Stochastik

Heinrich Voß

Technische Universität Hamburg–Harburg
Arbeitsbereich Mathematik

1999

Inhaltsverzeichnis

Kapitel 1

Grundbegriffe

Bei vielen Vorgängen des täglichen Lebens, deren Ausgang nicht vorhersagbar ist, verwendet man den Begriff des "Zufalls", um die Ergebnisse zu erklären (oder die eigene Unwissenheit zu begründen).

Dies ist z.B. bei der Zahl der Unfälle aller Versicherten eines Unternehmens in einem festen Zeitraum, bei der Schadenshöhe dieser Unfälle, bei der Laufzeit einer Maschine, aber auch bei Glücksspielen (Lotto, Roulette, Würfelspiele) der Fall.

Eine große Zahl von Beobachtungsreihen mit solchen vom Zufall beeinflussten Vorgängen hat gezeigt, dass auch der Zufall Gesetzmäßigkeiten genügt. Auch wenn in einer Krankenversicherung nicht vorhergesagt werden kann, welcher Versicherte zu welchem Zeitpunkt erkrankt und welche Kosten damit verbunden sind, so können doch die Gesamtkosten in einem Jahr im voraus geschätzt werden.

Die **Stochastik** ist diejenige mathematische Disziplin, die versucht, Gesetzmäßigkeiten bei vom Zufall beeinflussten Vorgängen mathematisch zu beschreiben (häufig mit dem Ziel, Vorhersagen zu ermöglichen). Die vom Zufall gesteuerten Vorgänge nennen wir **Zufallsexperimente**. Sie besitzen die folgenden charakteristischen Merkmale:

- sie sind (prinzipiell) unbeschränkt wiederholbar
- der Ausgang der einzelnen Experimente ist nicht vorhersagbar.

Definition 1.1. *Die Menge aller möglichen Ergebnisse des Zufallsexperiments heißt* **Stichprobenraum**. *Er wird meistens mit Ω bezeichnet.*

Beispiel 1.2.

1. Werfen einer Münze: $\Omega_1 = \{\text{Kopf, Zahl}\}$
2. Werfen eines Würfels: $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$
3. Werfen von zwei Würfeln: $\Omega_3 = \Omega_2 \times \Omega_2$
4. Anzahl der Anrufe in einer Telefonzentrale : $\Omega_4 = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$
5. Laufzeit einer Maschine: $\Omega_5 = \mathbb{R}_+ = [0, \infty)$. □

Definition 1.3. *Ein* **Ereignis** *ist eine Teilmenge A der Menge Ω (mit noch zu besprechenden Eigenschaften). Ist $A = \Omega$ bzw. $A = \emptyset$, so sprechen wir von dem* **sicheren** *bzw. von dem* **unmöglichen Ereignis** *. Die Teilmengen von Ω, die genau ein Element enthalten, heißen* **Elementarereignisse**.

Für die Mengenoperationen und -relationen verwendet man in der Stochastik auch die folgenden Sprechweisen:

$A \subset B$	:	das Ereignis A zieht das Ereignis B nach sich.
$A \cup B$	:	Summe der Ereignisse A und B (es tritt A oder B ein)
$A \cap B$	:	Produkt der Ereignisse A und B (sowohl A als auch B tritt ein)
$A \cap B = \emptyset$	:	sich ausschließende Ereignisse
$A \setminus B$	:	Differenz der Ereignisse A und B (es tritt A aber nicht B ein)
$\bar{A} := \Omega \setminus A$	:	das zu A komplementäre Ereignis (es tritt A nicht ein)

Wir ordnen nun Ereignissen Wahrscheinlichkeiten für ihr Auftreten zu. Eine natürliche Vorgehensweise ist die folgende: Wir betrachten eine Folge von gleichartigen Zufallsexperimenten. Wir bezeichnen mit p_n die Anzahl der Versuche unter den ersten n Experimenten, bei denen das Ereignis A eintrat. Dann ist p_n/n der Anteil der Erfolge unter den ersten n Versuchen. Wächst n über alle Grenzen, so wird dieser Quotient (hoffentlich) konvergieren, und man ordnet den Grenzwert dem Ereignis A als "Wahrscheinlichkeit" zu:

$$H(A) = \lim_{n\to\infty} \frac{p_n}{n} \tag{1.1}$$

$H(A)$ heißt **relative Häufigkeit** von A.

Beispiel 1.4. Abbildung 1.1 und Abbildung 1.2 enthalten die Quotienten p_n/n für $A := \{1, 2\}$ in Beispiel 2 für 100 Würfe bzw. 800 Würfe. Wie erwartet konvergiert p_n/n gegen 1/3. □

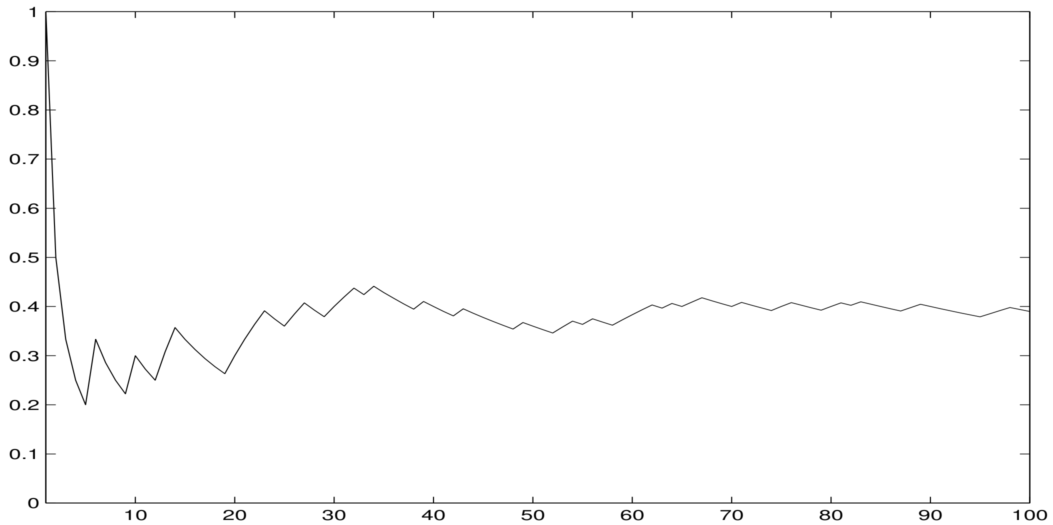

Abbildung 1.1: Beispiel 1.4. mit 100 Würfen

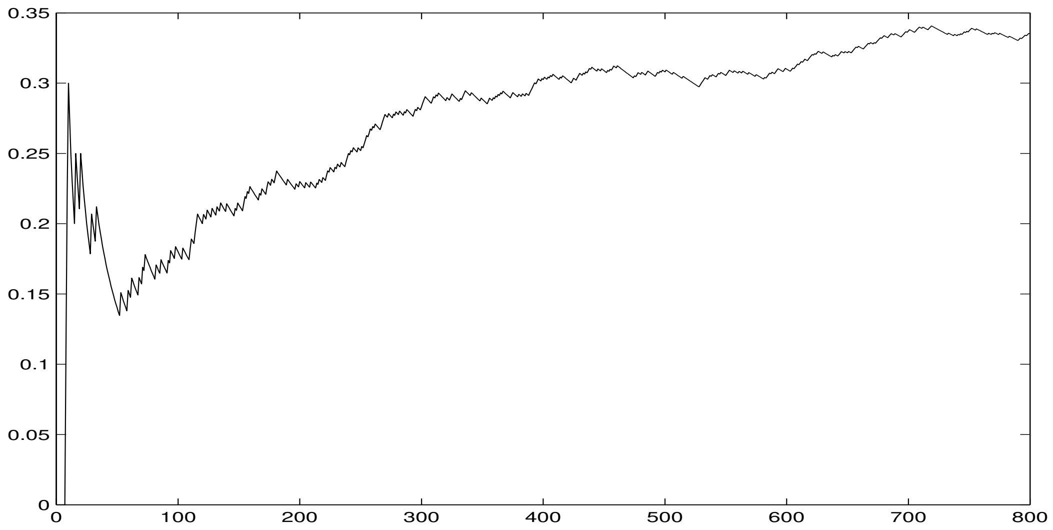

Abbildung 1.2: Beispiel 1.4. mit 800 Würfen

Die relative Häufigkeit besitzt die folgenden Eigenschaften:

1. Wegen $0 \leq p_n \leq n$ gilt $0 \leq p_n/n \leq 1$, d.h. $0 \leq H(A) \leq 1$.

2. Für $A = \Omega$ gilt sicher $p_n = n$ für alle n, d.h. $H(\Omega) = 1$.

3. Es gelte $A \cap B = \emptyset$. Ist p_n bzw. q_n bzw. r_n die Anzahl der Versuche unter den ersten n Experimenten, bei denen A bzw. B bzw. $A \cup B$ eingetreten ist, so gilt $r_n = p_n + q_n$, und daher
$$H(A \cup B) = H(A) + H(B).$$

4. Sind $A_1, \ldots, A_n$ endlich viele Ereignisse, die einander paarweise ausschließen (d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$), so erhält man durch Induktion
$$H\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} H(A_i).$$

Wir definieren umgekehrt

Definition 1.5. *Eine Abbildung P (wie probability), die jedem Ereignis eine reelle Zahl zuordnet, heißt* **Wahrscheinlichkeitsmaß**, *falls gilt:*

(i) $0 \leq P(A) \leq 1$ *für alle Ereignisse*

(ii) $P(\Omega) = 1$

(iii) Ist A_n eine Folge von Ereignissen, die einander paarweise ausschließen, so gilt
$$P\left(\bigcup_{n} A_n\right) = \sum_{n} P(A_n),$$
(dies ist eine Verschärfung der dritten Eigenschaft von H)

Hiermit kann man umgekehrt zeigen, dass bei geeigneter Definition der Begriffe "Konvergenz" und "Folge identischer Experimente" (1.1) gilt.

Der Übergang von endlichen zu abzählbaren Vereinigungen in (iii) ist natürlich, wie das folgende Beispiel zeigt.

Beispiel 1.6. Wir betrachten die zufallsabhängige Anzahl von Teilchen, die in einer Zeiteinheit von einem radioaktiven Präparat in die Zählkammer eines Geigerzählers eintreten.

Wir wählen als Stichprobenraum $\Omega := \mathbb{N}_0$ (die Anzahl ist zwar endlich, aber es ist unbequem, eine obere Schranke hierfür festzulegen), wobei $\omega \in \Omega$ der Anzahl der beobachteten Teilchen entspricht. $P(\omega)$ sei die Wahrscheinlichkeit dafür, dass genau ω Teilchen eintreten.

Für alle $n \in N$ folgt aus der endlichen Additivität

$$\begin{aligned} 1 &= P(\Omega) = P(\{0,\dots,n\} \cup \{n+1, n+2, \dots\}) \\ &= P(\{0,\dots,n\}) + P(\{n+1, n+2, \dots\}) \\ &= \sum_{i=0}^{n} P(\{i\}) + P(\{n+1, n+2 \dots\}) =: s_n + P(\{n+1, n+2, \dots\}). \end{aligned}$$

Offensichtlich gilt $0 \le s_0 \le s_1 \le s_2 \le \dots \le s_n \le \dots 1$, daher gibt es ein $s \in [0,1]$ mit $\lim_{n\to\infty} s_n = s$, und es folgt

$$\lim_{n\to\infty} P(\{n+1, n+2, \dots\}) = \lim_{n\to\infty} (1 - s_n) = 1 - s.$$

Nehmen wir an, dass $s < 1$ gilt, so erhalten wir wegen $s_n \le s$ für alle n

$$P(\{n+1, n+2, \dots\}) = 1 - s_n \ge 1 - s > 0,$$

d.h. für beliebig große $n \in \mathbb{N}$ ist die Wahrscheinlichkeit dafür, dass mehr als n Teilchen in die Kammer eintreten, größer als die positive Zahl $1-s$. Dies widerspricht unserer Vorstellung. Es ist also sinnvoll, über die endliche Additivität hinaus

$$\sum_{n=0}^{\infty} P(\{n\}) = 1 = P\left(\bigcup_{n=0}^{\infty} \{n\}\right),$$

oder (im Falle allgemeinerer Räume) (iii) zu fordern. □

Wir stellen nun einige Eigenschaften eines Wahrscheinlichkeitsmaßes zusammen.

Lemma 1.7. *(i)* $P(\bar{A}) = 1 - P(A)$ *für alle Ereignisse* A,

(ii) $P(\emptyset) = 0$,

(iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *für alle Ereignisse* A *und* B,

(iv) $A \subset B \implies P(A) \le P(B)$.

Beweis: (i) Wegen $A \cup \bar{A} = \Omega$, und $A \cap \bar{A} = \emptyset$ gilt nach Definition 1.5. (ii) und (iii)

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

(ii) Wegen $\emptyset \cup \emptyset = \emptyset$, $\emptyset \cap \emptyset = \emptyset$ gilt

$$P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset) \implies P(\emptyset) = 0.$$

(iii) Es gilt $A \cup B = A \cup (\bar{A} \cap B)$ und $B = (A \cap B) \cup (\bar{A} \cap B)$, wobei die Vereinigungen auf der rechten Seite jeweils disjunkt sind. Daher ist

$$P(A \cup B) = P(A) + P(\bar{A} \cap B), \quad P(B) = P(A \cap B) + P(\bar{A} \cap B),$$

und hieraus erhält man die Behauptung.

(iv) Es gilt $B = A \cup (B \cap \bar{A})$ (disjunkt), und $P(B \cap \bar{A}) \geq 0$ liefert

$$P(B) = P(A) + P(B \cap \bar{A}) \geq P(A). \qquad \blacksquare$$

Besonders einfach ist der Fall, dass die Menge Ω endlich oder abzählbar unendlich ist. Sind dann die Wahrscheinlichkeiten der Elementarereignisse bekannt, so kann man wegen (iii) die Wahrscheinlichkeit jedes Ereignisses bestimmen. Sei $P(\{\omega\}) = p_\omega$, $\omega \in \Omega$, und $A \subset \Omega$. Dann folgt aus (iii)

$$P(A) = \sum_{j \in A} p_j.$$

Ein so definiertes Wahrscheinlichkeitsmaß P heißt **diskret** .

Noch einfacher ist der Fall, dass Ω endlich ist und dass alle Elementarereignisse gleich wahrscheinlich sind (z.B. beim Werfen eines fairen Würfels). Dann gilt $p_j = \frac{1}{n}$ für alle j, und für eine beliebige Teilmenge A von Ω ist $P(A) = \frac{k}{n}$, falls A genau k Elemente besitzt.

Dies ist die klassische Definition der Wahrscheinlichkeit, die auf Laplace (1623-1662) zurückgeht. Da sich anfangs die Wahrscheinlichkeitstheorie vornehmlich mit Glücksspielen beschäftigte und da das Ereignis A häufig den günstigen Fall bezeichnete, bei dem man eine Auszahlung erhielt, findet man in der Literatur hierfür auch die Bezeichnung

$$P(A) = \frac{\text{Anzahl der für das Eintreten von } A \text{ günstigen Ereignisse}}{\text{Anzahl der möglichen Ereignisse}}.$$

Im Falle einer abzählbar unendlichen Menge Ω ist es natürlich unmöglich, dass alle Elementarereignisse gleiche Wahrscheinlichkeit besitzen. (beachte: $P(\Omega) = \sum\limits_{j=1}^{\infty} p_j = 1$).

Beispiel 1.8. Wir betrachten das Experiment, eine Münze so oft zu werfen, bis erstmals "Zahl" erscheint. Dies kann man beschreiben durch

$$\Omega := \mathbb{N}, \quad p_j := P(\text{erste "Zahl" erscheint beim } j\text{-ten Wurf}).$$

Da es genau 2^j (gleichwahrscheinliche) Möglichkeiten gibt, eine Münze j-mal zu werfen und nur bei einer von ihnen genau der j-te Wurf Zahl zeigt, gilt $p_j = 2^{-j}$, $j \in \mathbb{N}$.

Tatsächlich gilt

$$P(\Omega) = \sum_{j=1}^{\infty} p_j = \sum_{j=1}^{\infty} 2^{-j} = \frac{1}{1-0.5} - 1 = 1.$$

Ist A das Ereignis, dass die erste Zahl in einem Wurf mit ungerader Nummer auftritt, so gilt

$$P(A) = \sum_{j=1}^{\infty} 2^{-(2j-1)} = 2\sum_{j=1}^{\infty} \left(\frac{1}{4}\right)^j = 2\left(\frac{1}{1-0.25} - 1\right) = \frac{2}{3}. \qquad \square$$

Im Falle eines diskreten Wahrscheinlichkeitsmaßes ist es vernünftig, als Ereignisse alle Teilmengen von Ω zuzulassen. Ist Ω eine überabzählbare Menge, so ist die Potenzmenge (= Menge aller Teilmengen von Ω) als Menge aller Ereignisse häufig zu großḊie Menge $\mathcal{A}$ aller Ereignisse muss aber die folgenden Eigenschaften haben.

Definition 1.9. *Es sei Ω eine beliebige Menge. Ein System $\mathcal{A}$ von Teilmengen von Ω heißt σ-**Algebra** in Ω, wenn gilt*

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

(ii) $A \in \mathcal{A} \;\Rightarrow\; \bar{A} \in \mathcal{A}$

(iii) $A_j \in \mathcal{A},\ j \in \mathbb{N} \;\Rightarrow\; \bigcup\limits_{j=1}^{\infty} A_j \in \mathcal{A}$
(die abzählbare Vereinigung von Ereignissen ist ein Ereignis)

Bemerkung 1.10. Ist $\mathcal{A}$ eine σ-Algebra, so liegen auch die abzählbaren Durchschnitte von Elementen aus $\mathcal{A}$ in $\mathcal{A}$, denn

$$\bigcap_{n=1}^{\infty} A_n = \overline{\bigcup_{n=1}^{\infty} \bar{A}_n},$$

und es liegen Mengendifferenzen in $\mathcal{A}$, denn $B \setminus A = B \cap \bar{A}$. $\square$

Beispiel 1.11. Die wichtigste σ-Algebra (die nicht die Potenzmenge der Grundmenge ist) ist die σ-Algebra $\mathcal{B}$ der **Borelschen Mengen** in $\mathbb{R}$. Dies ist die kleinste σ-Algebra in $\mathbb{R}$, die alle Intervalle des Typs $(-\infty, a]$, $a \in \mathbb{R}$, enthält.

Die folgenden Gleichungen zeigen, dass $\mathcal{B}$ alle Intervalle in $\mathbb{R}$ enthält.

$$\begin{aligned} (a,b] &= (-\infty,b] \setminus (-\infty,a] \in \mathcal{B}, \quad (a,b) = \bigcup_{n\in\mathbb{N}} (a, b-\frac{1}{n}] \in \mathcal{B}, \\ [a,b] &= \bigcap_{n\in\mathbb{N}} (a-\frac{1}{n}, b] \in \mathcal{B}, \quad [a,b) = \bigcap_{n\in\mathbb{N}} (a-\frac{1}{n}, b) \in \mathcal{B}. \qquad \Box \end{aligned}$$

Bisher haben wir bei der Berechnung der Wahrscheinlichkeit eines Ereignisses A nur vorausgesetzt, dass der Ausgang des Experiments in der Menge Ω liegt. Wir nehmen nun an, dass wir die zusätzliche Information haben, dass der Ausgang des Experiments in der Teilmenge B von Ω liegt. Dann kann diese zusätzliche Kenntnis die Wahrscheinlichkeit für das Eintreten von A verändern. Wir nennen sie die bedingte Wahrscheinlichkeit von A unter der Bedingung B. (in Zeichen: $P(A\,|\,B)$).

Um die nachfolgenden Definition zu motivieren, betrachten wir ein diskretes Wahrscheinlichkeitsmaß. Dann ist die bedingte Wahrscheinlichkeit eines Elementarereignisses $\{j\} \subset \Omega$ unter der Bedingung B offenbar

$$P(\{j\}\,|\,B) = \begin{cases} \dfrac{p_j}{P(B)} & \text{falls} \quad j \in B \\ 0 & \text{falls} \quad j \in \bar{B}, \end{cases}$$

und daher

$$\begin{aligned} P(A\,|\,B) &= \sum_{j\in A} P(\{j\}\,|\,B) \\ &= \sum_{j\in A\cap B} P(\{j\}\,|\,B) + \sum_{j\in A\cap \bar{B}} P(\{j\}\,|\,B) \\ &= \sum_{j\in A\cap B} \frac{p_j}{P(B)} = \frac{P(A\cap B)}{P(B)}. \end{aligned}$$

Wir definieren daher allgemein

Definition 1.12. *Es seien $A, B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt*

$$P(A\,|\,B) := \frac{P(A\cap B)}{P(B)} \tag{1.2}$$

die **bedingte Wahrscheinlichkeit** *von A unter der Bedingung B . Im Falle $P(B) = 0$ ist $P(A\,|\,B)$ nicht definiert.*

Beispiel 1.13. Von 10000 IC Chips wurden 3000 von einer Firma X und 7000 von der Firma Y gefertigt. 10 % der Chips, die von X hergestellt wurden, sind defekt und 5 % der von Y gefertigten. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig gewählter Chip, der defekt ist, von der Firma X gefertigt wurde?

Sei

$$A := \{\text{Chip wurde von der Firma } X \text{ hergestellt}\}, \quad B := \{\text{Chip ist defekt}\}.$$

Dann gilt

$$P(A) = \frac{3000}{10000} = 0.3, \quad P(B) = \frac{650}{10000} = 0.065, \quad P(A \cap B) = \frac{300}{10000} = 0.03,$$

und daher

$$P(A \,|\, B) = \frac{0.03}{0.065} = \frac{30}{65}. \qquad \square$$

Löst man die Gleichung (1.2) nach $P(A \cap B)$ auf, so folgt

$$P(A \cap B) = P(A \,|\, B) \cdot P(B), \quad \text{falls } P(B) > 0 \tag{1.3}$$

und genauso

$$P(A \cap B) = P(B \,|\, A) \cdot P(A), \quad \text{falls } P(A) > 0. \tag{1.4}$$

(1.3) bzw. (1.4) heißen **Multiplikationsregel** für Wahrscheinlichkeiten .

Aus (1.3) und (1.4) folgt auch

$$\frac{P(A \,|\, B)}{P(B \,|\, A)} = \frac{P(A)}{P(B)}, \quad \text{falls } P(A) \cdot P(B) > 0.$$

In Beispiel 1.13. ist offensichtlich $P(B \,|\, A) = 0.1$ und daher

$$\frac{P(A \,|\, B)}{P(B|A)} = \frac{30}{65} \cdot 10 = \frac{P(A)}{P(B)}.$$

Beispiel 1.14. Ein Übertragungskanal zur Übermittlung binär (0,1) kodierter Daten arbeite (verursacht durch stochastisches Rauschen im Kanal) unzuverlässig.

Aus Untersuchungen weiß man, dass

93% aller “1” richtig übertragen werden,

92% aller “0” richtig übertragen werden.

Ferner ist bekannt, dass 40% aller übertragenen Zeichen “0” sind.

Wie groß ist die Wahrscheinlichkeit dafür, dass ein Übertragungsfehler auftritt?

Wir modellieren das Zufallsexperiment "Übertragen eines Bits" mit dem Stichprobenraum $\Omega := \{0,1\}^2$. Dabei tritt $(a,b) \in \Omega$ als Versuchsausgang auf, wenn a gesendet und b empfangen wurde. Da Ω endlich ist, wählen wir als σ-Algebra die Potenzmenge von Ω.

Über das Wahrscheinlichkeitsmaß P wissen wir

$$\begin{aligned} P(\text{“0” gesendet}) = P(\{(0,0),(0,1)\}) &= 0.4, \\ P(\text{“1” gesendet}) = P(\{(1,0),(1,1)\}) &= 0.6, \\ P(\text{“1” empfangen} \mid \text{“1” gesendet}) &= 0.93, \\ P(\text{“0” empfangen} \mid \text{“0” gesendet}) &= 0.92. \end{aligned}$$

Gesucht ist

$$p := P(\text{“Fehlübertragung”}) = P(\{(0,1),(1,0)\}).$$

Nach der Multiplikationsregel für Wahrscheinlichkeiten gilt

$$\begin{aligned} p &= P(\text{“0” empfangen} \mid \text{“1” gesendet}) \cdot P(\text{“1” gesendet}) \\ &\quad + P(\text{“1” empfangen} \mid \text{“0” gesendet}) \cdot P(\text{“0” gesendet}) \\ &= 0.07 \cdot 0.6 + 0.08 \cdot 0.4 = 0.074. \end{aligned}$$

Wir haben dabei verwendet, dass gilt

$$P(\text{“0” empfangen} \mid \text{“1” gesendet}) = 1 - P(\text{“1” empfangen} \mid \text{“1” gesendet}).$$

Dies ist zulässig, denn für jede Menge $B \in \mathcal{A}$ ist $P(\cdot \mid B)$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. □

Wir betrachten nun ein vollständiges System A_i, $i = 1, \ldots, n$, von Ereignissen, d.h. $A_i \in \mathcal{A}$, $i = 1, \ldots, n$, mit $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup\limits_{i=1}^n A_i = \Omega$, und es gelte $P(A_i) > 0$ für alle i.

Es sei $B \in \mathcal{A}$. Dann gilt

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^n A_i \right) = \bigcup_{i=1}^n (B \cap A_i),$$

und die Ereignisse $B \cap A_i$, $i = 1, \ldots, n$ schließen einander aus. Daher folgt

$$P(B) = P\left(\bigcup_{i=1}^n (B \cap A_i) \right) = \sum_{i=1}^n P(B \cap A_i), \tag{1.5}$$

und wegen $P(A_i) > 0$ und (1.3)

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i). \tag{1.6}$$

(1.6) heißt Formel für die **totale Wahrscheinlichkeit** . Sie ist offenbar auch für den Fall $n = \infty$ richtig.

Eine typische Anwendung dieser Formel zeigt das folgende Beispiel.

Beispiel 1.15. In einem Betrieb wird ein Produkt von drei Maschinen gefertigt. Die folgende Tabelle zeigt den Anteil jeder Maschine an der Gesamtproduktion und den Ausschussanteil

Maschine	Anteil(%)	Ausschuss(%)
1	50	1
2	30	2
3	20	5

Wie groß ist die Wahrscheinlichkeit, dass ein zufällig dem Lager entnommenes Erzeugnis nicht den Qualitätsansprüchen genügt?

Sei

$$A_i := \{\text{Produkt wurde mit Maschine } i \text{ erzeugt}\},\ i = 1, 2, 3$$

$$B := \{\text{Produkt ist Ausschuss}\}.$$

Dann bilden die A_i ein vollständiges System von Ereignissen, und die Formel für die totale Wahrscheinlichkeit liefert

$$\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3) \\ &= 0.01 \cdot 0.5 + 0.02 \cdot 0.3 + 0.05 \cdot 0.2 \ = \ 0.021. \end{aligned}$$

Wir fragen nun nach der Wahrscheinlichkeit, dass ein dem Lager entnommenes, defektes Produkt mit der Maschine i gefertigt wurde, d.h. $P(A_i \mid B)$.

Ausgehend von $P(A_i \mid B) \cdot P(B) = P(B \mid A_i)P(A_i)$ erhält man

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)},$$

d.h.

$$P(A_1 \mid B) = \frac{0.01 \cdot 0.5}{0.021} = \frac{5}{21}, \quad P(A_2 \mid B) = \frac{0.02 \cdot 0.3}{0.021} = \frac{6}{21},$$

$$P(A_3 \mid B) = \frac{0.05 \cdot 0.2}{0.021} = \frac{10}{21}. \qquad \square$$

Ist allgemein A_i, $i = 1, \dots, n, (\leq \infty)$ ein vollständiges System von Ereignissen, so gilt genauso für jedes Ereignis B

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{P(B)}, \quad i = 1, \dots, n,$$

und unter Benutzung der Formel für die totale Wahrscheinlichkeit

$$P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{\sum_{k=1}^{n} P(B \mid A_k) P(A_k)}. \tag{1.7}$$

(1.7) heißt die **Formel von Bayes**.

Beispiel 1.16. Wir setzen Beispiel 1.14. fort und fragen, wie groß die Wahrscheinlichkeit dafür ist, dass eine "1" gesendet worden ist, wenn eine "0" empfangen wird?

Nach der Formel von Bayes gilt

$$\begin{aligned}
&P(\text{“1” gesendet} \mid \text{“0”empfangen}) = P(\{(1,0),(1,1)\} \mid \{(0,0),(1,0)\}) \\
&= \frac{P(\text{“0” empfangen} \mid \text{“1”gesendet}) \cdot P(\text{“1”gesendet})}{P(\text{“0” em.} \mid \text{“1”gs.}) \cdot P(\text{“1”gs.}) + P(\text{“0” em.} \mid \text{“0”gs.}) \cdot P(\text{“0”gs.})} \\
&= \frac{0.07 \cdot 0.6}{0.07 \cdot 0.6 + 0.92 \cdot 0.4} \approx 0.102. \Box
\end{aligned}$$

Beispiel 1.17. Ein Test zur Erkennung einer infektiösen Krankheit habe eine Zuverlässigkeit von 98%, er bescheinigt also 98% aller Gesunden, dass sie gesund sind, und 98% aller Kranken, dass sie krank sind.

Es sei bekannt, dass ein Patient einer Risikogruppe angehört, in der 0.1% aller Individuen die Krankheit haben. Wir groß ist die Wahrscheinlichkeit, dass dieser Patient tatsächlich erkrankt ist, wenn der Test bei ihm positiv verlaufen ist.

Die Formel von Bayes liefert

$$\begin{aligned}
P(\text{krank}|\text{positiv}) &= \frac{P(\text{positiv}|\text{krank}) \cdot P(\text{krank}))}{P(\text{positiv}|\text{gesund}) P(\text{gesund}) + P(\text{positiv}|\text{krank}) \cdot P(\text{krank})} \\
&= \frac{0.98 \cdot 0.001}{0.02 \cdot 0.999 + 0.98 \cdot 0.001} = 0.047. \qquad \Box
\end{aligned}$$

Es ist möglich, dass $P(A \mid B)$ gegenüber $P(A)$ wächst (z.B. $B = A$, $P(A) < 1$) oder fällt (z.B. $B = \bar{A}$, $P(A) > 0$) oder unverändert bleibt. Im letzten Fall wird man die Ereignisse A und B als unabhängig betrachten.

