] &= \sum_{x,y} xy\mathbf{P}(X = x, Y = y) = \sum_{x,y} xy\mathbf{P}(X = x)\mathbf{P}(Y = y) \\ &= \sum_x x\mathbf{P}(X = x) \cdot \sum_y y\mathbf{P}(Y = y) = \mathbf{E}[X]\mathbf{E}[Y]. \end{aligned}$$

Haben X, Y die gemeinsame Dichte f und $f(x, y) = f_X(x)f_Y(y)$ wegen der Unabhängigkeit (siehe Proposition 4.3), so folgt

$$\begin{aligned} \mathbf{E}[XY] &= \int xyf(x, y)dxdy = \int xf_X(x)yf_Y(y)dxdy = \int xf_X(x)dx \cdot \int yf_Y(y)dy \\ &= \mathbf{E}[X]\mathbf{E}[Y]. \end{aligned}$$

\square

Beispiel 4.14. Seien X, Y die Ergebnisse zweier unabhängiger Würfelwürfe eines fairen Würfels. Dann ist

$$\begin{aligned} \mathbf{P}(XY = 1) &= \frac{1}{36}, & \mathbf{P}(XY = 2) &= \frac{2}{36}, & \mathbf{P}(XY = 3) &= \frac{2}{36}, & \mathbf{P}(XY = 4) &= \frac{3}{36}, \\ \mathbf{P}(XY = 5) &= \frac{2}{36}, & \mathbf{P}(XY = 6) &= \frac{4}{36}, & \mathbf{P}(XY = 8) &= \frac{2}{36}, & \mathbf{P}(XY = 9) &= \frac{1}{36}, \\ \mathbf{P}(XY = 10) &= \frac{2}{36}, & \mathbf{P}(XY = 12) &= \frac{4}{36}, & \mathbf{P}(XY = 15) &= \frac{2}{36}, & \mathbf{P}(XY = 16) &= \frac{1}{36}, \\ \mathbf{P}(XY = 18) &= \frac{2}{36}, & \mathbf{P}(XY = 20) &= \frac{2}{36}, & \mathbf{P}(XY = 24) &= \frac{2}{36}, & \mathbf{P}(XY = 25) &= \frac{1}{36}, \\ \mathbf{P}(XY = 30) &= \frac{2}{36}, & \mathbf{P}(XY = 36) &= \frac{1}{36}, & & & & \end{aligned}$$

und damit

$$\begin{aligned} \mathbf{E}[XY] &= \frac{1}{36} + \frac{4}{36} + \frac{6}{36} + \frac{12}{36} + \frac{10}{36} + \frac{24}{36} \\ &\quad + \frac{16}{36} + \frac{9}{36} + \frac{20}{36} + \frac{48}{36} + \frac{30}{36} + \frac{16}{36} + \frac{36}{36} + \frac{40}{36} + \frac{48}{36} + \frac{25}{36} + \frac{60}{36} + \frac{39}{36} = \frac{49}{4}. \end{aligned}$$

Bekanntlich ist auch

$$\mathbf{E}[X] = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}, \quad \text{also} \quad \mathbf{E}[X]\mathbf{E}[Y] = \frac{49}{4}.$$

Wir sehen, dass die Berechnung von $\mathbf{E}[XY]$ als Produkt $\mathbf{E}[X]\mathbf{E}[Y]$ deutlich einfacher ist.

4.2 Die Varianz und die Kovarianz

Will man die Verteilung einer Zufallsvariable mit zwei Parametern beschreiben, so kommt neben einem Lageparameter meist noch ein Streuungsparameter (oder Skalierungsparameter) hinzu. Dieser gibt an, wie eng X um den Lageparameter herum verteilt ist. Der wichtigste Streuungsparameter ist dabei die Varianz einer Zufallsvariable.

Definition 4.15 (Varianz und Kovarianz). *Seien X, Y eine $S \subseteq \mathbb{R}$ -wertige Zufallsvariable mit endlichem Erwartungswert.*

1. Die Varianz von X ist

$$\sigma^2 := \mathbf{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2].$$

Weiter ist die Standardabweichung von X gegeben durch

$$\sigma := \sqrt{\mathbf{Var}[X]}.$$

2. Die Kovarianz von X, Y ist

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

Ist $\mathbf{Cov}[X, Y] = 0$, so heißen X, Y unkorreliert.

Bemerkung 4.16 (k -tes zentriertes Moment). Allgemeiner als die Varianz ist das k -te zentrierte Moment einer Zufallsvariable X , gegeben als $\mathbf{E}[(X - \mathbf{E}[X])^k]$, $k = 2, 3, \dots$

Lemma 4.17 (Rechenregeln für die Varianz und die Kovarianz). *Seien X, Y, Z, X_1, \dots, X_n reellwertige Zufallsvariable mit endlichem Erwartungswert. Es gilt*

1.
$$\mathbf{Var}[X] = \mathbf{Cov}[X, X] = \mathbf{E}[X(X - 1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

2.
$$\mathbf{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

3.
$$\mathbf{Cov}[aX + bY, Z] = a\mathbf{Cov}[X, Z] + b\mathbf{Cov}[Y, Z].$$

4.
$$\mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{Var}[X_i] + \sum_{1 \leq i \neq j \leq n} \mathbf{Cov}[X_i, X_j].$$

5. Sind X, Y unabhängig, so ist $\mathbf{Cov}[X, Y] = 0$.

Beispiel 4.18 (Unkorrelierte Zufallsvariable). Wir betrachten noch ein Beispiel zu Lemma 4.17.5. Seien (X, Y) Zufallsvariable, die jeweils mit Wahrscheinlichkeit $\frac{1}{4}$ die Zustände $(1, 0), (0, 1), (-1, 0)$ und $(0, -1)$ annehmen. Dann ist $\mathbf{E}[XY] = \mathbf{E}[X] = \mathbf{E}[Y] = 0$. Insbesondere sind X, Y unkorreliert.

Wir haben zwar gerade gezeigt, dass unabhängige Zufallsvariable immer unkorreliert sind. In obigem Beispiel ist allerdings

$$\mathbf{P}(X = 1, Y = 1) = 0 \neq \frac{1}{16} = \mathbf{P}(X = 1) \cdot \mathbf{P}(Y = 1),$$

also sind X, Y zwar unkorreliert, jedoch nicht unabhängig.

Beweis von Lemma 4.17. 2. Es gilt wegen der Linearität des Erwartungswertes

$$\begin{aligned}\mathbf{Cov}[X, Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[XY] - \mathbf{E}[X\mathbf{E}[Y]] - \mathbf{E}[\mathbf{E}[X]Y] + \mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

1. Die erste Gleichheit liest man direkt aus der Definition ab. Die dritte Gleichheit folgt aus 1. und der ersten Gleichheit. Die zweite Gleichheit folgt aus der dritten Gleichheit und der Linearität des Erwartungswertes.

3. ist eine einfache Anwendung von 2.

4. Wir schreiben mit 3.

$$\mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \mathbf{Cov}\left[\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right] = \sum_{i=1}^n \mathbf{Cov}[X_i, X_i] + \sum_{1 \leq i \neq j \leq n} \mathbf{Cov}[X_i, X_j]$$

und die Behauptung folgt mit 1.

5. folgt aus Lemma 4.13. \square

Beispiel 4.19 (Runs in Münzwürfen). Wir betrachten nochmal das Beispiel 4.7, wollen nun jedoch die Varianz der Anzahl von Runs mit Hilfe von Lemma (4.17).4 berechnen. Ganz formal schreiben wir

$$\mathbf{Var}[Y] = \mathbf{Var}\left[\sum_{i=2}^n 1_{E_i}\right] = \sum_{i=2}^n \mathbf{Var}[1_{E_i}] + \sum_{i \neq j} \mathbf{Cov}[1_{E_i}, 1_{E_j}].$$

Zunächst ist

$$\mathbf{V}[1_{E_i}] = \mathbf{E}[1_{E_i}^2] - (2pq)^2 = \mathbf{E}[1_{E_i}] - (2pq)^2 = 2pq(1 - 2pq).$$

Weiter sind $1_{E_i}, 1_{E_j}$ für $|i - j| > 1$ unabhängig, und für $2 \leq i \leq n - 1$ ist

$$\begin{aligned}\mathbf{Cov}[1_{E_i}, 1_{E_{i+1}}] &= \mathbf{E}[1_{E_i} \cdot 1_{E_{i+1}}] - (2pq)^2 = \mathbf{P}(X_{i+1} \neq X_i \neq X_{i-1}) - (2pq)^2 \\ &= pqp + qpq - (2pq)^2 = pq(1 - 4pq).\end{aligned}$$

Setzen wir dies zusammen, so ist

$$\begin{aligned}\mathbf{Var}[Y] &= \sum_{i=2}^n \mathbf{Var}[1_{E_i}] + 2 \sum_{i=2}^{n-1} \mathbf{Cov}[1_{E_i}, 1_{E_{i+1}}] \\ &= 2pq(1 - 2pq)(n - 1) + 2pq(1 - 4pq)(n - 2).\end{aligned}$$

Beispiel 4.20 (Fixpunkte in Permutationen). Wir betrachten die Situation aus Beispiel 4.6, sei also $\Sigma \sim U(\mathbb{S}(n))$ und

$$X = \sum_{i=1}^n X_i, \quad X_i = 1_{\Sigma(i)=i}.$$

Ziel ist es, $\mathbf{Var}[X]$ zu bestimmen. Wir berechnen

$$\begin{aligned}\mathbf{P}(X_i = 1) &= \frac{1}{n}, \\ \mathbf{P}(X_i = X_j = 1) &= \frac{1}{n(n-1)}.\end{aligned}$$

Daraus ergibt sich für $i \neq j$

$$\begin{aligned}\mathbf{Var}[X_i] &= \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}, \\ \mathbf{Cov}[X_i, X_j] &= \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)},\end{aligned}$$

und insgesamt

$$\mathbf{Var}[X] = n \frac{n-1}{n^2} + n(n-1) \frac{1}{n^2(n-1)} = \frac{n-1}{n} + \frac{1}{n} = 1.$$

Wir formulieren noch die Cauchy-Schwarz-Ungleichung im Kontext von Varianzen und Kovarianzen.

Korollar 4.21 (Cauchy-Schwarz-Ungleichung). *Seien X, Y reellwertige Zufallsvariable mit endlichen Varianzen. Dann gilt*

$$\mathbf{Cov}[X, Y]^2 \leq \mathbf{Var}[X] \mathbf{Var}[Y].$$

Insbesondere ist

$$-1 \leq \rho(X, Y) \leq 1$$

für den Korrelationskoeffizienten

$$\rho(X, Y) := \frac{\mathbf{Cov}[X, Y]}{\sqrt{\mathbf{Var}[X]} \sqrt{\mathbf{Var}[Y]}}.$$

Beweis. Die Behauptungen folgen direkt aus Lemma 4.12, wenn wir dieses Lemma auf $X - \mathbf{E}[X]$ und $Y - \mathbf{E}[Y]$ anwenden. \square

4.3 Die erzeugende Funktion und die Laplace-Transformierte

Während Erwartungswert und Varianz nur zwei Zahlen sind, die die Verteilung einer Zufallsvariablen X beschreiben, geben die erzeugende Funktion (oder die Laplace-Transformierte) gleich eine ganze Funktion an, die die Verteilung von X beschreiben. Wie sich herausstellt, sind diese Funktionen so informativ, dass sie die Verteilung von X eindeutig beschreiben; siehe Proposition 4.24 und Bemerkung 4.25.

Definition 4.22 (Erzeugende Funktion und Laplace-Transformierte). *Sei X eine reellwertige Zufallsvariable. Die Abbildungen*

$$\chi_X : \begin{cases} [0, 1] & \rightarrow [0, \infty), \\ s & \mapsto \mathbf{E}[s^X] \end{cases}, \quad \mathcal{L}_X : \begin{cases} [0, \infty) & \rightarrow [0, \infty), \\ t & \mapsto \mathbf{E}[e^{-tX}] \end{cases}$$

heißen (Wahrscheinlichkeits-)Erzeugende Funktion und Laplace-Transformierte von X .

Bemerkung 4.23 (Zusammenhang zwischen erzeugenden Funktionen und Laplace-Transformierten). Klar ist, dass

$$\mathcal{L}_X(t) = \chi_X(e^{-t}) \quad \text{sowie} \quad \chi_X(s) = \mathcal{L}_X(-\log s),$$

d.h. dass sich beide Funktionen schnell ineinander umrechnen lassen.

Wir beginnen mit einer Aussage im Fall diskreter Zufallsvariablen, die den Zusammenhang zwischen erzeugender Funktion und Verteilung einer Zufallsvariablen beschreibt.

Proposition 4.24 (Berechnung der Verteilung aus der erzeugenden Funktion). *Sei X eine \mathbb{N}_0 -wertige, diskrete Zufallsvariable mit erzeugender Funktion χ . Dann gilt für $k = 0, 1, 2, \dots$*

$$\mathbf{P}(X = k) = \frac{1}{k!} \frac{d^k}{ds^k} \chi(s) \Big|_{s=0}.$$

Insbesondere bestimmt χ die Verteilung von X eindeutig.

Beweis. Wir berechnen, da alle Summanden positiv sind und die Reihe damit absolut konvergiert,

$$\frac{d^k}{ds^k} \chi(s) \Big|_{s=0} = \frac{d^k}{ds^k} \sum_{i=0}^{\infty} s^i \mathbf{P}(X = i) \Big|_{s=0} = \sum_{i=k}^{\infty} \frac{i!}{(i-k)!} s^{i-k} \mathbf{P}(X = i) \Big|_{s=0} = k! \mathbf{P}(X = k).$$

□

Bemerkung 4.25 (Laplace-Transformierte bestimmt die Verteilung eindeutig). Das letzte Resultat lässt sich mit maßtheoretischen Methoden auch auf stetige Zufallsvariable übertragen. Es gilt: Sei X eine \mathbb{R} -wertige, kontinuierliche Zufallsvariable mit Laplace-Transformierten \mathcal{L} , die zumindest auf \mathbb{R}_+ existiert. Dann ist die Dichte f_X von X eindeutig durch \mathcal{L} bestimmt.