Diese Definition ist nur sinnvoll, falls $P(B) > 0$ gilt. Aus (1.3) folgt dann $P(A \cap B) = P(A) \cdot P(B)$, und diese Gleichung ergibt auch im Fall $P(B) = 0$ einen Sinn. Wir definieren daher

Definition 1.18. *Zwei Ereignisse A und B heißen* **unabhängig** *(auch stochastisch unabhängig), wenn gilt*

$$P(A \cap B) = P(A) \cdot P(B). \tag{1.8}$$

Beispiel 1.19. Eine Münze werde dreimal geworfen. Es sei

$$\begin{aligned} A &:= \{\text{erster Wurf zeigt Kopf}\} = \{KKK, KKZ, KZK, KZZ\}, \\ B &:= \{\text{zweiter Wurf zeigt Kopf}\} = \{KKK, ZKK, KKZ, ZKZ\} \\ C &:= \{\text{genau zwei nacheinander ausgeführte Würfe zeigen Kopf}\} \\ &= \{KKZ, ZKK\}. \end{aligned}$$

Dann gilt

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{2}, \quad P(C) = \frac{1}{4},$$

und wegen

$$P(A \cap B) = P(\{KKK, KKZ\}) = \frac{1}{4} = P(A)P(B)$$

sind A und B unabhängig, wegen

$$P(A \cap C) = P(\{KKZ\}) = \frac{1}{8} = P(A)P(C)$$

sind A und C unabhängig, und wegen

$$P(B \cap C) = P(\{KKZ, ZKK\}) = \frac{1}{4} \neq P(B)P(C)$$

sind B und C nicht unabhängig. □

Definition 1.20. *Die Ereignisse $A_1, \ldots, A_n$ heißen* **unabhängig**, *wenn für jede Teilmenge $\{i_1, \ldots, i_k\}$ der Indexmenge $\{1, \ldots, n\}$*

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}) \quad \textit{gilt} \;,$$

sie heißen **paarweise unabhängig**, *wenn*

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \textit{für alle } i, j \in \{1, \ldots, n\}, \; i \neq j,$$

gilt.

Aus der paarweisen Unabhängigkeit von Ereignissen folgt i.a. nicht die Unabhängigkeit, wie das folgende Beispiel zeigt.

Beispiel 1.21. Ein Paar von Münzen werde geworfen. Sei

$$\begin{aligned} A &:= \{\text{erste Münze zeigt Kopf}\} = \{KK, KZ\}, \\ B &:= \{\text{zweite Münze zeigt Kopf}\} = \{KK, ZK\}, \\ C &:= \{\text{genau eine Münze zeigt Kopf}\} = \{KZ, ZK\}. \end{aligned}$$

Dann gilt

$$P(A) = P(B) = P(C) = \frac{1}{2},$$

und wegen

$$P(A \cap B) = \frac{1}{4} = P(A)P(B), \quad P(A \cap C) = \frac{1}{4} = P(A)P(C)$$

$$\text{und} \qquad P(B \cap C) = \frac{1}{4} = P(B)P(C)$$

sind die Ereignisse paarweise unabhängig.

Wegen $A \cap B \cap C = \emptyset$ gilt

$$0 = P(A \cap B \cap C) \neq P(A)P(B)P(C),$$

und sie sind nicht unabhängig. $\square$

Kapitel 2

Zufallsvariable

2.1 Motivation und Definition

Bei wahrscheinlichkeitstheoretischen Fragen interessieren häufig nicht alle Ereignisse $A \in \mathcal{A}$.

Beispiel 2.1. Wir würfeln mit zwei Würfeln und fragen, mit welcher Wahrscheinlichkeit die Summe der Augenzahlen größer als 8 ist.

Naheliegender Wahrscheinlichkeitsraum zur Beschreibung dieses Experiments ist

$$\Omega = \{(\omega_1, \omega_2) : \omega_i \in \{1, 2, 3, 4, 5, 6\}, i = 1, 2\}$$

mit dem Laplaceschen Wahrscheinlichkeitsmaß.

Gesucht ist $P(A)$ für $A := \{(\omega_1, \omega_2) \,:\, \omega_1 + \omega_2 > 8\}$.

Die Menge A kann man mit der Abbildung

$$X := \begin{cases} \Omega & \to \quad \mathbb{R} \\ (\omega_1, \omega_2) & \to \quad \omega_1 + \omega_2 \end{cases}$$

beschreiben als das Urbild der Menge $B := \{9, 10, 11, 12\}$ bzgl. X oder kurz durch

$$A = X^{-1}(B) =: [X > 9]. \qquad \square$$

Allgemeiner sind Abbildungen eines Wahrscheinlichkeitsraumes in die reellen Zahlen nützlich bei der Beschreibung von Versuchen. Wir haben Ereignisse bisher häufig verbal charakterisiert, z.B. "eine geworfene Münze zeigt Kopf" oder "eine mehrfach geworfene Münze zeigt im j-ten Wurf erstmals Kopf".

Diese Ereignisse lassen sich leichter mit einer Abbildung $X : \Omega \to \mathbb{R}$ als Urbildmengen beschreiben, z.B. $X(\text{Kopf}) = 0$, $X(\text{Zahl}) = 1$. Dann gilt

$$\begin{aligned} &\text{“eine geworfene Münze zeigt Kopf”} \\ = \quad & X^{-1}(\{0\}) =: [X = 0] = X^{-1}((-\infty, 0.5]) =: [X < 0.5]. \end{aligned}$$

Mit der Abbildung

$$X(\text{mehrfach geworfene Münze zeigt genau im } j\text{-ten Wurf erstmals Kopf}) = j$$

gilt

$$\begin{aligned} &\text{“eine mehrfach geworfene Münze zeigt spätestens im } k\text{-ten Wurf Kopf”} \\ = \quad & X^{-1}((-\infty, k]) \quad =: \quad [X \leq k]. \end{aligned}$$

Wir betrachten nur Abbildungen, die als Urbilder von Borelschen Mengen (dies sind Mengen, die man durch (abzählbare) Vereinigung, Durchschnitts- und Differenzbildung aus Intervallen erhält) Ereignisse besitzen. Hierzu genügt es zu fordern, dass $[X \leq a] \in \mathcal{A}$ für alle $a \in \mathbb{R}$ gilt (vgl. Beispiel 1.11.).

Definition 2.2. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X : \Omega \to \mathbb{R}$ heißt* **Zufallsvariable***, falls gilt*

$$[X \leq a] \in \mathcal{A} \quad \textit{für alle } a \in \mathbb{R}.$$

Bemerkung 2.3. Der Begriff Zufallsvariable ist missverständlich, hat sich aber eingebürgert. Besser wäre Zufallsfunktion oder Zufallsabbildung oder (was bisweilen auch benutzt wird) Zufallsgröße. □

2.2 Verteilungsfunktionen

Definition 2.4. *Sei X eine Zufallsvariable. Dann heißt die reelle Funktion*

$$F : \mathbb{R} \to \mathbb{R}, \quad F(t) := P(X \leq t)$$

Verteilungsfunktion *von X.*

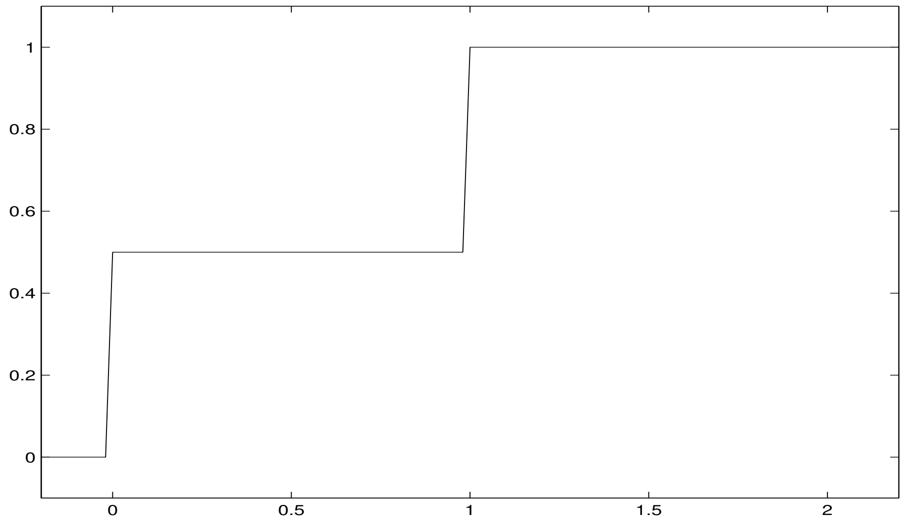

Abbildung 2.1: Verteilungsfunktion zu Beispiel 2.5.

Beispiel 2.5. Es sei

$$\Omega = \{\text{Kopf}, \text{Zahl}\}, \quad P(\text{Kopf}) = P(\text{Zahl}) = \frac{1}{2}$$

und

$$X : \Omega \to \mathbb{R}, \quad X(\text{Kopf}) = 0, \quad X(\text{Zahl}) = 1.$$

Dann gilt

$$[X \le t] = \begin{cases} \emptyset & \text{für } t < 0 \\ \{\text{Kopf}\} & \text{für } 0 \le t < 1 \\ \Omega & \text{für } 1 \le t \end{cases} \qquad \Rightarrow \qquad F(t) = \begin{cases} 0 & \text{für } t < 0 \\ 0.5 & \text{für } 0 \le t < 1 \\ 1 & \text{für } 1 \le t \end{cases}$$

Die Verteilungsfunktion F (vgl. Abbildung 2.1) ist also eine Treppenfunktion, die rechtsseitig stetig ist. □

In dem folgenden Satz fassen wir Eigenschaften von Verteilungsfunktionen zusammen.

Satz 2.6. *Sei X eine Zufallsvariable mit der Verteilungsfunktion F. Dann gilt*

(i) $0 \le F(t) \le 1$ *für alle* $t \in \mathbb{R}$

(ii) $F(t_1) \le F(t_2)$ *für alle* $t_1, t_2 \in \mathbb{R}$ *mit* $t_1 < t_2$, *d.h.* F *ist monoton nicht fallend.*

(iii) $\lim_{t\to\infty} F(t) = 1, \quad \lim_{t\to-\infty} F(t) = 0$

(iv) $\lim_{t\to t_0+0} F(t) = F(t_0)$ *für alle* $t_0 \in \mathbb{R}$, *d.h.* F *ist rechtsseitig stetig in* $\mathbb{R}$.

Beweis: (i) ist wegen $F(t) = P(X \leq t)$ klar.

(ii) Für $t_1 < t_2$ gilt $[X \leq t_1] \subset [X \leq t_2]$, und daher

$$F(t_1) = P(X \leq t_1) \leq P(X \leq t_2) = F(t_2).$$

(iii) und (iv) erhält man sofort aus Lemma 2.7. ∎

Lemma 2.7. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Dann gilt*

(i) (Ausschöpfung von innen; Stetigkeit von unten).
Ist $\{A_n\}$ eine isotone Folge von Ereignissen, d.h. $A_1 \subset A_2 \subset A_3 \subset \ldots$, so gilt

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

(ii) (Überdeckung von außen, Stetigkeit von oben).
Ist $\{A_n\}$ eine antitone Folge von Ereignissen, d.h. $A_1 \supset A_2 \supset A_3 \supset \ldots$, so gilt

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n).$$

Beweis: (i) Wegen $P(A_n) \leq P(A_{n+1}) \leq \ldots \leq 1$ existiert der Grenzwert

$$\lim_{n\to\infty} P(A_n).$$

Wegen $A_0 := \emptyset \subset A_1 \subset A_2 \subset \ldots$ gilt

$$\bigcup_{n=1}^{\infty} A_n = (A_1 - A_0) \cup (A_2 - A_1) \cup \ldots = \bigcup_{n=1}^{\infty} (A_n - A_{n-1}),$$

und die Vereinigung auf der rechten Seite ist disjunkt. Daher folgt

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} (A_n - A_{n-1})\right) = \sum_{n=1}^{\infty} P(A_n - A_{n-1}) \\ &= \lim_{m\to\infty} \sum_{n=1}^{m} (P(A_n) - P(A_{n-1})) = \lim_{m\to\infty} P(A_m). \end{aligned}$$

(ii) Wegen $A_1 \supset A_2 \supset \ldots$ erfüllt die Folge $\{A_1 - A_n\}_n$ die Voraussetzungen des Teils (i). Daher gilt

$$P\left(\bigcup_{n=1}^{\infty} (A_1 - A_n)\right) = \lim_{n\to\infty} P(A_1 - A_n) = P(A_1) - \lim_{n\to\infty} P(A_n).$$

Andererseits gilt

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty}(A_1 - A_n)\right) &= P\left(A_1 \cap \bigcup_{n=1}^{\infty} \bar{A}_n\right) = P\left(A_1 - \bigcap_{n=1}^{\infty} A_n\right) \\ &= P(A_1) - P\left(\bigcap_{n=1}^{\infty} A_n\right), \end{aligned}$$

zusammengenommen also die Behauptung. ∎

Bemerkung 2.8. Beispiel 2.5. zeigt, dass Verteilungsfunktionen i.a. nicht linksseitig stetig sind. □

Bemerkung 2.9. Für $t \in \mathbb{R}$ gilt wegen Lemma 2.7.(ii)

$$\begin{aligned} P(X = t) &= \lim_{n\to\infty} P\left(t - \frac{1}{n} < X \le t\right) \\ &= \lim_{n\to\infty} P\left([X \le t] - \left[X \le t - \frac{1}{n}\right]\right) \\ &= \lim_{n\to\infty} \left(P(X \le t) - P\left(X \le t - \frac{1}{n}\right)\right) \\ &= F(t) - \lim_{n\to\infty} F\left(t - \frac{1}{n}\right) = F(t) - F(t-0), \end{aligned}$$

wobei $F(t-0)$ den linksseitigen Grenzwert von F an der Stelle t bezeichnet (er existiert wegen der Monotonie von F).

Insbesondere gilt $P(X = t) = 0$ für jede Stetigkeitsstelle t von F. □

Bemerkung 2.10. Ist P ein Wahrscheinlichkeitsmaß auf der σ-Algebra der Borelschen Mengen $\mathcal{B}$, so hat die durch $F(t) := P((-\infty, t])$ definierte Funktion F die in Satz 2.6. enthaltenen Eigenschaften.

Umgekehrt kann man zeigen, dass es zu jeder monoton nicht fallenden, rechtsseitig stetigen Funktion $F : \mathbb{R} \to \mathbb{R}$ mit $\lim_{t\to-\infty} F(t) = 0$ und $\lim_{t\to\infty} F(t) = 1$ genau ein Wahrscheinlichkeitsmaß P auf der σ-Algebra der Borelschen Mengen gibt mit $P((-\infty, t]) = F(t)$ für alle $t \in \mathbb{R}$.

Die Verteilungsfunktionen und die Wahrscheinlichkeitsmaße auf $\mathcal{B}$ entsprechen einander also umkehrbar eindeutig, und es genügt daher, die Verteilungsfunktionen zu untersuchen. □

Beispiel 2.11.

$$F(t) := \begin{cases} 0 & , \text{ für } \; t < 0 \\ 1 - \exp(-\alpha t) & , \text{ für } \; t \ge 0 \end{cases}$$

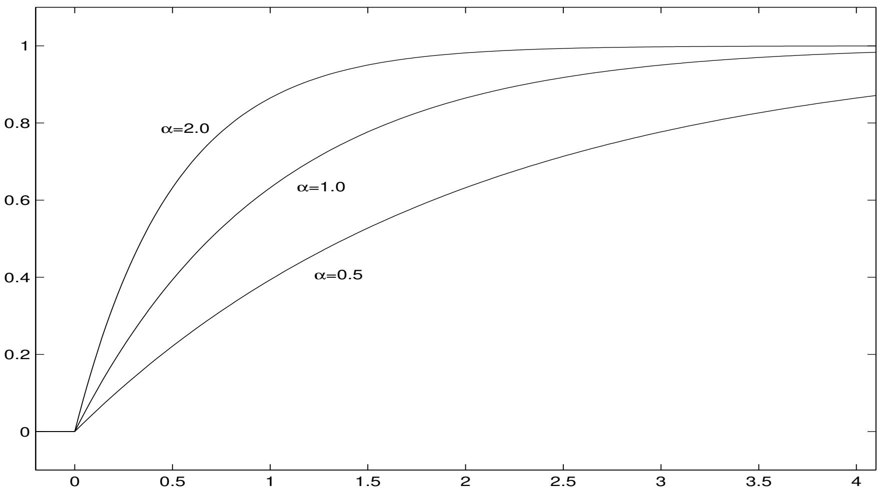

Abbildung 2.1: Exponentialverteilung

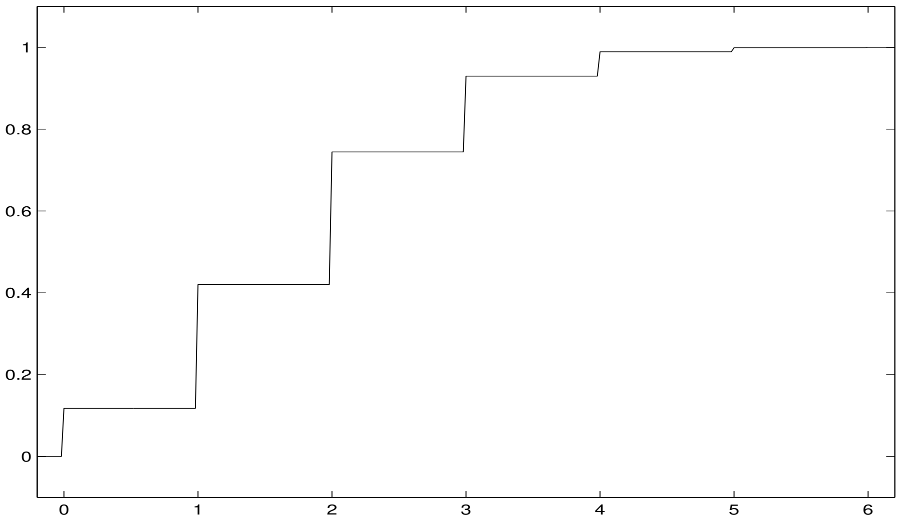

Abbildung 2.3: Binomialverteilung $b(6, 0.3)$

ist für jedes $\alpha > 0$ eine Verteilungsfunktion, die **Exponentialverteilung** zum Parameter α.

Durch sie wird die Lebensdauer von Stoffen beschrieben, die (zumindest angenähert) keinem Alterungsprozess unterliegen. □

Beispiel 2.12. Es sei $p \in [0, 1]$ und $n \in \mathbb{N}$.

Dann ist

$$F(t) := \sum_{j \le t} \binom{n}{j} p^j (1-p)^{n-j}$$

eine Verteilungsfunktion, die **Binomialverteilung** zu den Parametern n und p.

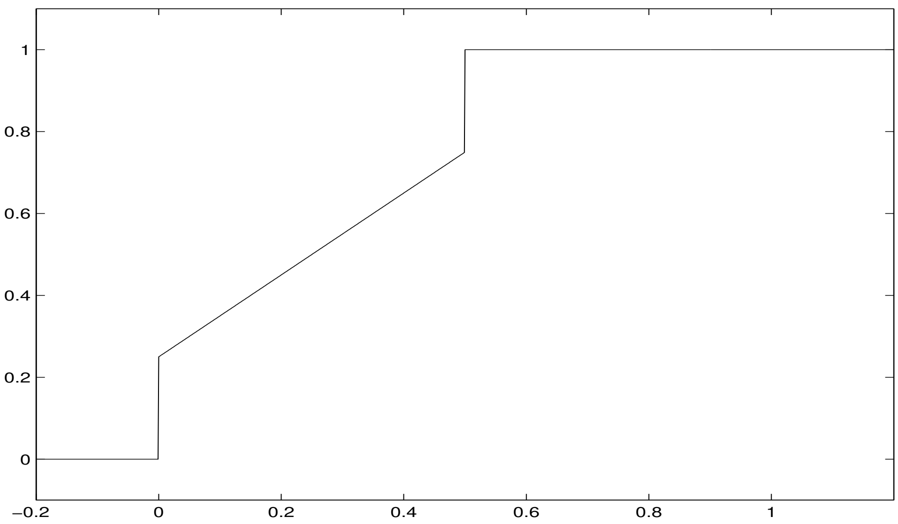

Abbildung 2.4: Verteilungsfunktion von Beispiel 2.13.

Diese tritt in der folgenden Situation auf. Bei der Produktion von Transistoren wird bei der Endkontrolle ein zufällig ausgewählter Transistor mit der Wahrscheinlichkeit $p \in [0,1]$ defekt sein. Die Kontrolle einer Einheit kann man beschreiben durch den diskreten Wahrscheinlichkeitsraum $\Omega_1 = \{0,1\}$ (0 $\hat{=}$defekt, 1$\hat{=}$nicht defekt), $P(\{0\}) = p$, $P(\{1\}) = 1-p$.

Die Entnahme von n Transistoren lässt sich beschreiben durch $\Omega = \{(\omega_1,\ldots,\omega_n) : \omega_i \in \{0,1\}\}$. Unter den Elementen von Ω gibt es $\binom{n}{k}$, die in genau k Komponenten eine 0 enthalten. Daher ist die Wahrscheinlichkeit, dass genau k Transistoren in einer entnommenen Stichprobe von n Transistoren defekt sind, $\binom{n}{k}p^k(1-p)^{n-k}$. Die Zufallsvariable

$$X : \Omega \to \mathbb{R}, \quad X(\omega) = \sum_{i=1}^{n} \omega_i,$$

besitzt also eine Binomialverteilung. □

Beispiel 2.13.

$$F(t) = \begin{cases} 0 & \text{für} & & & t < 0 \\ 0.25 + t & \text{für} & 0 & \leq & t < 0.5 \\ 1 & \text{für} & 0.5 & \leq & t \end{cases}$$

ist offenbar eine Verteilungsfunktion.

Die Verteilungsfunktion in Beispiel 2.11. ist stetig, in Beispiel 2.12. stückweise konstant. Beispiel 2.13. kann geschrieben werden als Konvexkombination

$F(t) = \frac{1}{2}(F_1(t) + F_2(t))$, wobei

$$F_1(t) = \begin{cases} 0 & \text{für} \quad t < 0 \\ 2t & \text{für} \quad 0 \leq t < 0.5 \\ 1 & \text{für} \quad 0.5 \leq t \end{cases}, F_2(t) = \begin{cases} 0 & \text{für} \quad t < 0 \\ 0.5 & \text{für} \quad 0 \leq t < 0.5, \\ 1 & \text{für} \quad 0.5 \leq t \end{cases}$$

also $F_1(t)$ stetig und $F_2(t)$ stückweise konstant ist. □

Allgemein kann man zeigen, dass jede Verteilungsfunktion als Konvexkombination

$$F(t) = \alpha_1 F_1(t) + \alpha_2 F_2(t) + \alpha_3 F_3(t), \quad 0 \leq \alpha_1, \alpha_2, \alpha_3, \quad \alpha_1 + \alpha_2 + \alpha_3 = 1,$$

geschrieben werden kann, wobei $F_1(t)$ eine stetige Verteilungsfunktion ist, $F_2(t)$ eine stückweise konstante Verteilungsfunktion (mit endlich oder abzählbar vielen Sprüngen) und $F_3(t)$ eine "singuläre" Verteilungsfunktion ist.

Da singulär Verteilungsfunktionen keine praktische Bedeutung besitzen, beschränken wir uns auf den Fall $\alpha_3 = 0$.

Definition 2.14. *Ist $\alpha_2 = 0$ und $\alpha_3 = 0$, also $F(t)$ stetig, so nennen wir die Zufallsvariable X* **stetig** *, ist $\alpha_1 = 0$, $\alpha_2 = 1$, also F stückweise konstant, so heißt die Zufallsvariable X* **diskret** *, und im Falle $0 < \alpha_1, \alpha_2 < 1$, $\alpha_1 + \alpha_2 = 1$, heißt die Zufallsvariable X* **vom gemischten Typ**.

2.3 Zähldichten

Wir betrachten nun eine diskrete Zufallsvariable X. Dann gibt es endlich oder abzählbar unendlich viele Punkte $t_0, t_1, t_2, \ldots$, so dass alle Ereignisse mit positiver Wahrscheinlichkeit in der Vereinigung der Mengen der Gestalt $[X = t_i]$, $i = 0, 1, 2, \ldots$, enthalten sind.

Definition 2.15. *Es sei X eine diskrete Zufallsvariable. Dann heißt die reelle Funktion*

$$f(t) := P(X = t) = F(t) - F(t - 0)$$

die **Zähldichte** *(oder auch nur Dichte) von X.*

Die Zähldichte ist nur in den Punkten t_i, $i = 0, 1, \ldots$, von Null verschieden. Mit der Verteilungsfunktion besteht der folgende Zusammenhang.

$$F(t) = \sum_{t_i \le t} f(t_i).$$

Wir geben nun die Zähldichten der wichtigsten diskreten Verteilungen an.

Beispiel 2.16. Für die Binomialverteilung ist $f(t) \neq 0$ genau für $t = i$, $i = 0, \ldots, n$, und die Zähldichte ist

$$f(i) =: b(i; n, p) = \binom{n}{i} p^i (1-p)^{n-i}.$$

Ist n groß so ist die Berechnung der $b(i; n, p)$ sehr aufwendig. Für $n = 2$ bis $n = 49$ wurde sie vom **National Bureau of Standards** (1950) und für $n = 50$ bis $n = 100$ von **H.G. Romig** ($50 - 100$ Binomial Tables, Wiley 1953) tabelliert. □

Beispiel 2.17. Wir betrachten das folgende Problem: Aus einer Gesamtheit von N Produkten, von denen genau a defekt seien, werde eine Stichprobe vom Umfang n gezogen. Wir fragen nach der Wahrscheinlichkeit dafür, dass genau i der ausgewählten Produkte defekt sind.

Die i defekten Produkte können auf genau $\binom{a}{i}$ Weisen gezogen werden, die $n - i$ nicht defekten Produkte auf genau $\binom{N-a}{n-i}$ Arten. Es können also die i defekten und die $n - i$ nicht defekten Produkte auf genau $\binom{a}{i}\binom{N-a}{n-i}$ Arten ausgewählt werden.

Da n Objekte aus einer Gesamtheit von N Objekten auf genau $\binom{N}{n}$ Weisen gezogen werden können, ist die gesuchte Wahrscheinlichkeit

$$f(i) =: h(i; n, a, N) = \frac{\binom{a}{i}\binom{N-a}{n-i}}{\binom{N}{n}}, \quad i = 0, \ldots, n,\ i \le a,\ n - i \le N - a.$$

Die Zähldichte h definiert die **hypergeometrische Verteilung** .

Man kann zeigen, dass bei wachsenden N und a, so dass $\frac{a}{N}$ gegen ein $p \in [0, 1]$ konvergiert, die hypergeometrische Verteilung $h(i; n, a, N)$ gegen die Binomialverteilung $b(i; n, p)$ konvergiert. Ist also N groß gegen n, so kann man $h(i; n, a, N)$ durch $b\left(i; n, \frac{a}{N}\right)$ ersetzen. Schon im Falle $n \le \frac{N}{10}$ erhält man hiermit gute Ergebnisse. □

Beispiel 2.18. Die Binomialverteilung kann man wiederum für große n durch die **Poissonverteilung** approximieren, die durch die Zähldichte

$$\pi(i; \lambda) = e^{-\lambda} \cdot \frac{\lambda^i}{i!}, \quad i \in \mathbb{N}_0,$$

definiert ist, wobei $\lambda > 0$ ein Parameter ist.

Wegen
$$\sum_{i=0}^{\infty} \pi(i;\lambda) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$
ist $\pi(i,\lambda)$ eine Zähldichte. □

Genauer gilt

Satz 2.19. *Sei $\{b(n,p_n)\}_{n\in\mathbb{N}}$ eine Folge von Binomialverteilungen mit $\lambda_n := n \cdot p_n \to \lambda \geq 0$. Dann gilt*
$$\lim_{n\to\infty} b(i;n,p_n) = \pi(i;\lambda) \quad \textit{für alle } i \in \mathbb{N}_0.$$

Beweis: durch vollständige Induktion: Für $i = 0$ gilt
$$b(0;n,p_n) = \binom{n}{0} p_n^0 (1-p_n)^n = \left(1 - \frac{\lambda_n}{n}\right)^n \to e^{-\lambda}.$$
Ist die Behauptung für ein $i \in \mathbb{N}_0$ bewiesen, so folgt wegen
$$\frac{b(i+1;n,p_n)}{b(i;n,p_n)} = \frac{\binom{n}{i+1} p_n^{i+1}(1-p_n)^{n-i-1}}{\binom{n}{i} p_n^i (1-p_n)^{n-i}} = \frac{n-i}{i+1} \frac{\lambda_n/n}{1-\lambda_n/n} \to \frac{\lambda}{i+1}$$
die Behauptung
$$\lim_{n\to\infty} b(i+1;n,p_n) = \frac{\lambda}{i+1} \lim_{n\to\infty} b(i;n,p_n) = \frac{\lambda^{i+1}}{(i+1)!} e^{-\lambda}. \qquad \blacksquare$$

Zu den angesprochenen Approximationseigenschaften betrachten wir

Beispiel 2.20. Bei der Produktion einer Ware sei 25% Ausschuss. Aus einer Gesamtheit von N Produkten wird eine Stichprobe vom Umfang 10 zufällig ausgewählt. Wie groß ist die Wahrscheinlichkeit W dafür, dass höchstens 2 Produkte der Stichprobe defekt sind?

Im Falle $N = 20$ bzw. $N = 100$ bzw. $N = 1000$ erhält man
$$\begin{array}{rcl} W &=& h(0;10,5,20) + h(1;10,5,20) + h(2;10,5,20) &=& 0.5000 \\ W &=& h(0;10,25,100) + h(1;10,25,100) + h(2;10,25,100) &=& 0.5217 \\ W &=& h(0;10,250,1000) + \ldots &=& 0.5252. \end{array}$$

Wegen $\frac{a}{N} = 0.25 = p$ erhält man in allen drei Fällen mit der Binomialverteilung als Näherung

$$W \approx b(0; 10, 0.25) + b(1; 10, 0.25) + b(2; 10, 0.25) = 0.5256.$$

Mit der Poissonverteilung erhält man hierfür als Approximation

$$W \approx \pi(0; 2.5) + \pi(1; 2.5) + \pi(2; 2.5) = 0.5438. \qquad \square$$

Beispiel 2.21. Wir betrachten eine Folge von Bernoulliexperimenten (d.h. Versuche mit zwei möglichen Ergebnissen, wobei der "Erfolg" mit Wahrscheinlichkeit p und der "Misserfolg" mit Wahrscheinlichkeit $q := 1 - p$ eintritt). Führt man diesen Versuch genau n mal durch, so ist die Anzahl der Erfolge natürlich binomialverteilt.

Wir führen die Experimente nun bis zum ersten Erfolg durch. Dann ist die Wahrscheinlichkeit dafür, dass genau im i-ten Versuch sich der erste Erfolg einstellt, offenbar

$$f(i) = q^{i-1}p.$$

Dies ist eine Zähldichte auf $\mathbb{N}$, denn

$$\sum_{i=1}^{\infty} f(i) = p \sum_{i=0}^{\infty} q^i = p \frac{1}{1 - (1 - p)} = 1.$$

Sie definiert die **geometrische Verteilung** .

Die geometrische Verteilung spielt eine Rolle bei Warteschlangenproblemen. Betrachtet man z.B. ein time-sharing System eines Computers mit fester Zeitintervallzuteilung und ist die Wahrscheinlichkeit dafür, dass die Ausführung eines Programmes am Ende eines Intervalls beendet ist, gleich p, so ist die Anzahl der Zeitintervalle bis zum Programmende geometrisch verteilt. $\square$

Bemerkung 2.22. Neben der geometrischen Verteilung betrachtet man die **modifizierte geometrische Verteilung** mit der Zähldichte

$$f(i) := q^i p, \quad i \in \mathbb{N}_0,$$

die man durch Umnummerierung aus der geometrischen Verteilung erhält. $\square$

2.4 Dichten

Ist die Zufallsvariable X stetig, ist also die zugehörige Verteilungsfunktion F stetig, so gilt $P(X = t) = F(t) - F(t-0) = 0$ für alle $t \in \mathbb{R}$, und anders als bei diskreten Zufallsvariablen ist diese Funktion nicht zur Beschreibung der Verteilung von X geeignet.

Für diskrete Verteilungen ist der Zusammenhang zwischen der Verteilungsfunktion und der Zähldichte gegeben durch

$$F(t) = \sum_{t_i \le t} f(t_i).$$

Eine natürliche Verallgemeinerung hiervon für stetige Zufallsvariable führt auf die folgende Definition.

Definition 2.23. *Es sei X eine stetige Zufallsvariable mit der Verteilungsfunktion F. Existiert eine integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit*

$$F(t) = \int_{-\infty}^{t} f(s)\, ds,$$

so heißt f **Verteilungsdichte** *oder kurz* **Dichte** *von X.*

Ist f eine Dichte von X und ist f stetig in $t \in \mathbb{R}$, so gilt

$$\lim_{h \to 0} \frac{1}{h}(F(t+h) - F(t)) = \lim_{h \to 0} \frac{1}{h} \int_{t}^{t+h} f(s)\, ds = f(t),$$

d.h. F ist differenzierbar in t und $F'(t) = f(t)$.

Da F monoton nicht fallend ist, gilt $f(t) \geq 0$ in jedem Stetigkeitspunkt t von f. Ferner ist

$$\int_{-\infty}^{\infty} f(s)\, ds = \lim_{t \to \infty} F(t) = 1,$$

und für jedes Intervall $(a, b] \subset \mathbb{R}$

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(s)\, ds$$

(und wegen der Stetigkeit von F ist dies auch gleich $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b)$).

Ist umgekehrt $f : \mathbb{R} \to \mathbb{R}$ integrierbar mit

(i) $f(s) \geq 0$ für alle $s \in \mathbb{R}$

(ii) $\int\limits_{-\infty}^{\infty} f(s)\, ds = 1$,

so wird durch $F(t) := \int\limits_{-\infty}^{t} f(s)\, ds$ eine stetige Verteilungsfunktion definiert.

Wir geben in den folgenden Beispielen die Dichten der wichtigsten stetigen Verteilungen an.

Beispiel 2.24. Eine stetige Zufallsvariable X heißt **gleichverteilt** auf dem Intervall $[a, b]$, wenn X die Dichte

$$f(s) = \begin{cases} \dfrac{1}{b-a} & \text{für} \quad a \leq s \leq b \\ 0 & \text{sonst} \end{cases}$$

besitzt. Die Verteilungsfunktion F ist dann offenbar

$$F(t) = \int\limits_{-\infty}^{t} f(s)\, ds = \begin{cases} 0 & \text{für} \quad t \leq a \\ \dfrac{t-a}{b-a} & \text{für} \quad a \leq t \leq b \\ 1 & \text{für} \quad b \leq t. \end{cases}$$

Auf dem Intervall $[a, b]$ gleichverteilte Zufallsvariable eignen sich offensichtlich zur Beschreibung von Beobachtungen, bei denen die Ergebnisse nur Zahlen des Intervalls $[a, b]$ sind und für die die Wahrscheinlichkeit dafür, dass das Ergebnis in einem Teilintervall $[c, d]$ von $[a, b]$ liegt, nur proportional zur Länge $d - c$ ist. □

Beispiel 2.25. Die Exponentialverteilung mit Parameter α besitzt die Verteilungsfunktion

$$F(t) = \begin{cases} 0 & \text{für} \quad t < 0 \\ 1 - \exp(-\alpha t) & \text{für} \quad t \geq 0, \end{cases}$$

und damit die Dichte

$$f(s) = \begin{cases} 0 & \text{für} \quad s < 0 \\ \alpha \exp(-\alpha s) & \text{für} \quad s \geq 0. \end{cases} \qquad \square$$

Wir haben bereits erwähnt, dass durch exponentialverteilte Zufallsvariable Wartezeiten für Prozesse beschrieben werden, bei denen Alterung keine Rolle spielt, z.B.

- die Zeit zwischen zwei eingehenden Anrufen in einer Telefonzentrale bzw. bis zum ersten Anruf
- die Lebensdauer eines elektronischen Bauteils
- die Reparaturzeit einer defekten Komponente eines Systems.

Diesen Zufallsvariablen ist gemeinsam, dass sie kein Gedächtnis haben. Wir demonstrieren dies an dem folgenden Beispiel.

Beispiel 2.26. Es sei die Lebensdauer X eines Transistors exponentialverteilt. Es sei $X > t$ für ein $t > 0$, und es sei Y die verbleibende Lebensdauer des Transistors, d.h. $Y := X - t$. Wir zeigen, dass die Verteilungsfunktion G von Y mit der von X übereinstimmt.

Es sei $y \geq 0$. Dann gilt, wenn f die Dichte von X bezeichnet,

$$\begin{aligned}
G(y) &= P(Y \leq y | X > t) = P(X - t \leq y | X > t) \\
&= \frac{P(t < X \leq t + y)}{P(X > t)} = \frac{\int_t^{y+t} f(s)\, ds}{\int_t^{\infty} f(s)\, ds} \\
&= \frac{\int_t^{y+t} \alpha \exp(-\alpha s)\, ds}{\int_t^{\infty} \alpha \exp(-\alpha s)\, ds} = \frac{\exp(-\alpha t)(1 - \exp(-\alpha y)}{\exp(-\alpha t)} \\
&= 1 - \exp(-\alpha y).
\end{aligned}$$

Für $y < 0$ ist $G(y) = P(X \leq y + t | X > t) = 0$, da dann

$$[X \leq y + t] \cap [X > t] = \emptyset.$$

Ist umgekehrt X eine nichtnegative stetige Zufallsvariable “ohne Gedächtnis”, so ist X exponentialverteilt, denn für $y > 0, t > 0$ folgt aus

$$\frac{P(t < X \leq y + t)}{P(X > t)} = P(X \leq y) = P(0 < X \leq y)$$

mit der Verteilungsfunktion F von X

$$F(y + t) - F(t) = (1 - F(t))(F(y) - F(0)).$$

Wegen $F(0) = 0$ erhält man hieraus

$$\frac{F(y+t) - F(y)}{t} = \frac{F(t) - F(0)}{t}(1 - F(y)),$$

und für $t \to 0$

$$F'(y) = F'(0+0)(1 - F(y)).$$

Durch Trennung der Variablen erhält man als Lösung dieser Differentialgleichung

$$1 - F(y) = C \cdot \exp(-F'(0+0)y)\ ,\ C \in \mathbb{R}.$$

Die Anfangsbedingung $F(0) = 0$ liefert $C = 1$, und mit $\alpha := F'(0+0)$ erhält man

$$F(y) = 1 - \exp(-\alpha y), \quad y > 0. \qquad \square$$

Beispiel 2.27. Die Exponentialverteilung ist ein Spezialfall der **Erlang Verteilung** $E(n, \alpha)$ mit den Parametern $n \in \mathbb{N}$ und $\alpha > 0$, die definiert ist durch die Verteilungsfunktion

$$F(t) := \begin{cases} 0 & , \ t \le 0 \\ 1 - \exp(-\alpha t) \sum\limits_{k=0}^{n-1} \dfrac{(\alpha t)^k}{k!} & , \ t \ge 0 \end{cases}$$

und die Dichte

$$f(s) = \begin{cases} 0 & , \ s < 0 \\ \alpha \exp(-\alpha s) \dfrac{(\alpha s)^{n-1}}{(n-1)!} & , \ s \ge 0. \end{cases}$$

Wie wir noch sehen werden, wird durch $E(n, \alpha)$ die Wartezeit bis zum Eintreten des n-ten Ereignisses in einem Prozess ohne Gedächtnis beschrieben (z.B. Wartezeit bis zum n-ten Anruf in einer Telefonzentrale). $\square$

Beispiel 2.28. Eine weitere Verallgemeinerung der Exponentialverteilung ist die **Weibull Verteilung**, durch die die Lebensdauer unter Berücksichtigung von Alterung beschrieben wird. Die Verteilungsfunktion der Weibull Verteilung ist

$$F(t) = \begin{cases} 0 & , \ t < 0 \\ 1 - \exp(-\alpha t^\lambda) & , \ t \ge 0 \end{cases}$$

mit positiven Parametern α, λ, ihre Dichte ist

$$f(s) = \begin{cases} 0 & , \ s < 0 \\ \alpha \lambda s^{\lambda - 1} \exp(-\alpha s^\lambda) & , \ s \ge 0. \end{cases}$$

Abbildung 2.5 zeigt den Graphen der Dichte der Weibull Verteilung für $\alpha = 1$ und für $\lambda = 0.5$, 1 und 2. Für $\lambda = 1$ hat man natürlich die Exponentialverteilung. $\square$

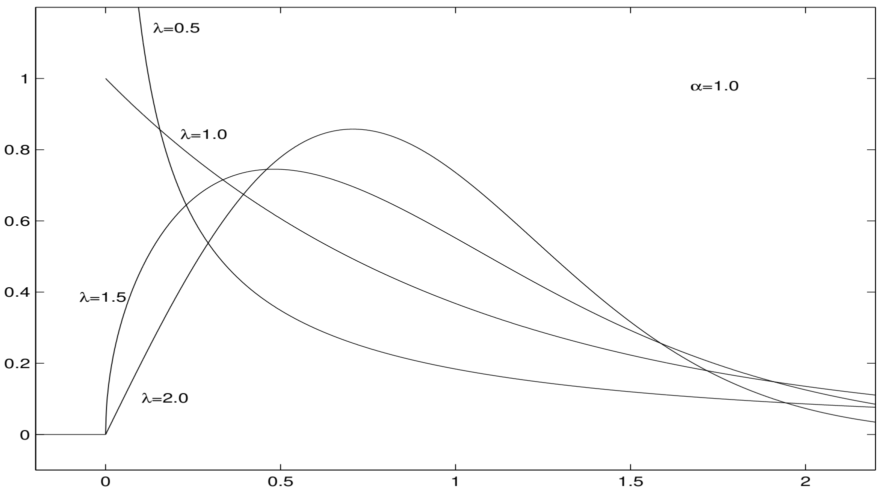

Abbildung 2.5: Dichte der Weibull Verteilung

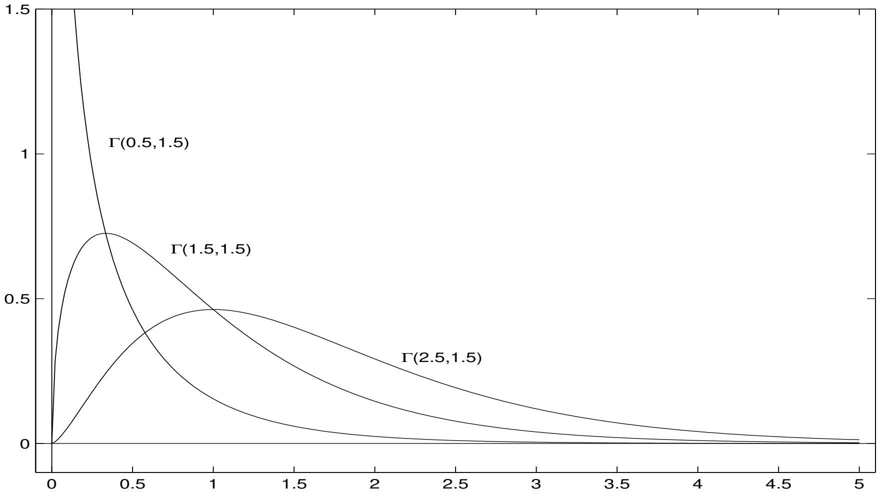

Abbildung 2.6: Dichte der Gamma Verteilung

Beispiel 2.29. Eine dritte Verallgemeinerung der Exponentialverteilung ist die **Gamma Verteilung**, die definiert ist durch die Dichte

$$f(s) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s} & , \quad s > 0 \\ 0 & \text{sonst,} \end{cases}$$

wobei $\alpha > 0$ und $\beta > 0$ Parameter sind.

Dabei ist $\Gamma(\alpha)$ der Wert der Gamma-Funktion an der Stelle α, die definiert ist durch

$$\Gamma(\alpha) := \int_0^\infty s^{\alpha-1} e^{-s}\, ds.$$

Offensichtlich gilt $\Gamma(1) = \int_0^\infty e^{-s}\, ds = 1$, und durch partielle Integration folgt $\Gamma(\alpha) =$

$(\alpha-1)\Gamma(\alpha-1)$ für alle $\alpha > 1$. Durch vollständige Induktion erhält man hieraus $\Gamma(n) = (n-1)!$ für jede natürliche Zahl n. Die Gamma-Funktion ist also eine verallgemeinerte Fakultät.

Für $\alpha = 1$ erhält man als Spezialfall der Gamma Verteilung die Exponentialverteilung mit dem Parameter β.

Ein weiterer wichtiger Spezialfall ist $\beta = \frac{1}{2}$, $\alpha = \frac{n}{2}$, $n \in \mathbb{N}$, die χ^2**-Verteilung** mit n Freiheitsgraden. □

Beispiel 2.30. Die wohl wichtigste stetige Verteilung der Wahrscheinlichkeitstheorie und der mathematischen Statistik ist die Normalverteilung. Als normalverteilt können Zufallsvariable angesehen werden, die Summen von vielen unabhängigen (also einander nicht beeinflussenden) Zufallsvariablen mit identischer Verteilung sind. Wir werden diesen Sachverhalt später noch in Form des zentralen Grenzwertsatzes präzisieren. Beispiele für normalverteilte Zufallsvariable sind

— der Stromverbrauch einer Großstadt während einer festen Zeitspanne.

— zufällige Mess- oder Beobachtungsfehler

— zufällige Abweichung vom Nennmaß bei der Produktion von Werkstücken.

Eine stetige Zufallsvariable X heißt **normalverteilt** mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$ — kurz $N(\mu, \sigma^2)$-verteilt —, wenn ihre Dichte gegeben ist durch

$$f(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right), \; s \in \mathbb{R}.$$

Die zugehörige Verteilungsfunktion

$$F(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{t} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds$$

ist nicht in geschlossener Form angebbar. Daher müssen Wahrscheinlichkeiten des Typs $P(a \le X \le b)$ numerisch bestimmt werden oder mit Hilfe von Tafelwerken.

Ist die Zufallsvariable X $N(\mu, \sigma^2)$-verteilt, so gilt für $Z := (X-\mu)/\sigma$ (mit der Variablentransformation: $S := (s-\mu)/\sigma$)

$$\begin{aligned} P(Z \le t) &= P(\frac{X-\mu}{\sigma} \le t) = P(X \le \sigma t + \mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\sigma t+\mu} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{1}{2}S^2\right) dS, \end{aligned}$$

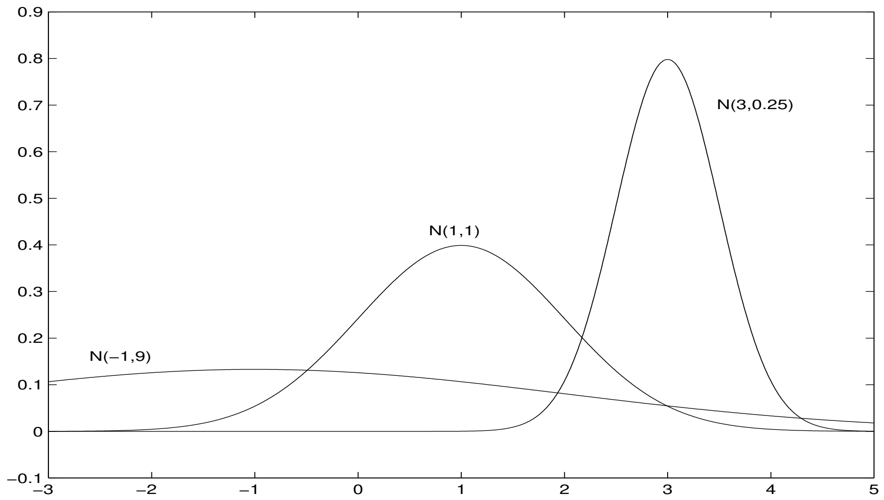

Abbildung 2.7: Dichte der Normalverteilung

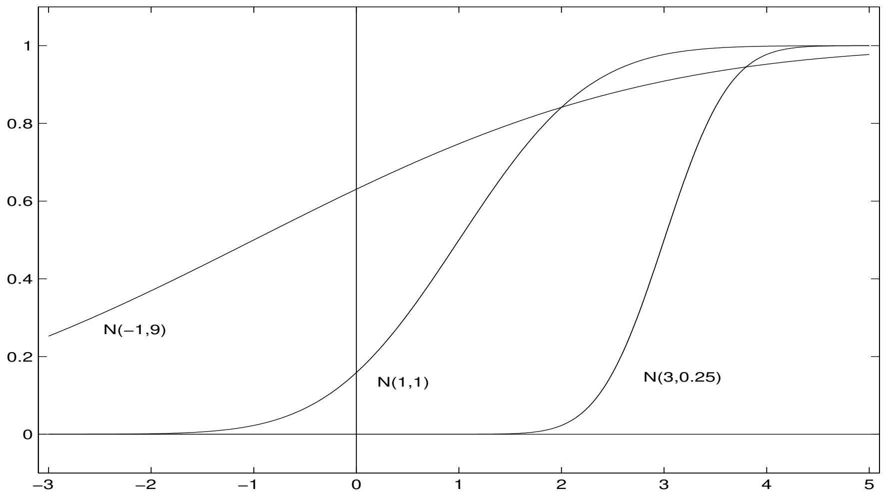

Abbildung 2.8:Verteilungsfunktion der Normalverteilung

d.h. Z ist $N(0,1)$-verteilt. Z heißt die zu X gehörende **standardisierte Zufallsvariable** .

Da die Dichte der standardisierten Zufallsvariable gerade ist, gilt für die zugehörige Verteilungsfunktion

$$\begin{aligned}\Phi(-t) &= \int_{-\infty}^{-t} f(s)\,ds = \int_{t}^{\infty} f(-s)\,ds = \int_{t}^{\infty} f(s)\,ds \\ &= 1 - \int_{-\infty}^{t} f(s)\,ds = 1 - \Phi(t).\end{aligned}$$

Es genügt daher, die Verteilungsfunktion nur für die standardisierte Normalverteilung nur für positive t zu tabellieren. □

Beispiel 2.31. Ein analoges Signal (gemessen in Mikrovolt), an einem Detektor zu einem festen Zeitpunkt könne als $N(240,256)$-verteilte Zufallsvariable modelliert werden. Wie groß ist die Wahrscheinlichkeit dafür, dass das Signal größer als 272 Mikrovolt ist?

Ist Φ die Verteilungsfunktion der $N(0,1)$-Verteilung und Z die standardisierte Zufallsvariable, so gilt

$$\begin{aligned}P(X>272) &= P(\sigma Z+\mu>272) = P(16Z+240>272) \\ &= P(Z>2) = 1-P(Z\le 2) = 1-\Phi(2) = 0.0228.\end{aligned}$$ □

Bezeichnet X eine gemessene Größe in einem Experiment, so ist

$$P(\mu-k\sigma\le X\le \mu+k\sigma)$$

ein Maß für den Messfehler. Es gilt, wenn Φ wieder die Verteilungsfunktion der $N(0,1)$-Verteilung bezeichnet,

$$\begin{aligned}P(\mu-k\sigma\le X\le\mu+k\sigma) &= F(\mu+k\sigma)-F(\mu-k\sigma) \\ &= \Phi(k)-\Phi(-k) \\ &= \frac{2}{\sqrt{2\pi}}\int_0^k \exp\left(-\frac{s^2}{2}\right)\,ds.\end{aligned}$$

Die **Fehlerfunktion** (oder das Fehlerintegral) ist definiert durch

$$e(u) := \frac{2}{\sqrt{\pi}}\int_0^u \exp(-y^2)\,dy.$$

Hiermit gilt

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = e\left(\frac{k}{\sqrt{2}}\right).$$

In vielen Büchern ist nicht die Verteilungsfunktion Φ der $N(0,1)$-Verteilung, sondern die Fehlerfunktion e tabelliert. In MATLAB ist das Fehlerintegral als Funktion 'erf' aufrufbar.

Viele physikalische Experimente führen auf nicht negative Zufallsvariable, während eine normalverteilte Zufallsvariable auch negative Werte annehmen kann. In dieser Situation verwendet man die **gestutzte Normalverteilung** mit der Dichte

$$f(s) = \begin{cases} 0 & , \quad s < 0 \\ \dfrac{\alpha}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(s-\mu)^2}{2\sigma^2}\right) & , \quad s \geq 0 \end{cases}$$

mit

$$\alpha^{-1} := \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds.$$

Wegen

$$\alpha^{-1} = \frac{1}{\sqrt{2\pi}} \int_{-\mu/\sigma}^\infty \exp\left(-\frac{t^2}{2}\right) dt = \Phi\left(\frac{\mu}{\sigma}\right)$$

gilt $1 > \alpha^{-1} > 0.9987$ für $\mu > 3\sigma$, und der Korrekturfaktor α kann für die meisten praktischen Zwecke vernachlässigt werden.

2.5 Erwartungswert

Wir haben gesehen, dass die Verteilung einer Zufallsvariable vollständig beschrieben wird durch die Verteilungsfunktion oder auch durch die Dichte. Häufig ist man mit einem Teil dieser Information zufrieden, der leichter fassbar ist als die Funktionen, da er nur durch eine oder wenige reelle Zahlen beschrieben wird. Man trifft dann auf Formulierungen wie "die mittlere Laufzeit einer Maschine ..." oder "der durchschnittliche Gewinn bei einem Glücksspiel ...".

Wir betrachten ein Spiel, bei dem ein Spieler mit der Wahrscheinlichkeit p_i den Gewinn x_i erzielen kann, $i = 1, \ldots, n$. Werden insgesamt N (N sehr groß) Runden gespielt, so wird der Spieler in etwa $N \cdot p_i$ Runden den Betrag $x_i, i = 1, \ldots, n$, gewinnen. Sein Gesamtgewinn wird also

$$N \cdot p_1 \cdot x_1 + N \cdot p_2 \cdot x_2 + \ldots + N \cdot p_n \cdot x_n = N \sum_{i=1}^n x_i \cdot p_i$$

sein, d.h. er kann erwarten, dass er je Runde einen durchschnittlichen Gewinn

$$\sum_{i=1}^{n} x_i p_i$$

erzielt. Diese Überlegung, die auf J. Bernoulli zurückgeht, motiviert die folgende Definition.

Definition 2.32. *Es sei X eine diskrete Zufallsvariable, die die Werte $t_i \in \mathbb{R}$, $i = 1, \dots, N \leq \infty$, annimmt, und es sei $p_i := P(X = t_i)$, $i = 1, \dots, N$, die zugehörige Zähldichte. Ist N endlich oder konvergiert die Reihe $\sum_{i=1}^{N} |t_i| \cdot p_i$, so heißt*

$$E(X) := \sum_{i=1}^{N} t_i p_i$$

der **Erwartungswert** *von X.*

Bemerkung 2.33. Konvergiert (im Falle $N = \infty$) die Reihe $\sum_{i=1}^{N} |t_i| p_i$ nicht, so existiert der Erwartungswert von X nicht. □

Bemerkung 2.34. Die Forderung der absoluten Konvergenz der Reihe wird eingeführt, um zu sichern, dass im Falle $N = \infty$ die Summe unabhängig von der Summationsreihenfolge ist. □

Bemerkung 2.35. Deutet man die Verteilung von X als ein Massensystem, wobei im Punkte t_i die Masse p_i liegt, $i = 1, \dots, N$, so ist $E(X)$ gerade der Schwerpunkt. □

Beispiel 2.36. Für die $b(1, p)$-Verteilung, die definiert ist durch

$$p_0 := P(X = 0) = 1 - p, \quad p_1 := P(X = 1) = p, \quad 0 \leq p \leq 1,$$

gilt

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

Für $p = 0.5$ ("Werfen einer fairen Münze") ist also der Erwartungswert gleich 0.5. Man sieht hieran schon, dass der Erwartungswert i.a. nicht im Wertebereich der Zufallsvariable X liegt. □

Beispiel 2.37. Ist X gleichverteilt auf $\{1,\dots,N\}$,

$$P(X=i)=\frac{1}{N},\quad i=1,\dots,N,$$

so gilt

$$E(X)=\sum_{i=1}^{N} i\frac{1}{N}=\frac{1}{N}\sum_{i=1}^{N} i=\frac{1}{2}(N+1). \qquad \square$$

Beispiel 2.38. Für die Binomialverteilung $b(n,p)$ gilt

$$\begin{aligned}
E(X) &= \sum_{i=0}^{n} i\binom{n}{i}p^i(1-p)^{n-i} = \sum_{i=1}^{n} i\frac{n!}{i!(n-i)!}p^i(1-p)^{n-i}\\
&= n\sum_{i=1}^{n}\frac{(n-1)!}{(i-1)!(n-i)!}p^i(1-p)^{n-i}\\
&= np\sum_{i=0}^{n-1}\frac{(n-1)!}{i!(n-1-i)!}p^i(1-p)^{n-1-i} = np. \qquad \square
\end{aligned}$$

Beispiel 2.39. Ist X Poisson-verteilt mit dem Parameter λ, so gilt

$$E(X)=\sum_{i=0}^{\infty} ie^{-\lambda}\cdot\frac{\lambda^i}{i!}=e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^i}{(i-1)!}=\lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=\lambda. \qquad \square$$

Beispiel 2.40. Besitzt X eine geometrische Verteilung,

$$P(X=i)=p(1-p)^{i-1},\quad i\in\mathbb{N},$$

so gilt

$$E(X)=\sum_{i=1}^{\infty} ip(1-p)^{i-1}.$$

Um diese Summe auszuwerten, betrachten wir für $z\in\mathbb{C}$ die Funktion

$$g(z):=\sum_{i=0}^{\infty}P(X=i)z^i=p\sum_{i=1}^{\infty}(1-p)^{i-1}z^i=\frac{p}{1-p}\left(\sum_{i=0}^{\infty}((1-p)z)^i-1\right).$$

Dann ist $g(z)$ für $(1-p)|z|<1$, insbesondere also in $z=1$, eine holomorphe Funktion, und die Reihe kann gliedweise differenziert werden.

$$g'(z)=\frac{p}{1-p}\sum_{i=1}^{\infty} i(1-p)^i z^{i-1}=\sum_{i=1}^{\infty} ip(1-p)^{i-1}z^{i-1}.$$

Es gilt also $g'(1)=E(X)$.

Andererseits hat man

$$g(z)=\frac{p}{1-p}\left(\frac{1}{1-(1-p)z}-1\right)=\frac{pz}{1-(1-p)z}.$$

Es folgt

$$g'(z) = \frac{p(1-(1-p)z)+(1-p)pz}{(1-(1-p)z)^2} = \frac{p}{(1-(1-p)z)^2},$$

und man erhält schließlich

$$E(X) = g'(1) = \frac{1}{p}. \qquad \square$$

Die hier am Beispiel der geometrischen Verteilung dargestellte Methode ist allgemeiner anwendbar.

Definition 2.41. *Es sei X eine diskrete Zufallsvariable mit Werten in $\mathbb{N}_0$. Dann heißt die Funktion*

$$g(z) := \sum_{i=0}^{\infty} P(X=i)\, z^i$$

die **erzeugende Funktion** *oder* **z-Transformierte** *von X.*

Satz 2.42. *Die diskrete Zufallsvariable X mit Werten in $\mathbb{N}_0$ besitze die erzeugende Funktion g.*

Der Erwartungswert $E(X)$ ist genau dann definiert, wenn der Grenzwert $g'(1-0) = \lim_{z\to 1-0} g'(z)$ existiert. In diesem Fall ist

$$E(X) = g'(1-0). \tag{2.1}$$

Beweis: g wird durch eine Potenzreihe definiert. Wegen

$$g(1) = \sum_{i=0}^{\infty} P(X=i) = 1 < \infty$$

ist g holomorph in der offenen Kreisscheibe $E = \{z : |z| < 1\}$, als reelle Funktion $g : [0,1] \to \mathbb{R}$ also beliebig oft differenzierbar in $[0,1)$ und stetig in $z=1$.

Wegen

$$g'(z) = \sum_{i=1}^{\infty} i\, P(X=i)\, z^{i-1}$$

besitzt X genau dann einen Erwartungswert, wenn $g'(1-0)$ existiert, und es gilt (2.1). ∎

Beispiel 2.43. Die erzeugende Funktion der Binomialverteilung $b(n,p)$ ist

$$g(z) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} z^i = (pz+1-p)^n.$$

Tatsächlich gilt, wie wir schon gesehen haben,

$$E(X) = g'(1) = np. \qquad \square$$

Für stetige Zufallsvariable definieren wir:

Definition 2.44. *Ist X eine stetige Zufallsvariable mit der Dichte f, so heißt*

$$E(X) = \int_{-\infty}^{\infty} s f(s)\, ds$$

der **Erwartungswert** *von X, falls das uneigentliche Integral absolut konvergiert, d.h.*

$$\int_{-\infty}^{\infty} |s| f(s)\, ds < \infty.$$

Ist diese Bedingung nicht erfüllt, so existiert kein Erwartungswert von X.

Beispiel 2.45. Ist X gleichverteilt auf dem Intervall $[a, b]$, so gilt

$$E(X) = \int_{-\infty}^{\infty} s f(s)\, ds = \frac{1}{b-a} \int_a^b s\, ds = \frac{1}{2}(a+b),$$

d.h. $E(X)$ ist der Mittelwert des Intervalls $[a, b]$. □

Beispiel 2.46. Ist X exponentialverteilt mit Parameter α, so gilt

$$E(X) = \int_0^{\infty} \alpha s \exp(-\alpha s)\, ds = \frac{1}{\alpha}.$$

Genauso erhält man für eine $E(n, \alpha)$-verteilte Zufallsvariable X

$$E(X) = \frac{n}{\alpha}. \qquad \square$$

Beispiel 2.47. Ist X $N(\mu, \sigma^2)$-verteilt, so gilt mit der Variablentransformation $t := (s - \mu)/\sigma$

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} s \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (t\sigma + \mu) \exp\left(-\frac{1}{2}t^2\right) dt \;=\; \mu, \end{aligned}$$

da $t \to t \exp(-t^2/2)$ eine ungerade Funktion ist, also $\int_{-\infty}^{\infty} t \exp(-t^2/2)\, dt = 0$, und $\int_{-\infty}^{\infty} \exp(-t^2/2)\, dt = \sqrt{2\pi}$ gilt. □

Definition 2.48. *Ist X eine gemischte Zufallsvariable, so besitzt X eine (eindeutige) konvexe Zerlegung $X = \alpha_1 X_1 + \alpha_2 X_2$, wobei X_1 eine diskrete und X_2 eine stetige Zufallsvariable ist. Existieren die Erwartungswerte $E(X_1)$ und $E(X_2)$, so heißt*

$$E(X) = \alpha_1 E(X_1) + \alpha_2 E(X_2)$$

der **Erwartungswert** *von X.*

Ist X eine Zufallsvariable über dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und ist $\phi : \mathbb{R} \to \mathbb{R}$ eine stückweise stetige Funktion, so ist auch $\phi \circ X : \Omega \to \mathbb{R}$ eine Zufallsvariable über $(\Omega, \mathcal{A}, P)$, denn die Borelschen Mengen werden auch von dem System der offenen Mengen in $\mathbb{R}$ erzeugt. Ist $A \subset \mathbb{R}$ offen und ϕ stetig, so ist auch $\phi^{-1}(A)$ eine offene Menge in $\mathbb{R}$, und daher gilt

$$[\phi \circ X \in A] = [X \in \phi^{-1}(A)] \in \mathcal{A}.$$

Ist ϕ nur stückweise stetig, so muss man die Sprungstellen von ϕ in A gesondert behandeln. Ihre Urbilder sind halboffene Intervalle, also ebenfalls Borelsche Mengen.