Erzeugende Funktionen und Laplace-Transformierte sind äußerst praktisch, um Erwartungswerte und Varianzen zu berechnen.

Proposition 4.26 (Momente und erzeugende Funktionen). *Sei X eine reellwertige, diskrete Zufallsvariable mit erzeugender Funktion χ . Dann gilt*

$$\mathbb{E}[X] = \frac{d}{ds} \chi(s) \Big|_{s=1}, \quad \mathbb{E}[X(X-1)] = \frac{d^2}{ds^2} \chi(s) \Big|_{s=1},$$

falls die Ableitungen existieren. Außerdem gilt für eine reellwertige Zufallsvariable Y mit Laplace-Transformierter \mathcal{L}

$$\mathbb{E}[X] = -\frac{d}{dt} \mathcal{L}(t) \Big|_{t=0}, \quad \mathbb{E}[X^2] = \frac{d^2}{dt^2} \mathcal{L}(t) \Big|_{t=0}$$

falls die Ableitungen existieren.

Bemerkung 4.27 (Berechnung der Varianz). Für die Varianz einer Zufallsvariable X gilt insbesondere (siehe Lemma 4.17)

$$\begin{aligned} \mathbf{Var}[X] &= \chi''_X(1) + \chi'_X(1) - (\chi'_X(1))^2, \\ \mathbf{Var}[X] &= \mathcal{L}''_X(0) - (\mathcal{L}'_X(0))^2. \end{aligned}$$

Beweis von Proposition 4.26. Sei $S = \{k_0, k_1, \dots\}$ der Zielbereich von S . Wir schreiben für die erzeugende Funktion

$$\begin{aligned} \frac{d}{ds} \chi(s) \Big|_{s=1} &= \frac{d}{ds} \sum_{i=0}^{\infty} s^{k_i} \mathbf{P}(X = k_i) \Big|_{s=1} = \sum_{i=0}^{\infty} k_i s^{k_i-1} \mathbf{P}(X = k_i) \Big|_{s=1} = \sum_{i=0}^{\infty} k_i \mathbf{P}(X = k_i) = \mathbf{E}[X], \\ \frac{d^2}{ds^2} \chi(s) \Big|_{s=1} &= \sum_{i=0}^{\infty} k_i(k_i-1) s^{k_i-2} \mathbf{P}(X = k_i) \Big|_{s=1} = \sum_{i=0}^{\infty} k_i(k_i-1) \mathbf{P}(X = k_i) = \mathbf{E}[X(X-1)]. \end{aligned}$$

Bei der Laplace-Transformierten erhalten wir für diskrete Zufallsvariable mit Bemerkung 4.23

$$-\frac{d}{dt}\mathcal{L}(t)\Big|_{t=0} = \chi'(1),$$

$$\frac{d^2}{dt^2}\mathcal{L}(t)\Big|_{t=0} = -\frac{d}{dt}\chi'(e^{-t})e^{-t}\Big|_{t=0} = \chi''(1) + \chi'(1) = \mathbf{E}[X(X-1)] + \mathbf{E}[X] = \mathbf{E}[X^2].$$

Für Verteilungen mit Dichte erinnern wir an parameterabhängige Integrale aus Analysis 2 und berechnen, falls X die Dichte f besitzt,

$$-\frac{d}{dt}\mathcal{L}(t)\Big|_{t=0} = -\frac{d}{dt}\int e^{-tx}f(x)dx\Big|_{t=0} = -\int\frac{d}{dt}e^{-tx}\Big|_{t=0}f(x)dx = \int xf(x)dx = \mathbf{E}[X],$$

$$\frac{d^2}{dt^2}\mathcal{L}(t)\Big|_{t=0} = -\int\frac{d^2}{dt^2}e^{-tx}\Big|_{t=0}f(x)dx = \int x^2f(x)dx = \mathbf{E}[X^2].$$

□

4.4 Kenngrößen wichtiger Verteilungen

Wir sammeln nun Erwartungswerte, Varianzen und erzeugende Funktionen der uns bekannten diskreten Verteilungen. Einige davon sind bereits aus den Übungen bekannt.

Theorem 4.28 (Kenngrößen wichtiger diskreter Verteilungen). *Für X diskret gilt:*

	$\mathbf{P}(X = k)$	$\mathbf{E}[X]$	$\mathbf{Var}[X]$	$\chi_X(s)$
$X \sim U([1 : n])$	$\frac{1}{n}1_{1 \leq k \leq n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{1}{n}\frac{s(1-s^n)}{1-s}$
$X \sim B(n, p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	$(1+ps-p)^n$
$X \sim Hyp(n, N, K)$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n\frac{K}{N}$	$n\frac{K}{N}\frac{N-K}{N}\left(1-\frac{n-1}{N-1}\right)$	$\sum_{k=0}^n \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}s^k$
$X \sim Poi(\lambda)$	$e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	$e^{-(1-s)\lambda}$
$X \sim Geo(p)$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{ps}{1-(1-p)s}$

Beweis. Für $X \sim U([1 : n])$ erinnern wir an die (vermutlich) aus der Analysis 1 bekannten Identitäten

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Wir berechnen

$$\mathbf{E}[X] = \frac{1}{n}\sum_{k=1}^n k = \frac{n(n+1)}{2n} = \frac{n+1}{2},$$

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{1}{n}\left(\sum_{k=1}^n k^2\right) - \frac{(n+1)^2}{4} = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{(n+1)(4n+2-3n-3)}{12} = \frac{n^2-1}{12},$$

$$\chi_X(s) = \frac{1}{n}\sum_{k=1}^n s^k = \frac{1}{n}s\sum_{k=0}^{n-1} s^k = \frac{1}{n}\frac{s(1-s^n)}{1-s}.$$

Für $X \sim B(n, p)$ berechnen wir zunächst die erzeugende Funktion mit Hilfe des binomischen Lehrsatzes

$$\begin{aligned}\chi_X(s) &= \mathbf{E}[s^X] = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} s^k = \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} = (1+ps-p)^n, \\ \chi'_X(s) &= np(1+ps-p)^{n-1}, \\ \chi''_X(s) &= n(n-1)p^2(1+ps-p)^{n-2}.\end{aligned}$$

Daraus ergeben sich mittels Proposition 4.26 und Bemerkung 4.27

$$\begin{aligned}\mathbf{E}[X] &= \chi'_X(1) = np, \\ \mathbf{Var}[X] &= \chi''_X(1) + \chi'_X(1) - (\chi'_X(1))^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p).\end{aligned}$$

Für $X \sim \text{Hyp}(n, N, K)$ verwenden wir $X = Z_1 + \dots + Z_n$ für eine Urne mit den N Kugeln, wovon K weiß sind und

$$Z_i := 1_{\{i\text{-te gezogene Kugel ist weiß}\}}.$$

Klar ist, dass $\mathbf{P}[Z_i = 1] = \frac{K}{N}$, $i = 1, \dots, n$ und damit

$$\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[Z_i] = n \frac{K}{N}.$$

Weiter gilt für $i \neq j$

$$\mathbf{Cov}[Z_i, Z_j] = \mathbf{P}[Z_i = Z_j = 1] - \left(\frac{K}{N}\right)^2 = \frac{K}{N} \left(\frac{K-1}{N-1} - \frac{K}{N}\right) = -\frac{K}{N} \frac{N-K}{N(N-1)},$$

also nach Lemma 4.17.4 und $\mathbf{Var}[Z_i] = \frac{K}{N} \frac{N-K}{N}$

$$\begin{aligned}\mathbf{Var}[X] &= \sum_{i=1}^n \mathbf{Var}[Z_i] + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}[Z_i, Z_j] \\ &= n \frac{K}{N} \frac{N-K}{N} - n(n-1) \frac{K}{N} \frac{N-K}{N(N-1)} = n \frac{K}{N} \frac{N-K}{N} \left(1 - \frac{n-1}{N-1}\right).\end{aligned}$$

Für $X \sim \text{Poi}(\lambda)$ ist

$$\begin{aligned}\chi_X(s) &= \mathbf{E}[s^X] = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-(1-s)\lambda}, \\ \chi'_X(s) &= \lambda e^{-(1-s)\lambda}, \\ \chi''_X(s) &= \lambda^2 e^{-(1-s)\lambda}.\end{aligned}$$

Daraus ergeben sich mittels Proposition 4.26 und Bemerkung 4.27

$$\begin{aligned}\mathbf{E}[X] &= \chi'_X(1) = \lambda, \\ \mathbf{Var}[X] &= \chi''_X(1) + \chi'_X(1) - (\chi'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

Für $X \sim \text{Geo}(p)$ ist

$$\begin{aligned}\chi_X(s) &= \mathbf{E}[s^X] = p \sum_{k=1}^{\infty} (1-p)^{k-1} s^k = ps \sum_{k=0}^{\infty} ((1-p)s)^k = \frac{ps}{1 - (1-p)s}, \\ \chi'_X(s) &= \frac{p(1 - (1-p)s) + sp(1-p)}{(1 - (1-p)s)^2} = \frac{p}{(1 - (1-p)s)^2}, \\ \chi''_X(s) &= \frac{2p(1-p)}{(1 - (1-p)s)^3}.\end{aligned}$$

Daraus ergeben sich mittels Proposition 4.26 und Bemerkung 4.27

$$\begin{aligned}\mathbf{E}[X] &= \chi'_X(1) = \frac{1}{p}, \\ \mathbf{Var}[X] &= \chi''_X(1) + \chi'_X(1) - (\chi'_X(1))^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{(1-p)}{p^2}.\end{aligned}$$

□

Theorem 4.29 (Kenngrößen wichtiger kontinuierlicher Verteilungen). *Für kontinuierliche Zufallsvariablen X gilt:*

	$f_X(x)$	$\mathbf{E}[X]$	$\mathbf{Var}[X]$	$\mathcal{L}_X(t)$
$X \sim U([a, b])$	$\frac{1}{b-a} 1_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{-tb} - e^{-ta}}{t(b-a)}$
$X \sim \text{Exp}(\lambda)$	$\lambda e^{-\lambda x} 1_{x \geq 0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda+t}$
$X \sim N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\frac{t^2}{2}\sigma^2 + t\mu}$

Beweis. Für $X \sim U([a, b])$ ist

$$\begin{aligned}\mathbf{E}[X] &= \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{1}{2} (b^2 - a^2) = \frac{b+a}{2}, \\ \mathbf{Var}[X] &= \frac{1}{b-a} \int_a^b x^2 dx - \left(\frac{b+a}{2}\right)^2 = \frac{a^2 + ab + b^2}{3} - \frac{(b+a)^2}{4} \\ &= \frac{4a^2 + 4ab + 4b^2 - 3b^2 - 6ab - 3a^2}{12} = \frac{(b-a)^2}{12}, \\ \mathcal{L}_X(t) &= \mathbf{E}[e^{-tX}] = \frac{1}{b-a} \int_a^b e^{-tx} dx = \frac{e^{-tb} - e^{-ta}}{t(b-a)}.\end{aligned}$$

Für $X \sim \text{Exp}(\lambda)$ ist

$$\begin{aligned}\mathcal{L}_X(t) &= \mathbf{E}[e^{-tX}] = \int_0^{\infty} \lambda e^{-\lambda x} e^{-tx} dx = \lambda \int_0^{\infty} e^{-(\lambda+t)x} dx = \frac{\lambda}{\lambda+t}, \\ \mathcal{L}'_X(t) &= -\frac{\lambda}{(\lambda+t)^2}, \\ \mathcal{L}''_X(t) &= \frac{2\lambda}{(\lambda+t)^3}.\end{aligned}$$

Daraus ergeben sich mittels Proposition 4.26 und Bemerkung 4.27

$$\begin{aligned}\mathbf{E}[X] &= -\mathcal{L}'_X(0) = \frac{1}{\lambda}, \\ \mathbf{Var}[X] &= \mathcal{L}''_X(0) - (\mathcal{L}'_X(0))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.\end{aligned}$$

Sei zunächst $Z \sim N(0, 1)$. Dann ist

$$\mathcal{L}_Z(t) = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{z^2}{2}} e^{-tz} dz = e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(z^2+2tz+t^2)} dz = e^{\frac{t^2}{2}}.$$

Damit ist wegen $X := \mu + \sigma Z \sim N(\mu, \sigma^2)$

$$\begin{aligned}\mathcal{L}_X(t) &= \mathbf{E}[e^{-t(\mu+\sigma Z)}] = e^{\frac{t^2\sigma^2}{2}-t\mu}, \\ \mathcal{L}'_X(t) &= (t\sigma^2 - \mu)e^{\frac{t^2\sigma^2}{2}-t\mu}, \\ \mathcal{L}''_X(t) &= ((t\sigma^2 - \mu)^2 + \sigma^2)e^{\frac{t^2\sigma^2}{2}-t\mu}.\end{aligned}$$

Wieder ergeben sich mittels Proposition 4.26 und Bemerkung 4.27

$$\begin{aligned}\mathbf{E}[X] &= -\mathcal{L}'_X(0) = \mu, \\ \mathbf{Var}[X] &= \mathcal{L}''_X(0) - (\mathcal{L}'_X(0))^2 = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2.\end{aligned}$$

□

5 Approximationssätze

Wie üblich ist ein Approximationssatz ein Grenzwertresultat. Da wir es in der Stochastik mit Zufallsvariablen zu tun haben, werden wir also Grenzwertsätze über Zufallsvariablen (was ja spezielle Funktionen sind) betrachten. Speziell ist hier, dass wir nicht nur über die Konvergenz dieser Funktionen reden können (wie etwa im Gesetz der Großen Zahlen, Theorem 5.6), sondern auch über die Konvergenz ihrer Verteilungen (wie etwa im Zentralen Grenzwertsatz, Theorem 5.8).

5.1 Summen unabhängiger Zufallsvariablen

Wir beschäftigen uns nun mit (der Verteilung von)

$$S := X_1 + \dots + X_n,$$

wobei X_1, \dots, X_n unabhängige (und oft auch identisch verteilte) Zufallsvariable sind. Wir betrachten zunächst zwei spezielle Fälle normalverteilter und Poisson-verteilter Zufallsvariable.