Der folgende Satz gibt Auskunft über den Erwartungswert von $\phi \circ X$.

Satz 2.49. *Sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, und sei $\phi : \mathbb{R} \to \mathbb{R}$ eine stückweise stetige Funktion.*

Ist

$$F(t) := \alpha F_d(t) + (1 - \alpha) F_s(t)$$

die konvexe Zerlegung der Verteilungsfunktion F von X in den diskreten Anteil F_d und den stetigen Anteil F_s und bezeichnet f_d die Zähldichte zu F_d und f_s die Verteilungsdichte von F_s, so gilt für den Erwartungswert von $\phi \circ X$

$$E(\phi \circ X) = \alpha \sum_{y \in \mathbb{R}} \phi(y) f_d(y) + (1 - \alpha) \int_{-\infty}^{\infty} \phi(y) f_s(y)\, dy.$$

Beweis: Die Aussage folgt unmittelbar aus einem (sehr allgemeinen) Transformationssatz für Integrale, der hier nicht zur Verfügung steht.

Wir betrachten nur den Fall, dass X eine diskrete Zufallsvariable mit der Zähldichte f_d ist. Dann nimmt X nur abzählbar viele Werte t_j, $j \in \mathbb{N}$, an. Für $y \in \mathbb{R}$ gilt

$$[\phi \circ X = y] = \{\omega \in \Omega : \phi \circ X(\omega) = y\} = \{\omega \in \Omega : X(\omega) \in \phi^{-1}(y)\}.$$

Da für $y \neq \tilde{y}$ die Mengen $[\phi \circ X = y]$ und $[\phi \circ X = \tilde{y}]$ disjunkt sind und der Wertebereich von X abzählbar ist, gibt es (höchstens) abzählbar viele Werte y_j, $j \in \mathbb{N}$, mit $[\phi \circ X = y_j] \neq \emptyset$. Es ist also auch die Zufallsvariable $\phi \circ X$ diskret.

Für die Zähldichte gilt

$$g(y_j) := P([\phi \circ X = y_j]) = P([X \in \phi^{-1}(y_j)]) = \sum_{\phi(t_k)=y_j} f_d(t_k),$$

und daher ist der Erwartungswert

$$\begin{aligned} E(\phi \circ X) &= \sum_j y_j g(y_j) = \sum_j y_j \sum_{\phi(t_k)=y_j} f_d(t_k) \\ &= \sum_j \sum_{\phi(t_k)=y_j} y_j f_d(t_k) = \sum_j \sum_{\phi(t_k)=y_j} \phi(t_k) f_d(t_k) \\ &= \sum_k \phi(t_k) f_d(t_k). \end{aligned}$$

Der Beweis lässt sich nicht auf den Fall einer stetigen Zufallsvariable X übertragen, da dann $\phi \circ X$ nicht notwendig stetig ist. ∎

Beispiel 2.50. Es sei X gleichverteilt auf $[a, b]$ und $Y := X^2$. Dann gilt

$$E(Y) = \int_a^b x^2 \frac{1}{b-a} \, dx = \frac{b^3 - a^3}{3(b-a)} = \frac{1}{3} (b^2 + ab + a^2). \qquad \square$$

2.6 Varianz

Man kann den Erwartungswert einer Zufallsvariable X als mittleren Wert von X auffassen, um den X schwankt. Über die Größe dieser Schwankung gibt der Erwartungswert keinerlei Auskunft. Es liegt nahe, als Maß für die Größe dieser Schwankung die mittlere quadratische Abweichung der Zufallsvariable von ihrem Erwartungswert einzuführen.

Definition 2.51. *Es sei X eine Zufallsvariable, und es existiere $E(X)$. Dann ist auch $(X - E(X))^2$ eine Zufallsvariable, die die quadratische Abweichung von X von $E(X)$ beschreibt.*

Existiert

$$V(X) = E((X - E(X))^2),$$

so heißt $V(X)$ die **Varianz** *von X (auch Dispersion von X). Die Größe $\sqrt{V(X)}$ heißt* **Standardabweichung** *von X.*

Bemerkung 2.52. Ist X eine diskrete Zufallsvariable, die den Erwartungswert $E(X) = \mu$ besitzt und die Werte t_i, $i = 0, \dots, N \le \infty$ annimmt, so gilt also

$$V(X) = \sum_{i=0}^{N} (t_i - \mu)^2 P(X = t_i).$$

Ist X eine stetige Zufallsvariable mit dem Erwartungswert μ und der Dichte f, so gilt

$$V(X) = \int_{-\infty}^{\infty} (s - \mu)^2 f(s)\, ds. \qquad \square$$

Beispiel 2.53. Ist X $b(1,p)$-verteilt, so gilt $E(X) = p$, und daher

$$V(X) = E((X - p)^2) = (0 - p)^2 (1 - p) + (1 - p)^2 \cdot p = p(1 - p). \qquad \square$$

Für diskrete Zufallsvariable kann man die Varianz wieder mit Hilfe der erzeugenden Funktion berechnen.

Satz 2.54. *Die diskrete Zufallsvariable X mit Werten in $\mathbb{N}_0$ besitze die erzeugende Funktion g, und es existiere*

$$g'(1 - 0) = \mu = E(X).$$

Die Varianz $V(X)$ von X existiert genau dann ,wenn der Grenzwert $g''(1 - 0) := \lim_{z \to 1-0} g''(z)$ existiert. In dem Fall gilt

$$V(X) = g''(1 - 0) + g'(1 - 0) - (g'(1 - 0))^2.$$

Beweis: Mit $p_i := P(X = i)$ ist

$$g''(z) = \sum_{i=2}^{\infty} i(i - 1) p_i z^{i-2},$$

und daher

$$\begin{aligned} V(X) &= \sum_{i=0}^{\infty} (i - \mu)^2 p_i \\ &= \sum_{i=0}^{\infty} i(i-1)p_i + \sum_{i=0}^{\infty} i p_i - 2\mu \underbrace{\sum_{i=0}^{\infty} i p_i}_{=\mu} + \mu^2 \underbrace{\sum_{i=0}^{\infty} p_i}_{=1} \\ &= g''(1 - 0) + g'(1 - 0) - (g'(1 - 0))^2, \end{aligned}$$

und hieraus liest man die Behauptung ab. ∎

Beispiel 2.55. Ist X $b(n,p)$-verteilt, so besitzt X die erzeugende Funktion

$$g(z) = (pz + 1 - p)^n$$

mit $g'(1) = np$ und $g''(1) = n(n-1)p^2$. Es folgt also

$$V(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p). \qquad \square$$

Beispiel 2.56. Ist X geometrisch verteilt, so gilt

$$\begin{aligned} g'(z) &= p(1-(1-p)z)^{-2}, \\ g''(z) &= 2p(1-p)(1-(1-p)z)^{-3}, \end{aligned}$$

und daher

$$V(X) = 2p^{-2}(1-p) + p^{-1} - p^{-2} = (1-p)p^{-2}. \qquad \square$$

Beispiel 2.57. Die Poisson Verteilung besitzt die erzeugende Funktion

$$g(z) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} z^i = e^{\lambda(z-1)},$$

und daher besitzt eine Poisson verteilte Zufallsvariable X die Varianz

$$V(X) = g''(1-0) + g'(1-0) - (g'(1-0))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \qquad \square$$

Beispiel 2.58. Ist X exponentialverteilt, so erhält man durch partielle Integration

$$V(X) = \int_0^{\infty} \left(s - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha s}\, ds = \frac{1}{\alpha^2},$$

und genauso für $E(n,\alpha)$-verteiltes X

$$V(X) = \frac{n}{\alpha^2}. \qquad \square$$

Beispiel 2.59. Ist X $N(\mu,\sigma^2)$-verteilt, so gilt mit $t := (s-\mu)/\sigma$

$$\begin{aligned} V(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (s-\mu)^2 \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 t^2 \exp\left(-\frac{1}{2}t^2\right) dt = \sigma^2, \end{aligned}$$

denn durch partielle Integration erhält man

$$\int_{-\infty}^{\infty} t \cdot t \exp\left(-\frac{1}{2}t^2\right) dt = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}t^2\right) dt = \sqrt{2\pi}. \qquad \square$$

Bevor wir in Satz 2.61. einige wichtige Eigenschaften des Erwartungswerts und der Varianz zusammenstellen, beweisen wir zunächst

Lemma 2.60. *Es sei X eine stetige, nichtnegative Zufallsvariable mit endlichem Erwartungswert. Dann gilt mit der Verteilungsfunktion F von X*

$$E(X) = \int_0^\infty (1 - F(x))\, dx.$$

Beweis: Durch partielle Integration erhält man

$$\begin{aligned} E(X) &= \int_0^\infty x f(x)\, dx \\ &= \lim_{t\to\infty} \left\{ [xF(x)]_0^t - \int_0^t F(x)\, dx \right\} \\ &= \lim_{t\to\infty} \left\{ tF(t) - \int_0^t F(x)\, dx \right\} \\ &= \lim_{t\to\infty} \left\{ -t(1 - F(t)) + \int_0^t (1 - F(x))\, dx \right\}. \end{aligned}$$

Da der Erwartungswert von X endlich ist, gilt

$$\begin{aligned} t(1 - F(t)) &= t\left(1 - \int_0^t f(x)\, dx\right) = t\int_t^\infty f(x)\, dx \\ &\le \int_t^\infty x f(x)\, dx \to 0 \quad \text{für } t \to \infty, \end{aligned}$$

und daher

$$E(X) = \int_0^\infty (1 - F(x))\, dx. \qquad \blacksquare$$

Satz 2.61. *Es seien X und Y Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Existieren die Erwartungswerte von X und Y und die Varianz von X, so existieren für alle $a, b \in \mathbb{R}$ auch $E(aX + bY)$ und $V(aX + b)$, und es gilt*

(i) $E(aX + bY) = aE(X) + bE(Y)$ *(Linearität)*

(ii) Aus $X(\omega) \le Y(\omega)$ für alle $\omega \in \Omega$ folgt $E(X) \le E(Y)$ *(Monotonie)*

(iii) $V(X) = E(X^2) - (E(X))^2$

(iv) $V(aX + b) = a^2 V(X)$

Beweis: Den Beweis von (i) müssen wir zurückstellen, bis uns die gemeinsame Verteilung von Zufallsvariablen in Kapitel 3 zur Verfügung steht.

Die Monotonie (ii) zeigen wir nur für nichtnegative, stetige Zufallsvariable. Aus $X(\omega) \leq Y(\omega)$ für alle ω folgt $[X \leq x] \supset [Y \leq x]$. Daher ist $P(X \leq x) \geq P(Y \leq x)$, d.h. $F_X(x) \geq F_Y(x)$, für alle x, und aus Lemma 2.60. folgt

$$E(X) = \int_0^\infty (1 - F_X(x))\, dx \leq \int_0^\infty (1 - F_Y(x))\, dx = E(Y).$$

(iii) gilt wegen

$$\begin{aligned} V(X) &= E((X - E(X))^2) = E(X^2 - 2X \cdot E(X) + E(X)^2) \\ &= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2 \end{aligned}$$

und (iv) sieht man so ein: Wegen (i) gilt $E(aX + b) = aE(X) + b$, und daher

$$\begin{aligned} V(aX + b) &= E(((aX + b) - E(aX + b))^2) = E((aX + b - aE(X) - b)^2) \\ &= E(a^2(X - E(X))^2) = a^2 E((X - E(X))^2) = a^2 V(X). \end{aligned}$$ ∎

Bemerkung 2.62. Ist X eine Zufallsvariable, für die $E(X)$ und $V(X)$ existieren mit $V(X) \neq 0$, so gilt für die Zufallsvariable $Y := (X - E(X))/\sqrt{V(X)}$ wegen (i) und (iv) $E(Y) = 0$, $V(Y) = 1$. Y heißt die **Standardisierung von** X. Für den Spezialfall einer normalverteilten Zufallsvariable haben wir die Standardisierung schon in Beispiel 2.30. betrachtet. □

Die Aussage (iv) von Satz 2.61. besagt, dass die Varianz bei Translation einer Zufallsvariablen unverändert bleibt. Es gibt Situationen, in denen die Größe der Schwankungen relativ zu den angenommenen Werten wichtig ist.

Definition 2.63. *Es sei X eine nichtnegative Zufallsvariable mit endlichem Erwartungswert $E(X) \neq 0$. Dann heißt*

$$C_X := \frac{\sqrt{V(X)}}{E(X)}$$

Variationskoeffizient *von X. C_X^2 wird als* **quadrierter Variationskoeffizient** *bezeichnet.*

Beispiel 2.64. Ist X exponentialverteilt mit dem Parameter α, so gilt

$$E(X) = \frac{1}{\alpha} \quad \text{und} \quad V(X) = \frac{1}{\alpha^2},$$

und daher sind der Variationskoeffizient und der quadrierte Variationskoeffizient beide gleich 1. □

Neben dem Erwartungswert und der Varianz betrachtet man weitere Parameter von Zufallsvariablen.

Definition 2.65. *Ist X eine Zufallsvariable, so ist auch X^k für alle $k \in \mathbb{N}$ eine Zufallsvariable. Ihr Erwartungswert*

$$m_k := E(X^k) = \begin{cases} \sum\limits_{i=1}^{\infty} t_i^k P(X = t_i) & , \text{ falls } X \text{ diskret} \\ \int\limits_{-\infty}^{\infty} s^k f(s)\, ds & , \text{ falls } X \text{ stetig,} \end{cases}$$

heißt k**-tes Moment von** X.

Das k-te Moment der Zufallsvariable $X - m_1$,

$$\mu_k := E((X - m_1)^k),$$

heißt das k**-te zentrale Moment** *von X.*

Bemerkung 2.66. Der Erwartungswert $E(X)$ ist das erste Moment von X. Die Varianz $V(X)$ ist das zweite zentrale Moment von X. □

Bemerkung 2.67. Die zentralen Momente existieren genau dann, wenn die gewöhnlichen Momente existieren, und sie lassen sich durch die gewöhnlichen Momente ausdrücken (und umgekehrt). Es gilt z.B. (vgl. Satz 2.61., (iii)) $\mu_2 = m_2 - m_1^2$. □

Bemerkung 2.68. Es kann gezeigt werden (vgl. **Feller** [12]), dass unter sehr allgemeinen Bedingungen die folgende Aussage gilt: Existieren alle Momente einer Zufallsvariable X, so ist durch sie die Verteilung von X eindeutig festgelegt. Besitzen also X und Y dieselbe Folge von Momenten, so gilt für ihre Verteilungsfunktionen $F_X = F_Y$. □

Das folgende Beispiel zeigt, dass eine Zufallsvariable keine Momente zu besitzen braucht.

Beispiel 2.69. Sei X eine stetige Zufallsvariable mit der Dichte

$$f(s) = \begin{cases} 0 & \text{für } s < 1 \\ s^{-2} & \text{für } s \geq 1. \end{cases}$$

Dann gilt

$$\int_{-\infty}^{\infty} f(s)\,ds = \int_{1}^{\infty} s^{-2}\,ds = -\left[\frac{1}{s}\right]_1^{\infty} = 1,$$

also ist f eine Dichte, aber das k-te Moment

$$\int_{-\infty}^{\infty} s^k f(s)\,ds = \int_{1}^{\infty} s^{k-2}\,ds, \quad k \geq 1,$$

existiert nicht. □

Neben dem Erwartungswert wird als Charakterisierung des "Zentrums" einer Zufallsvariable auch der **Median** verwendet. Das ist jede Zahl $m \in \mathbb{R}$ mit

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

Der Median ist i.a. nicht eindeutig bestimmt. Ist z.B. X $b(1, 0.5)$-verteilt, so ist jedes $m \in (0, 1)$ Median von X.

Ist X eine symmetrisch verteilte Zufallsvariable, d.h.

$$P(X < m_1 - t) = P(X > m_1 + t) \quad \text{für alle } t,$$

so gilt insbesondere

$$P(X < m_1) = P(X > m_1) \leq \frac{1}{2},$$

und $m_1 = E(X)$ ist Median von X. Es kann aber, wie wir eben gesehen haben, noch weitere Mediane geben.

Kapitel 3

Zufallsvektoren

3.1 Gemeinsame Verteilung von Zufallsvariablen

Bei vielen Experimenten ist man nicht nur an einer, sondern an mehreren Größen gleichzeitig interessiert (z.B. Größe, Gewicht, Alter und Geschlecht einer zufällig ausgewählten Person). Man hat also mehrere Zufallsvariablen auf demselben Stichprobenraum zu betrachten, die wir zu einem Vektor zusammenfassen.

Definition 3.1. *Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ heißt* **Zufallsvektor***, falls $\boldsymbol{X}^{-1}(B) \in \mathcal{A}$ für alle Borelschen Mengen $B \in \mathcal{B}$ gilt, wobei (wie in $\mathbb{R}$) $\mathcal{B}$ die kleinste σ-Algebra ist, die alle n-dimensionalen Intervalle $(\boldsymbol{a}, \boldsymbol{b}] := \{\boldsymbol{x} \in \mathbb{R}^n : a_i < x_i \leq b_i,\ i = 1, \ldots, n\}$ enthält.*

Ist $\boldsymbol{X}$ ein Zufallsvektor, X_i die i-te Komponente von $\boldsymbol{X}$, $i \in \{1, \ldots, n\}$, und $\tilde{B} \subset \mathbb{R}$ eine (eindimensionale) Borelsche Menge, so ist $B := \mathbb{R}^{i-1} \times \tilde{B} \times \mathbb{R}^{n-i}$ eine Borelsche Menge in $\mathbb{R}^n$, und daher gilt

$$\{\omega : X_i(\omega) \in \tilde{B}\} = \{\omega : \boldsymbol{X}(\omega) \in B\} \in \mathcal{A},$$

d.h. jede Komponente ist eine Zufallsvariable.

Definition 3.2. *Es sei $\boldsymbol{X}$ ein Zufallsvektor auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann wird durch*

$$P_X(B) := P(\{\omega \in \Omega : \boldsymbol{X}(\omega) \in B\})$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{B}$ definiert. P_X heißt die **gemeinsame Verteilung** *der Zufallsvariablen $X_1, \ldots, X_n$.*

In der gemeinsamen Verteilung P_X steckt die gesamte Information über das gemeinsame zufällige Verhalten von $X_1, \ldots, X_n$.

Ist P_X bekannt, so kennt man jede der **Randverteilungen** (auch **marginalen Verteilungen**)

$$\begin{aligned} P_{X_i}(\tilde{B}) &= P(\{\omega : X_i(\omega) \in \tilde{B}\}) \\ &= P(\{\omega : X_i(\omega) \in \tilde{B},\ X_j(\omega)\ \in \mathbb{R},\ j \neq i\}) \\ &= P_X(\mathbb{R}^{i-1} \times \tilde{B} \times \mathbb{R}^{n-i}) \end{aligned}$$

Umgekehrt ist jedoch die gemeinsame Verteilung nicht durch die Randverteilungen P_{X_i}, $i = 1, \ldots, n$, festgelegt.

Die gemeinsame Verteilung P_X ist (wie im eindimensionalen Fall) durch die **gemeinsame Verteilungsfunktion**

$$F_X : \mathbb{R}^n \to \mathbb{R}, \quad F_X(t) := P\left(\bigcap_{j=1}^{n} [X_j \leq t_j]\right),$$

vollständig bestimmt.

Die folgenden Eigenschaften der gemeinsamen Verteilungsfunktion zeigt man ähnlich wie im eindimensionalen Fall.

Lemma 3.3. *Ist F_X die gemeinsame Verteillungsfunktion von $X_1, \ldots, X_n$, so gilt*

(i) $0 \leq F_X(\boldsymbol{t}) \leq 1$ *für alle* $\boldsymbol{t} \in \mathbb{R}^n$

(ii) $F_X(\boldsymbol{t}) \leq F_X(\boldsymbol{s})$, *falls* $t_j \leq s_j$, $j = 1, \ldots, n$ *(Monotonie)*

(iii) $\lim_{j\to\infty} F_X(\boldsymbol{t}^{(j)}) = 1$ *für jede Folge* $\{\boldsymbol{t}^{(j)}\} \subset \mathbb{R}^n$ *mit* $t_i^{(j)} \leq t_i^{(j+1)}$ *und* $\lim_{j\to\infty} t_i^{(j)} = \infty$, $i = 1, \ldots, n$.

(iv) $\lim_{j\to\infty} F_X(\boldsymbol{t}^{(j)}) = 0$ *für jede Folge* $\{\boldsymbol{t}^{(j)}\} \subset \mathbb{R}^n$ *mit* $t_i^{(j)} \geq t_i^{(j+1)}$, $i = 1, \ldots, n$, *und* $\lim_{j\to\infty} t_k^{(j)} = -\infty$ *für ein* $k \in \{1, \ldots, n\}$.

(v) F_X *ist rechtsseitig stetig, d.h.* $\lim_{t_j \to \hat{t}_j + 0} F_X(\boldsymbol{t}) = F_X(\hat{\boldsymbol{t}})$

Die angegebenen Bedingungen (i) - (v) genügen noch nicht, um eine gemeinsame Verteilungsfunktion zu charakterisieren.

Beispiel 3.4.

$$F(t_1, t_2) = \begin{cases} 1 - \exp(-t_1 - t_2) & , \quad 0 \leq t_j < \infty, \; j = 1, 2 \\ 0 & , \quad \text{sonst} \end{cases}$$

erfüllt die Bedingungen (i) - (v).

Wir nehmen an, dass F gemeinsame Verteilungsfunktion von (X_1, X_2) ist.

Es sei

$$\begin{aligned} A &:= [X_1 \leq 1, X_2 \leq 2], \\ B &:= [X_1 \leq 2, X_2 \leq 1], \\ C &:= [X_1 \leq 2, X_2 \leq 2]. \end{aligned}$$

Dann gilt für

$$D := C \setminus (A \cup B)$$

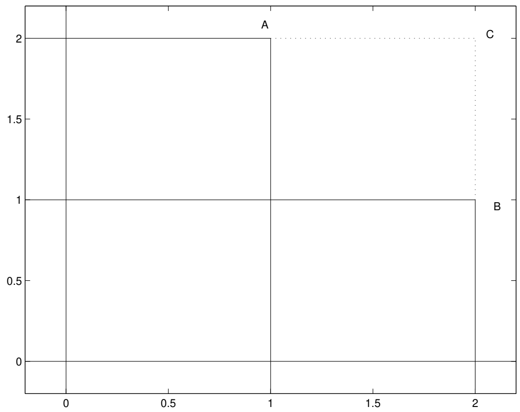

Abbildung 3.1

$$\begin{aligned} P_X(D) &= P_X(C) - P_X(A \cup B) = F(2,2) - (F(1,2) + F(2,1) - F(1,1)) \\ &= 1 - e^{-4} - (1 - e^{-3} + 1 - e^{-3} - (1 - e^{-2})) \\ &= -e^{-2} + 2e^{-3} - e^{-4} \; = \; -(e^{-1} - e^{-2})^2 < 0. \qquad \square \end{aligned}$$

Fordert man zusätzlich zu (i) - (v) für die Funktion F_X, dass gilt

$$\text{(vi)} \quad \sum_{(k_1,\ldots,k_n) \in \{0,1\}^n} (-1)^{n-k_1-\ldots-k_n} \, F_X(b_1^{k_1} a_1^{1-k_1}, \ldots, b_n^{k_n} a_n^{1-k_n}) \geq 0$$
$$\text{für alle } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n \text{ mit } a_i \leq b_i, i = 1, \ldots, n,$$

so kann man zeigen, dass es zu F_X ein eindeutiges Wahrscheinlichkeismaß P_X auf $\mathcal{B}$ gibt mit

$$P_X((\boldsymbol{a}, \boldsymbol{b}]) = F_X(\boldsymbol{b}) - F_X(\boldsymbol{a}) \text{ für } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n \text{ mit } \boldsymbol{a} \leq \boldsymbol{b}.$$

Die Bedingung (vi) besagt gerade, dass die Intervalle $(\boldsymbol{a}, \boldsymbol{b}]$ nichtnegative Wahrscheinlichkeit besitzen.

Für $n = 2$ lautet (vi)

$$F_X(b_1, b_2) - F_X(b_1, a_2) - F_X(a_1, b_2) + F_X(a_1, a_2) \geq 0$$

$$\text{für alle } a_1 \leq b_1,\ a_2 \leq b_2.$$

Für die Verteilungsfunktionen der Komponenten X_i (die marginalen Verteilungsfunktionen von $\boldsymbol{X}$) gilt natürlich

$$\begin{aligned} F_{X_i}(t_i) &= P([X_i \leq t_i]) = P(\{\omega : X_i(\omega) \leq t_i\}) \\ &= \lim_{t_j \to \infty, j \neq i} F_X(t_1, \ldots, t_{i-1}, t_i, t_{i+1}, \ldots, t_n). \end{aligned}$$

Definition 3.5. *Sind die Komponenten X_i des Zufallsvektors $\boldsymbol{X}$ diskrete Zufallsvariable, so heißt $\boldsymbol{X}$* **diskreter Zufallsvektor** *.*

In diesem Fall kann $\boldsymbol{X}$ nur abzählbar unendlich viele Werte $\boldsymbol{t}^{(j)} \in \mathbb{R}^n$ annehmen und F_X ist eine stückweise konstante Funktion.

$$f_X(\boldsymbol{t}) := P([\boldsymbol{X} = \boldsymbol{t}])$$

heißt dann die **gemeinsame Verteilungsdichte** *oder* **gemeinsame Zähldichte** *von $X_1, \ldots, X_n$.*

Wie im eindimensionalen Fall ist durch sie die gemeinsame Verteilung P_X schon vollständig bestimmt:

$$P_X(A) = P([\boldsymbol{X} \in A]) = \sum_{\boldsymbol{t} \in A} f_X(\boldsymbol{t})$$

wobei natürlich wieder nur über die höchstens abzählbar vielen $\boldsymbol{t} \in A$ summiert wird, die im Wertebereich von $\boldsymbol{X}$ liegen.

Zwischen der gemeinsamen Verteilungsdichte von $\boldsymbol{X}$ und den marginalen Verteilungsdichten der X_i besteht natürlich der folgende Zusammenhang:

$$f_{X_i}(t_i) = \sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{i-1}=-\infty}^{\infty} \sum_{t_{i+1}=-\infty}^{\infty} \cdots \sum_{t_n=-\infty}^{\infty} f_X(t_1, \ldots, t_{i-1}, t_i, t_{i+1}, \ldots, t_n),$$

wobei die Summe $\sum_{t_j=-\infty}^{\infty}$ über den Wertebereich von X_j erstreckt wird.

Bemerkung 3.6. Mit der letzten Formel sieht man (jedenfalls für diskrete Zufallsvariable) die Additivität des Erwartungswerts leicht ein. Es sei f_X die gemeinsame Zähldichte der diskreten Zufallsvariablen $X_1, \ldots, X_n$. Dann gilt

$$\begin{aligned} E(X_1 + \ldots X_n) &= \sum_{t_1} \cdots \sum_{t_n} (t_1 + \ldots + t_n)\, f_X(t_1, \ldots, t_n) \\ &= \sum_{i=1}^{n} \Big(\sum_{t_i} t_i \sum_{t_1} \cdots \sum_{t_{i-1}} \sum_{t_{i+1}} \cdot \sum_{t_n} F_X(t_1, \ldots, t_n) \\ &= \sum_{i=1}^{n} \sum_{t_i} t_i\, f_{X_i}(t_i) = \sum_{i=1}^{n} E(X_i). \end{aligned}$$

Für stetige Zufallsvariable schließt man ähnlich mit der Formel, mit der man die Randverteilungen aus der gemeinsamen Verteilung bestimmt. □

Beispiel 3.7. Wir betrachten eine Folge von N Zufallsexperimenten, von denen jedes n mögliche Ergebnisse habe, die mit den Wahrscheinlichkeiten $p_1, \ldots, p_n$ mit $\sum_{j=1}^n p_j = 1$ eintreten. Es sei X_j die Anzahl der Experimente, bei denen das j-te Ergebnis eintritt. Dann ist die gemeinsame Verteilungsdichte von $\boldsymbol{X} = (X_1, \ldots, X_n)^T$

$$f_X(k_1, \ldots, k_n) = \frac{N!}{k_1! \ldots k_n!} p_1^{k_1} \ldots p_n^{k_n},$$

wobei $k_1, \ldots, k_n \geq 0$ und $\sum_{j=1}^n k_j = N$ gilt.

Für die marginalen Verteilungsdichten der X_i gilt

$$\begin{aligned} f_{X_i}(k_i) &= \sum{}' \frac{N!}{k_1! \ldots k_n!} p_1^{k_1} \ldots p_n^{k_n} \\ &= \frac{N!}{(N-k_i)!k_i!} p_i^{k_i} \sum{}' \frac{(N-k_i)!k_i!}{k_1! \ldots k_n!} p_1^{k_1} \ldots p_{i-1}^{k_{i-1}} p_{i+1}^{k_{i+1}} \ldots p_n^{k_n} \\ &= \binom{N}{k_i} p_i^{k_i} (p_1 + \ldots + p_{i-1} + p_{i+1} + \ldots + p_n)^{N-k_i} \\ &= \binom{N}{k_i} p_i^{k_i} (1-p_i)^{N-k_i}, \end{aligned}$$

wobei die Summe $\sum{}'$ über alle $k = (k_1, \ldots, k_n)$ erstreckt wird, für die $\sum\limits_{j \neq i} k_j = N - k_i$ gilt.

Die X_i sind also wie zu erwarten $b(N, p_i)$ verteilt. □

Definition 3.8. *Ist die gemeinsame Verteilungsfunktion F_X des Zufallsvektors $\boldsymbol{X}$ stetig, so heißt $\boldsymbol{X}$* **stetiger Zufallsvektor** *.*

Wir betrachten wieder nur den Spezialfall, dass eine integrierbare Funktion $f_X : \mathbb{R}^n \to \mathbb{R}$ existiert mit

$$F_X(\boldsymbol{t}) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_n} f_X(s_1, \ldots, s_n)\, ds_n \ldots ds_1. \tag{3.1}$$

Dann heißt f_X die **gemeinsame Dichte** *von X.*

Die gemeinsame Dichte f_X besitzt die folgenden Eigenschaften:

(i) $f_X(\boldsymbol{s}) \geq 0$ für alle $\boldsymbol{s} \in \mathbb{R}^n$

(ii) $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_X(s_1, \ldots, s_n)\, ds_1 \ldots ds_n = 1.$

Ist umgekehrt f_X eine integrierbare Funktion mit den Eigenschaften (i) und (ii), so wird durch (3.1) eine stetige Verteilungsfunktion F_X auf $\mathbb{R}^n$ definiert.

Die gemeinsame Dichte f_X bestimmt die gemeinsame Verteilung des Zufallsvektors $\boldsymbol{X}$ vollständig, d.h. für jede (Borelsche) Menge $A \subset \mathbb{R}^n$ gilt

$$P_X(A) = P([\boldsymbol{X} \in A]) = \int\limits_A f_X(\boldsymbol{s})\, ds_1 \dots ds_n.$$

Ist die gemeinsame Dichte f_X in einem Punkt $\boldsymbol{t} \in \mathbb{R}^n$ stetig, so sieht man wie im Falle $n = 1$, dass für die zugehörige gemeinsame Verteilungsfunktion F_X die Ableitung $\frac{\partial^n}{\partial t_1 \dots \partial t_n} F_X(t)$ existiert und dass

$$\frac{\partial^n}{\partial t_1 \dots \partial t_n} F_X(\boldsymbol{t}) = f_X(\boldsymbol{t})$$

gilt.

Aus der Beziehung zwischen der gemeinsamen Verteilungsfunktion F_X und den marginalen Verteilungsfunktionen F_{X_i} folgt schließlich, dass für die Dichte f_{X_i} der i-ten Komponente X_i von $\boldsymbol{X}$ gilt

$$f_{X_i}(s_i) = \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} f_X(s_1, \dots, s_{i-1}, s_i, s_{i+1}, \dots, s_n) ds_1, \dots ds_{i-1} ds_{i+1} \dots ds_n.$$

Entsprechend Bemerkung 3.6. erhält man hieraus für stetige Zufallsvariable die Additivität des Erwartungswerts.

Beispiel 3.9. Um die Zuverlässigkeit eines Gerätes zu erhöhen, wird bei besonders störanfälligen Komponenten zusätzlich eine "Stand-by Komponente" eingebaut, die bei Ausfall der Originalkomponente deren Aufgabe sofort übernimmt.

Es sei X_1 bzw. X_2 die Zeit bis zum Ausfall der Original bzw. Stand-by Komponente. Dann ist die gemeinsame Dichte von $\boldsymbol{X} = (X_1, X_2)$

$$f_X(s) = \begin{cases} \lambda\mu \exp(-\mu s_1 - \lambda(s_2 - s_1)) & \text{für} \quad 0 \le s_1 \le s_2 \\ 0 & \text{sonst} \end{cases}$$

mit $\lambda, \mu > 0$ und $\lambda \neq \mu$.

Für die marginalen Dichten gilt für $s_1 \geq 0$

$$\begin{aligned} f_{X_1}(s_1) &= \int\limits_{-\infty}^{\infty} f_X(s_1, s_2)\, ds_2 = \int\limits_{s_1}^{\infty} \lambda\mu \exp(-\mu s_1 - \lambda(s_2 - s_1))\, ds_2 \\ &= [-\mu \exp(-\mu s_1 - \lambda(s_2 - s_1))]_{s_1}^{\infty} = \mu \exp(-\mu s_1), \end{aligned}$$

und für $s_1 < 0$ ist $f_{X_1}(s_1) = 0$, d.h. X_1, die Ausfallzeit des Gerätes ohne Stand-by Komponente, ist exponentialverteilt.

Für $s_2 < 0$ gilt $f_{X_2}(s_2) = 0$ und für $s_2 \geq 0$

$$\begin{aligned} f_{X_2}(s_2) &= \int_0^{s_2} \lambda\mu \exp((\lambda-\mu)s_1 - \lambda s_2)\, ds_1 = \left[\frac{\lambda\mu}{\lambda-\mu}\exp((\lambda-\mu)s_1 - \lambda s_2)\right]_0^{s_2} \\ &= \frac{\lambda\mu}{\lambda-\mu}(\exp(-\mu s_2) - \exp(-\lambda s_2)). \end{aligned}$$

Als mittlere Laufzeit des Gerätes mit Stand-by Komponente erhält man daher

$$E(X_2) = \int_{-\infty}^{\infty} s_2 f_{X_2}(s_2)\, ds_2 = \int_0^{\infty} s_2 \frac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu s_2} - e^{-\lambda s_2}\right) ds_2 = \frac{\lambda+\mu}{\lambda\cdot\mu}.$$

Wegen $E(X_1) = \frac{1}{\mu}$ wird also die mittlere Laufzeit bei Benutzung einer Stand-by Komponente um den Faktor $1 + \frac{\mu}{\lambda}$ erhöht. □

Wir geben nun weitere wichtige mehrdimensionale Verteilungen an.

Definition 3.10. *Der stetige Zufallsvektor $\boldsymbol{X}$ heißt* **gleichverteilt** *auf der (Borelschen) Menge $B \subset \mathbb{R}^n$, wenn die gemeinsame Dichte gegeben ist durch*

$$f_X(\boldsymbol{s}) = \begin{cases} \frac{1}{\mu(B)} & \text{für} \quad \boldsymbol{s} \in B \\ 0 & \text{für} \quad \boldsymbol{s} \notin B, \end{cases}$$

wobei $\mu(B)$ das elementargeometrische Maß (=Länge für $n = 1$, Fläche für $n = 2$, Volumen für $n = 3$) bezeichnet.

Offenbar ist ein Zufallsvektor genau dann gleichverteilt, wenn für jede Teilmenge A von B die Wahrscheinlichkeit $P([\boldsymbol{X} \in A])$ dem Maß $\mu(A)$ proportional ist.

Definition 3.11. *Es seien $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$ und $\boldsymbol{\Sigma} := (\sigma_{ik}) \in \mathbb{R}^{(n,n)}$ eine symmetrische, positiv definite Matrix (d.h. $\sigma_{ik} = \sigma_{ki}$ für alle i, k und $\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} > 0$ für alle $\boldsymbol{t} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$). Ein Zufallsvektor $\boldsymbol{X}$ heißt bf normalverteilt mit den Parametern $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$, kurz $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-verteilt, wenn $\boldsymbol{X}$ die gemeinsame Dichte*

$$f_X(\boldsymbol{s}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{s}-\boldsymbol{\mu})\right\}.$$

besitzt.

Bemerkung 3.12. Da eine positiv definite Matrix nur positive Eigenwerte hat und $\det\boldsymbol{\Sigma}$ gleich dem Produkt der Eigenwerte von $\boldsymbol{\Sigma}$ ist, ist f_X definiert. □

Bemerkung 3.13. Mit der Variablentransformation $\boldsymbol{t} := \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{s}-\boldsymbol{\mu})$ gilt

$$\int\limits_{\mathbb{R}^n} f_X(\boldsymbol{s})\, d\boldsymbol{s} = \frac{1}{(2\pi)^{n/2}} \int\limits_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}\sum_{j=1}^n t_j^2\right\} dt = 1.$$

Wegen $f_X(\boldsymbol{s}) \geq 0$ für alle $\boldsymbol{s} \in \mathbb{R}^n$ ist f_X also eine Dichte. □

Satz 3.14. *Besitzt der Zufallsvektor $\boldsymbol{X}$ ein $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$-Verteilung, so ist die i-te Komponente X_i $N(\mu_i,\sigma_{ii})$-verteilt.*

Beweis: Wir zeigen die Aussage nur für den Fall $n = 2$. Es sei $\boldsymbol{\Sigma}^{-1} =: \boldsymbol{A} = (a_{ij})$. Dann gilt (mit den Variablentransformationen $t_i := s_i - \mu_i$, $i = 1,2$ und $r := \sqrt{a_{22}}(t_2 + a_{12}t_1/a_{22})$)

$$\begin{aligned}
\int\limits_{-\infty}^{\infty} f_X(s_1,s_2)\, ds_2 &= \frac{1}{2\pi\sqrt{\det\boldsymbol{\Sigma}}}\int\limits_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}a_{11}t_1^2 - a_{12}t_1t_2 - \frac{1}{2}a_{22}t_2^2\right\} dt_2 \\
&= \frac{1}{2\pi\sqrt{\det\boldsymbol{\Sigma}}}\exp\left\{-\frac{1}{2}(a_{11}-\frac{a_{12}^2}{a_{22}})t_1^2\right\}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}a_{22}\left(t_2+\frac{a_{12}}{a_{22}}t_1\right)^2\right\} dt_2 \\
&= \frac{1}{2\pi\sqrt{\det\boldsymbol{\Sigma}}\sqrt{a_{22}}}\exp\left\{-\frac{1}{2}\sigma_{11}^{-1}t_1^2\right\}\int\limits_{-\infty}^{\infty}\exp\left(-\frac{1}{2}r^2\right) dr \\
&= \frac{1}{\sqrt{2\pi\sigma_{11}}}\exp\left\{-\frac{(s_1-\mu_1)^2}{2\sigma_{11}}\right\}. \qquad \blacksquare
\end{aligned}$$

Bemerkung 3.15. Die positiv definite Matrix $\boldsymbol{\Sigma}$ kann man schreiben als

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \dots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \dots & \sigma_2\sigma_n\rho_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_1\sigma_n\rho_{1n} & \sigma_2\sigma_n\rho_{2n} & \dots & \sigma_n^2 \end{pmatrix}$$

mit $\sigma_i^2 = \sigma_{ii}$ und $|\rho_{ij}| < 1$ für alle $i,j \in \{1,\dots,n\}$.

Wegen Satz 3.14. gilt $\sigma_i^2 = V(X_i)$. Auf die Bedeutung der ρ_{ij} gehen wir später ein. □

Bemerkung 3.16. Der Graph der Dichte f_X (vgl. Abbildung 3.2) hat (im Falle $n = 2$) die folgenden Eigenschaften.

(i) Das (eindeutige) Maximum liegt in (μ_1,μ_2).

(ii) Schnitte parallel zur (s_1,s_2)-Ebene sind Ellipsen mit dem gemeinsamen Mittelpunkt (μ_1,μ_2).

(iii) Schnitte parallel zur (s_1,f)-Ebene und zur (s_2,f)-Ebene sind nichtnormierte "Glockenkurven". □

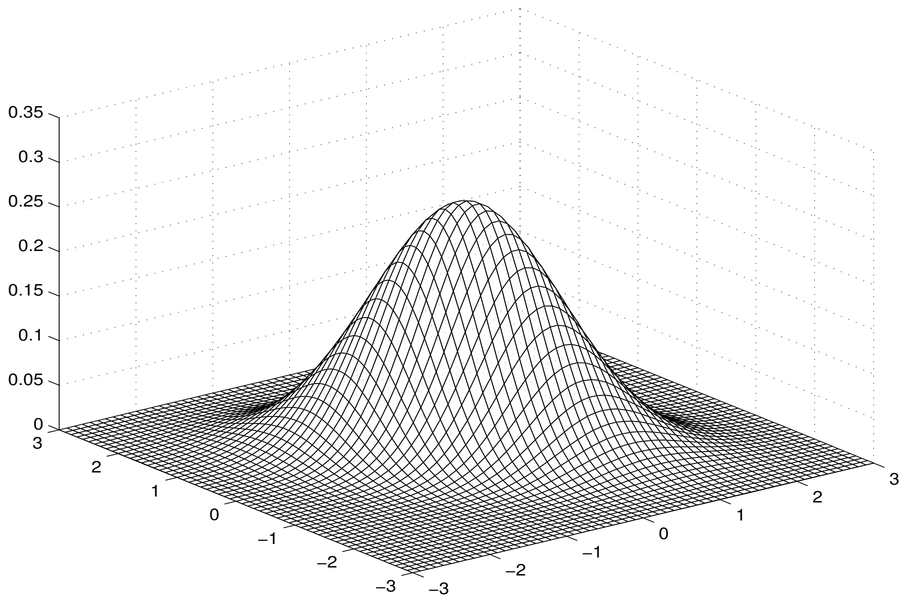

Abbildung 3.2: Dichte der zweidimensionalen Normalverteilung

3.2 Unabhängigkeit von Zufallsvariablen

In Kapitel 1 haben wir Ereignisse $A_1, \ldots, A_n$ unabhängig genannt, wenn für jede Teilmenge $\{i_1, \ldots, i_k\}$ der Indexmenge $\{1, \ldots, n\}$

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

gilt. Wir übertragen diese Definition auf Zufallsvariable.

Definition 3.17. *Es seien $X_1, \ldots, X_n$ Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. $X_1, \ldots, X_n$ heißen* **unabhängig** *, wenn für alle Borelschen Mengen $A_j \subset \mathbb{R}$, $j = 1, \ldots, n$, gilt*

$$P([X_j \in A_j,\, j = 1, \ldots, n]) = \prod_{j=1}^{n} P([X_j \in A_j]). \tag{3.2}$$

Bemerkung 3.18. Anders als in der Definition der Unabhängigkeit von Ereignissen müssen wir hier nicht beliebige Durchschnitte von $k \leq n$ Ereignissen betrachten, denn sind $i_1, \ldots, i_k \in \{1, \ldots, n\}$ verschiedene Indizes und sind A_{i_ν} für $\nu = 1, \ldots, k$ Borelsche Mengen, so gilt mit $A_{i_\nu} = \mathbb{R}$ für $i_\nu \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$

$$P([X_j \in A_j,\, j = 1, \ldots, n]) = P([X_{i_\nu} \in A_{i_\nu},\, \nu = 1, \ldots, k]),$$

und daher

$$P([X_{i_\nu} \in A_{i_\nu},\, \nu = 1, \ldots, k]) = \prod_{j=1}^{n} P([X_j \in A_j]) = \prod_{\nu=1}^{k} P([X_{i_\nu} \in A_{i_\nu}]) \quad \square$$

.

Bemerkung 3.19. Man beachte, dass aus der Unabhängigkeit von $X_1, \dots, X_n$ zwar folgt, dass je zwei Zufallsvariable X_i und X_j unabhängig sind, dass aber umgekehrt nicht die paarweise Unabhängigkeit von $X_1, \dots, X_n$ bereits die Unabhängigkeit von $X_1, \dots, X_n$ impliziert (vgl. auch Kapitel 1). □

Man kann die Unabhängigkeit von Zufallsvariablen mit Hilfe von Verteilungsfunktionen und von Dichten charakterisieren.

Satz 3.20. *Es sei F_X die gemeinsame Verteilungsfunktion des Zufallsvektors $\boldsymbol{X} := (X_1, \dots, X_n)$, und es seien F_{X_j}, $j = 1, \dots, n$, die marginalen Verteilungsfunktionen der Komponenten X_j.*

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

$$F_X(t_1, \dots, t_n) = \prod_{j=1}^{n} F_{X_j}(t_j) \quad \textit{für alle } \boldsymbol{t} \in \mathbb{R}^n \tag{3.3}$$

gilt

Beweis: Aus der Unabhängigkeit der X_j folgt für alle $\boldsymbol{t} \in \mathbb{R}^n$

$$F_X(\boldsymbol{t}) = P([X_j \le t_j,\ j = 1, \dots, n]) = \prod_{j=1}^{n} P([X_j \le t_j]) = \prod_{j=1}^{n} F_{X_j}(t_j).$$

Diese Gleichung zeigt ebenfalls, dass aus der Gültigkeit von (3.3)

$$P([X_j \in A_j,\, j = 1, \dots, n]) = \prod_{j=1}^{n} P([X_j \in A_j])$$

für alle Intervalle $A_j = (-\infty, t_j]$, $t_j \in \mathbb{R}$, folgt. Da diese Intervalle die σ-Algebra der Borelschen Mengen erzeugen, erhält man (mit einem hier nicht bekannten Satz über die Fortsetzung von Maßen), dass (3.3) auch hinreichend für die Unabhängigkeit der X_j ist. ∎

Satz 3.21. *$X_1, \dots, X_n$ seien diskrete Zufallsvariable mit der gemeinsamen Zähldichte f_X und den marginalen Zähldichten f_{X_j}, $j = 1, \dots, n$.*

$X_1, \dots, X_n$ sind genau dann unabhängig, wenn

$$f_X(\boldsymbol{t}) = \prod_{j=1}^{n} f_{X_j}(t_j) \quad \textit{für alle } \boldsymbol{t} \in \mathbb{R}^n \tag{3.4}$$

gilt.

Beweis: Sind $X_1, \ldots, X_n$ unabhängig, so folgt aus (3.2) für alle $\boldsymbol{t} \in \mathbb{R}^n$

$$f_X(\boldsymbol{t}) = P([X_j = t_j,\, j = 1, \ldots, n]) = \prod_{j=1}^{n} P([X_j = t_j]) = \prod_{j=1}^{n} f_{X_j}(t_j),$$

und umgekehrt erhält man aus (3.4) wegen

$$P([X_j \in A_j]) = \sum_{t_j \in A_j} f_{X_j}(t_j), \quad j = 1, \ldots, n,$$

auch (3.2). ∎

Satz 3.22. *Es seien $X_1, \ldots, X_n$ stetige Zufallsvariable mit der gemeinsamen Verteilungsdichte f_X und den marginalen Dichten f_{X_j}, $j = 1, \ldots, n$.*

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

$$f_X(\boldsymbol{t}) = \prod_{j=1}^{n} f_{X_j}(t_j) \quad \textit{für alle } \boldsymbol{t} \in \mathbb{R}^n \tag{3.5}$$

gilt.

Beweis: Wir zeigen die Aussage nur für Punkte $\boldsymbol{t} \in \mathbb{R}^n$, in denen f_X stetig ist. In diesen gilt, wenn $X_1, \ldots, X_n$ unabhängig sind, wegen Satz 3.20.

$$\begin{aligned} f_X(\boldsymbol{t}) &= \frac{\partial^n}{\partial t_1 \ldots \partial t_n} F_X(\boldsymbol{t}) = \frac{\partial^n}{\partial t_1 \ldots \partial t_n} \left(\prod_{j=1}^{n} F_{X_j}(t_j) \right) \\ &= \prod_{j=1}^{n} \frac{\partial}{\partial t_j} F_{X_j}(t_j) = \prod_{j=1}^{n} f_{X_j}(t_j), \end{aligned}$$

und aus (3.5) folgt umgekehrt

$$F_X(\boldsymbol{t}) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_n} f_X(\boldsymbol{s})\, ds_1 \ldots ds_n = \prod_{j=1}^{n} \int_{-\infty}^{t_j} f_{X_j}(s_j)\, ds_j = \prod_{j=1}^{n} F_{X_j}(t_j).$$ ∎

Beispiel 3.23. Ist der Zufallsvektor $\boldsymbol{X}$ normalverteilt und sind die marginalen Verteilungen $N(\mu_i, \sigma_{ii})$-verteilt, so gilt mit $\sigma := \prod_{i=1}^{n} \sigma_{ii}$

$$\prod_{i=1}^{n} f_{X_i}(s_i) = \frac{1}{(2\pi)^{n/2}\sqrt{\sigma}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(s_i - \mu_i)^2}{\sigma_{ii}} \right\}.$$

$X_1, \ldots, X_n$ sind also genau dann unabhängig, wenn alle Nichtdiagonalelemente von $\boldsymbol{\Sigma}$ Null sind. □

Beispiel 3.24. In Beispiel 3.9. ("Stand-by Komponente") sind X_1 und X_2 nicht unabhängig. Dies ist auch anschaulich klar. □

Beispiel 3.25. Der Zufallsvektor $X = (X_1, X_2)$ sei gleichverteilt auf $B \subset \mathbb{R}^2$, d.h.

$$f_X(\boldsymbol{s}) = \begin{cases} \dfrac{1}{\mu(B)} & , \text{ falls } \boldsymbol{s} \in B \\ 0 & , \text{ falls } \boldsymbol{s} \notin B. \end{cases}$$

Dann sind die marginalen Dichten

$$\begin{aligned} f_{X_1}(s_1) &= \int_{-\infty}^{\infty} f_X(s_1, s_2)\, ds_2 = \frac{1}{\mu(B)} \text{ "Länge von } \{s_2 : (s_1, s_2) \in B\}\text{"} \\ f_{X_2}(s_2) &= \frac{1}{\mu(B)} \text{ "Länge von } \{s_1 : (s_1, s_2) \in B\}\text{"}, \end{aligned}$$

und X_1, X_2 sind unabhängig, wenn X_1 und X_2 gleichverteilt sind. Dies ist der Fall, wenn B ein Rechteck ist.

Für $B := \{\boldsymbol{s} \in \mathbb{R}^2 : s_1^2 + s_2^2 \leq 1\}$ gilt

$$f_{X_1}(s_1) = \begin{cases} \frac{2}{\pi}\sqrt{1 - s_1^2} & , \quad |s_1| \leq 1 \\ 0 & \quad \text{sonst} \end{cases}$$

$$f_{X_2}(s_2) = \begin{cases} \frac{2}{\pi}\sqrt{1 - s_2^2} & , \quad |s_2| \leq 1 \\ 0 & \quad \text{sonst} \end{cases}$$

und X_1 und X_2 sind nicht unabhängig. □

Bemerkung 3.26. Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariable und $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ (stückweise) stetige Funktionen, so sind auch $g_1 \circ X_1, \ldots, g_n \circ X_n$ unabhängige Zufallsvariable, denn für beliebige Borelsche Mengen $B_1, \ldots, B_n \subset \mathbb{R}$ sind auch $g_1^{-1}(B_1), \ldots, g_n^{-1}(B_n)$ Borelsche Mengen, und aus der Unabhängigkeit von $X_1, \ldots, X_n$ folgt

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} [g_i \circ X_i \in B_i]\right) &= P\left(\bigcap_{i=1}^{n} [X_i \in g_i^{-1}(B_i)]\right) = \prod_{i=1}^{n} P([X_i \in g_i^{-1}(B_i)]) \\ &= \prod_{i=1}^{n} P([g_i \circ X_i \in B_i]). \end{aligned}$$

Sind also $X_1, \ldots, X_n$ unabhängig, so z.B. (mit $g_i(t) := t^2$) auch $X_1^2, \ldots, X_n^2$. □

Bemerkung 3.27. Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariable und ist die Funktion $g : \mathbb{R}^k \to \mathbb{R}$ stetig, so sind die Zufallsvariablen $g(X_1, \ldots, X_k), X_{k+1}, \ldots, X_n$ ebenfalls unabhängig. Um dies einzusehen, benötigt man, dass die Borelschen Mengen in $\mathbb{R}^k$ erzeugt werden von dem System

$$\{B_1 \times \ldots \times B_k \ : \ B_j \subset \mathbb{R} \text{ Borelsche Menge}\}.$$

Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariable, so sind also z.B. auch die $n-k+1$ Zufallsvariablen $X_1 + X_2 + \ldots + X_k, X_{k+1}, \ldots, X_n$ unabhängig. □

Beispiel 3.28. Es seien X_1 und X_2 unabhängige Zufallsvariable. Dann sind auch $Y_1 := \min(X_1, X_2)$ und $Y_2 := \max(X_1, X_2)$ Zufallsvariable. Diese besitzen die Verteilungsfunktionen

$$\begin{aligned} F_{Y_2}(t) &= P([Y_2 \leq t]) = P([X_1 \leq t] \cap [X_2 \leq t]) \\ &= P([X_1 \leq t])P([X_2 \leq t]) = F_{X_1}(t) \cdot F_{X_2}(t) \end{aligned}$$

und

$$\begin{aligned} F_{Y_1}(t) &= P([Y_1 \leq t]) = 1 - P([Y_1 > t]) \\ &= 1 - P([X_1 > t] \cap [X_2 > t]) \\ &= 1 - P([X_1 > t]) \cdot P([X_2 > t]) \\ &= 1 - (1 - P([X_1 \leq t]))(1 - P([X_2 \leq t])) \\ &= 1 - (1 - F_{X_1}(t))(1 - F_{X_2}(t)) \\ &= F_{X_1}(t) + F_{X_2}(t) - F_{X_1}(t)F_{X_2}(t). \qquad \square \end{aligned}$$

Wir betrachten nun allgemeiner den Fall von n unabhängigen Zufallsvariablen $X_1, \ldots, X_n$, setzen aber voraus, dass diese stetig und identisch verteilt sind (Jedes X_i besitze die Verteilungsfunktion F). Es seien $Y_1, \ldots, Y_n$ die Zufallsvariablen, die man durch Anordnung der X_i der Größe nach erhält, z.B.

$$Y_1 = \min\{X_1, \ldots, X_n\} \ , \ Y_n := \max\{X_1, \ldots, X_n\}.$$

Die Zufallsvariable Y_k heißt dann k**-te Ordnungsstatistik** .

Es sei z.B. X_i die Lebensdauer der i-ten Komponente eines Systems mit n Komponenten. Ist das System in Serie geschaltet, so ist Y_1 die Lebensdauer des gesamten Systems. Ist das System parallel geschaltet, so ist Y_n die Lebensdauer des Systems.

Ist das System funktionsfähig, wenn wenigstens m Komponenten intakt sind, so ist Y_{n-m+1} die Lebensdauer des Systems. ("m-aus-n System").

Wir bestimmen nun die Verteilungsfunktion von Y_k. Die Wahrscheinlichkeit, dass genau j der X_i einen Wert in der Menge $(-\infty, t]$ annehmen und $n-j$ einen Wert in (t,∞) (beachte: Da die X_i stetige Zufallsvariable sind, gilt mit Wahrscheinlichkeit 1 $Y_1 < Y_2 < \ldots < Y_n$!) ist

$$\binom{n}{j} F^j(t)(1-F(t))^{n-j}$$

(Binomialverteilung mit Parametern n und $p = F(t)$!).

Daher gilt

$$\begin{aligned} F_{Y_k}(t) &= P([Y_k \le t]) \\ &= P(\text{“Wenigstens } k \text{ der } X_i \text{ liegen in } (-\infty,t]\text{”}) \\ &= \sum_{j=k}^{n} \binom{n}{j} F^j(t)(1-F(t))^{n-j}. \end{aligned}$$

Insbesondere sind die Verteilungsfunktionen von Y_1 und Y_n gegeben durch

$$\begin{aligned} F_{Y_n}(t) &= (F(t))^n \\ F_{Y_1}(t) &= 1-(1-F(t))^n. \end{aligned}$$

Die Wahrscheinlichkeit $R(t)(t>0)$, dass die Lebensdauer eines Systems wenigstens t ist, heißt **Zuverlässigkeit** (reliability) des Systems. Bezeichnet X die Lebensdauer des Systems und F die Verteilungsfunktion, so gilt offenbar

$$R(t) = P(X > t) = 1 - F(t).$$

Die obigen Überlegungen zeigen, dass für die Zuverlässigkeit eines Systems mit n identisch verteilten, unabhängigen Komponenten

$$\begin{aligned} R_{\text{Serie}}(t) &= R_{Y_1}(t) = 1 - F_{Y_1}(t) = (1-F(t))^n = R(t)^n, \\ R_{\text{parallel}}(t) &= R_{Y_n}(t) = 1 - F_{Y_n}(t) = 1 - F(t)^n = 1-(1-R(t))^n, \end{aligned}$$

gilt, wobei $R(t)$ die Zuverlässigkeit der einzelnen Komponenten bezeichnet.

Beide Formeln können leicht auf den Fall verallgemeinert werden, dass die X_i nicht identisch verteilt sind. Ist $R_i(t)$ die Zuverlässigkeit von X_i, so gilt

$$\begin{aligned} R_{\text{Serie}}(t) &= \prod_{i=1}^{n} R_i(t) \\ R_{\text{parallel}}(t) &= 1 - \prod_{i=1}^{n}(1-R_i(t)). \end{aligned}$$

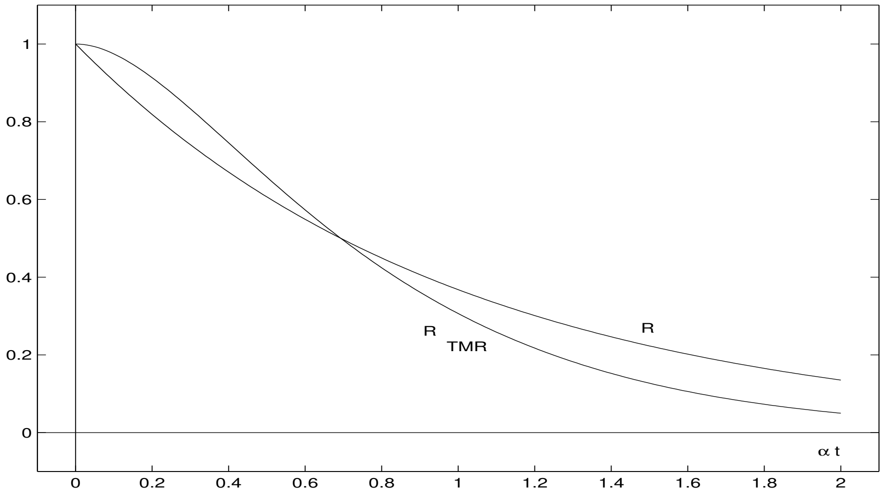

Abbildung 3.3: Zuverlässigkeit des TMR-Systems

Beispiel 3.29. Es seien X_i, $i = 1, \ldots, n$, exponentialverteilt mit Parameter α_i. Dann gilt bei Serienschaltung

$$R_{\text{Serie}}(t) = \exp\left(-\sum_{i=1}^{n} \alpha_i t\right), \quad t \geq 0,$$

d.h. die Lebensdauer des gesamten Systems ist ebenfalls exponentialverteilt mit dem Parameter $\sum\limits_{i=1}^{n} \alpha_i$. □

Beispiel 3.30. Wir betrachten ein TMR-System ("triple modular redundant"; 2-aus-3-System) mit identischen Komponenten. Dann ist Y_2 die Lebensdauer dieses Systems. Für die Zuverlässigkeit gilt (wenn R wieder die Zuverlässigkeit der einzelnen Komponenten bezeichnet)

$$\begin{aligned}
R_{TMR}(t) &= 1 - F_{Y_2}(t) = 1 - \sum_{j=2}^{3} \binom{3}{j} F^j(t)(1-F(t))^{3-j} \\
&= 1 - 3F^2(t)(1-F(t)) - F^3(t) = 1 - 3F^2(t) + 2F^3(t) \\
&= 3R^2(t) - 2R^3(t).
\end{aligned}$$

Wir vergleichen diese Funktion mit der Zuverlässigkeit der einzelnen Komponenten. Da $R_{TMR} = R$ genau für $R = 1$ und $R = 0.5$ gilt, besitzt das TMR-System für $0.5 < R(t) < 1$ eine höhere, für $0 < R(t) < 0.5$ eine niedrigere Zuverlässigkeit als die einzelnen Komponenten.