Proposition 5.1 (Summen unabhängiger normal- und Poisson-verteilter Zufallsvariablen).

1. Sind X_1, \dots, X_n unabhängig und $X_i \sim Poi(\lambda_i)$, $i = 1, \dots, n$. Dann ist $S = X_1 + \dots + X_n \sim Poi\left(\sum_{i=1}^n \lambda_i\right)$.

2. Sind X_1, \dots, X_n unabhängig und $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$. Dann ist $S = X_1 + \dots + X_n \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.

Beweis. 1. Da X_1, \dots, X_n unabhängig sind, berechnen wir die erzeugende Funktion von S , nämlich

$$\chi_S(s) = \mathbf{E}[s^{X_1 + \dots + X_n}] = \mathbf{E}[s^{X_1}] \dots \mathbf{E}[s^{X_n}] = e^{-(1-s)\lambda_1} \dots e^{-(1-s)\lambda_n} = e^{-(1-s)(\lambda_1 + \dots + \lambda_n)},$$

und da die erzeugende Funktion die Verteilung von S eindeutig festlegt, ist $S \sim \text{Poi}\left(\sum_{i=1}^n \lambda_i\right)$.

2. Auch im Falle normalverteilter Zufallsvariablen berechnen wir mit Hilfe der Laplace-Transformierten

$$\begin{aligned} \mathcal{L}_S(t) &= \mathbf{E}[e^{-tX_1} \dots e^{-tX_n}] = \mathbf{E}[e^{-tX_1}] \dots \mathbf{E}[e^{-tX_n}] \\ &= e^{\frac{t^2}{2}\sigma_1^2 - t\mu_1} \dots e^{\frac{t^2}{2}\sigma_n^2 - t\mu_n} = e^{\frac{t^2}{2}(\sigma_1^2 + \dots + \sigma_n^2) - t(\mu_1 + \dots + \mu_n)}, \end{aligned}$$

wobei die rechte Seite die Laplace-Transformierte einer $N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$ -Verteilung ist. Da diese die Verteilung eindeutig bestimmt (siehe Bemerkung 4.25) folgt die Behauptung. \square

Definition 5.2 (Faltung). Seien X, Y unabhängige Zufallsvariablen. Dann heißt die Verteilung von $X + Y$ die Faltung der Verteilungen von X und Y .

Proposition 5.3 (Faltung diskreter und kontinuierlicher Zufallsvariable). Seien X, Y unabhängige Zufallsvariablen.

1. Sind X, Y diskret mit Zielbereich \mathbb{Z} , so ist die Faltung von X, Y gegeben durch

$$\mathbf{P}(X + Y = k) = \sum_{\ell=-\infty}^{\infty} \mathbf{P}(X = \ell) \mathbf{P}(Y = k - \ell).$$

Sind χ_X und χ_Y die erzeugenden Funktionen von X und Y , so ist $\chi_X \chi_Y$ die erzeugende Funktion von $X + Y$.

2. Sind X, Y kontinuierlich mit Zielbereich \mathbb{R} und Dichten f_X und f_Y , so besitzt die Faltung X, Y ebenfalls eine Dichte, nämlich

$$f_{X+Y}(z) = \int f_X(x) f_Y(z - x) dx.$$

Sind \mathcal{L}_X und \mathcal{L}_Y die Laplace-Transformierten von X und Y , so ist $\mathcal{L}_X \mathcal{L}_Y$ die Laplace-Transformierte von $X + Y$.

Beweis. Die Formeln für die Zähldichte und Dichte der Faltung sind klar. Die Multiplikativität von erzeugender Funktion und Laplace-Transformierter haben wir bereits im Beweis von Proposition 5.1 gesehen und verwendet. \square

Bemerkung 5.4 (\sqrt{n} -Gesetz). Sind X_1, \dots, X_n identisch verteilt und $S = X_1 + \dots + X_n$, so ist (siehe Proposition 4.5)

$$\mathbf{E}[S] = n\mathbf{E}[X_1].$$

Sind weiterhin X_1, \dots, X_n unabhängig, so gilt (siehe Lemma 4.17)

$$\mathbf{Var}[S] = n\mathbf{Var}[X_1].$$

Da die Varianz die mittlere *quadratische* Abweichung misst, typische Abweichungen von S um den Erwartungswert also im Bereich $\sqrt{\mathbf{Var}[S]}$ liegen sollten, kann man bereits folgendes für große n ablesen: eine Zufallsvariable, die die Summe von n unabhängigen Zufallsvariablen ist, streut ihre Werte typischerweise in einem Bereich der Größenordnung \sqrt{n} um ihren Erwartungswert. Dies werden wir mit dem Zentralen Grenzwertsatz, Theorem 5.8 (der auf dem Gesetz der großen Zahlen, Theorem Theorem 5.6 aufbaut), präzisieren.

5.2 Das schwache Gesetz der großen Zahlen

Bevor wir zum Hauptresultat dieses Abschnittes kommen, benötigen wir zwei Ungleichungen.

Proposition 5.5 (Die Markov- und Chebyshev-Ungleichung). *1. Markov-Ungleichung: Sei X eine \mathbb{R}_+ -wertige Zufallsvariable. Dann gilt für alle $\varepsilon > 0$*

$$\mathbf{P}(X \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbf{E}[X].$$

2. Chebyshev-Ungleichung: Sei X eine \mathbb{R} -wertige Zufallsvariable mit endlichem Erwartungswert. Dann gilt für alle $\varepsilon > 0$

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

Beweis. 1. Es gilt $X \geq \varepsilon \cdot 1_{X \geq \varepsilon}$ und damit

$$\varepsilon \cdot \mathbf{P}(X \geq \varepsilon) = \mathbf{E}[\varepsilon \cdot 1_{X \geq \varepsilon}] \leq \mathbf{E}[X].$$

2. Wir wenden die Markov-Ungleichung auf die Zufallsvariable $(X - \mathbf{E}[X])^2$ an. Dann gilt

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq \varepsilon) = \mathbf{P}((X - \mathbf{E}[X])^2 \geq \varepsilon^2) \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{\varepsilon^2}$$

und die Behauptung folgt. □

Theorem 5.6 (Schwaches Gesetz der großen Zahlen (von Jacob Bernoulli)). *Seien X_1, X_2, \dots unabhängige und identisch verteilte, \mathbb{R} -wertige Zufallsvariable mit endlichem Erwartungswert und endlicher Varianz. Dann gilt für alle $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbf{E}[X_1]\right| \geq \varepsilon\right) = 0. \quad (5.1)$$

Beweis. Es gilt

$$\begin{aligned} \mathbf{E}\left[\frac{X_1 + \dots + X_n}{n}\right] &= \frac{\mathbf{E}[X_1] + \dots + \mathbf{E}[X_n]}{n} = \mathbf{E}[X_1], \\ \mathbf{Var}\left[\frac{X_1 + \dots + X_n}{n}\right] &= \frac{\mathbf{Var}[X_1] + \dots + \mathbf{Var}[X_n]}{n^2} = \frac{\mathbf{Var}[X_1]}{n}. \end{aligned}$$

Wendet man nun die Chebyshev-Ungleichung auf $(X_1 + \dots + X_n)/n$ an, so folgt die Behauptung aus

$$\mathbf{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbf{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{\mathbf{Var}[X_1]}{\varepsilon^2 n} \xrightarrow{n \rightarrow \infty} 0.$$

□

Bemerkung 5.7 (Starkes Gesetz der großen Zahlen). Es gilt auch folgende, stärkere Version des Gesetzes der großen Zahlen, die wir ohne Beweis angeben:

Seien X_1, X_2, \dots unabhängige und identisch verteilte, \mathbb{R} -wertige Zufallsvariable mit endlichem Erwartungswert. Dann gilt

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mathbf{E}[X_1]\right) = 1. \quad (5.2)$$

Auf zwei Dinge möchten wir jedoch hinweisen:

1. Im starken Gesetz der großen Zahlen kommt das Ereignis $\{\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mathbf{E}[X_1]\}$ vor. Um diesem Ereignis sinnvoll eine Wahrscheinlichkeit zuordnen zu können, bedarf es der Maßtheorie.
2. Die Tatsache, dass das starke Gesetz wirklich stärker als das schwache Gesetz ist, liegt daran, dass (5.1) aus (5.2) folgt. Insbesondere gilt das schwache Gesetz der großen Zahlen auch dann, wenn die Varianz von X_1 unendlich ist. (Dies ist ja im oben formulierten schwachen Gesetz gefordert, im starken Gesetz jedoch nicht.)

5.3 Der zentrale Grenzwertsatz

Der Zentrale Grenzwertsatz gibt die Fluktuationen im Gesetz der großen Zahlen (im Falle endlicher Varianzen) an.

Theorem 5.8 (Zentraler Grenzwertsatz). *Seien X_1, X_2, \dots unabhängige und identisch verteilte, \mathbb{R} -wertige Zufallsvariable mit endlichem Erwartungswert μ und endlicher Varianz σ^2 . Dann gilt für alle $-\infty \leq c < d \leq \infty$*

$$\lim_{n \rightarrow \infty} \mathbf{P}\left(c \leq \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq d\right) = \mathbf{P}(c \leq Z \leq d),$$

wobei $Z \sim N(0, 1)$.

Bemerkung 5.9 (Einfache Fälle). 1. Klar ist, dass

$$\begin{aligned} \mathbf{E}\left[\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}}\right] &= 0, \\ \mathbf{Var}\left[\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}}\right] &= 1. \end{aligned}$$

Deshalb stimmen schon mal Erwartungswert und Varianz von $\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}}$ und Z überein. Der Zentrale Grenzwertsatz erweitert dies auf die ganze Verteilung.

2. Generell spricht man für eine Zufallsvariable X von ihrer Standardisierung $\frac{X - \mathbf{E}[X]}{\sqrt{\mathbf{Var}[X]}}$.
3. Klar ist der Zentrale Grenzwertsatz in dem Fall, wenn $X_1, X_2, \dots \sim N(\mu, \sigma^2)$. Dann nämlich ist $X_1 + \dots + X_n \sim N(n\mu, n\sigma^2)$, also

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}} \sim N(0, 1)$$

und der Satz gilt sogar ohne die Grenzwertbildung.

4. Sei $X_1, X_2, \dots \sim B(1, p)$, also $S_n := X_1 + \dots + X_n \sim B(n, p)$. Dann gilt nach dem Zentralen Grenzwertsatz für $c < d$

$$\mathbf{P}\left(c \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq d\right) \xrightarrow{n \rightarrow \infty} \mathbf{P}(c \leq Z \leq d).$$

Dies kann man bereits anhand der Laplace-Transformierten erahnen. Es gilt nämlich mit $S_n^* := \frac{S_n - np}{\sqrt{np(1-p)}}$

$$\begin{aligned} \log \mathcal{L}_{S_n^*}(t) &= \log \mathbf{E}[e^{-tS_n^*}] = \log \mathbf{E}[e^{t(np - S_n)/\sqrt{np(1-p)}}] \\ &= t\sqrt{np/(1-p)} + \log \mathbf{E}[e^{-\frac{t}{\sqrt{np(1-p)}} S_n}] = t\sqrt{np/(1-p)} + \log \chi_{S_n}(e^{-\frac{t}{\sqrt{np(1-p)}}}) \\ &= t\sqrt{np/(1-p)} + n \log(1 - p(1 - e^{-\frac{t}{\sqrt{np(1-p)}}})) \\ &\approx t\sqrt{np/(1-p)} - np(1 - e^{-\frac{t}{\sqrt{np(1-p)}}}) - np^2(1 - e^{-\frac{t}{\sqrt{np(1-p)}}})^2/2 \\ &\approx \frac{t^2}{2(1-p)} - \frac{t^2 p}{2(1-p)} = \frac{t^2}{2} = \log \mathcal{L}_Z(t). \end{aligned}$$

In diesem Fall (wenn $X_i \sim B(1, p)$) heißt der zentrale Grenzwertsatz auch Satz von deMoivre-Laplace.

5. Eine Anwendung des Satzes von deMoivre Laplace kann etwa so aussehen: Sei X eine $B(n, p)$ verteilte Zufallsvariable mit n groß, npq groß, sowie $c, d \in \mathbb{Z}$. Dann gilt approximativ, mit Verteilungsfunktion $\Phi(z) = \mathbf{P}(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$ von $Z \sim N(0, 1)$

$$\begin{aligned} \mathbf{P}(c \leq X \leq d) &= \mathbf{P}\left(\frac{c - np}{\sqrt{npq}} \leq \frac{X - np}{\sqrt{npq}} \leq \frac{d - np}{\sqrt{npq}}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{(c - \frac{1}{2} - np)/\sqrt{npq}}^{(d + \frac{1}{2} - np)/\sqrt{npq}} \exp\left(-\frac{z^2}{2}\right) dz = \Phi\left(\frac{d + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{c - \frac{1}{2} - np}{\sqrt{npq}}\right). \end{aligned}$$

Bevor wir zum Beweis des Zentralen Grenzwertsatzes kommen, benötigen wir ein Lemma.

Lemma 5.10. Seien X_1, X_2, \dots wie im Zentralen Grenzwertsatz mit $\mu = 0$ und $\sigma^2 = 1$. Weiter sei $h : \mathbb{R} \rightarrow \mathbb{R}$ eine 3-mal stetig differenzierbare Funktion mit beschränkten ersten drei Ableitungen und $Z \sim N(0, 1)$ verteilt. Dann gilt

$$\mathbf{E}\left[h\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right)\right] \xrightarrow{n \rightarrow \infty} \mathbf{E}[h(Z)].$$

Beweis. Seien $Z_1, Z_2, \dots \sim N(0, 1)$ unabhängig und unabhängig von X_1, X_2, \dots . Wir schreiben

$$U_i := \frac{X_1 + \dots + X_{i-1} + Z_{i+1} + \dots + Z_n}{\sqrt{n}}$$

und

$$\begin{aligned} h\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) &- h\left(\frac{Z_1 + \dots + Z_n}{\sqrt{n}}\right) \\ &= \sum_{i=1}^n h\left(U_i + \frac{X_i}{\sqrt{n}}\right) - h\left(U_i + \frac{Z_i}{\sqrt{n}}\right) \end{aligned} \tag{5.3}$$

Weiter ist mit der Taylor-Formel

$$\begin{aligned} & h\left(U_i + \frac{X_i}{\sqrt{n}}\right) - h\left(U_i + \frac{Z_i}{\sqrt{n}}\right) \\ &= h'(U_i) \frac{X_i - Z_i}{\sqrt{n}} + h''(U_i) \frac{X_i^2 - Z_i^2}{2n} \\ & \quad + \underbrace{\left(h''(V_i) - h''(U_i)\right) \frac{X_i^2}{2n} + \left(h''(U_i) - h''(W_i)\right) \frac{Z_i^2}{2n}}_{=: R_{in}} \end{aligned}$$

mit Zufallsvariablen V_i, W_i , so dass

$$|V_i - U_i| \leq \frac{|X_i|}{\sqrt{n}}, \quad |W_i - U_i| \leq \frac{|Z_i|}{\sqrt{n}}.$$

Weiter ist

$$\begin{aligned} |R_{in}| &= |R_{in}|1_{|X_i| \leq k} + |R_{in}|1_{|X_i| > k} \\ &\leq c''' \frac{k^3}{n^{3/2}} 1_{|X_i| \leq k} + c'' \frac{X_i^2}{n} 1_{|X_i| > k} + c''' \frac{|Z_i|^3}{n^{3/2}} \end{aligned}$$

mit $c'' = \sup_{x \in \mathbb{R}} h''(x)$, $c''' := \sup_{x \in \mathbb{R}} h'''(x)$ und beliebigem k . Da U_i, X_i, Z_i unabhängig sind, folgt

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

Wegen (5.3) und da $\frac{Z_1 + \dots + Z_n}{\sqrt{n}} \sim Z$, ist nun

$$\begin{aligned} \left| h\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) - h(Z) \right| &\leq c''' \frac{k^3 + \mathbf{E}[|Z_i|^3]}{n^{1/2}} + c'' \mathbf{E}[X_1^2 1_{|X_1| > k}] \\ &\xrightarrow{n \rightarrow \infty} c'' \mathbf{E}[X_1^2 1_{|X_1| > k}] \xrightarrow{k \rightarrow \infty} 0. \end{aligned}$$

□

Beweis von Theorem 5.8. OBdA ist $\mu = 0$ und $\sigma^2 = 1$, ansonsten gehen wir zu $\frac{X_i - \mu}{\sqrt{\sigma^2}}$ über. Wir setzen

$$Y_n^* := \frac{X_1 + \dots + X_n}{\sqrt{n}}.$$

Wähle für $\varepsilon > 0$ beliebig h_1, h_2 dreimal stetig differenzierbar mit beschränkten ersten drei Ableitungen und

$$1_{c+\varepsilon \leq x \leq d-\varepsilon} \leq h_1(x) \leq 1_{c \leq x \leq d} \leq h_2(x) \leq 1_{c-\varepsilon \leq x \leq d+\varepsilon}.$$

Dann gilt

$$\begin{aligned} \mathbf{P}(c + \varepsilon \leq Z \leq d - \varepsilon) &\leq \mathbf{E}[h_1(Z)] = \liminf_{n \rightarrow \infty} \mathbf{E}[h_1(Y_n^*)] \\ &\leq \liminf_{n \rightarrow \infty} \mathbf{P}(c \leq Y_n^* \leq d) \leq \limsup_{n \rightarrow \infty} \mathbf{P}(c \leq Y_n^* \leq d) \\ &\leq \limsup_{n \rightarrow \infty} \mathbf{E}[h_2(Y_n^*)] = \mathbf{E}[h_2(Z)] \leq \mathbf{P}(c - \varepsilon \leq Z \leq d + \varepsilon). \end{aligned}$$

Die Behauptung folgt nun, da $|\mathbf{P}(c \pm \varepsilon \leq Z \leq d \pm \varepsilon) - \mathbf{P}(c \leq Z \leq d)| \xrightarrow{\varepsilon \rightarrow 0} 0$. \square

Ähnlich wie im Zentralen Grenzwertsatz kann man auch eine Approximation von großen Fakultäten erhalten.

Korollar 5.11 (Stirling Formel). *Es gilt*

$$\lim_{n \rightarrow \infty} \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1.$$

Beweis. Sei $t^+ := \max(t, 0)$. Für $\varepsilon > 0$ sei h eine dreimal stetig differenzierbare Funktion mit beschränkten Ableitungen und

$$t^+ \leq h(t) \leq t^+ + \varepsilon.$$

Seien $X_1, X_2, \dots \sim \text{Poi}(1)$ unabhängig. Nach Proposition 5.1 ist damit $X_1 + \dots + X_n \sim \text{Poi}(n)$. Sei weiter $Z \sim N(0, 1)$. Mit Hilfe von Lemma 5.10 folgt (mit einem ähnlichen Approximationsargument wie im Beweis des Zentralen Grenzwertsatzes)

$$\mathbf{E}\left[\left(\frac{X_1 + \dots + X_n - n}{\sqrt{n}}\right)^+\right] \xrightarrow{n \rightarrow \infty} \mathbf{E}[Z^+] = \frac{1}{\sqrt{2\pi}} \int_0^\infty z e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Big|_0^\infty = \frac{1}{\sqrt{2\pi}} \quad (5.4)$$

Andererseits ist auch

$$\begin{aligned} &\mathbf{E}\left[\left(\frac{X_1 + \dots + X_n - n}{\sqrt{n}}\right)^+\right] \\ &= \frac{1}{\sqrt{n}} e^{-n} \sum_{k=n+1}^\infty (k-n) \frac{n^k}{k!} \\ &= \frac{e^{-n}}{\sqrt{n}} \sum_{k=n+1}^\infty \left(\frac{n^k}{(k-1)!} - \frac{n^{k+1}}{k!}\right) = \frac{e^{-n}}{\sqrt{n}} \frac{n^{n+1}}{n!}. \end{aligned}$$

Zusammen mit (5.4) folgt die Behauptung. \square

Bemerkung 5.12 (Ein einfacher Beweis). Man kann die Stirling-Formel auch fast mit ganz einfachen Methoden beweisen, nämlich

$$\log(n!) = \sum_{k=1}^n \log k = \int_1^n \log x + o(n) = x \log x - x \Big|_1^n + o(n) = n \log n - n + o(n).$$

Allerdings zeigt dies nur, dass

$$\frac{\log(n!)}{\log\left(\left(\frac{n}{e}\right)^n\right)} \xrightarrow{n \rightarrow \infty} 1,$$

jedoch nicht $\log(n!) - \log\left(\left(\frac{n}{e}\right)^n\right) \xrightarrow{n \rightarrow \infty} 0$ (was nach Korollar 5.11 auch gar nicht stimmt).

6 Abhängige Zufallsvariable

Stochastisch unabhängige Zufallsvariablen haben wir bereits in Abschnitt 2.4 kennen gelernt. Nun geht es um das Gegenteil, nämlich (voneinander) abhängige Zufallsvariablen.

6.1 Bedingte Wahrscheinlichkeiten

Zufallsvariablen können auf viele verschiedene Arten voneinander abhängig sein. Um diese Vielfalt in einer uns bekannten Form beschreiben zu können, benötigen wir bedingte Wahrscheinlichkeiten.

Definition 6.1 (Bedingte Wahrscheinlichkeit und bedingte Verteilung). *Seien X und Y Zufallsvariable mit Zielbereichen S und T .*

1. Die bedingte Wahrscheinlichkeit von $\{Y \in B\}$, gegeben $\{X \in A\}$ ist gegeben durch

$$\mathbf{P}(Y \in B|X \in A) := \frac{\mathbf{P}(X \in A, Y \in B)}{\mathbf{P}(X \in A)}.$$

Ist $\mathbf{P}(X \in A) = 0$, so definieren wir die rechte Seite als 0.

2. Die Abbildung $B \mapsto \mathbf{P}(Y \in B|X \in A)$ heißt die bedingte Verteilung von Y , gegeben $\{X \in A\}$.
3. Ist X diskret, so ist die bedingte Verteilung von Y gegeben X die Abbildung

$$\mathbf{P}(Y \in B|X) : \begin{cases} S & \rightarrow [0; 1] \\ x & \mapsto \mathbf{P}(Y \in B|X = x). \end{cases}$$

4. Haben X und Y die gemeinsame Dichte f , so ist

$$\mathbf{P}(Y \in B|X) : \begin{cases} S & \rightarrow [0; 1] \\ x & \mapsto \frac{\int_B f(x, y) dy}{\int f(x, y) dy}. \end{cases}$$

Die Abbildung

$$y \mapsto f_Y(y|X = x) := \frac{f(x, y)}{\int f(x, y) dy}$$

heißt auch bedingte Dichte von Y unter X .

Bemerkung 6.2 (Rechenregeln für bedingte Wahrscheinlichkeiten). Mit bedingten Verteilungen kann man genauso rechnen wie gewohnt, z.B. in Falle von diskreten Verteilungen

$$\mathbf{P}(Y \in B|X \in A) = \sum_{y \in B} \mathbf{P}(Y = y|X \in A).$$

Lemma 6.3 (Bedingte Wahrscheinlichkeiten und Unabhängigkeit). *Die Zufallsvariablen X und Y sind genau dann unabhängig, wenn*

$$\mathbf{P}(Y \in B | X \in A) = \mathbf{P}(Y \in B)$$

für alle (messbaren) A, B gilt. Dies ist dies genau dann der Fall, wenn

$$\mathbf{P}(Y \in B | X) = \mathbf{P}(Y \in B)$$

für alle B gilt.

Beweis. Alle Aussagen folgen direkt aus der Definition der bedingten Wahrscheinlichkeit. \square

Beispiel 6.4 (Gedächtnislosigkeit der geometrischen Verteilung und der Exponentialverteilung).

1. Sei $X \sim \text{Geo}(p)$. Dann gilt für $i, j = 0, 1, 2, \dots$

$$\mathbf{P}(X > i + j | X > i) = \mathbf{P}(X > j).$$

Die Interpretation ist die folgende: wenn i Versuche erfolglos verliefen, ist die Wahrscheinlichkeit für mindestens weitere j erfolglose Versuche genauso groß wie am Anfang. Deshalb spricht man auch von der Gedächtnislosigkeit der geometrischen Verteilung. Der Beweis ist ganz einfach:

$$\begin{aligned} \mathbf{P}(X > i + j | X > i) &= \frac{\mathbf{P}(X > i + j, X > i)}{\mathbf{P}(X > i)} = \frac{(1-p)^{i+j}}{(1-p)^i} = (1-p)^j \\ &= \mathbf{P}(X > j). \end{aligned}$$

2. Analog ist im Fall $X \sim \text{Exp}(\lambda)$ für $s, t \geq 0$

$$\begin{aligned} \mathbf{P}(X > s + t | X > s) &= \frac{\mathbf{P}(X > s + t, X > s)}{\mathbf{P}(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= \mathbf{P}(X > t). \end{aligned}$$

Also ist auch die Exponentialverteilung gedächtnislos.

Beispiel 6.5 (Summen zweier Zufallsvariablen). Seien X, Y zwei unabhängige Zufallsvariable mit Zielbereich \mathbb{Z} . Es gilt

$$\mathbf{P}(X + Y \in A | X = x) = \frac{\mathbf{P}(X + Y \in A, X = x)}{\mathbf{P}(X = x)} = \mathbf{P}(x + Y \in A_2),$$

d.h. die bedingte Verteilung von $X + Y$, gegeben $X = x$ ist dieselbe wie die der Zufallsvariable $Y + x$. Andersherum berechnen wir

$$\mathbf{P}(X = x | X + Y = z) = \frac{\mathbf{P}(X = x) \cdot \mathbf{P}(Y = z - x)}{\mathbf{P}(X + Y = z)}.$$

Ist etwa $X, Y \sim \text{Geo}(p)$ unabhängig, so ist

$$\mathbf{P}(X + Y = z) = (z - 1)p^2(1 - p)^{z-2},$$

da genau im z -ten Versuch der zweite Erfolg stattfinden muss. Damit ist

$$\mathbf{P}(X = x | X + Y = z) = \frac{(1-p)^{x-1} p (1-p)^{z-x-1} p}{(z-1)p^2(1-p)^{z-2}} = \frac{1}{z-1}.$$

Also ist X gegeben $X + Y$ gerade $U([1 : (X + Y - 1)])$ -verteilt.

Beispiel 6.6 (Summen zweier Exponentialverteilungen). Sei $X, Y \sim \text{Exp}(\lambda)$ **unabhängig**. Analog zum letzten Beispiel betrachten wir die gemeinsame Verteilung von X und $X + Y$, also die Verteilung mit Dichte

$$f(x, z) = \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)} \mathbf{1}_{0 \leq x \leq z} = \lambda^2 e^{-\lambda z} \mathbf{1}_{0 \leq x \leq z}.$$

Damit ist die bedingte Dichte von X gegeben $X + Y$

$$\frac{f(x, z)}{\int f(x, z) dx} = \frac{\lambda^2 e^{-\lambda z} \mathbf{1}_{0 \leq x \leq z}}{\int \lambda^2 e^{-\lambda z} \mathbf{1}_{0 \leq x \leq z} dx} = \frac{1}{z} \mathbf{1}_{0 \leq x \leq z},$$

also ist die bedingte Verteilung uniform auf $[0, X + Y]$.

Proposition 6.7 (Formel für die totale Wahrscheinlichkeit). Seien X, Y Zufallsvariable mit Zielbereichen S und T .