Abbildung 3.3 enthält die Zuverlässigkeiten des TMR-Systems und des einfachen Systems für exponentialverteiltes X. □

Beispiel 3.31. Wir betrachten ein System von zwei in Reihe geschalteten unabhängigen Komponenten, deren Lebensdauer X bzw. Y exponentialverteilt sei mit dem Parameter α_1 bzw. α_2. Wir fragen nach der Wahrscheinlichkeit, dass beim Ausfall des Systems die zweite Komponente die Ursache ist.

Wegen der Unabhängigkeit der Komponenten gilt

$$\begin{aligned} P(X \geq Y) &= \int_0^\infty \int_y^\infty f(x,y)\,dx\,dy \\ &= \int_0^\infty \int_y^\infty f_X(x) f_Y(y)\,dx\,dy \\ &= \int_0^\infty (1 - F_X(y)) f_Y(y)\,dy \\ &= \int_0^\infty e^{-\alpha_1 y} \alpha_2 e^{-\alpha_2 y}\,dt = \frac{\alpha_2}{\alpha_1 + \alpha_2}\,. \end{aligned}$$

Genauso erhält man natürlich für ein System von n unabhängigen, in Reihe geschalteten, $\exp(\alpha_j)$-verteilten Komponenten die Wahrscheinlichkeit, dass bei einem Ausfall des Systems die j-te Komponente der Grund für das Versagen ist, als

$$\alpha_j \Big/ \sum_{k=1}^n \alpha_k\,. \qquad \square$$

Beispiel 3.32. Wir verzichten auf die Unabhängigkeit der Komponenten. Wir betrachten ein System mit zwei in Reihe geschalteten Komponenten. Für den Ausfall kann es drei Ursachen geben: Beim Typ 1 wird die erste Komponente zerstört, beim Typ 2 die zweite und beim Typ 3 beide Komponenten. Die Wartezeit U_j bis zum Auftreten des Ereignisses von Typ j sei exponentialverteilt mit Parameter α_j, und es seien die U_j unabhängig.

Für die Lebensdauer X der ersten Komponente gilt dann $X = \min(U_1, U_3)$, und X ist exponentialverteilt mit dem Parameter $\alpha_1 + \alpha_3$ (vgl. Beispiel 3.29.). Entsprechend ist die Lebensdauer $Y = \min(U_2, U_3)$ exponentialverteilt mit dem Parameter $\alpha_2 + \alpha_3$. Für die marginalen Dichten gilt also

$$\begin{aligned} f_X(x) &= (\alpha_1 + \alpha_3) \exp(-(\alpha_1 + \alpha_3)x), \quad x > 0, \\ f_Y(y) &= (\alpha_2 + \alpha_3) \exp(-(\alpha_2 + \alpha_3)y), \quad y > 0. \end{aligned}$$

Wir bestimmen um die gemeinsame Verteilungsfunktion. Dazu berechnen wir zunächst (beachte: U_1, U_2, U_3 sind unabhängig)

$$R(x,y) = P(X > x, Y > y)$$

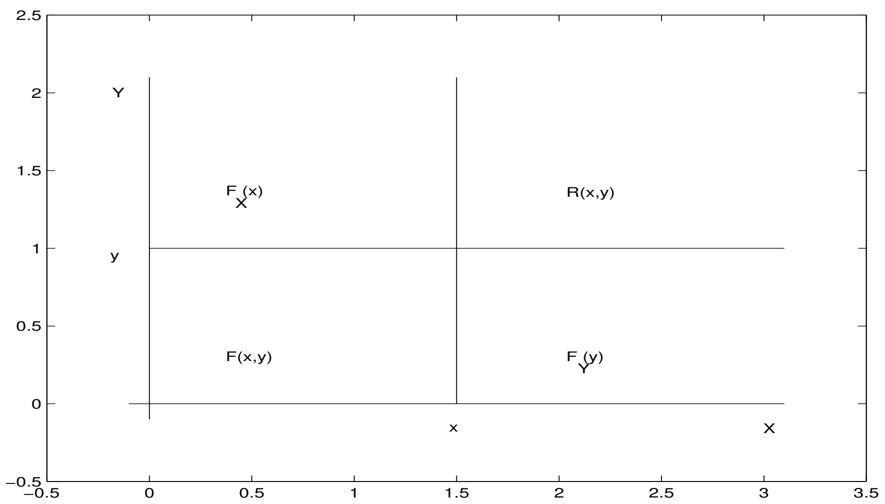

Abbildung 3.4: Zu Beispiel 3.32.

$$\begin{aligned}
&= P(\min(U_1, U_3) > x, \min(U_2, U_3) > y) \\
&= P(U_1 > x, U_2 > y, U_3 > \max(x, y)) \\
&= P(U_1 > x)P(U_2 > y)P(U_3 > \max(x, y)) \\
&= \exp(-\alpha_1 x - \alpha_2 y - \alpha_3 \max(x, y)), \quad x > 0,\ y > 0,
\end{aligned}$$

und hieraus erhält man (vgl. Abbildung 3.4)

$$\begin{aligned}
F(x, y) &= R(x, y) + F_X(x) + F_Y(y) - 1 \\
&= 1 + \exp(-\alpha_1 x - \alpha_2 y - \alpha_3 \max(x, y)) \\
&\quad - \exp(-(\alpha_1 + \alpha_3)x) - \exp(-(\alpha_2 + \alpha_3)y).
\end{aligned}$$

Insbesondere gilt $F(x, y) \neq F_X(x)F_Y(y)$, d.h. X und Y sind tatsächlich abhängig.

Für die gemeinsame Dichte erhält man

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \begin{cases} \alpha_1(\alpha_2 + \alpha_3) \exp(-\alpha_1 x - \alpha_2 y - \alpha_3 y) & , \quad x < y \\ \alpha_2(\alpha_1 + \alpha_3) \exp(-\alpha_1 x - \alpha_2 y - \alpha_3 x) & , \quad x > y. \end{cases} \qquad \square$$

Sind $X_1, \ldots, X_n$ Zufallsvariable mit der gemeinsamen (Zähl-)Dichte f und ist $\phi : \mathbb{R}^n \to \mathbb{R}$ stückweise stetig, so besitzt die Zufallsvariable $Y := \phi(X_1, \ldots, X_n)$ den Erwartungswert

$$E(Y) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \phi(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n,$$

falls die X_i stetig sind, und

$$E(Y) = \sum_{x_1} \ldots \sum_{x_n} \phi(x_1, \ldots, x_n) f(x_1, \ldots, x_n),$$

falls die X_i diskrete Zufallsvariable sind. Dabei ist vorausgesetzt, dass die Ausdrücke auf der rechten Seite existieren.

Beispiel 3.33. Wir betrachten einen Plattenspeicher mit beweglichem Schreib/Lese-Kopf, dessen innerer Zylinder den Radius a und dessen äußerer Zylinder den Radius b hat. Wir nehmen an, dass die Zahl der Zylinder sehr groß ist und dass die Zylinder sehr eng liegen, so dass wir ein Kontinuum von Zylindern annehmen können. Es seien X bzw. Y die gegenwärtige bzw. gewünschte Position des Kopfes. Wir nehmen an, dass X und Y unabhängig sind und dass beide gleichverteilt seien auf $[a, b]$.

Die Zufallsvariable $Z := |X - Y|$ beschreibt die Weglänge, die der Kopf bei einem Suchvorgang zurücklegt. Wir bestimmen die mittlere Suchlänge. Wegen der Unabhängigkeit von X und Y ist die gemeinsame Dichte von X und Y

$$f(x, y) = \frac{1}{(b-a)^2}, \quad a \le x \ , \ y \le b.$$

Daher folgt

$$\begin{aligned}
E(Z) &= \int_a^b \int_a^b |x-y| f(x,y)\, dx\, dy \\
&= \frac{1}{(b-a)^2} \left(\int_a^b \left(\int_a^x (x-y)\, dy + \int_x^b (y-x)\, dy \right) dx \right. \\
&= \frac{1}{(b-a)^2} \int_a^b \left(\left[-\frac{1}{2}(x-y)^2 \right]_a^x + \left[\frac{1}{2}(y-x)^2 \right]_x^b \right) dx \\
&= \frac{1}{(b-a)^2} \int_a^b \left(\frac{1}{2}(x-a)^2 + \frac{1}{2}(b-x)^2 \right) dx \\
&= \frac{1}{(b-a)^2} \left(\left[\frac{1}{6}(x-a)^3 \right]_a^b + \left[-\frac{1}{6}(b-x)^3 \right]_a^b \right) \\
&= \frac{b-a}{3}.
\end{aligned}$$

Die mittlere Suchlänge ist also $\frac{1}{3}$ (nicht die Hälfte!) der maximalen Suchlänge. In der Praxis ist die mittlere Suchlänge i.a. wesentlich kleiner, da aufeinanderfolgende Lese- oder Schreibvorgänge i.a. nicht unabhängig sind. □

3.3 Kovarianz von Zufallsvariablen

Wir haben bereits gesehen, dass der Erwartungswert der Summe von Zufallsvariablen gleich der Summe der Erwartungswerte ist. Für das Produkt gilt eine entsprechende Regel i.a. nicht (Hat X den Erwartungswert $E(X) = 0$, so gilt $E(X^2) = (E(X))^2$ nur im Falle $V(X) = 0$). Sind X und Y unabhängig, so gilt jedoch (im stetigen Fall)

$$\begin{aligned} E(X \cdot Y) &= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x \cdot y f(x,y)\, dy\, dx = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} xy f_X(x) f_Y(y)\, dy\, dx \\ &= \int\limits_{-\infty}^{\infty} x f_X(x)\, dx \int\limits_{-\infty}^{\infty} y f_Y(y)\, dy = E(X) \cdot E(Y) \end{aligned}$$

(und im diskreten Fall entsprechend).

Dieses Resultat kann man sofort auf Produkte von endlich vielen unabhängigen Zufallsvariablen verallgemeinern. Da mit $X_1, \ldots, X_n$ auch $g_1 \circ X_1, \ldots, g_n \circ X_n$ für stückweise stetige g_i unabhängig sind, hat man:

Satz 3.34. *Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariable und $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ stückweise stetig, so gilt*

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

Wir betrachten nun die Varianz der Summe zweier Zufallsvariablen X und Y. Es gilt

$$\begin{aligned} V(X+Y) &= E(((X - E(X)) + (Y - E(Y)))^2) \\ &= E((X - E(X))^2 + 2(X - E(X))(Y - E(Y)) + (Y - E(Y))^2) \\ &= V(X) + V(Y) + 2E((X - E(X))(Y - E(Y))), \end{aligned}$$

und weiter

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

Definition 3.35. *Es seien X und Y reelle Zufallsvariable. Dann heißt*

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

die **Kovarianz** *von X und Y.*

Die Kovarianz von X und Y misst in einem schwachen Sinne die Unabhängigkeit von X und Y, denn wegen Satz 3.34. gilt $\mathrm{Cov}(X,Y)=0$, falls X und Y unabhängig sind. Insbesondere hat man also

Satz 3.36. *Sind X und Y unabhängige Zufallsvariable, so gilt*

$$V(X+Y)=V(X)+V(Y).$$

Umgekehrt folgt aber aus $\mathrm{Cov}(X,Y)=0$ nicht, dass X und Y unabhängig sind. Ist nämlich X gleichverteilt auf $(-1,1)$ und $Y:=X^2$, so sind X und Y sicher nicht unabhängig, aber es gilt

$$E(X)=0 \quad \text{und} \quad E(X\cdot Y)=\frac{1}{2}\int\limits_{-1}^{1} x^3\,dx=0,$$

d.h.

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

In Verallgemeinerung der Cauchy Schwarzschen Ungleichung im $\mathbb{R}^n$ gilt (s. **Behnen & Neuhaus** [2, p. 223])

Satz 3.37. (Cauchy Schwarzsche Ungleichung) *Sind X und Y reelle Zufallsvariable, für die $E(X^2)$ und $E(Y^2)$ existiert, so existiert auch $E(X\cdot Y)$, und es gilt*

$$(E(X\cdot Y))^2\le E(X^2)\cdot E(Y^2).$$

Beispiel 3.38. Der Zufallsvektor (X,Y) sei gleichverteilt auf der Menge

$$\{(x_1,y_1),\dots,(x_n,y_n)\}\subset\mathbb{R}^2,$$

und es gelte $x_i\neq x_j$ und $y_i\neq y_j$ für $i\neq j$.

Dann ist X gleichverteilt auf $\{x_1,\dots,x_n\}$ und Y ist gleichverteilt auf $\{y_1,\dots,y_n\}$, und die Cauchy Schwarzsche Ungleichung besagt

$$\left(\sum_{j=1}^n x_jy_j\,\frac{1}{n}\right)^2\le\sum_{j=1}^n x_j^2\,\frac{1}{n}\cdot\sum_{j=1}^n y_j^2\,\frac{1}{n},$$

d.h.

$$\left(\sum_{j=1}^n x_jy_j\right)^2\le\sum_{j=1}^n x_j^2\sum_{j=1}^n y_j^2,$$

und dies ist die aus der Vorlesung Mathematik I bekannte Cauchy Schwarzsche Ungleichung. □

Aus der Cauchy Schwarzschen Ungleichung folgt

$$\begin{aligned}(\mathrm{Cov}(X,Y))^2 &= (E((X-E(X))\cdot(Y-E(Y))))^2\\ &\le E((X-E(X))^2)\cdot E((Y-E(Y))^2) = V(X)\cdot V(Y),\end{aligned}$$

und daher erhält man

$$-1 \le \rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)\cdot V(Y)}} \le 1.$$

Definition 3.39. *Es seien X und Y Zufallsvariable. Dann heißt*

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)\cdot V(Y)}}$$

Korrelationskoeffizient *von X und Y. Ist $\rho(X,Y) = 0$, so heißen X und Y* **unkorreliert**.

Beispiel 3.40. Ist $Y := aX + b$, $a, b, \in \mathbb{R}$, so gilt

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E((X-E(X))(aX+b-(aE(X)+b)))\\ &= aE((X-E(X))^2) = \mathrm{sign}(a)\sqrt{V(X)\cdot V(Y)},\end{aligned}$$

d.h. $\rho(X,Y) = \pm 1$. □

Beispiel 3.41. Es sei der Zufallsvektor (X,Y) normalverteilt mit der gemeinsamen Dichte

$$f(x,y) = \frac{\sqrt{\det \boldsymbol{A}}}{2\pi}\exp\left\{-\frac{1}{2}a_{11}x^2 - a_{12}xy - \frac{1}{2}a_{22}y^2\right\},$$

wobei $\boldsymbol{A} = (a_{ij})$ positiv definit ist. Wir nehmen also $E(X) = E(Y) = 0$ an. Dies geschieht nur, um den Schreibaufwand zu verkleinern. Dann gilt (mit der Variablentransformation $z := \sqrt{a_{22}}(y + a_{12}x/a_{22})$)

$$\begin{aligned}
&\mathrm{Cov}(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y)\\
&= \frac{\sqrt{\det \boldsymbol{A}}}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\exp\left\{-\frac{1}{2}a_{11}x^2 - a_{12}xy - \frac{1}{2}a_{22}y^2\right\}dy\,dx\\
&= \frac{\sqrt{\det \boldsymbol{A}}}{2\pi}\int_{-\infty}^{\infty} x\exp\left\{-\frac{1}{2}\left(a_{11}-\frac{a_{12}^2}{a_{22}}\right)x^2\right\}\int_{-\infty}^{\infty} y\exp\left\{-\frac{1}{2}a_{22}\left(y+\frac{a_{12}}{a_{22}}x\right)^2\right\}dy\,dx\\
&= \frac{\sqrt{\det \boldsymbol{A}}}{2\pi}\int_{-\infty}^{\infty} x\exp\left\{-\frac{1}{2}\frac{\det \boldsymbol{A}}{a_{22}}x^2\right\}\int_{-\infty}^{\infty}\left(\frac{z}{\sqrt{a_{22}}} - \frac{a_{12}}{a_{22}}x\right)\frac{1}{\sqrt{a_{22}}}\exp\left\{-\frac{1}{2}z^2\right\}dz\,dx.
\end{aligned}$$

Da $z \to z \cdot \exp\left\{-\frac{1}{2}z^2\right\}$ eine ungerade Funktion ist, gilt weiter

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= -\sqrt{\frac{\det \boldsymbol{A}}{2\pi}} \frac{a_{12}}{a_{22}\sqrt{a_{22}}} \int_{-\infty}^{\infty} x^2 \exp\left\{-\frac{1}{2}\frac{\det \boldsymbol{A}}{a_{22}}x^2\right\} dx \\
&= -\sqrt{\frac{\det \boldsymbol{A}}{2\pi}} \frac{a_{12}}{a_{22}\sqrt{a_{22}}} \int_{-\infty}^{\infty} \frac{a_{22}}{\det \boldsymbol{A}} t^2 \exp\left\{-\frac{1}{2}t^2\right\} \sqrt{\frac{a_{22}}{\det \boldsymbol{A}}}\, dt \\
&= -\frac{a_{12}}{\det \boldsymbol{A}} = \sigma_{12}.
\end{aligned}$$

Mit den Bezeichnungen aus Bemerkung 3.15. ist also

$$\operatorname{Cov}(X,Y) = \rho_{12}\sqrt{V(X)\cdot V(Y)}.$$

Genauso sind im allgemeinen Fall eines normalverteilten, n dimensionalen Zufallsvektors $\boldsymbol{X} = (X_1,\ldots,X_n)^T$ die ρ_{ij} die Korrelationskoeffizienten von X_i und X_j.

Ist X ein normalverteilter, n-dimensionaler Zufallsvektor, so sind die Komponenten X_i von X_j genau dann unabhängig, wenn sie unkorreliert sind. In diesem Fall kann man $\rho(X_i, X_j)$ als ein Maß für die Abhängigkeit von X_i und X_j betrachten. □

3.4 Faltung von Zufallsvariablen

Wir betrachten die unabhängigen Zufallsvariablen X und Y und fragen nach der Verteilung von $X + Y$.

Seien zunächst X und Y diskret mit Werten in $\mathbb{N}_0$ und den Zähldichten f_X und f_Y. Dann gilt für $i \in \mathbb{N}_0$ wegen der Unabhängigkeit

$$P(X+Y=i) = \sum_{j=0}^{i} P(X=j, Y=i-j) = \sum_{j=0}^{i} P(X=j)P(Y=i-j).$$

Die Zähldichte f_Z von $Z := X + Y$ ist also gegeben durch

$$f_Z(i) = \sum_{j=0}^{i} f_X(j) f_Y(i-j).$$

Definition 3.42. *$f_Z(i) := \sum_{j=0}^{i} f_X(j)f_Y(i-j)$ heißt die* **Faltung** *der Zähldichten f_X und f_Y.*

Wir wenden dieses Resultat an, um die erzeugende Funktion der Summe von unabhängigen Zufallsvariablen zu ermitteln.

Satz 3.43. *Es sein $X_1, \ldots, X_n$ unabhängige diskrete Zufallsvariable mit Werten in $\mathbb{N}_0$ und mit den erzeugenden Funktionen g_{X_j}, $j = 1, \ldots, n$. Dann ist*

$$g_{X_1+\ldots+X_n}(z) = \prod_{j=1}^{n} g_{X_j}(z)$$

die erzeugende Funktion von $Z := X_1 + \ldots + X_n$.

Beweis: Sind X und Y unabhängige Zufallsvariable und ist $Z := X + Y$, so gilt

$$\begin{aligned} g_Z(z) &= \sum_{i=0}^{\infty} P(Z = i) z^i = \sum_{i=0}^{\infty} z^i \sum_{j=0}^{i} f_X(j) f_Y(i-j) \\ &= \sum_{j=0}^{\infty} f_X(j) z^j \sum_{i=j}^{\infty} f_Y(i-j) z^{i-j} \\ &= \sum_{j=0}^{\infty} f_X(j) z^j \sum_{k=0}^{\infty} f_Y(k) z^k = g_X(z) g_Y(z). \end{aligned}$$

Da nach Bemerkung 3.27. auch die Zufallsvariablen $X_1 + \ldots + X_k, X_{k+1}, \ldots, X_n$ unabhängig sind, erhält man genauso durch vollständige Induktion die Behauptung.
∎

Als Folgerung erhält man speziell

Korollar 3.44. *Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariable und sei $Z := \sum_{i=1}^{n} X_i$.*

(i) Ist X_i $b(n_i, p)$-verteilt, so besitzt Z die $b\left(\sum_{i=1}^{n} n_i, p\right)$-Verteilung.

(ii) Ist X_i $\pi(\lambda_i)$-verteilt, so besitzt Z die $\pi\left(\sum_{i=1}^{n} \lambda_i\right)$-Verteilung.

Beweis:

(i) Die erzeugende Funktion von X_i ist

$$g_{X_i}(z) = (pz + 1 - p)^{n_i},$$

und daher gilt

$$g_Z(z) = \prod_{i=1}^{n} g_{X_i}(z) = \prod_{i=1}^{n} (pz + 1 - p)^{n_i} = (pz - 1 - p)^{\sum_{i=1}^{n} n_i}.$$

(ii) Die erzeugende Funktion von X_i ist $g_{X_i}(z) = e^{\lambda_i(z-1)}$, und es folgt

$$g_Z(z) = \prod_{i=1}^{n} g_{X_i}(z) = \exp((z-1) \sum_{i=1}^{n} \lambda_i). \qquad ∎$$

Es seien nun X und Y unabhängig und geometrisch verteilt. Dann kann man $Z := X + Y$ als Wartezeit bis zum zweiten Erfolg in einer Folge von $b(1, p)$-verteilten

Zufallsexperimenten deuten. Wegen $P(X=i)=P(Y=i)=p(1-p)^{i-1}$, $i\geq 1$, und $P(X=0)=P(Y=0)=0$ ist die Zähldichte von Z

$$\begin{aligned} f_Z(i) &= \sum_{j=0}^{i} f_X(j)f_Y(i-j) = \sum_{j=1}^{i-1} f_X(j)f_Y(i-j) \\ &= p^2(1-p)^{i-2}(i-1), \quad i\geq 2, \\ f_Z(0) &= f_Z(1) = 0. \end{aligned}$$

Sind allgemeiner $X_1,\ldots,X_n$ unabhängig und geometrisch verteilt, so erhält man (durch Induktion) für die Zähldichte von $Z:=\sum_{i=1}^n X_i$

$$f_Z(i)=\binom{i-1}{n-1}p^n(1-p)^{i-n}, \quad i\geq n, \qquad f_Z(i)=0, \quad i<n. \tag{3.6}$$

Definition 3.45. *Eine diskrete Zufallsvariable mit der Zähldichte (3.6) heißt* **negativ binomial** *verteilt mit den Parametern n und p (kurz $Nb(n,p)$-verteilt).*

Bemerkung 3.46. Offensichtlich ist die Wartezeit bis zum n-ten Erfolg in einer Folge von Bernoulliexperimenten $Nb(n,p)$-verteilt. □

Es seien nun X und Y stetige Zufallsvariable mit der gemeinsamen Dichte f. Dann gilt für die Verteilungsfunktion von $Z:=X+Y$ (mit der Variablentransformation $y=s-x$)

$$\begin{aligned} F_Z(t) &= \iint\limits_{x+y\leq t} f(x,y)\,d(x,y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{t-x} f(x,y)\,dy\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{t} f(x,s-x)\,ds\,dx \\ &= \int_{-\infty}^{t}\int_{-\infty}^{\infty} f(x,s-x)\,dx\,ds, \end{aligned}$$

d.h. Z hat die Dichte

$$f_Z(s)=\int_{-\infty}^{\infty} f(x,s-x)\,dx\ ,\ \ s\in\mathbb{R}.$$

Sind X und Y unabhängig, so gilt $f(x,y)=f_X(x)f_Y(y)$, und man erhält speziell

$$f_Z(s)=\int_{-\infty}^{\infty} f_X(x)f_Y(s-x)\,dx\ ,\ \ s\in\mathbb{R}, \tag{3.7}$$

sowie

$$F_Z(t)=\int_{-\infty}^{\infty} f_X(x)F_Y(t-x)\,dx\ ,\ \ t\in\mathbb{R}. \tag{3.8}$$

Definition 3.47. *Es seien $f_X, f_Y : \mathbb{R} \to \mathbb{R}_+$ Dichten der stetigen Zufallsvariablen X und Y. Dann heißt die Dichte f_Z die* **Faltung** *von f_X und f_Y.*

Bemerkung 3.48. Sind X und Y nichtnegative Zufallsvariable, so gilt für die Faltung

$$f_Z(s) = \int_0^s f_X(x) f_Y(s-x)\,dx\ ,\ 0 < s < \infty. \qquad \square$$

Beispiel 3.49. Es seien X und Y unabhängige und auf $[a,b]$ gleichverteilte Zufallsvariable. Dann gilt für die Dichte von $Z := X + Y$ (wegen $f_X(x) = 0$ für alle $x \notin (a,b)$)

$$f_Z(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x)\,dx = \frac{1}{b-a}\int_a^b f_Y(s-x)\,dx.$$

Für $s \le 2a$ gilt $s - x \le a$ für alle $x \in (a,b)$, d.h. $f_Y(s-x) = 0$ für alle $x \in (a,b)$ und daher $f_Z(s) = 0$. Genauso gilt $f_Z(s) = 0$ für $s \ge 2b$.

Für $2a < s \le a + b$ gilt $s - x \le s - a \le b$ für $x \in (a,b)$, d.h.

$$f_Z(s) = \frac{1}{b-a}\int_a^{s-a} f_Y(s-x)\,dx = \frac{1}{(b-a)^2}(s-2a),$$

und genauso erhält man für $a + b \le s < 2b$

$$f_Z(s) = \frac{1}{(b-a)^2}(2b - s).$$

Eine stetige Zufallsvariable mit der Dichte f_Z heißt **dreiecksverteilt**. Den Graphen von f_Z findet man in Abbildung 3.5. $\square$

Beispiel 3.50. Es seien nun $X_1, \ldots, X_n$ unabhängige, exponentialverteilte Zufallsvariable zum Parameter α, und es sei $Y_i := \sum_{j=1}^{i} X_j$.

Z.B. sei X_i die Wartezeit zwischen dem $(i-1)$-ten und dem i-ten Anruf in einer Telefonzentrale und Y_i die Wartezeit bis zum i-ten Anruf, oder es sei in einem System mit i identischen Stand-by-Komponenten X_j die Lebensdauer der j-ten Komponente und Y_i die Ausfallzeit des gesamten Systems.

Wir zeigen (durch Induktion), dass dann Y_i Erlang verteilt ist mit den Parametern i und α.

Für $i = 1$ ist $Y_1 = X_1$, und die Aussage ist trivial.

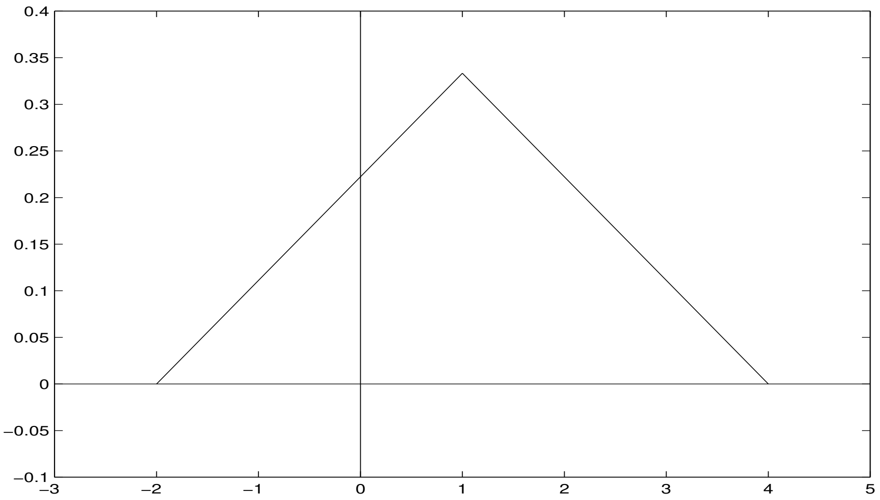

Abbildung 3.5: Dichte der Dreiecksverteilung

Es sei für $i > 1$ die Zufallsvariable Y_{i-1} Erlang verteilt mit der Dichte

$$f_{Y_{i-1}}(s) = \begin{cases} 0 & , \quad s < 0 \\ \alpha \exp(-\alpha s) \dfrac{(\alpha s)^{i-2}}{(i-2)!} & , \quad s \geq 0 \ . \end{cases}$$

Dann gilt wegen $f_{X_1}(s-x) = 0$ für $s < x$ und der Unabhängigkeit von Y_{i-1} und X_i (vgl. Bemerkung 3.27.)

$$\begin{aligned} f_{Y_i}(s) &= \int_0^\infty f_{Y_{i-1}}(x) f_{X_i}(s-x)\, dx \\ &= \int_0^s \alpha \exp(-\alpha x) \frac{(\alpha x)^{i-2}}{(i-2)!} \cdot \alpha \cdot \exp(-\alpha(s-x))\, dx \\ &= \alpha^2 \exp(-\alpha s) \int_0^s \frac{(\alpha x)^{i-2}}{(i-2)!}\, dx \\ &= \alpha \exp(-\alpha s) \frac{(\alpha s)^{i-1}}{(i-1)!} \ . \qquad \square \end{aligned}$$

Beispiel 3.51. Sind $X_1, \ldots, X_n$ unabhängig und identisch normalverteilt mit den Parametern μ und σ, so ist auch $Y := \sum_{i=1}^{n} X_i$ normalverteilt mit den Parametern $n\mu$ und $\sqrt{n}\sigma$ (und daher ist das arithmetische Mittel $\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$ normalverteilt mit dem Erwartungswert $E(\bar{X}) = \mu$ und der Varianz $V(\bar{X}) = \frac{\sigma^2}{n}$). Dieses Ergebnis erhält man durch n-fache Faltung der Dichte der $N(\mu, \sigma^2)$-Verteilung. $\square$

3.5 Charakteristische Funktion

Ein analytisches Hilfsmittel zur Berechnung der Faltung ist die charakteristische Funktion. Mit ihr kann man ebenfalls auf einfache Weise die Momente einer Zufallsvariable berechnen (ähnlich wie mit der erzeugenden Funktion für diskrete Zufallsvariable).

Definition 3.52. *Für eine Zufallsvariable X heißt*

$$\phi_X : \mathbb{R} \to \mathbb{C}, \quad \phi_X(s) := E(e^{isX})$$

die **charakteristische Funktion**.

Ist X eine diskrete Zufallsvariable mit den Werten x_k, $k \in \mathbb{N}$, so ist die charakteristische Funktion gegeben durch

$$\phi_X(s) = \sum_{k=1}^{\infty} e^{isx_k} P(X = x_k).$$

Ist X eine stetige Zufallsvariable mit der Dichte f_X, so ist die charakteristische Funktion

$$\phi_X(s) = \int_{-\infty}^{\infty} e^{isx} f_X(x)\, dx,$$

wobei $i := \sqrt{-1}$ die imaginäre Einheit bezeichnet.