1. Sind X, Y diskret, so gilt

$$\mathbf{P}(Y \in B) = \sum_{x \in S} \mathbf{P}(Y \in B | X = x) \cdot \mathbf{P}(X = x).$$

2. Haben X, Y die gemeinsame Dichte f , so gilt

$$f_Y(y) = \int f_Y(y | X = x) \cdot f_X(x) dx.$$

Beweis. 1. Die Formel ergibt sich durch Einsetzen der Definition von $\mathbf{P}(Y \in B | X = x)$ in

$$\mathbf{P}(Y \in B) = \sum_{x \in S} \mathbf{P}(Y \in B, X = x) = \sum_{x \in S} \mathbf{P}(Y \in B | X = x) \cdot \mathbf{P}(X = x).$$

2. Ebenfalls sieht man hier direkt

$$f_Y(y) = \int_S f(x, y) dy = \int_S \frac{f(x, y)}{\int_S f(x, z) dz} \left(\int f(x, z) dz \right) dy = \int f(y | X = x) \cdot f_X(x) dx.$$

□

Beispiel 6.8 (Münzwurf mit zufälliger Erfolgswahrscheinlichkeit). Sei $U \sim U([0, 1])$. Gegeben $\{U = u\}$ sei $Z = (Z_1, \dots, Z_n)$ ein u -Münzwurf und $X = Z_1 + \dots + Z_n$ die Anzahl der Erfolge. Damit hat X gegeben $\{U = u\}$ eine $B(n, u)$ -Verteilung. Wir berechnen nun

$$\mathbf{P}(X = k) = \frac{1}{n+1}, \quad k = 0, \dots, n.$$

Denn für $0 < k < n$ ist

$$\begin{aligned} \mathbf{P}(X = k) &= \int \mathbf{P}(X = k|U = u) f_U(u) du = \int \underbrace{\binom{n}{k} u^k}_{=f'} \underbrace{(1-u)^{n-k}}_{=g} du \\ &= \int_0^1 \frac{n-k}{k+1} \binom{n}{k} u^{k+1} (1-u)^{n-k-1} du \\ &= \int_0^1 \binom{n}{k+1} u^{k+1} (1-u)^{n-(k+1)} du = \mathbf{P}(X = k+1). \end{aligned}$$

Damit ist die Behauptung gezeigt.

Theorem 6.9 (Formel von Bayes). *Seien X, Y diskrete Zufallsvariable und A so, dass $\mathbf{P}(X \in A) > 0$. Dann gilt*

$$\mathbf{P}(X \in A|Y \in B) = \frac{\mathbf{P}(Y \in B|X \in A) \cdot \mathbf{P}(X \in A)}{\sum_x \mathbf{P}(Y \in B|X = x) \cdot \mathbf{P}(X = x)}.$$

Haben X, Y eine gemeinsame Dichte f , so gilt analog

$$f_X(x|Y = y) = \frac{f_Y(y|X = x) f_X(x)}{\int f_Y(x|X = x) f_X(x) dx}.$$

Beweis. 1. Der Zähler ist gleich $\mathbf{P}(X \in A, Y \in B)$ nach Definition der bedingten Wahrscheinlichkeit, und der Nenner ist

$$\sum_x \mathbf{P}(Y \in B|X = x) \cdot \mathbf{P}(X = x) = \sum_x \mathbf{P}(Y \in B, X = x) = \mathbf{P}(Y \in B).$$

2. folgt analog. □

Beispiel 6.10 (Reihenuntersuchungen). Bei einer Reihenuntersuchung werden nicht nur kranke, sondern manchmal auch gesunde Personen positiv getestet. In einer Population sind insgesamt 0.8% der Personen erkrankt. Eine kranke Person wird in 90% der Fälle positiv getestet, eine gesunde Person mit 7%. Wie groß ist die Wahrscheinlichkeit, dass eine positiv getestete Person wirklich krank ist?

Sei hierzu X eine $\{e, \bar{e}\}$ -wertige Zufallsvariable, die angibt, ob die zufällig gezogene Person erkrankt ist und Y eine $\{p, n\}$ -wertige Zufallsvariable, die angibt, ob die Person positiv oder negativ getestet wird. Gesucht ist also $\mathbf{P}(X = e|Y = p)$. Gegeben ist

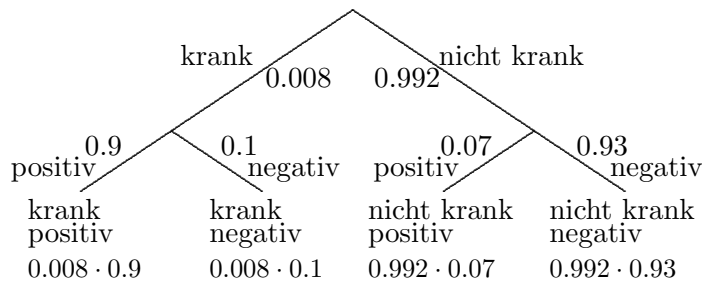
$$\mathbf{P}(X = e) = 0.008, \quad \mathbf{P}(Y = p|X = e) = 0.9, \quad \mathbf{P}(Y = p|X = \bar{e}) = 0.07.$$

Wir berechnen mit der Formel von Bayes

$$\begin{aligned} \mathbf{P}(X = e|Y = p) &= \frac{\mathbf{P}(Y = p|X = e) \cdot \mathbf{P}(X = e)}{\mathbf{P}(Y = p|X = e) \cdot \mathbf{P}(X = e) + \mathbf{P}(Y = p|X = \bar{e}) \cdot \mathbf{P}(X = \bar{e})} \\ &= \frac{0.9 \cdot 0.008}{0.9 \cdot 0.008 + 0.07 \cdot 0.992} \approx 0.0939 \end{aligned}$$

In der Tat bedeutet also ein positives Testergebnis in nicht einmal einem von zehn Fällen wirklich eine Erkrankung.

Dieses Beispiel lässt sich auch gut anhand eines Wahrscheinlichkeitsbaumes illustrieren:



6.2 Bedingte Erwartungen

Wir kommen nun zu einem Thema, das in der Wahrscheinlichkeitstheorie eine große Rolle spielen wird.

Definition 6.11 (Bedingte Erwartung). Seien X, Y Zufallsvariable mit Zielbereichen S und T . Weiter sei $h : S \times T \rightarrow \mathbb{R}$, so dass $\mathbf{E}[h(X, Y)]$ existiert.

1. Sind X, Y diskret, so definieren wir die bedingte Erwartung von $h(X, Y)$, gegeben X , durch

$$\mathbf{E}[h(X, Y)|X] = \sum_{y \in T} h(X, y) \cdot \mathbf{P}(Y = y|X).$$

2. Besitzen X, Y die gemeinsame Dichte f , so definieren wir die bedingte Erwartung von $h(X, Y)$, gegeben X , durch

$$\mathbf{E}[h(X, Y)|X] = \int_T h(X, y) \cdot f_Y(y|X) dy.$$

Proposition 6.12 (Turmeigenschaft). Seien X, Y Zufallsvariable mit Zielbereichen S und T , sowie $h : S \times T \rightarrow \mathbb{R}$ so dass $\mathbf{E}[h(X, Y)]$ existiert. Dann gilt

$$\mathbf{E}[h(X, Y)] = \mathbf{E}[\mathbf{E}[h(X, Y)|X]].$$

Beweis. Im diskreten Fall berechnen wir direkt

$$\begin{aligned} \mathbf{E}[\mathbf{E}[h(X, Y)|X]] &= \sum_{x \in S, y \in T} h(x, y) \cdot \mathbf{P}(Y = y|X = x) \cdot \mathbf{P}(X = x) \\ &= \sum_{x \in S, y \in T} h(x, y) \cdot \mathbf{P}(Y = y, X = x) = \mathbf{E}[h(X, Y)]. \end{aligned}$$

Für Zufallsvariable mit Dichte f ist

$$\begin{aligned} \mathbf{E}[\mathbf{E}[h(X, Y)|X]] &= \int_S \int_T h(x, y) \cdot f_Y(y|X = x) f_X(x) dy dx \\ &= \int_S \int_T h(x, y) \cdot f(x, y) dy dx = \mathbf{E}[h(X, Y)]. \end{aligned}$$

□

Beispiel 6.13 (Münzwurf mit zufälliger Erfolgswahrscheinlichkeit). Wir betrachten dieselbe Situation wie in Beispiel 6.8. Wir wissen, dass $Z \sim U([0 : n])$, insbesondere ist $\mathbf{E}[U] = \frac{n}{2}$. dies lässt sich auch ohne Kenntnis der Verteilung von U berechnen, da nämlich Z gegeben U eine $B(n, U)$ -Verteilung besitzt, also

$$\mathbf{P}(Z = k|U) = \binom{n}{k} U^k (1 - U)^{n-k},$$

und damit

$$\begin{aligned} \mathbf{E}[Z] &= \mathbf{E}[\mathbf{E}[Z|U]] = \mathbf{E}\left[\sum_{k=0}^n k \cdot \mathbf{P}(Z = k|U)\right] = \mathbf{E}\left[\sum_{k=0}^n k \cdot \binom{n}{k} U^k (1 - U)^{n-k}\right] \\ &= \mathbf{E}[nU] = \frac{1}{2}n. \end{aligned}$$

Analog zur bedingten Erwartung gibt es auch die bedingte Varianz.

Definition 6.14 (Bedingte Varianz). Seien X, Y Zufallsvariable Zielbereichen S und T , sowie $h : S \times T \rightarrow \mathbb{R}$. Wir setzen, falls $\mathbf{Var}[h(X, Y)] < \infty$

$$\mathbf{Var}[h(X, Y)|X] := \mathbf{E}[(h(X, Y) - \mathbf{E}[h(X, Y)|X])^2|X].$$

Analog zu Lemma 4.17.1 gilt dann

$$\mathbf{Var}[h(X, Y)|X] := \mathbf{E}[h(X, Y)^2|X] - \mathbf{E}[h(X, Y)|X]^2.$$

Proposition 6.15 (Varianzzerlegung). Seien X, Y Zufallsvariable mit Zielbereichen $S, T \subseteq \mathbb{R}$, sowie $h : S \rightarrow \mathbb{R}$. Sind $\mathbf{Var}[Y]$, $\mathbf{Var}[h(X)] < \infty$, so gilt

$$\mathbf{E}[(Y - h(X))^2] = \mathbf{E}[\mathbf{Var}[Y|X]] + \mathbf{E}[(\mathbf{E}[Y|X] - h(X))^2].$$

Insbesondere ist für $h(X) = \mathbf{E}[Y]$

$$\mathbf{Var}[Y] = \mathbf{E}[\mathbf{Var}[Y|X]] + \mathbf{Var}[\mathbf{E}[Y|X]].$$

Beweis. Da man mit bedingten Verteilungen wie üblich rechnen kann, ist

$$\begin{aligned} \mathbf{E}[(Y - h(X))^2|X] &= \mathbf{Var}[Y - h(X)|X] + (\mathbf{E}[Y - h(X)|X])^2 \\ &= \mathbf{Var}[Y|X] + (\mathbf{E}[Y|X] - h(X))^2 \end{aligned}$$

Bilden des Erwartungswertes ergibt

$$\mathbf{E}[(Y - h(X))^2] = \mathbf{E}[\mathbf{Var}[Y|X]] + \mathbf{E}[(\mathbf{E}[Y|X] - h(X))^2].$$

□

Beispiel 6.16 (Zufällige Anzahl unabhängiger Summanden). Sei N eine Zufallsvariable mit Zielbereich \mathbb{N} und Z_1, Z_2, \dots unabhängig und identisch verteilte Zufallsvariable mit Zielbereich $S \subseteq \mathbb{R}$ und $\mu := \mathbf{E}[Z_1]$, $\sigma^2 = \mathbf{Var}[Z_1]$. Wir untersuchen

$$Y := \sum_{i=1}^N Z_i.$$

Da

$$\mathbf{E}[Y|N] = N\mu, \quad \mathbf{Var}[Y|N] = N\sigma^2,$$

gilt

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[\mathbf{E}[Y|N]] = \mathbf{E}[N] \cdot \mu, \\ \mathbf{Var}[Y] &= \mathbf{E}[\mathbf{Var}[Y|N]] + \mathbf{Var}[\mathbf{E}[Y|N]] = \mathbf{E}[N] \cdot \sigma^2 + \mathbf{Var}[N] \cdot \mu^2. \end{aligned}$$

6.3 Mehrstufige Experimente

Bisher hatten wir nur höchstens zweistufige Experimente behandelt, etwa hatten wir zunächst eine Erfolgswahrscheinlichkeit U bestimmt, mit der dann eine Münze geworfen wird (Beispiele 6.8 und 6.13). In diesem Abschnitt behandeln wir nun exemplarisch mehrstufige Experimente, vor allem das Beispiel der Pólya-Urne.

Beispiel 6.17 (Pólya-Urne mit zwei Farben). Wir starten mit einer Urne, die zwei Kugeln enthält, nämlich eine grüne (abgekürzt durch 0) und eine schwarze (abgekürzt durch 1). In jedem Zug wählen wir eine Kugel aus der Urne zufällig und legen sie mit einer neuen derselben Farbe zurück. Nun bezeichnen wir mit

- $Z_1 \in \{0, 1\}$ das Ergebnis des ersten Zuges,
- $Z_2 \in \{0, 1\}$ das Ergebnis des zweiten Zuges (dessen Verteilung von Z_1 abhängt),
- $Z_3 \in \{0, 1\}$ das Ergebnis des dritten Zuges (dessen Verteilung von Z_1, Z_2 abhängt),
- ...

Wir zeigen nun, dass für $a_1, \dots, a_n \in \{0, 1\}$

$$\mathbf{P}(Z_{n+1} = a_{n+1} | Z_1 = a_1, \dots, Z_n = a_n) = \frac{k+1}{n+2}$$

mit

$$k := \#\{j \leq n : a_j = a_{n+1}\}.$$

Denn: Nach Hinzufügen der n -ten Kugel liegen insgesamt $n+2$ Kugeln in der Urne. Davon haben $1+k$ Kugeln die Farbe a_{n+1} , nämlich eine Kugel, die bereits am Anfang in der Urne liegt, und k hinzugelegte Kugeln.

Weiter zeigen wir nun noch für $X = \sum_{i=1}^n Z_i$, dass

$$\mathbf{P}(X = k) = \frac{1}{n+1},$$

also ist die Gesamtzahl schwarzer Kugeln in der Urne nach $n-1$ -maligem Hinzufügen von Kugeln gerade uniform verteilt.

Dies berechnet man wie folgt: zunächst gilt etwa für $(a_1, \dots, a_8) = (1, 1, 0, 1, 0, 0, 1, 1)$, dass

$$\mathbf{P}(Z_1 = a_1, \dots, Z_8 = a_8) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{2}{6} \cdot \frac{3}{7} \cdot \frac{4}{8} \cdot \frac{5}{9} = \frac{5!3!}{9!}$$

Wir berechnen allgemein für (a_1, \dots, a_n) mit $\sum_{i=1}^n a_i = k$

$$\mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n) = \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1} \frac{1}{\binom{n}{k}}.$$

Da es $\binom{n}{k}$ Folgen $(a_1, \dots, a_n) \in \{0, 1\}^n$ mit $\sum_{i=1}^n a_i = k$ gibt, folgt dass

$$\mathbf{P}(X = k) = \frac{1}{n+1}, \quad k = 0, \dots, n.$$

Beispiel 6.18 (Pólya Urne mit r Farben). Wir behandeln nun noch den Fall, in dem die Pólya-Urne mit r Kugeln mit r verschiedenen Farben, die mit $1, \dots, r$ nummeriert sind, startet. Ansonsten ist alles wie im letzten Beispiel: in jedem Zug ziehen wir eine Kugel aus der Urne und legen eine mit gleicher Farbe mit hinein. Nun ist $Z_i = j$, falls die i -te hinzugefügte Kugel Farbe j hat, $j \in \{1, \dots, r\}$. Die Anzahl der bis zum n -ten Zug hinzugefügten Kugeln der Farbe j ist

$$X_j := \sum_{i=1}^n 1_{Z_i=j}.$$

Sei $(a_1, \dots, a_n) \in \{1, \dots, r\}^n$ mit $\sum_{i=1}^n 1_{a_i=j} = k_j, j = 1, \dots, r$ und $k_1 + \dots + k_r = n$. Es gibt $\binom{n}{k_1 \dots k_r}$ solche Folgen. Wie für $r = 2$ gilt für jede dieser Folgen

$$\mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n) = \frac{k_1! \cdots k_r! \cdot (r-1)!}{(n+r-1)!} = \frac{1}{\binom{n+r-1}{r-1}} \frac{1}{\binom{n}{k_1 \dots k_r}}$$

und

$$\mathbf{P}(X_1 = k_1, \dots, X_r = k_r) = \frac{1}{\binom{n+r-1}{r-1}}.$$

Die Pólya-Urne liefert also uniform verteilte Besetzungen der r Farben.

7 Markov-Ketten

Die Vorlesung hat sich bisher mit Zufallsvariablen und deren Verteilung beschäftigt. Speziell haben wir bereits die Situation von mehreren abhängigen Zufallsvariablen kennen gelernt, die etwa bei zwei- oder mehrstufigen Experimenten auftauchen. In diesem Abschnitt werden wir stochastische Prozesse kennen lernen, also (fast) beliebige Familien von Zufallszahlen. Wie der Name schon sagt, stellt man sich dabei vor, dass die Zufallszahlen nacheinander, dynamisch als Prozess realisiert werden, genau wie bei der Pólya-Urne aus dem letzten Abschnitt. Dabei beschränken wir uns auf den Fall von diskreten Zeitschritten, betrachten also Familien von Zufallsvariablen $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$.

Eine besondere Form der Abhängigkeit von Zufallsvariablen tritt in Markov-Ketten zu Tage. Hier werden der Reihe nach Zufallsvariablen X_0, X_1, \dots realisiert, und zwar so, dass X_{t+1} nur von X_t abhängt. Man sagt auch, X_{t+1} ist unabhängig von X_1, \dots, X_{t-1} , wenn X_t bekannt ist. Diese Form der Abhängigkeit wird häufig zur stochastischen Modellierung verwendet. Beispielsweise könnte X_0, X_1, \dots der Preis einer Aktie an Tagen $0, 1, \dots$ sein. Die Markov-Eigenschaft besagt in diesem Fall, dass die Verteilung der Kursänderungen am Tag $t+1$ nur davon abhängt, wie der Kurs X_t am Tag t war.

7.1 Grundlegendes

Wir beginnen mit der Definition von Markov-Ketten. Wir werden vor allem den Fall von zeitlich homogenen Markov-Ketten behandeln. Bei solchen gibt es eine sich zeitlich nicht

ändernde stochastische Übergangsvorschrift, wie die Verteilung des Zustandes X_{t+1} ist, wenn der Zustand X_t der Kette zur Zeit t bekannt ist. Diese Vorschrift wird mit Hilfe einer Matrix zusammengefasst, der Übergangsmatrix.

Definition 7.1 (Stochastischer Prozess und Markov-Ketten). 1. Seien I und S Mengen. Ein (S -wertiger) stochastischer Prozess (mit Indexmenge I) ist eine Familie von Zufallsvariablen $\mathcal{X} = (X_t)_{t \in I}$ mit Wertebereich S . Die Menge S heißt auch Zustandsraum von \mathcal{X} .

2. Sei $I = \{0, 1, 2, \dots\}$, S abzählbar und $\mathcal{X} = (X_t)_{t \in I}$ ein S -wertiger stochastischer Prozess. Falls

$$\mathbf{P}(X_{t+1} = i \mid X_1, \dots, X_t) = \mathbf{P}(X_{t+1} = i \mid X_t) \quad (7.1)$$

für alle $i \in S$, so heißt \mathcal{X} eine Markov-Kette. Sie heißt endlich, falls S endlich ist.

3. Sei $\mathcal{X} = (X_t)_{t \in I}$ eine S -wertige Markov-Kette. Existiert eine Matrix $P = (P_{ij})_{i,j \in S}$ mit

$$P_{ij} := \mathbf{P}(X_{t+1} = j \mid X_t = i)$$

für alle t , so heißt \mathcal{X} zeitlich homogen und P heißt Übergangsmatrix von \mathcal{X} .

Bemerkung 7.2. 1. Die Eigenschaft (7.1) bedeutet in Worten: die zukünftige Entwicklung von \mathcal{X} nach t hängt von X_1, \dots, X_t nur durch den aktuellen Zustand X_t ab.

2. Sei P die Übergangsmatrix einer homogenen Markov-Kette \mathcal{X} mit Zustandsraum S . Dann gilt

$$\begin{aligned} 0 \leq P_{ij} \leq 1, & \quad i, j \in S, \\ \sum_{j \in S} P_{ij} = 1, & \quad i \in S. \end{aligned} \quad (7.2)$$

Die erste Eigenschaft ist klar, da die Einträge in P Wahrscheinlichkeiten sind. Außerdem ist

$$1 = \mathbf{P}(X_{t+1} \in S \mid X_t = i) = \sum_{j \in S} \mathbf{P}(X_{t+1} = j \mid X_t = i) = \sum_{j \in S} P_{ij}.$$

Matrizen P mit den Eigenschaften (7.2) heißen *stochastische Matrizen*.

3. Sei \mathcal{X} eine homogene Markov-Kette mit Übergangsmatrix P . Definiere einen gewichteten, gerichteten Graphen (E, K, W) wie folgt: die Menge der Knoten ist E , die Menge der (gerichteten) Kanten ist $K := \{(i, j) : P_{ij} > 0\}$. Das Gewicht der Kante (ij) ist $w_{(ij)} := P_{ij}$ und $W = (w_{(ij)})_{(ij) \in K}$. Der Graph (E, K, W) heißt *Übergangsgraph* von \mathcal{X} .

4. Für stochastische Prozesse mit überabzählbarem I bedarf es einiger maßtheoretischer Arbeit, um die Verteilung der Familie \mathcal{X} zu definieren und zu verstehen, durch was diese eindeutig gegeben ist. Diese fortgeschrittene Theorie stochastischer Prozesse wird in der Vorlesung *Stochastischer Prozesse* eingeführt.

Beispiel 7.3 (Irrfahrt im Dreieck). Betrachte eine homogene Markov-Kette \mathcal{X} mit Zustandsraum $\{1, 2, 3\}$ und Übergangsmatrix

$$P = \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix} \quad (7.3)$$

nur:

Mit welcher Wahrscheinlichkeit gewinnt Spieler 1?

Wir bezeichnen mit

$$p_n := \mathbf{P}(X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n)$$

die Wahrscheinlichkeit, dass Spieler 1 gewinnt. Wir werden zeigen, dass

$$p_n = \begin{cases} \frac{n}{N}, & p = q = \frac{1}{2}, \\ \frac{1 - (q/p)^n}{1 - (q/p)^N}, & \text{sonst.} \end{cases}$$

Zunächst gilt für $n = 1, \dots, N - 1$

$$\begin{aligned} p_n &= q \cdot \mathbf{P}(X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n, X_1 = n - 1) + p \cdot \mathbf{P}(X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n, X_1 = n + 1) \\ &= q \cdot \mathbf{P}(X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n - 1) + p \cdot \mathbf{P}(X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n + 1) \\ &= q \cdot p_{n-1} + p \cdot p_{n+1}, \end{aligned}$$

also mit $\Delta p_n := p_n - p_{n-1}$

$$q\Delta p_n = p\Delta p_{n+1}.$$

Im Fall $p = q = \frac{1}{2}$ folgt mit $\sum_{m=1}^N \Delta p_m = p_N - p_0 = 1$ daraus bereits

$$p_n = \frac{\sum_{m=1}^n \Delta p_m}{\sum_{m=1}^N \Delta p_m} = \frac{n\Delta p_1}{N\Delta p_1} = \frac{n}{N}.$$

Im Fall $p \neq q$ setzen wir $u := \frac{q}{p}$ und berechnen iterativ

$$\Delta p_n = u\Delta p_{n-1} = u^2\Delta p_{n-2} \cdots = u^{n-1}\Delta p_1 = u^{n-1}p_1.$$

Weiter ist

$$1 = \sum_{m=1}^N \Delta p_m = p_1 \sum_{m=0}^{N-1} u^m = p_1 \frac{1 - u^N}{1 - u}.$$

Also

$$p_n = \sum_{m=1}^n \Delta p_m = p_1 \sum_{m=1}^n u^{m-1} = \frac{1 - u}{1 - u^N} \frac{1 - u^n}{1 - u} = \frac{1 - u^n}{1 - u^N}$$

und die Behauptung ist gezeigt.

Beispiel 7.5 (Ehrenfest'sche Urne). Betrachte folgendes Urnenmodell: In einer Urne gibt es zwei durch eine Trennwand getrennte Kammern. Insgesamt liegen n Kugeln in den beiden Kammern. Wir ziehen eine Kugel rein zufällig aus der Urne und legen sie anschließend in die andere Kammer zurück. In Abbildung 7.3 findet sich eine Illustration.

Wir betrachten den $\{0, \dots, n\}$ -wertigen stochastischen Prozess $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$, wobei X_t die Anzahl der Kugeln in der linken Kammer nach dem t -Schritt darstellt. Dann ist \mathcal{X}

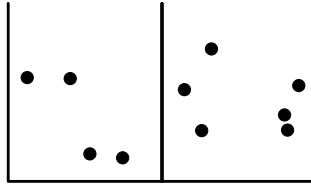


Abbildung 7.3: Die Ehrenfest'sche Urne mit $n = 10$ Kugeln aus Beispiel 7.5. Im nächsten Schritt wird mit Wahrscheinlichkeit $\frac{4}{10}$ eine Kugel von der linken Kammer in die rechte und mit Wahrscheinlichkeit $\frac{6}{10}$ von der rechten in die linke Kammer gelegt.

eine homogene Markov-Kette, denn der Ausgang des Schrittes $t + 1$ hängt nur von X_t ab. Die Übergangsmatrix von \mathcal{X} ist gegeben durch

$$P_{ij} = \begin{cases} \frac{i}{n}, & j = i - 1, \\ \frac{n-i}{n}, & j = i + 1, \\ 0, & \text{sonst.} \end{cases} \quad (7.4)$$

□

Wir kommen nun zu einem wichtigen Zusammenhang zwischen der Verteilung einer Markov-Ketten zum Zeitpunkt t und Potenzen der Übergangsmatrix.

Theorem 7.6. Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ eine homogene Markov-Kette mit Übergangsmatrix P und $\mu^{(t)} = (\mu^{(t)}(i))_{i \in S}$,

$$\mu^{(t)}(i) = \mathbf{P}(X_t = i)$$

für $t = 0, 1, 2, \dots$. Dann gilt

$$\mu^{(t)} = \mu^{(0)} P^t, \quad (7.5)$$

für $t = 0, 1, 2, \dots$, wobei die rechte Seite als Multiplikation des Zeilenvektors $\mu^{(0)}$ und der Matrix $P^t = \underbrace{P \cdots P}_{t \text{ mal}}$ zu verstehen ist.

Beweis. Der Beweis geht mittels Induktion über t . Für $t = 0$ ist die Aussage klar. Ist (7.5) für t gezeigt, so gilt

$$\begin{aligned} \mu^{(t+1)}(j) &= \mathbf{P}(X_{t+1} = j) = \sum_{i \in S} \mathbf{P}(X_{t+1} = j \mid X_t = i) \cdot \mathbf{P}(X_t = i) \\ &= \sum_{i \in S} \mu^{(t)}(i) \cdot P_{ij} = \sum_{i \in S} (\mu^{(0)} P^t)_i \cdot P_{ij} = (\mu^{(0)} P^{t+1})_j. \end{aligned} \quad \square$$

Beispiel 7.7 (Irrfahrt im Dreieck). Betrachte die Markov-Kette \mathcal{X} aus Beispiel 7.3 und Übergangsmatrix P aus (7.3). Sei $X_0 = 1$, die Markov-Kette startet also in 1, d.h. $\mu^{(0)} =$

$(1, 0, 0)$. Weiter berechnen wir

$$P = \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 2pq & q^2 & p^2 \\ p^2 & 2pq & q^2 \\ q^2 & p^2 & 2pq \end{pmatrix} \quad (7.6)$$

und damit

$$\mu^{(2)} = (2pq, q^2, p^2).$$

Also ist etwa $\mathbf{P}(X_2 = 1) = 2pq$. Dies ist klar, weil $X_2 = 1$ genau dann, wenn die ersten beiden Schritte im Übergangsgraphen aus Abbildung 7.1 einmal mit und einmal entgegen dem Uhrzeigersinn gingen. Es ist $X_2 = 3$ genau dann, wenn zwei Schritte im Uhrzeigersinn realisiert wurden, was Wahrscheinlichkeit p^2 hat.