Bemerkung 3.53. Für stetige Zufallsvariable ist also die charakteristische Funktion gerade (bis auf einen konstanten Faktor) die Fouriertransformierte der zugehörigen Dichtefunktion f_X. □

Bemerkung 3.54. Da $|e^{isx}| = |\cos(sx) + i\sin(sx)| = 1$ gilt, besitzt jede Zufallsvariable eine charakteristische Funktion, denn die definierende Summe bzw. das definierende Integral konvergiert absolut. □

Beispiel 3.55. Es sei X Poisson-verteilt mit dem Parameter λ. Dann gilt

$$\begin{aligned} \phi_X(s) &= \sum_{k=0}^{\infty} e^{isk} \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{is})^k}{k!} \\ &= e^{-\lambda} \exp(\lambda e^{is}) = \exp(\lambda(e^{is} - 1)) = g_X(e^{is}). \end{aligned}$$

Dieser Zusammenhang zwischen der charakteristischen Funktion und der erzeugenden Funktion einer diskreten Zufallsvariable mit Werten in $\mathbb{N}_0$ gilt stets. □

Beispiel 3.56. Sei X normalverteilt mit $E(X) = 0$ und $V(X) = 1$. Dann ist die charakteristische Funktion von X

$$\begin{aligned}\phi_X(s) &= \int_{-\infty}^{\infty} e^{isx} f_X(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(isx - \frac{x^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x-is)^2\right) \cdot \exp\left(-\frac{1}{2}s^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}s^2\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x-is)^2\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}s^2\right) \int_{\Gamma} \exp\left(-\frac{1}{2}y^2\right) dy\ ,\end{aligned}$$

wobei das Integral über die Gerade $\Gamma : x - is, -\infty < x < \infty$, zu erstrecken ist.

Nach dem Cauchyschen Integralsatz gilt für jedes $R > 0$

$$\left(\int_{-R-is}^{R-is} + \int_{R-is}^{R} + \int_{R}^{-R} + \int_{-R}^{-R-is} \right) \exp\left(-\frac{1}{2}y^2\right) dy = 0\ ,$$

und wegen

$$\left| \int_{R-is}^{R} \exp\left(-\frac{1}{2}y^2\right) dy \right|\ ,\ \left| \int_{-R}^{-R-is} \exp\left(-\frac{1}{2}y^2\right) dy \right| \le |s| \cdot \exp\left(\frac{1}{2}(s^2 - R^2)\right)$$

erhält man für $R \to \infty$

$$\int_{\Gamma} \exp\left(-\frac{1}{2}y^2\right) dy = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right) dx = \sqrt{2\pi}\ ,$$

d.h.

$$\phi_X(s) = \exp\left(-\frac{1}{2}s^2\right). \qquad \Box$$

Beispiel 3.57. Ist Y $N(\mu, \sigma^2)$-verteilt, so gilt mit der zugehörigen standardisierten Zufallsvariablen X (die $N(0,1)$-verteilt ist) $Y = \sigma X + \mu$, und daher

$$\begin{aligned}\phi_Y(s) &= E(e^{isY}) = E(e^{is(\sigma X + \mu)}) \\ &= e^{is\mu} E(e^{is\sigma X}) = e^{is\mu} \phi_X(\sigma s) \\ &= \exp(is\mu - \frac{1}{2}\sigma^2 s^2). \qquad \Box\end{aligned}$$

Beispiel 3.58. Es sei X Gamma-verteilt mit den Parametern α und β. Dann gilt

$$f_X(s) = \begin{cases} 0 & ,\ s \le 0 \\ \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s} & ,\ s > 0, \end{cases}$$

und daher ist die charakteristische Funktion (mit der Variablentransformation $z := (\beta - is)x$)

$$\begin{aligned}
\phi_X(s) &= \frac{\beta^a}{\Gamma(\alpha)} \int_0^\infty e^{isx} x^{\alpha-1} e^{-\beta x}\,dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-is)x}\,dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\beta-is)^\alpha} \int_\Gamma z^{\alpha-1} e^{-z}\,dz\ ,
\end{aligned}$$

wobei das Integral über die Halbgerade $\Gamma : (\beta - is)x, 0 < x < \infty$, erstreckt wird. Ähnlich wie in Beispiel 3.56. kann man mit dem Cauchyschen Integralsatz zeigen, dass

$$\int_\Gamma z^{\alpha-1} e^{-z}\,dz = \int_0^\infty x^{\alpha-1} e^{-x}\,dx = \Gamma(\alpha)$$

gilt, und daher folgt

$$\phi_X(s) = \left(\frac{\beta}{\beta - is}\right)^\alpha . \qquad \square$$

Wir berechnen nun mit Hilfe der charakteristischen Funktion Momente von Zufallsvariablen. Wir beschränken uns auf den stetigen Fall (für diskrete Zufallsvariable können wir die erzeugende Funktion benutzen).

Differenziert man

$$\phi_X(s) = \int_{-\infty}^\infty e^{isx} f_X(x)\,dx$$

nach s (ohne sich um die Rechtfertigung der Vertauschung von Integration und Differentiation zu kümmern), so erhält man

$$\phi_X'(s) = \int_{-\infty}^\infty ixe^{isx} f_X(x)\,dx\ ,$$

und dieses Integral existiert, wenn der Erwartungswert $E(X)$ von X existiert.

Für $s = 0$ erhält man

$$\phi_X'(0) = i \int_{-\infty}^\infty x f_X(x)\,dx\ ,$$

d.h. für den Erwartungswert

$$E(X) = -i \cdot \phi_X'(0).$$

Genauso erhält man durch Induktion für die höheren Momente

$$E(X^k) = (-i)^k \phi_X^{(k)}(0)\ ,\ \ k = 1, 2, \ldots,$$

und damit insbesondere für die Varianz

$$V(X) = E(X^2) - (E(X))^2 = -\phi_X''(0) + (\phi_X'(0))^2\ .$$

Die Vertauschung der Differentiation und Integration ist gerechtfertigt. Den genauen Zusammenhang zwischen der Differenzierbarkeit der charakteristischen Funktion ϕ_X einer reellen Zufallsvariable und der Existenz ihrer Momente beschreibt der folgende Satz.

Satz 3.59. *Sei X eine reelle Zufallsvariable mit der charakteristischen Funktion ϕ und sei $k \in \mathbb{N}$. Dann gilt:*

(i) Existiert das k-te Moment von X, so ist ϕ k-mal differenzierbar, und es gilt

$$\phi^{(k)}(s) = i^k \int_{-\infty}^{\infty} x^k e^{ixs} f_X(x)\, dx.$$

Speziell ist $E(X^k) = (-i)^k \phi^{(k)}(0)$.

(ii) Ist ϕ in einer Umgebung von $s = 0$ k-mal differenzierbar und ist k gerade, so existiert das k-te Moment.

Beweis: s. **Hinderer** *[16, p. 193].*

Bemerkung 3.60. Ist ϕ für ein ungerades k in einer Umgebung von $s = 0$ k-mal differenzierbar, so folgt aus (ii), dass das $(k-1)$-te Moment existiert, es kann aber $E(|X|^k) = \infty$ sein. $\square$

Beispiel 3.61. Es sei X $N(\mu, \sigma^2)$-verteilt. Dann besitzt X die charakteristische Funktion

$$\phi_X(s) = \exp\left(i\mu s - \frac{1}{2}\sigma^2 s^2\right)$$

mit den Ableitungen

$$\begin{aligned}
\phi_X'(s) &= (i\mu - \sigma^2 s) \exp\left(i\mu s - \frac{1}{2}\sigma^2 s^2\right) \\
\phi_X''(s) &= ((i\mu - \sigma^2 s)^2 - \sigma^2) \exp\left(i\mu s - \frac{1}{2}\sigma^2 s^2\right)\ ,
\end{aligned}$$

und es folgt

$$\begin{aligned}
E(X) &= -i\phi_X'(0) = \mu \\
V(X) &= -\phi_X''(0) + (\phi_X'(0))^2 = -(-\mu^2 - \sigma^2) - \mu^2 = \sigma^2. \qquad \square
\end{aligned}$$

Beispiel 3.62. Es sei X Gamma-verteilt mit den Parametern α und β. Die charakteristische Funktion von X ist dann

$$\phi_X(s) = \left(\frac{\beta}{\beta - is}\right)^{\alpha}.$$

Es gilt

$$\begin{aligned} \phi'_X(s) &= i\alpha\beta^{\alpha}(\beta - is)^{-\alpha-1} \\ \phi''_X(s) &= -\alpha(\alpha+1)\beta^{\alpha}(\beta - is)^{-\alpha-2}, \end{aligned}$$

und daher

$$\begin{aligned} E(X) &= -i\phi'_X(0) = \frac{\alpha}{\beta}, \\ V(X) &= -\phi''_X(0) + (\phi'_X(0))^2 = \frac{\alpha}{\beta^2}. \qquad \square \end{aligned}$$

Außer bei der Berechnung von Momenten hat die charakteristische Funktion auch bei der Bestimmung der Faltung von Dichten eine große Bedeutung.

Satz 3.63. *Es seien X und Y unabhängige Zufallsvariable mit den charakteristischen Funktionen ϕ_X und ϕ_Y. Dann ist die charakteristische Funktion ϕ_Z der Zufallsvariable $Z := X + Y$ das Produkt von ϕ_X und ϕ_Y*

$$\phi_Z(s) = \phi_X(s) \cdot \phi_Y(s).$$

Beweis: Da X und Y unabhängig sind, gilt für die Dichte f_Z

$$f_Z(x) = \int_{-\infty}^{\infty} f_X(y) f_Y(x - y)\, dy,$$

und daher folgt

$$\begin{aligned} \phi_Z(s) &= \int_{-\infty}^{\infty} e^{isx} f_Z(x)\, dx \\ &= \int_{-\infty}^{\infty} e^{isx} \int_{-\infty}^{\infty} f_X(y) f_Y(x-y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} e^{isy} f_X(y) \int_{-\infty}^{\infty} e^{is(x-y)} f_Y(x-y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} e^{isy} f_X(y) \int_{-\infty}^{\infty} e^{ist} f_Y(t)\, dt\, dy \\ &= \phi_X(s) \cdot \phi_Y(s). \qquad \blacksquare \end{aligned}$$

Bemerkung 3.64. Es ist klar, dass sich Satz 3.63. auf die Summe von endlich vielen unabhängigen Zufallsvariablen verallgemeinern lässt. □

Beispiel 3.65. Es seien X ($N(\mu_X, \sigma_X^2)$-verteilt) und Y ($N(\mu_Y, \sigma_Y^2)$ -verteilt) unabhängige Zufallsvariable und $Z := X + Y$. Dann besitzt Z die charakteristische Funktion

$$\begin{aligned}
\phi_Z(s) &= \phi_X(s) \cdot \phi_Y(s) \\
&= \exp(i\mu_X s - \frac{1}{2}\sigma_X^2 s^2) \cdot \exp(i\mu_Y s - \frac{1}{2}\sigma_Y^2 s) \\
&= \exp(i(\mu_X + \mu_Y)s - \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)s^2).
\end{aligned}$$

Mit $\mu_Z := \mu_X + \mu_Y$ und $\sigma_Z^2 := \sigma_X^2 + \sigma_Y^2$ hat also Z die charakteristische Funktion der $N(\mu_Z, \sigma_Z^2)$-Verteilung. Nach dem folgenden Eindeutigkeitssatz ist daher Z $N(\mu_Z, \sigma_Z^2)$-verteilt.

Bemerkung 3.66. Auf ähnliche Weise wie in Beispiel 3.65. sieht man ein, dass das arithmetische Mittel $\bar{X}$ von n unabhängigen, identisch $N(\mu, \sigma^2)$-verteilten Zufallsvariablen normalverteilt ist mit den Parametern

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{1}{n}\sigma^2.$$

Zum Beweis benötigt man zusätzlich, dass für jede Zufallsvariable X und für alle $\alpha > 0$

$$\phi_{\alpha X}(s) = \phi_X(\alpha s)$$

gilt (Aufgabe!). □

Satz 3.67. (Eindeutigkeitssatz) *Verschiedene Verteilungen auf* $\mathbb{R}$ *besitzen verschiedene charakteristische Funktionen*

Beweis: s. **Hinderer** [16, p. 191].

Beispiel 3.68. Eine Zufallsvariable Y hat eine χ^2-Verteilung mit n Freiheitsgraden, wenn Y dargestellt werden kann als

$$Y = X_1^2 + \ldots + X_n^2 \ ,$$

wobei $X_1, \ldots, X_n$ unabhängige Zufallsvariable sind, die alle $N(0,1)$-verteilt sind.

Wir haben bereits bemerkt, dass die χ^2-Verteilung eine Gamma-Verteilung mit den Parametern $\alpha = \frac{n}{2}, \beta = \frac{1}{2}$ ist. Wir können diese Aussage nun nachweisen.

Dazu leiten wir zunächst die Verteilung von $Z := X^2$ her, wenn X $N(0,1)$-verteilt ist. Für die Verteilungsfunktion F_Z gilt für $t \leq 0$ sicher $F_Z(t) = 0$ und für $t > 0$

$$\begin{aligned} F_Z(t) &= P([Z \leq t]) = P([X^2 \leq t]) \\ &= P(-\sqrt{t} \leq X \leq \sqrt{t}) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) \\ &= 2F_X(\sqrt{t}) - 1 \end{aligned}$$

(wegen der Symmetrie $F_X(-t) = 1 - F_X(t)$).

Für die Dichte folgt daher wegen

$$f_X(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)$$

$$\begin{aligned} f_Z(t) &= \frac{d}{dt} F_Z(t) = 2F_X'(\sqrt{t}) \cdot \frac{1}{2\sqrt{t}} \\ &= \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{t}{2}\right) \quad \text{für } t > 0 \end{aligned}$$

und $f_Z(t) = 0$ für $t < 0$, d.h. Z ist Gamma-verteilt mit den Parametern $\alpha = \frac{1}{2}, \beta = \frac{1}{2}$.

Die charakteristische Funktion von Z ist daher

$$\phi_Z(s) = \sqrt{\frac{1}{1-2is}},$$

nach Satz 3.63. ist die charakteristische Funktion von Y

$$\phi_Y(s) = (\phi_Z(s))^n = \sqrt{\left(\frac{1}{1-2is}\right)^n},$$

und nach Beispiel 3.58. und dem Eindeutigkeitssatz (Satz 3.67.) besitzt Y eine $\Gamma(\alpha, \beta)$ Verteilung mit den Parametern $\alpha = \frac{n}{2}$ und $\beta = \frac{1}{2}$. □

Die χ^2-Verteilung besitzt in der mathematischen Statistik eine große Bedeutung. Es gilt (ohne Beweis)

Satz 3.69. *Sind $X_1, \ldots, X_n$ unabhängige und identisch normal-verteilte Zufallsvariable mit den Parametern μ und σ und bezeichnet $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ das arithmetische Mittel der X_i, so besitzt*

$$Y := \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

eine χ^2-Verteilung mit $(n-1)$ Freiheitsgraden.

Bemerkung 3.70. Sind $X_1, \dots, X_n, \bar{X}$ und Y wie in Satz 3.69. definiert, so ist $E(Y) = n-1$ der Erwartungswert von Y, und daher kann man die Varianz von $X_1, \dots, X_n$ schätzen als

$$\sigma^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2. \qquad \Box$$

3.6 Bedingte Verteilungen

Sind zwei Zufallsvariable X und Y unabhängig, so ist die gemeinsame Verteilung von (X, Y) eindeutig durch die marginalen Verteilungen festgelegt. Im Falle abhängiger Zufallsvariablen benötigt man bedingte Verteilungen.

Wir erinnern an die Definition der bedingten Wahrscheinlichkeit

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{falls } P(B) > 0 \ .$$

Definition 3.71. *Es seien zunächst X und Y diskrete Zufallsvariable, f die gemeinsame Zähldichte und f_X und f_Y die marginalen Zähldichten.*

Dann heißt

$$f_{Y|X}(y|x) := P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{f(x,y)}{f_X(x)}$$

die **bedingte Zähldichte** *von Y unter der Bedingung $X = x$.*

Die bedingte Zähldichte ist für alle x definiert, für die $f_X(x) \neq 0$ ist.

Genauso ist

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad f_Y(y) \neq 0 \ ,$$

die bedingte Zähldichte von X unter der Bedingung $Y = y$.

Die beiden letzten Gleichungen kann man auch schreiben als

$$f(x,y) = f_{Y|X}(y|x) f_X(x) = f_{X|Y}(x|y) f_Y(y) \ .$$

Sind X und Y unabhängig, so gilt $f(x,y) = f_X(x) \cdot f_Y(y)$, und daher

$$f_{Y|X}(y|x) = f_Y(y) \quad \text{für alle } x.$$

Umgekehrt impliziert diese Gleichung natürlich auch die Unabhängigkeit von X und Y.

$f_{Y|X}(\cdot|x)$ bzw. $f_{X|Y}(\cdot|y)$ sind für jedes feste x bzw. y Zähldichten, denn wegen $f \geq 0$ und $f_X \geq 0$ gilt $f_{Y|X} \geq 0$ und

$$\sum_y f_{Y|X}(y|x) = \frac{1}{f_X(x)} \sum_y f(x,y) = 1.$$

Für die marginale Dichte f_Y hat man

$$f_Y(y) = \sum_x f(x,y) = \sum_x f_{Y|X}(y|x) \cdot f_X(x) \ .$$

Dies ist eine andere Gestalt der Formel für die totale Wahrscheinlichkeit (vgl. Kapitel 1).

Beispiel 3.72. Ein Rechenzentrum betreibe zwei Rechner A und B. Eingehende Jobs werden (unabhängig) mit der Wahrscheinlichkeit p auf den Rechner A und mit der Wahrscheinlichkeit $1-p$ auf den Rechner B geleitet. Die Anzahl X der eingehenden Jobs pro Zeiteinheit sei Poisson-verteilt mit dem Parameter α. Man bestimme die Verteilung der Anzahl Y der Jobs, die in Rechner A pro Zeiteinheit eingehen.

Da die Aufteilung von n eingehenden Jobs auf die beiden Rechner als Folge von n unabhängigen Bernoulliexperimenten betrachtet werden kann, ist die bedingte Verteilung von Y unter der Bedingung $[X = n]$ eine Binomialverteilung mit den Parametern n und p.

$$f_{Y|X}(k|n) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & , \quad 0 \leq k \leq n \\ 0 & , \quad \text{sonst} \end{cases} .$$

Daher folgt

$$\begin{aligned} f_Y(k) &= \sum_{n=0}^{\infty} f_{Y|X}(k|n) f_X(n) = \sum_{n=k}^{\infty} \tbinom{n}{k} p^k (1-p)^{n-k} \frac{\alpha^n}{n!} e^{-\alpha} \\ &= \frac{(\alpha p)^k e^{-\alpha}}{k!} \sum_{n=k}^{\infty} \frac{(\alpha(1-p))^{n-k}}{(n-k)!} = \frac{(\alpha p)^k e^{-\alpha}}{k!} e^{\alpha(1-p)} = \frac{(\alpha p)^k e^{-\alpha p}}{k!} \ . \end{aligned}$$

Es ist also auch Y Poisson-verteilt mit dem Parameter αp. Dies ist der Grund, warum man sagt, dass die Poisson-Verteilung unter zufälliger Aufteilung erhalten bleibt. □

Definition 3.73. *Sind X und Y stetige Zufallsvariable mit der gemeinsamen Dichte f und den marginalen Dichten f_X und f_Y, so definieren wir in Analogie zum diskreten Fall die bedingte Dichte von Y unter der Bedingung $X = x$ durch*

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \ .$$

Wie vorher ist dies natürlich wieder eine Dichte für jedes feste x, und es gilt $f_{Y|X}(\cdot|x) = f_Y$ für alle x genau dann, wenn X und Y unabhängig sind.

Im diskreten Fall haben wir $f_{Y|X}$ als bedingte Wahrscheinlichkeit interpretiert. Dies ist im stetigen Fall wegen $P(X = x) = 0$ für alle x nicht ohne zusätzliche Überlegungen möglich.

Für $x, y, h \in \mathbb{R}$ mit $h > 0$ und $f_X(x) \neq 0$ betrachten wir

$$\begin{aligned} F(y; x, h) &:= P(Y \leq y | x \leq X \leq x + h) \\ &= \frac{P(Y \leq y, x < X \leq x + h)}{P(x < X \leq x + h)} \\ &= \frac{\int_x^{x+h} \int_{-\infty}^{y} f(s,t)\,dt\,ds}{\int_x^{x+h} f_X(s)\,ds}. \end{aligned}$$

Hiermit definieren wir

Definition 3.74.

$$F_{Y|X}(y|x) := \lim_{h \to 0+0} F(y; x, h) \ .$$

heißt die **bedingte Verteilungsfunktion** *von Y unter der Bedingung $X = x$.*

Wir berechnen nun (unter Benutzung der Regel von de l'Hôpital) $F_{Y|X}(y|x)$ an einer Stetigkeitsstelle (x, y) von f. (Die Vertauschung des Grenzwerts und des Integrals kann wieder gerechtfertigt werden)

$$\begin{aligned} F_{Y|X}(y|x) &= \lim_{h \to 0+0} \frac{\int_x^{x+h} \int_{-\infty}^{y} f(s,t)\,dt\,ds}{\int_x^{x+h} f_X(s)\,ds} = \lim_{h \to 0+0} \frac{\int_{-\infty}^{y} f(x+h,t)\,dt}{f_X(x+h)} \\ &= \frac{\int_{-\infty}^{y} \lim_{h \to 0+0} f(x+h,t)\,dt}{\lim_{h \to 0+0} f_X(x+h)} = \frac{\int_{-\infty}^{y} f(x,t)\,dt}{f_X(x)} = \int_{-\infty}^{y} f_{Y|X}(t|x)\,dt. \end{aligned}$$

$f_{Y|X}$ ist also die zu $F_{Y|X}$ gehörige Dichte.

Für die marginale Dichte f_Y gilt

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\,dx \ . \tag{3.9}$$

Dies ist das stetige Analogon der Formel für die **totale Wahrscheinlichkeit** .

Gilt $f_Y(y) > 0$, so ist auch $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$ definiert. Setzt man hier den Ausdruck (3.9) für f_Y ein, so erhält man

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\,dx}.$$

Dies ist das stetige Analogon der **Formel von Bayes**.

Bemerkung 3.75. Die Definitionen und Resultate (auch die noch folgenden) lassen sich sofort auf den Fall übertragen, dass $\boldsymbol{X}$ ein k-dimensionaler und $\boldsymbol{Y}$ ein ℓ-dimensionaler Zufallsvektor ist. $\square$

Beispiel 3.76. Ist (X, Y) normalverteilt mit der gemeinsamen Dichte

$$f(x,y) = \frac{\sqrt{\det \boldsymbol{A}}}{2\pi} \exp\left\{-\frac{1}{2}(a_{11}(x-\mu_X)^2 + 2a_{12}(x-\mu_X)(y-\mu_Y) + a_{22}(y-\mu_Y)^2)\right\} ,$$

so gilt für die marginale Dichte

$$f_X(x) = \sqrt{\frac{\det \boldsymbol{A}}{2\pi a_{22}}} \exp\left\{-\frac{1}{2}\frac{\det \boldsymbol{A}}{a_{22}}(x-\mu_X)^2\right\} ,$$

und daher ist die bedingte Dichte von Y unter der Bedingung $X = x$

$$\begin{aligned} f_{Y|X}(y|x) &= \\ &\sqrt{\frac{a_{22}}{2\pi}} \exp\left\{-\frac{1}{2}\left((a_{11} - \frac{\det \boldsymbol{A}}{a_{22}})(x-\mu_X)^2 + 2a_{12}(x-\mu_X)(y-\mu_Y) + a_{22}(y-\mu_Y)^2\right)\right\} \\ &= \sqrt{\frac{a_{22}}{2\pi}} \exp\left\{-\frac{1}{2}a_{22}\left(y-\mu_Y + \frac{a_{12}}{a_{22}}(x-\mu_X)\right)^2\right\} . \end{aligned}$$

Für jedes x ist also die bedingte Verteilung von Y unter Bedingung $X = x$ eine Normalverteilung mit dem Erwartungswert

$$\mu_Y - \frac{a_{12}}{a_{22}}(x-\mu_X)$$

und der Varianz $\frac{1}{a_{22}}$.

$f_{Y|X}(\cdot|x)$ stimmt genau dann mit der marginalen Dichte f_Y überein, wenn $a_{12} = 0$ gilt (wenn also X und Y unabhängig sind) oder wenn X gleich dem Erwartungswert μ_X von X ist.

Aus Symmetriegründen ist natürlich auch $f_{X|Y}(\cdot|y)$ die Dichte der Normalverteilung mit den Parametern $\mu_X - \frac{a_{12}}{a_{11}}(y-\mu_Y)$ und $\frac{1}{a_{11}}$. $\square$

Wir haben bisher nur den Fall betrachtet, dass X und Y beide diskret oder beide stetig sind. Die Definition der bedingten Verteilung und der bedingten Dichte kann leicht auf den Fall übertragen werden, dass eine Zufallsvariable stetig und die andere diskret ist.

Beispiel 3.77. Es sei X, die Bearbeitungszeit für einen Job in einem Rechenzentrum, exponentialverteilt mit dem Parameter α, d.h. $f_X(x) = \alpha \exp(-\alpha x)$, $x > 0$.

Es sei die Anzahl der in einem Intervall der Länge $t > 0$ eingehenden Jobs Poisson-verteilt mit dem Parameter λt. Schließlich sei Y die Anzahl der Jobs, die im Rechenzentrum eingehen, während ein Job bearbeitet wird. Welche Verteilung besitzt Y?

Zunächst gilt

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{(\lambda x)^y}{y!} \exp(-\lambda x).$$

Daher erhält man für die gemeinsame Dichte

$$f(x,y) = f_X(x) f_{Y|X}(y|x) = \alpha \frac{(\lambda x)^y}{y!} \exp(-(\lambda + \alpha)x)$$

und durch Integration hieraus die marginale (Zähl-)Dichte

$$f_Y(y) = P(Y = y) = \int_0^\infty f(x,y)\,dx = \frac{\alpha}{y!} \int_0^\infty (\lambda x)^y \exp(-(\lambda + \alpha)x)\,dx\ .$$

Mit der Variablentransformation $w := (\alpha + \lambda)x$ folgt

$$f_Y(y) = \frac{\alpha \lambda^y}{y!(\alpha + \lambda)^{y+1}} \int_0^\infty w^y e^{-w}\,dw = \frac{\alpha \lambda^y}{(\alpha + \lambda)^{y+1}},$$

da das letzte Integral $\Gamma(y + 1) = y!$ ist. Mit $\rho := \dfrac{\lambda}{\alpha}$ hat man also

$$f_Y(y) = \frac{\rho^y}{(1+\rho)^{y+1}} = \left(\frac{\rho}{1+\rho}\right)^y \cdot \frac{1}{1+\rho}\ ,\ y \in \mathbb{N}_0\ ,$$

und Y besitzt eine modifizierte geometrische Verteilung, d.h. $Z(j) := Y(j - 1)$, $j \in \mathbb{N}$, ist geometrisch verteilt, mit dem Parameter $p := \dfrac{1}{1+\rho}$.

Der Erwartungswert ist

$$\begin{aligned} E(Y) &= \frac{1}{1+\rho} \sum_{y=1}^\infty y \left(\frac{\rho}{1+\rho}\right)^y = \frac{\rho}{1+\rho} \cdot \frac{1}{1+\rho} \sum_{y=1}^\infty y \left(\frac{\rho}{1+\rho}\right)^{y-1} \\ &= \frac{\rho}{1+\rho} E(Z) = \rho = \frac{\lambda}{\alpha}\ . \end{aligned}$$

Dieses Ergebnis ist auch anschaulich klar. Ist $E(Y) > 1$, so wird die Warteschlange im Rechner beliebig lang. Dies ist genau dann der Fall, wenn die Rate λ, mit der neue Jobs eingehen, größer ist als die Rate α, mit der Jobs abgearbeitet werden. □

Wir haben bereits gesehen, dass $f_{Y|X}(\cdot|x)$ (sowohl im diskreten als auch im stetigen Fall) eine Dichte ist. Wir können daher den Erwartungswert - oder allgemeiner das k-te Moment - bzgl. dieser Dichte definieren.

Definition 3.78. *Es seien X, Y reelle Zufallsvariable und $f_{Y|X}(y|x)$ die bedingte (Zähl-)Dichte von Y unter der Bedingung $X = x$. Dann heißt für $k \in \mathbb{N}$*

$$E(Y^k|X=x) := \begin{cases} \sum_j y_j^k f_{Y|X}(y_j|x) & , \text{ falls } Y \text{ diskret} \\ \int\limits_{-\infty}^{\infty} y^k f_{Y|X}(y|x)\,dy & , \text{ falls } Y \text{ stetig} \end{cases}$$

das **bedingte k-te Moment** *von Y unter der Bedingung $X = x$.*

Insbesondere heißt $E(Y|X = x)$ der **bedingte Erwartungswert** *von Y unter der Bedingung $X = x$.*

$E(Y^k|X = x)$ ist für alle x definiert, für die $f_X(x) \neq 0$ gilt. Wir setzen

$$E(Y^k|X=x) = 0 \quad \text{für alle } x \text{ mit } f_X(x) = 0.$$

Dann ist durch $m(x) := E(Y^k|X = x)$, $x \in \mathbb{R}$, für jedes $k \in \mathbb{N}$ eine Zufallsvariable definiert ($x \mapsto E(Y|X = x)$ heißt auch **Regressionsfunktion** von Y über X). Für die Momente von m gilt

Satz 3.79. (Gesetz über die totalen Momente bzw. totale Erwartung)
Es seien X, Y reelle Zufallsvariable. Dann besitzt die Zufallsvariable

$$m(x) := \begin{cases} E(Y^k|X=x) & , \text{ falls } \quad f_X(x) \neq 0 \\ 0 & , \text{ falls } \quad f_X(x) = 0 \end{cases}$$

den Erwartungswert

$$E(m) = E(Y^k).$$

Beweis: Wir beschränken uns auf den Fall, dass X und Y stetige Zufallsvariable sind mit der gemeinsamen Dichte f.

Es gilt für alle $x \in \mathbb{R}$ mit $f_X(x) \neq 0$

$$\begin{aligned} E(Y^k|X=x)f_X(x) &= \int\limits_{-\infty}^{\infty} y^k\, f_{Y|X}(y|x)\,dy\, f_X(x) \\ &= \int\limits_{-\infty}^{\infty} y^k f(x,y)\,dy, \end{aligned} \tag{3.10}$$

und da aus

$$f_X(x) = \int\limits_{-\infty}^{\infty} f(x,y)\,dy = 0$$

$f(x,y) = 0$ für alle $y \in \mathbb{R}$ folgt, ist (3.10) für alle $x \in \mathbb{R}$ richtig. Daher folgt

$$\begin{aligned} E(m) &= \int_{-\infty}^{\infty} E(Y^k|X=x) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} y^k \int_{-\infty}^{\infty} f(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} y^k f_Y(y)\,dy = E(Y^k)\ . \qquad \blacksquare \end{aligned}$$

Beispiel 3.80. Es seien $N, X_1, X_2, \ldots$ unabhängige Zufallsvariable. $X_1, X_2, \ldots$ seien identisch verteilt, und es sei N diskret mit Werten in $\mathbb{N}$. Wir berechnen den Erwartungswert und die Varianz von $Z := X_1 + X_2 + \ldots + X_N$. (Die Anzahl der Summanden von Z ist also zufällig).