7.2 Stationäre Verteilungen

Wir studieren nun Markov-Ketten, die schon lange gelaufen sind, also für die Verteilung von X_t für große t . Solche Markov-Ketten können (mehr dazu im nächsten Abschnitt) so beschaffen sein, dass sich die Verteilung nicht mehr viel ändert. Verteilungen auf dem Zustandsraum S , die invariant sind gegenüber Schritten der Markov-Kette, heißen stationär.

Definition 7.8 (Stationäre Verteilung). *Sei \mathcal{X} eine homogene Markov-Kette mit Übergangsmatrix P und π eine Verteilung auf S , gegeben durch einen Vektor $\pi = (\pi_i)_{i \in S}$.*

1. *Gilt*

$$\pi P = \pi,$$

so heißt π stationäre Verteilung von \mathcal{X} .

2. *Gilt*

$$\pi_i P_{ij} = \pi_j P_{ji}$$

für alle i, j , so heißt π reversible Verteilung.

Bemerkung 7.9 (Eigenschaften von stationären und reversiblen Verteilungen).

1. Sei π stationär und $\mathbf{P}(X_t = i) = \pi_i$. Dann gilt wegen Theorem 7.6, dass

$$\begin{aligned} \mathbf{P}(X_{t+1} = j) &= \sum_{i \in S} \mathbf{P}(X_t = i) \cdot \mathbf{P}(X_{t+1} = j \mid X_t = i) = \sum_{i \in S} \pi_i P_{ij} = (\pi P)_j = \pi_j \\ &= \mathbf{P}(X_t = j). \end{aligned}$$

Das bedeutet, dass X_t und X_{t+1} (und damit auch X_{t+2}, X_{t+3}, \dots) identisch verteilt sind.

2. Jede reversible Verteilung ist stationär. Ist nämlich π reversibel, so gilt

$$(\pi P)_j = \sum_{i \in S} \pi_i P_{ij} = \sum_{i \in S} \pi_j P_{ji} = \pi_j,$$

da P Zeilensumme 1 hat.

3. Sei π reversibel, $\mathbf{P}(X_t = i) = \pi_i$ und $i_t, \dots, i_{t+s} \in S$. Dann gilt

$$\begin{aligned} \mathbf{P}(X_t = i_t, \dots, X_{t+s} = i_{t+s}) &= \pi_{i_t} P_{i_t, i_{t+1}} \cdots P_{i_{t+s-1}, i_{t+s}} \\ &= \pi_{i_{t+1}} P_{i_{t+1}, i_{t+2}} \cdots P_{i_{t+s-1}, i_{t+s}} \cdot P_{i_{t+1}, i_t} = \cdots \\ &= \pi_{i_{t+s}} P_{i_{t+s}, i_{t+s-1}} \cdots P_{i_{t+1}, i_t} \\ &= \mathbf{P}(X_t = i_{t+s}, \dots, X_{t+s} = i_t). \end{aligned}$$

Das bedeutet, dass die Wahrscheinlichkeit, die Zustände i_t, \dots, i_{t+s} zu durchlaufen, dieselbe ist wie die, die Zustände in umgekehrter Reihenfolge (reversibel) zu durchlaufen.

Beispiel 7.10 (Irrfahrt im Dreieck). Betrachte die Irrfahrt auf dem Dreieck \mathcal{X} aus Beispiel 7.3 mit Übergangsmatrix P aus (7.3). Für $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ ist

$$\pi P = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix} = \pi.$$

Damit ist die Gleichverteilung auf $\{1, 2, 3\}$ stationäre Verteilung von \mathcal{X} . Allerdings ist π nicht reversibel (außer für $p = \frac{1}{2}$), denn es gilt etwa

$$\pi_1 P_{12} = \frac{1}{3}p \neq \frac{1}{3}q = \pi_2 P_{21}.$$

Letztere Beobachtung ist nicht erstaunlich: Sei etwa $p > \frac{1}{2}$. Dann erwarten wir, dass \mathcal{X} das Dreieck öfter im Uhrzeigersinn durchläuft als umgekehrt. Wäre nun π reversibel, würde daraus folgen (siehe Bemerkung 7.9.3), dass die Wahrscheinlichkeit für einen Durchlauf des Dreiecks gegen und im Uhrzeigersinn dieselbe ist.

Beispiel 7.11 (Ruinproblem). Im Ruinproblem aus Beispiel 7.4 gibt es zwar stationäre Verteilungen, diese sind jedoch nicht sonderlich interessant. Jede Verteilung $\pi = (p', 0, \dots, 0, q')$ mit $p' \in (0, 1)$ und $q' = 1 - p'$ ist stationäre (und nicht reversible) Verteilung, wie man leicht nachrechnet. Insbesondere sind $(1, 0, \dots, 0)$ und $(0, \dots, 0, 1)$ stationär. Das bedeutet, dass die Zustände $X_t = 0$ und $X_t = n$ von der Markov-Kette nicht mehr verlassen werden. Man sagt auch, 0 und n sind *Fallen* für \mathcal{X} .

Beispiel 7.12 (Ehrenfest'sche Urne). Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ die Anzahl der Kugeln in der linken Kammer einer Ehrenfest'schen Urne, wie in Beispiel 7.5. Hier ist die Binomialverteilung $\pi = B(n, \frac{1}{2})$ reversible Verteilung für \mathcal{X} . Es gilt nämlich mit (7.4) für $i = 1, \dots, n$

$$\pi_i P_{i, i-1} = \binom{n}{i} \frac{1}{2^n} \frac{i}{n} = \binom{n}{i-1} \frac{1}{2^n} \frac{n-i}{n} = \pi_{i-1} P_{i-1, i}.$$

Die Tatsache, dass $B(n, \frac{1}{2})$ stationär ist, interpretiert man am besten so: nach langer Zeit ist jede der Kugeln oft von der linken in die rechte Kammer und umgelegt worden, so dass jede Kugel mit Wahrscheinlichkeit $\frac{1}{2}$ jeweils in der linken und rechten Kammer liegt. Diese Verteilung ändert sich nicht mehr, weil in jedem Schritt nur eine der n Kugeln nochmals in die andere Kammer gelegt wird. \square

Stationäre Verteilungen existieren für die meisten Markov-Ketten, wie das nächste Resultat zeigt. Im nächsten Abschnitt werden wir dann betrachten, wann diese eindeutig sind.

Theorem 7.13 (Existenz von stationären Verteilungen). *Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ eine endliche, homogene Markov-Kette mit Zustandsraum S . Es gibt Folgen $t_n \rightarrow \infty$ so, dass*

$$\pi_i := \lim_{n \rightarrow \infty} \frac{1}{t_n} \sum_{s=1}^{t_n} \mathbf{P}(X_s = i) \quad (7.7)$$

für alle $i \in S$. Jeder solche Vektor $\pi = (\pi_i)_{i \in S}$ definiert eine Verteilung auf S . Diese ist stationäre Verteilung von \mathcal{X} .

Beweis. Sei $S = \{i_1, \dots, i_m\}$ und

$$\pi_{t,i} := \frac{1}{t} \sum_{s=1}^t \mathbf{P}(X_s = i).$$

Klar ist, dass für alle $i \in S$ die Folgen $(\pi_{t,i})_{t=1,2,\dots}$ Werte im kompakten Intervall $[0, 1]$ annehmen. Damit gibt es eine konvergente Teilfolge von $(\pi_{t,i_1})_{t=1,2,\dots}$. Von dieser Teilfolge gibt es eine weitere Teilfolge, so dass $(\pi_{t,i_1})_{t=1,2,\dots}$ und $(\pi_{t,i_2})_{t=1,2,\dots}$ entlang der Teilfolge konvergiert. Iteriert man dieses Verfahren, findet man eine Teilfolge, so dass (7.7) für alle $i \in S$ gilt.

Klar ist, dass $0 \leq \pi_i \leq 1$ für alle $i \in S$ gilt. Außerdem ist

$$\sum_{i \in S} \pi_i = \lim_{n \rightarrow \infty} \frac{1}{t_n} \sum_{s=1}^{t_n} \sum_{i \in S} \mathbf{P}(X_s = i) = 1,$$

und damit definiert π eine Verteilung auf S . Um zu sehen, dass π eine stationäre Verteilung ist, berechnen wir

$$\begin{aligned} (\pi P)_j &= \sum_{i \in S} \pi_i P_{ij} = \lim_{n \rightarrow \infty} \frac{1}{t_n} \sum_{s=1}^{t_n} \sum_{i \in S} \mathbf{P}(X_s = i) \cdot \mathbf{P}(X_{s+1} = j \mid X_s = i) \\ &= \lim_{n \rightarrow \infty} \frac{1}{t_n} \sum_{s=1}^{t_n} \mathbf{P}(X_{s+1} = j) \\ &= \frac{1}{t_n} (\mathbf{P}(X_{t_n+1} = j) - \mathbf{P}(X_1 = j)) + \lim_{n \rightarrow \infty} \frac{1}{t_n} \sum_{s=1}^{t_n} \mathbf{P}(X_s = j) \\ &= \pi_j \end{aligned}$$

und die Behauptung ist gezeigt. \square

7.3 Markov-Ketten-Konvergenzsatz

In diesem Abschnitt zeigen wir, dass jede endliche, homogene, *aperiodische* und *irreduzible* Markov-Kette konvergiert. Die Konvergenz ist dabei so zu verstehen, dass sich die Verteilung von X_t im Grenzwert großer t nicht mehr ändert, und auch nicht vom Anfangszustand X_0 abhängt. Wir beginnen mit Erklärungen der benötigten Begriffe.

Definition 7.14. *Sei \mathcal{X} eine S -wertige homogene Markov-Kette.*

1. Der Zustand $i \in S$ kommuniziert mit $j \in S$, falls es ein t gibt mit

$$\mathbf{P}(X_t = j \mid X_0 = i) > 0.$$

In diesem Fall schreiben wir $i \rightarrow j$. Falls $i \rightarrow j$ und $j \rightarrow i$, schreiben wir $i \leftrightarrow j$.

2. Falls $i \leftrightarrow j$ für alle $i, j \in S$, so heißt \mathcal{X} irreduzibel. Andernfalls heißt \mathcal{X} reduzibel.
3. Ein Zustand $i \in S$ heißt aperiodisch, falls

$$d(i) := \text{ggT}\{t : \mathbf{P}(X_t = i \mid X_0 = i) > 0\} = 1.$$

Andernfalls heißt i periodisch mit Periode $d(i)$.

4. Falls alle $i \in S$ aperiodisch sind, so heißt \mathcal{X} aperiodisch.

Bemerkung 7.15 (Irreduzibilität, Aperiodizität und die Übergangsmatrix).

1. Die Begriffe der Irreduzibilität und Aperiodizität lassen sich auch mittels der Übergangsmatrix P der Markov-Kette \mathcal{X} erklären. Es gilt etwa $i \rightarrow j$ genau dann, wenn es ein t gibt mit $(P^t)_{ij} > 0$. Außerdem ist $d(i) = \text{ggT}\{t : P_{ii}^t > 0\}$.
2. Es gibt einen einfachen Zusammenhang zwischen dem Begriff der kommunizierenden Zustände und dem Übergangsgraphen: ein Zustand i kommuniziert mit einem Zustand j , wenn man einen Pfad $i \rightarrow j$ im Übergangsgraphen der Markov-Kette findet. Das bedeutet, dass es eine endliche Folge $(i, i_1), (i_1, i_2), \dots, (i_{n-1}, i_n), (i_n, j)$ im Übergangsgraphen gibt, also der Zustand j von i aus durch eine endliche Folge von Kanten erreicht werden kann.
3. Sei \mathcal{X} eine reduzible Markov-Kette. Der Begriff der Reduzibilität erklärt sich so, dass sich die Markov-Kette \mathcal{X} auf einen Zustandsraum $S' \subseteq S$ reduzieren lässt. Das bedeutet, dass aus $X_0 \in S'$ folgt, dass $X_t \in S'$ für alle $t = 1, 2, \dots$

Beispiel 7.16 (Irrfahrt auf dem Dreieck). Sei \mathcal{X} die Irrfahrt auf dem Dreieck aus Beispiel 7.3. In Beispiel 7.7 haben wir ausgerechnet, dass $P^2 > 0$ ist. Weiter ist auch $P^3 > 0$, was bereits zeigt, dass \mathcal{X} sowohl irreduzibel als auch aperiodisch ist.

Beispiel 7.17 (Ruinproblem). Für die Markov-Kette \mathcal{X} aus dem Ruinproblem, Beispiel 7.4, gilt sicher, dass $0 \not\rightarrow i$ für $i = 1, \dots, n$, und $n \not\rightarrow i$ für $i = 0, \dots, n - 1$. Mit anderen Worten: Ist Spieler 1 pleite (d.h. $X_t = 0$ für ein t), so wird er nach den Spielregeln nie wieder Geld bekommen, d.h. alle Zustände $1, \dots, n$ sind für ihn unerreichbar. Das bedeutet also, \mathcal{X} ist reduzibel.

Beispiel 7.18 (Ehrenfest'sche Urne). Die Markov-Kette \mathcal{X} , die die Anzahl der Kugeln in der linken Kammer einer Ehrenfest'schen Urne beschreibt, ist irreduzibel. Betrachtet man nämlich Zustände i, j , etwa mit $i > j$, so genügt es ja, genau $i - j$ Kugeln nacheinander von der linken in die rechte Kammer zu legen. Allerdings ist die Markov-Kette periodisch. Sei etwa $X_0 = 2i$ gerade, dann folgt, dass X_1 ungerade (entweder $2i - 1$ oder $2i + 1$) ist. Damit kann nur zu geraden Zeiten t der Zustand $X_t = 2i$ eintreten und es ist $\text{ggT}\{t : \mathbf{P}(X_t = 2i \mid X_0 = 2i)\} = 2$.

Nun kommen wir also zum Markov-Ketten-Konvergenzsatz.

Theorem 7.19 (Markov-Ketten-Konvergenzsatz). Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ eine endliche, homogene, irreduzible und aperiodische Markov-Kette. Dann hat \mathcal{X} genau eine stationäre Verteilung π und es gilt für alle $i \in S$

$$\mathbf{P}(X_t = i) \xrightarrow{t \rightarrow \infty} \pi(i). \quad (7.8)$$

Bemerkung 7.20. Vor allem ist an dem Resultat bemerkenswert, dass (7.8) nicht von der Verteilung von X_0 abhängt. Das bedeutet, dass die Markov-Kette nach langer Zeit *vergisst*, in welchem Zustand X_0 sie zur Zeit 0 gestartet ist.

Beispiel 7.21 (Irrfahrt auf dem Dreieck). Von unseren drei Beispielen erfüllt (nur) die Irrfahrt auf dem Dreieck \mathcal{X} aus Beispiel 7.3 die Voraussetzungen des Satzes. Nach Beispiel 7.16 ist \mathcal{X} irreduzibel und aperiodisch. Nach Beispiel 7.10 und obigem Satz ist die Gleichverteilung auf $\{1, 2, 3\}$ die einzige stationäre Verteilung für \mathcal{X} . Außerdem gilt

$$\mathbf{P}(X_t = i) \xrightarrow{t \rightarrow \infty} \frac{1}{3}.$$

Bemerkung 7.22 (Beweistechnik Kopplung). Die Technik, mit der wir das Theorem beweisen werden heißt *Kopplung*. Sind \mathcal{V}, \mathcal{W} stochastische Prozesse mit Zustandsräumen S und S' , so heißt ein stochastischer Prozess $(\mathcal{Y}, \mathcal{Z})$ mit Zustandsraum $S \times S'$ Kopplung von \mathcal{V}, \mathcal{W} , wenn \mathcal{V} und \mathcal{Y} , sowie \mathcal{W} und \mathcal{Z} identisch verteilt sind. Wichtig ist hier, dass das Paar $(\mathcal{Y}, \mathcal{Z})$ eine gemeinsame Entwicklung definiert, die Prozesse \mathcal{V} und \mathcal{W} jedoch nicht.

Seien etwa $\mathcal{V} = (V_t)_{t=0,1,2,\dots}$ und $\mathcal{W} = (W_t)_{t=0,1,2,\dots}$ homogene Markov-Ketten mit Zustandsraum $\{0, 1\}$ mit Übergangsmatrix

$$P = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$$

mit $q := 1 - p$ und $V_0 = 0, W_0 = 1$. Sowohl \mathcal{V} als auch \mathcal{W} bleiben also jeweils mit Wahrscheinlichkeit p in ihrem Zustand und wechseln ihn mit Wahrscheinlichkeit q . Wir definieren die gekoppelte Markov-Kette $(\mathcal{Y}, \mathcal{Z})$ durch die Übergangsmatrix

$$P' = \begin{pmatrix} p & 0 & 0 & q \\ qp & p^2 & q^2 & pq \\ pq & q^2 & p^2 & qp \\ q & 0 & 0 & p \end{pmatrix},$$

wobei die Reihenfolge der Zustände in den Zeilen und Spalten durch $(0, 0), (0, 1), (1, 0), (1, 1)$ gegeben ist. Beispielsweise ist die Wahrscheinlichkeit eines Überganges $(0, 1)$ nach $(1, 0)$ und nach $(1, 1)$ gleich q^2 und pq . Daraus folgt auch, dass die Wahrscheinlichkeit, dass \mathcal{V} von 0 nach 1 springt, gleich q ist. Durch analoge Überlegungen lässt sich zeigen, dass \mathcal{V} und \mathcal{Y} , sowie \mathcal{W} und \mathcal{Z} , identisch verteilt sind. Auffällig ist weiterhin, dass die Menge $\{(0, 0), (1, 1)\}$ nicht verlassen wird. Die Markov-Kette $(\mathcal{Y}, \mathcal{Z})$ lässt sich so interpretieren, dass $\mathcal{Y} = (Y_t)_{t=0,1,2,\dots}$ und $\mathcal{Z} = (Z_t)_{t=0,1,2,\dots}$ sich so lange unabhängig entwickeln, bis $Y_t = Z_t$ gilt. Dann gilt weiterhin $Y_r = Z_r$ für $r \geq t$. Wie man leicht zeigt, gibt es mit Wahrscheinlichkeit 1 ein t mit $Y_t = Z_t$. Man spricht dann auch von einer *erfolgreichen Kopplung*.

Bevor wir den Markov-Ketten-Konvergenzsatz beweisen, benötigen wir noch ein paar Hilfsaussagen.

Lemma 7.23. Sei $A = \{s_1, s_2, \dots\} \subseteq \mathbb{Z}$ mit $\text{ggT}(A) = 1$.

1. Ist A abgeschlossen unter Addition und Subtraktion (d.h. aus $s, s' \in A$ folgt $s+s', s-s' \in A$), dann ist $1 \in A$.

2. Ist $A \subseteq \mathbb{N}$ und abgeschlossen unter Addition, dann gibt es ein $t < \infty$ mit $\{t, t+1, t+2, \dots\} \subseteq A$.

Beweis. 1. Sei a das kleinste positive Element von A . Wir zeigen zunächst, dass $A = \{ka : k \in \mathbb{Z}\}$ gilt. Ist nämlich $c \in A$, so lässt es sich eindeutig als $c = ka + r$ mit $0 \leq r < a$ schreiben. Es muss $r = 0$ gelten, da ansonsten $r = c - ka \in A$ ist und kleiner als a . Damit gilt, dass $c = ka$ für ein $k \in \mathbb{Z}$. Da a das kleinste positive Element von A ist, folgt

$$\text{ggT}(A) = \text{ggT}\{ka : k \in \mathbb{Z}\} = a,$$

woraus $a = 1$ folgt.

2. Wegen 1. gibt es $s_1, \dots, s_k \in A$ und $n_1, \dots, n_k \in \mathbb{Z}$ mit

$$1 = s_1 n_1 + \dots + s_k n_k.$$

Wenn wir die positiven von den negativen Termen trennen, können wir schreiben, dass $1 = m - p$ mit $m, p \in A$. Wir zeigen nun, dass die Aussage mit $t = p(p-1)$ gilt. Sei nämlich $c \geq t$ mit $c = ap + r$ mit $a \geq p-1$ und $r < p$. Dann gilt $c = ap + r(m-p) = (a-r)p + rm$. Da $m, p \in A$ und $a-r \geq p-1-r \geq 0$, folgt dass $c \in A$. \square

Proposition 7.24. *Sei \mathcal{X} eine endliche, homogene, aperiodische Markov-Kette mit Zustandsraum S und Übergangsmatrix P . Dann gibt es $t < \infty$, so dass*

$$\mathbf{P}(X_r = j \mid X_0 = i) > 0$$

für alle $i \in S$ und $r = t, t+1, t+2, \dots$ gilt.

Beweis. Wegen der Endlichkeit von S genügt es, die Behauptung für ein beliebiges Paar i, j zu zeigen. Sei zunächst $j = i$. Wir setzen

$$A := \{s : \mathbf{P}(X_s = i \mid X_0 = i) > 0\}.$$

Wegen der Aperiodizität von \mathcal{X} ist $\text{ggT}(A) = 1$. Sei weiter $s, s' \in A$. Dann gilt

$$\begin{aligned} \mathbf{P}(X_{s+s'} = i \mid X_0 = i) &= \sum_{j \in S} \mathbf{P}(X_{s+s'} = i \mid X_s = j) \cdot \mathbf{P}(X_s = j \mid X_0 = i) \\ &\geq \mathbf{P}(X_{s+s'} = i \mid X_s = i) \cdot \mathbf{P}(X_s = i \mid X_0 = i) \\ &= \mathbf{P}(X_{s'} = i \mid X_0 = i) \cdot \mathbf{P}(X_s = i \mid X_0 = i) > 0. \end{aligned}$$

Damit ist $s + s' \in A$ und aus Lemma 7.23 folgt nun die Behauptung im Falle $j = i$.

Falls $j \neq i$, können wir wegen der Irreduzibilität ein s so wählen, dass $\mathbf{P}(X_s = j \mid X_0 = i) > 0$ und s' so, dass $\mathbf{P}(X_{r'} = j \mid X_0 = j) > 0$ für $r' \geq s'$. Dann gilt mit $t = s + s'$ und $r \geq t$, dass

$$\begin{aligned} \mathbf{P}(X_r = j \mid X_0 = i) &\geq \mathbf{P}(X_r = j, X_s = j \mid X_0 = i) \\ &= \mathbf{P}(X_s = j \mid X_0 = i) \cdot \mathbf{P}(X_r = j \mid X_s = j) \\ &= \mathbf{P}(X_s = j \mid X_0 = i) \cdot \mathbf{P}(X_{r-s} = j \mid X_0 = j) > 0 \end{aligned}$$

und die Behauptung ist gezeigt. \square

Beweis von Theorem 7.19. Sei P die Übergangsmatrix der Markov-Kette. Nach Theorem 7.13 existiert eine stationäre Verteilung π . Es genügt nun, (7.8) zu zeigen. Würde nämlich eine weitere stationäre Verteilung $\pi' \neq \pi$ existieren, und würden wir \mathcal{X} so wählen, dass $\mathbf{P}(X_0 = i) = \pi'_i$. Dann würde aus (7.8) folgen, dass

$$0 = \lim_{t \rightarrow \infty} \mathbf{P}(X_t = i) - \pi(i) = \pi'(i) - \pi(i)$$

und damit $\pi' = \pi$.

Es bleibt also (7.8) zu zeigen. Der Beweis verwendet eine Kopplung, also einen stochastischen Prozess $(\mathcal{Y}, \mathcal{Z}) = (Y_t, Z_t)_{t=0,1,2,\dots}$ mit Wertebereich $S \times S$. Wir wählen $(\mathcal{Y}, \mathcal{Z})$ als homogene Markov-Kette mit $\mathbf{P}(Y_0 = i, Z_0 = j) = \mathbf{P}(X_0 = i) \cdot \pi_j$ und Übergangsmatrix $(Q_{ij,kl})_{i,j,k,l \in S^2}$,

$$Q_{ij,kl} = \begin{cases} P_{ik}P_{jl}, & i \neq j, \\ P_{ik}\delta_{kl}, & i = j. \end{cases}$$

Wir zeigen zunächst, dass sowohl \mathcal{Y} als auch \mathcal{Z} Markov-Ketten mit Übergangsmatrix P sind. Dazu berechnen wir

$$\begin{aligned} \mathbf{P}(Y_{t+1} = k \mid Y_t = i) &= \sum_{j,l \in S} \mathbf{P}(Y_{t+1} = k, Z_{t+1} = l \mid Y_t = i, Z_t = j) \cdot \mathbf{P}(Z_t = j) \\ &= \sum_{j \in S} \mathbf{P}(Z_t = j) \sum_{l \in S} Q_{ij,kl} = \sum_{j \in S} \mathbf{P}(Z_t = j) P_{ik} = P_{ik}. \end{aligned}$$

Genauso ergibt sich, dass \mathcal{Z} eine Markov-Kette mit Übergangsmatrix P ist.

Wie man aus der Übergangsmatrix Q ablesen kann, laufen in der Markov-Kette $(\mathcal{Y}, \mathcal{Z})$ die beiden Ketten \mathcal{Y} und \mathcal{Z} unabhängig voneinander, bis zu dem Zeitpunkt T , an dem $Y_T = Z_T$. Dann gilt, da $Q_{ij,kl} = P_{ik}\delta_{kl}$ für $i = j$, dass $Y_t = Z_t$ für $t \geq T$. Ein spezielles Beispiel für eine solche Kopplung haben wir bereits in Bemerkung 7.22 gesehen.

Da \mathcal{Y} dieselbe Startverteilung hat wie \mathcal{X} und \mathcal{Z} in der stationären Verteilung π startet, folgt, dass

$$\begin{aligned} |\mathbf{P}(X_s = i) - \pi(i)| &= |\mathbf{P}(Y_s = i) - \mathbf{P}(Z_s = i)| \\ &\leq \mathbf{E}[|1_{Y_s=i} - 1_{Z_s=i}|] \leq \mathbf{P}(Y_s \neq Z_s). \end{aligned} \tag{7.9}$$

Da $\mathbf{P}(Y_s \neq Z_s)$ fallend in s ist, genügt es, eine Teilfolge $s_n \rightarrow \infty$ zu finden mit $\mathbf{P}(Y_{s_n} \neq Z_{s_n}) \xrightarrow{n \rightarrow \infty} 0$. Nach Proposition 7.24 gibt es ein t mit

$$\alpha := \min\{(P^t)_{ik} : i, k \in S\} > 0.$$

Es gilt für jedes $k \in S$

$$\mathbf{P}(Y_t \neq Z_t) = 1 - \sum_{k \in S} \mathbf{P}(Y_t = Z_t = k) \leq 1 - \mathbf{P}(Y_t = k) \cdot \mathbf{P}(Z_t = k) \leq 1 - \alpha^2$$

Damit ist für jedes $n = 1, 2, \dots$

$$\begin{aligned} \mathbf{P}(Y_{nt} \neq Z_{nt}) &= \mathbf{P}(Y_t \neq Z_t) \cdot \mathbf{P}(Y_{2t} \neq Z_{2t} \mid Y_t \neq Z_t) \cdots \mathbf{P}(Y_{nt} \neq Z_{nt} \mid Y_{(n-1)t} \neq Z_{(n-1)t}) \\ &\leq (1 - \alpha^2)^n \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Mit (7.9) ist die Aussage gezeigt. □