Es sei $\mu := E(X_j), \sigma^2 := V(X_j)$. Dann gilt für die Zufallsvariable

$$m(n) := \begin{cases} E(X_1 + \ldots + X_N|N=n) & \text{falls} \quad P(N=n) > 0 \\ 0 & \text{falls} \quad P(N=n) = 0 \end{cases}$$

nach dem Gesetz über die totale Erwartung

$$\begin{aligned} E(Z) &= E(m) = \sum_{n=1}^{\infty} E(X_1 + \ldots + X_N|N=n)P(N=n) \\ &= \sum_{n=1}^{\infty} E(X_1 + \ldots + X_n)P(N=n) \\ &= \sum_{n=1}^{\infty} n\mu P(N=n) = \mu \sum_{n=1}^{\infty} nP(N=n) = \mu E(N)\ , \end{aligned}$$

und genauso

$$\begin{aligned} E(Z^2) &= \sum_{n=1}^{\infty} E((X_1 + \ldots + X_N)^2|N=n)P(N=n) \\ &= \sum_{n=1}^{\infty} E((X_1 + \ldots + X_n)^2)P(N=n)\ . \end{aligned}$$

Wegen der Unabhängigkeit der X_j folgt

$$\begin{aligned} E(Z^2) &= \sum_{n=1}^{\infty} \left(\sum_{j=1}^{n} E(X_j^2) + 2\sum_{j=1}^{n}\sum_{k=1}^{j-1} E(X_jX_k) \right) P(N=n) \\ &= \sum_{n=1}^{\infty} (nE(X_1^2) + n(n-1)\mu^2)P(N=n) \\ &= E(X_1^2)E(N) + \mu^2(E(N^2) - E(N))\ . \end{aligned}$$

Zusammen erhält man

$$\begin{aligned} V(Z) &= E(Z^2) - (E(Z))^2 \\ &= E(X_1^2)E(N) + \mu^2(E(N^2) - E(N)) - \mu^2(E(N))^2 \\ &= \sigma^2 E(N) + \mu^2 V(N) \qquad \square \end{aligned}$$

Beispiel 3.81. Über die Entwicklung einer Population treffen wir die folgende Voraussetzungen

(i) Die 0-Generation besteht aus genau einem Individuum.

(ii) Die Anzahl der Nachkommen jedes Individuums ist $\pi(\lambda)$-verteilt.

(iii) Die Individuen vermehren sich unabhängig voneinander.

Wir bestimmen den Erwartungswert der Anzahl der Individuen der n-ten Generation (=Menge aller direkten Nachkommen von Individuen der $(n-1)$-ten Generation) und die Wahrscheinlichkeit dafür, dass die Population ausstirbt.

Es sei N_i die Anzahl der Individuen der i-ten Generation und X_{ij} die Anzahl der Nachkommen des j-ten Individuums der i-ten Generation. Dann gilt

(i) X_{ij} ist $\pi(\lambda)$-verteilt

(ii) für alle $i \in \mathbb{N}$ sind $N_i, X_{i1}, \ldots, X_{iN_i}$ unabhängig

(iii) $N_0 = 1, N_1 = X_{01}, N_2 = X_{11} + \ldots + X_{1N_1}, \ldots,$
$N_n = X_{n-1,1} + \ldots + X_{n-1,N_{n-1}}$

Nach dem letzten Beispiel gilt

$$E(N_n) = E(X_{n-1,1} + \ldots + X_{n-1,N_{n-1}}) = E(X_{n-1,1})E(N_{n-1}) = \lambda E(N_{n-1}),$$

und durch vollständige Induktion erhält man hieraus

$$E(N_n) = \lambda^n.$$

Es sei nun A das Ereignis, dass die Population ausstirbt. Wegen $[N_1 = 0] \subset [N_2 = 0] \subset \ldots$ und der Stetigkeit von Wahrscheinlichkeitsmaßen (vgl. Lemma 2.7.) gilt

$$P(A) = P\left(\bigcup_{n=1}^{\infty} [n_n = 0]\right) = \lim_{n\to\infty} P([N_n = 0]) \ .$$

Nach der Formel für die totale Wahrscheinlichkeit gilt

$$\begin{aligned}
P([N_n = 0]) &= P(X_{n-1,1} + \ldots + X_{n-1,N_{n-1}} = 0) \\
&= \sum_{k=0}^{\infty} P(X_{n-1,1} + \ldots + X_{n-1,N_{n-1}} = 0 | N_{n-1} = k) P(N_{n-1} = k) \\
&= \sum_{k=0}^{\infty} P(X_{n-1,1} = 0) \cdot \ldots \cdot P(X_{n-1,k} = 0) P(N_{n-1} = k) \\
&= \sum_{k=0}^{\infty} e^{-\lambda k} P(N_{n-1} = k) = E(\exp(-\lambda N_{n-1})).
\end{aligned}$$

Da für eine $\pi(\alpha)$-verteilte Zufallsvariable X und $\alpha > 0$ stets

$$\begin{aligned}
E(e^{-\alpha X}) &= \sum_{j=0}^{\infty} e^{-\alpha j} P(X = j) = \sum_{j=0}^{\infty} e^{-\alpha j} e^{-\lambda} \cdot \frac{\lambda^j}{j!} \\
&= e^{-\lambda} \sum_{j=0}^{\infty} \frac{(e^{-\alpha} \cdot \lambda)^j}{j!} \;=\; \exp(-\lambda(1 - e^{-\alpha}))
\end{aligned}$$

gilt, hat man wegen der der Unabhängigkeit der $X_{n-1,j}$ für alle $\alpha > 0$ und für alle $n \in \mathbb{N}$

$$\begin{aligned}
&E(\exp(-\alpha N_n)) = E(\exp(-\alpha(X_{n-1,1} + X_{n-1,2} + \ldots + X_{n-1,N_{n-1}}))) \\
&= \sum_{k=0}^{\infty} E(\exp(-\alpha(X_{n-1,1} + \ldots + X_{n-1,N_{n-1}})) | N_{n-1} = k) P(N_{n-1} = k) \\
&= \sum_{k=0}^{\infty} E(\exp(-\alpha X_{n-1,1})) \cdot \ldots \cdot E(\exp(-\alpha X_{n-1,k})) P(N_{n-1} = k) \\
&= \sum_{k=0}^{\infty} \exp(-\lambda(1 - e^{-\alpha})k) P(N_{n-1} = k) \\
&= E(\exp(-\lambda(1 - e^{-\alpha}) N_{n-1})). \qquad (3.11)
\end{aligned}$$

Damit gilt alle $n \geq 2$

$$P(N_{n-1} = 0) = E(\exp(-\lambda N_{n-2})) \qquad (3.12)$$

und

$$P(N_n = 0) = E(\exp(-\lambda N_{n-1})) = E(\exp(-\lambda(1 - e^{-\lambda}) N_{n-2})). \qquad (3.13)$$

Durch vollständige Induktion erhält man für alle $\alpha > 0$ und alle $n \in \mathbb{N}_0$

$$E(\exp(-\lambda(1 - e^{-\alpha}) N_n)) = \exp(-\lambda(1 - E(\exp(-\alpha N_n)))), \qquad (3.14)$$

denn für $n = 0$ besagt (3.14) gerade wegen $N_0 = 1$

$$E(\exp(-\lambda(1 - e^{-\alpha}) N_0)) = \exp(-\lambda(1 - e^{-\alpha})) = \exp(-\lambda(1 - E(\exp(-\alpha N_0)))),$$

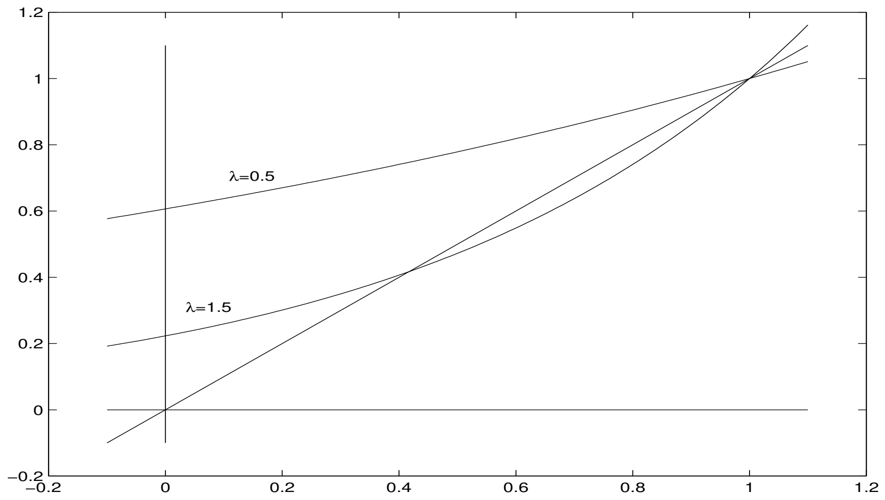

Abbildung 3.6: Zu Beispiel 3.81.

und ist (3.14) schon für $n-1$ gezeigt, so folgt aus (3.11) mit $\beta := \lambda(1-e^{-\alpha})$

$$\begin{aligned} E(\exp(-\lambda(1-e^{-\alpha})N_n)) &= E(\exp(-\beta N_n)) = E(\exp(-\lambda(1-e^{-\beta})N_{n-1})) \\ &= \exp(-\lambda(1-E(\exp(-\beta N_{n-1})))) \\ &= \exp(-\lambda(1-E(\exp(-\alpha N_n)))). \end{aligned}$$

Damit erhält man schließlich unter Benutzung von (3.13) und (3.12)

$$\begin{aligned} P(N_n = 0) &= E(\exp(-\lambda(1-e^{-\lambda})N_{n-2})) \\ &= \exp(-\lambda(1-E(\exp(-\lambda N_{n-2})))) \\ &= \exp(-\lambda(1-P(N_{n-1}=0))). \end{aligned}$$

$P(A) = \lim_{n\to\infty} P(N_n = 0)$ ist also Fixpunkt der Funktion (vgl. Abbildung 3.6)

$$\phi(x) = \exp(-\lambda(1-x)) \ .$$

Ist $\lambda \le 1$, so hat ϕ in dem Intervall $[0,1]$ nur den Fixpunkt $x = 1$. Besitzt also jedes Individuum der Population im Mittel nur $\lambda \le 1$ Nachkommen, so stirbt die Population mit Wahrscheinlichkeit 1 aus.

Ist $\lambda > 1$, so hat ϕ neben $x = 1$ einen weiteren Fixpunkt $\hat{x}$. Für diesen gilt

$$P(N_1 = 0) = e^{-\lambda} \le \exp(-\lambda(1-\hat{x})) = \hat{x} \ ,$$

und ist bereits $P(N_{n-1} = 0) \le \hat{x}$ gezeigt, so folgt

$$P(N_n = 0) = \exp(-\lambda(1-P(N_{n-1}=0))) \le \exp(-\lambda(1-\hat{x})) = \hat{x} \ ,$$

d.h. $P(N_n = 0) \le \hat{x}$ für alle $n \in \mathbb{N}$, und $\hat{x}$ ist die Aussterbewahrscheinlichkeit. Abbildung 3.7 enthält die Aussterbewahrscheinlichkeit in Abhängigkeit von λ. $\square$

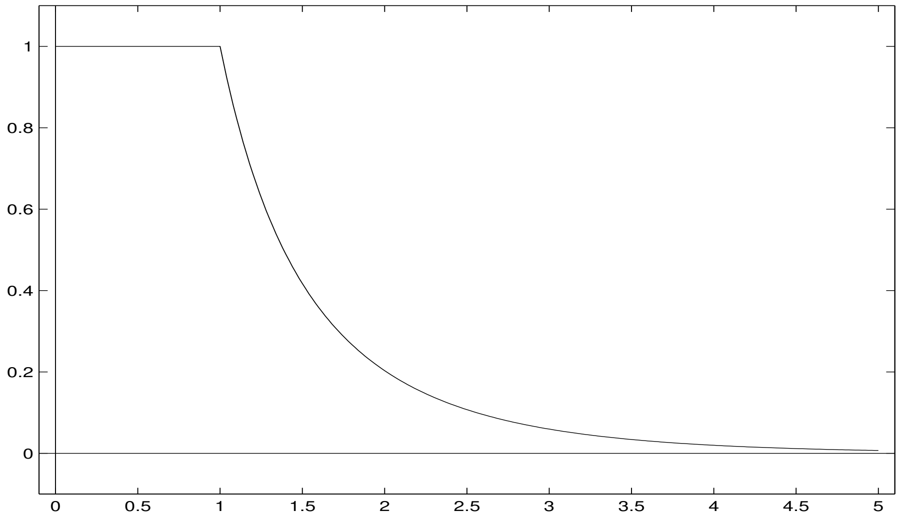

Abbildung 3.7: Aussterbewahrscheinlichkeit

Kapitel 4

Gesetz der großen Zahlen

4.1 Grenzwertsätze

Im Kapitel 1 haben wir den Begriff “Wahrscheinlichkeit” mit Hilfe der relativen Häufigkeit eingeführt. Wir haben bemerkt, dass “bei Wiederholung identischer Zufallsexperimente der Anteil der Erfolge unter den ersten n Versuchen gegen die Wahrscheinlichkeit des betrachteten Ereignisses konvergiert“. Wir wollen in diesem Abschnitt diesen Sachverhalt näher untersuchen (insbesondere den Konvergenzbegriff klären) und weitere Grenzwertsätze betrachten.

Ein nützliches Hilfsmittel bei diesen Untersuchungen ist:

Satz 4.1. *Es sei X eine reelle Zufallsvariable und $r > 0$. Dann gilt für alle $t > 0$*

$$P(|X| \geq t) \leq \frac{1}{t^r} E(|X|^r). \tag{4.1}$$

Besitzt X einen Erwartungswert $E(X)$, so gilt

$$P(|X - E(X)| \geq t) \leq \frac{1}{t^2} V(X). \tag{4.2}$$

Beweis: Wir zeigen die Aussagen nur für stetiges X, für das eine Dichte f existiert. Es gilt

$$\begin{aligned} P(|X| \geq t) &= \frac{1}{t^r} \left(\int_{-\infty}^{-t} + \int_{t}^{\infty} \right) t^r f(x)\, dx \leq \frac{1}{t^r} \left(\int_{-\infty}^{-t} + \int_{t}^{\infty} \right) |x|^r f(x)\, dx \\ &\leq \frac{1}{t^r} \int_{-\infty}^{\infty} |x|^r f(x)\, dx = \frac{1}{t^r} E(|X|^r)\ , \end{aligned}$$

also Ungleichung (4.1). Wenden wir diese Ungleichung auf die Zufallsvariable $Y := X - E(X)$ mit $r = 2$ an, so folgt (4.2)

$$P(|X - E(X)| \geq t) \leq \frac{E((X - E(X))^2)}{t^2} = \frac{V(X)}{t^2}. \qquad \blacksquare$$

Bemerkung 4.2. (4.1) heißt **Markovsche Ungleichung** , (4.2) heißt **Tschebyscheffsche Ungleichung** . □

Bemerkung 4.3. Eine andere, häufig benutzte Form der Tschebyscheffschen Ungleichung ist

$$P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2} \ ,$$

wobei σ die Standardabweichung von X bezeichnet. □

Bemerkung 4.4. Die größte Stärke der angegebenen Ungleichungen ist die allgemeine Abwendbarkeit (sie gelten für jede Verteilung). Diese Stärke ist zugleich ihre größte Schwäche, da die Schranke für die Wahrscheinlichkeit i.a. sehr grob ist. □

Beispiel 4.5. Es sei X (z.B. die Ausführungszeit eines Jobs in einem Rechenzentrum) exponentialverteilt mit $E(X) = 1/\alpha$ und $V(X) = 1/\alpha^2$. Wir fragen nach der Wahrscheinlichkeit dafür, dass $X \geq 2E(X)$ gilt.

Nach der Markovschen Ungleichung gilt wegen der Nichtnegativität von X (mit $r = 1$, $t = \frac{2}{\alpha}$)

$$P(X \geq \frac{2}{\alpha}) \leq \frac{1}{t} E(|X|) = \frac{1}{2} \ .$$

Die Tschebyscheffsche Ungleichung liefert

$$P(X \geq \frac{2}{\alpha}) = P(|X - \frac{1}{\alpha}| \geq \frac{1}{\alpha}) \leq \frac{1/\alpha^2}{1/\alpha^2} = 1 \ ,$$

also ein wertloses Resultat.

Tatsächlich gilt

$$P(X \geq \frac{2}{\alpha}) = 1 - F_X(\frac{2}{\alpha}) = e^{-2} \approx 0.135. \qquad \square$$

Beispiel 4.6. Es sei X Poisson-verteilt mit dem Parameter λ. Gesucht ist $P(X \geq 2\lambda)$.

Nach der Markovschen Ungleichung ist

$$P(X \geq 2\lambda) = P(|X| \geq 2\lambda) \leq \frac{E(|X|)}{2\lambda} = \frac{1}{2}.$$

Mit der Tschebyscheffschen Ungleichung erhält man

$$P(X \geq 2\lambda) = P(|X - E(X)| \geq \lambda) - P(X = 0) \leq \frac{V(X)}{\lambda^2} - e^{-\lambda}\frac{\lambda^0}{0!} = \frac{1}{\lambda} - e^{-\lambda}.$$

Tatsächlich gilt

$$P(X \geq 2\lambda) = e^{-\lambda} \sum_{j \geq 2\lambda} \frac{\lambda^j}{j!},$$

und damit

λ	$P(X \geq 2\lambda)$	Markov	Tschebyscheff
1	0.2642	0.5	0.6321
2	0.1429	0.5	0.3647
3	0.0839	0.5	0.2835
4	0.0511	0.5	0.2317
5	0.0318	0.5	0.1933

□

Die Ungleichungen lassen sich ohne zusätzliche Voraussetzungen nicht verbessern. Dies zeigt das folgende Beispiel.

Beispiel 4.7. Sei X eine diskrete Zufallsvariable mit den Werten $-1, 0, 1$ und der Zähldichte $f(1) = f(-1) = \frac{1}{8}, f(0) = \frac{3}{4}$. Dann gilt $E(X) = -1 \cdot \frac{1}{8} + 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{8} = 0$, $E(X^2) = 1 \cdot \frac{1}{8} + 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{8} = \frac{1}{4}$, und daher $V(X) = \frac{1}{4}$.

Nach der Tschebyscheffschen Ungleichung gilt

$$P(|X| \geq 1) = P(|X - E(X)| \geq 1) \leq V(X) = \frac{1}{4}\ ,$$

nach der Markovschen Ungleichung wegen $E(|X|^r) = \frac{1}{4}$ für alle $r > 0$

$$P(|X| \geq 1) \leq \frac{1}{4}\ .$$

Tatsächlich gilt

$$P(|X| \geq 1) = P([X = 1] \cup [X = -1]) = f(1) + f(-1) = \frac{1}{4}. \qquad \square$$

Bei vielen Anwendungen ist man (vgl. Beispiel 4.5.) an Wahrscheinlichkeiten des Typs $P(X > t)$ bzw. $P(X \leq t)$ interessiert. Hierfür gilt (vgl. **Cramer** [10, p. 256])

Satz 4.8. *Sei X eine reelle Zufallsvariable mit $E(X) = \mu$ und $V(X) = \sigma^2$. Dann gilt*

$$\begin{aligned} P(X \leq t) &\leq \frac{\sigma^2}{\sigma^2 + (t-\mu)^2} \quad \textit{für} \quad t < \mu\ , \\ P(X > t) &\leq \frac{\sigma^2}{\sigma^2 + (t-\mu)^2} \quad \textit{für} \quad t > \mu\ . \end{aligned}$$

In Beispiel 4.5. erhält man hiermit

$$P(X \geq \frac{2}{\alpha}) \leq \frac{1}{2}\ ,$$

in Beispiel 4.6.

$$\begin{aligned} P(X \geq 2\lambda) &= P(x > 2\lambda) + P(X = 2\lambda) \\ &\leq \frac{\lambda^2}{\lambda^2 + (2\lambda - \lambda)^2} + P(X = 2\lambda) = \frac{1}{2} + P(X = 2\lambda). \end{aligned}$$

Es sei A ein Ereignis, das mit der Wahrscheinlichkeit $p := P(A)$ eintrete. Wir führen eine Folge von unabhängigen Bernoulliexperimenten aus, wobei Erfolg das Eintreten von A bedeute. Es sei S_n die Anzahl der Erfolge unter den ersten n Versuchen. Dann ist S_n $b(n,p)$-verteilt mit $E(S_n) = np$, $V(S_n) = np(1-p)$.

Für den Mittelwert $\frac{1}{n}S_n$ (den relativen Anteil der Erfolge in den ersten n Versuchen) gilt dann

$$E\left(\frac{1}{n}S_n\right) = p, \quad V\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}V(S_n) = \frac{1}{n}p(1-p).$$

Daher folgt aus der Tschebyscheffschen Ungleichung für jedes $\varepsilon > 0$

$$P\left(\left|\frac{1}{n}S_n - P(A)\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\,P(A)(1 - P(A)). \tag{4.3}$$

(4.3) sagt aus, dass die Wahrscheinlichkeit dafür, dass die relative Häufigkeit und die Wahrscheinlichkeit eines Ereignisses A dem Betrage nach um mehr als ε voneinander abweichen, für jedes positive ε mit $n \to \infty$ gegen 0 strebt. Damit ist unsere Motivation des Wahrscheinlichkeitsbegriffs aus Kapitel 1 bestätigt.

Ist $X_1, X_2, \ldots$ eine Folge identisch verteilter, unabhängiger Zufallsvariablen mit endlicher Varianz, so bleibt (4.3) richtig, wenn man $S_n := \sum\limits_{i=1}^{n} X_i$ definiert und $P(A)$ durch $E(X_i) =: \mu$ und $P(A)(1-P(A))$ durch $V(X_i) := \sigma^2$ ersetzt. Es gilt also

$$P\left(\left|\frac{1}{n}S_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Man kann sogar zeigen (vgl **Kingman & Taylor** [20, p. 342]), dass auf die Existenz der Varianz der X_i verzichtet werden kann. Es gilt

Satz 4.9. (schwaches Gesetz der großen Zahlen von Chintschin)
Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $E(X_i) =: \mu < \infty$. Dann gilt

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) = 0$$

für jedes (feste) $\varepsilon > 0$.

Wir kehren nun zum Bernoulliexperiment zurück und fragen, wie groß die Anzahl n der Versuche wenigstens gewählt werden muss, um zu garantieren, dass bei gegebenem Wahrscheinlichkeitsniveau $1 - \delta$ der Fehler $\left|\frac{1}{n}S_n - P(A)\right|$ bei der Schätzung von $P(A)$ durch $\frac{1}{n}S_n$ kleiner als ε ist, d.h. bestimme $n \in \mathbb{N}$, so dass

$$P\left(\left|\frac{1}{n}S_n - P(A)\right| < \varepsilon\right) \geq 1 - \delta\ .$$

Da $p(1-p) \leq \frac{1}{4}$ für alle $p \in [0,1]$ gilt, folgt aus (4.3)

$$P\left(\left|\frac{1}{n}S_n - P(A)\right| < \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2},$$

woraus die gewünschte Ungleichung für $n \geq 1/(4\delta\varepsilon^2)$ folgt.

Beispiel 4.10. Wie groß sollte die Anzahl n der Bernoulliexperimente gewählt werden, um p durch $\frac{1}{n}S_n$ mit einer Toleranz von $\varepsilon = 0.01$ mit wenigstens der Wahrscheinlichkeit $1 - \delta = 0.95$ zu schätzen? Als Antwort erhalten wir nach unseren obigen Überlegungen $n \geq 1/(4 \cdot 0.05 \cdot 0.0001) = 50\,000$.

Zeigen die ersten 1000 Versuche, dass p ungefähr gleich 0.1 ist, so erhält man aus (4.3), dass vermutlich schon $n \geq 0.1 \cdot 0.9/(\delta\varepsilon^2) = 9000$ für die geforderten Schranken $\varepsilon = 0.01$ und $\delta = 0.05$ genügt. □

Man beachte, dass wir nicht die Konvergenz von $\frac{1}{n}S_n$ gegen $P(A)$ im Sinne der klassischen Analysis gezeigt haben, sondern den folgenden, wesentlich schwächeren Konvergenzbegriff verwendet haben.

Definition 4.11. *Eine Folge $\{Y_n\}$ von Zufallsvariablen heißt* **stochastisch konvergent** *(oder* **nach Wahrscheinlichkeit konvergent***) gegen eine Zufallsvariable Y, wenn gilt*

$$\lim_{n\to\infty} P(|Y_n - Y| \geq \varepsilon) = 0 \qquad \textit{für alle} \quad \varepsilon > 0\ .$$

Bemerkung 4.12. Anschaulich bedeutet die stochastische Konvergenz von Y_n gegen Y, dass für jedes $\varepsilon > 0$ die Wahrscheinlichkeitsmasse der Verteilung von $Y_n - Y$ mit wachsendem n in das Intervall um den Nullpunkt der Länge 2ε fließt. □

Definition 4.13. *Eine Folge $\{Y_n\}$ von Zufallsvariablen* **konvergiert fast sicher** *(oder* **mit Wahrscheinlichkeit 1***) gegen die Zufallsvariable Y, wenn die Menge $\{\omega : Y_n(\omega) \not\to Y(\omega)\}$ eine Nullmenge ist, d.h.*

$$P(\omega : \lim_{n\to\infty} |Y_n(\omega) - Y(\omega)| = 0) = 1 \ .$$

Bemerkung 4.14. Man kann zeigen (vgl. **Hinderer** [16, p. 226]), dass jede fast sicher gegen Y konvergente Folge von Zufallsvariablen auch stochastisch gegen Y konvergiert, dass es aber umgekehrt eine Folge von Zufallsvariablen gibt, die stochastisch aber nicht fast sicher konvergiert. Der Begriff der fast sicheren Konvergenz ist also stärker als der der stochastischen Konvergenz. □

Ferner konvergiert unter den Voraussetzungen von Satz 4.9. die Folge $\frac{1}{n}\sum_{i=1}^{n} X_i$ sogar fast sicher gegen μ (vgl. **Kingman & Taylor** [20, p. 344])

Satz 4.15. (starkes Gesetz der großen Zahlen von Kolmogorov)
Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $E(X_i) =: \mu < \infty$. Dann gilt

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

Wir kommen nun zu einem der wichtigsten Resultate der Stochastik, dem zentralen Grenzwertsatz. Durch ihn wird die besondere Bedeutung der Normalverteilung deutlich. In dem zentralen Grenzwertsatz wird ein weiterer Konvergenzbegriff verwendet.

Definition 4.16. *Es seien $X, X_1, X_2, \ldots$ reelle Zufallsvariable mit den Verteilungsfunktionen $F, F_1, F_2, \ldots$. Die Folge $\{X_n\}$* **konvergiert nach Verteilung** *gegen X, falls für jede Stetigkeitsstelle $t \in \mathbb{R}$ von F*

$$\lim_{n\to\infty} F_n(t) = F(t)$$

gilt.

Bemerkung 4.17. Es ist sinnvoll, die Konvergenz nur an den Stetigkeitsstellen von F zu fordern, denn sei X_n gleichverteilt in dem Intervall $\left[-\frac{1}{n}, \frac{1}{n}\right]$. Dann wird man von jedem vernünftigen Konvergenzbegriff fordern, dass für den Grenzwert X gilt $P(X=0)=1$, d.h.

$$F(t) = \begin{cases} 0 & , \quad t < 0 \\ 1 & , \quad t \geq 0 \end{cases}$$

Es gilt aber

$$\lim_{n\to\infty} F_n(0) = \begin{cases} 0 & , \quad t < 0 \\ 0.5 & , \quad t = 0 \\ 1 & , \quad t > 0 \end{cases} .$$

Dieses Beispiel zeigt auch, dass der punktweise Grenzwert von Verteilungsfunktionen keine Verteilungsfunktion zu sein braucht. □

Bemerkung 4.18. Konvergiert die Folge X_n stochastisch gegen X, so auch nach Verteilung (vgl. **Behnen & Neuhaus** [2, p. 358]). Die Konvergenz nach Verteilung ist also der schwächste der drei hier betrachteten Grenzwertbegriffe. □

Satz 4.19. (Zentraler Grenzwertsatz von Gnedenko)
Es seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte Zufallsvariable mit dem Erwartungswert $\mu = E(X_j)$ und der Varianz $\sigma^2 := V(X_j) > 0$. Es sei $S_n := \sum_{j=1}^{n} X_j$.

Dann gilt für alle $x < y$

$$\lim_{n\to\infty} P\left(x \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq y\right) = \Phi(y) - \Phi(x) ,$$

wobei Φ die Verteilungsfunktion der standardisierten Normalverteilung $N(0,1)$ bezeichnet.

Beweis: s. **Kingman & Taylor** [20, p. 348].

Bemerkung 4.20. Die betrachtete Zufallsvariable

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

ist wegen $E(S_n) = n\mu$ und $V(S_n) = n\sigma^2$ die Standardisierung von S_n. □

Bemerkung 4.21. Man beachte, dass Satz 4.19. für beliebige, also auch für diskrete Zufallsvariable gilt. □

Tabelle 4.1: $n = 10 \quad p = 0.5$

i	binomial $P(X \leq i)$	normal $P(0 \leq X \leq i)$	normal $P(X \leq i)$	*Satz* 4.23.
0	0.0010	0.0000	0.0008	0.0020
2	0.0547	0.0281	0.0289	0.0567
4	0.3770	0.2628	0.2635	0.3757
6	0.8281	0.7357	0.7365	0.8284
8	0.9893	0.9703	0.9711	0.9863
10	1.0000	0.9984	0.9992	0.9995

Tabelle 4.2: $n = 100 \quad p = 0.5$

i	binomial $P(X \leq i)$	normal $P(0 \leq X \leq i)$	normal $P(X \leq i)$	*Satz* 4.23.
35	0.0018	0.0013	0.0013	0.0019
40	0.0284	0.0228	0.0228	0.0287
45	0.1841	0.1587	0.1587	0.1841
50	0.5398	0.5000	0.5000	0.5398
55	0.8644	0.8413	0.8413	0.8643
60	0.9824	0.9772	0.9772	0.9821
65	0.9991	0.9987	0.9987	0.9990

Bemerkung 4.22. Wir haben in Satz 2.19. gesehen, dass die Binomialverteilung für große n durch die Poisson Verteilung approximiert wird. Satz 4.19. zeigt, dass die Binomialverteilung $b(n,p)$ auch durch die Normalverteilung $N(np, np(1-p))$ approximiert wird, denn ist X $b(n,p)$-verteilt, so gilt $X = \sum_{j=1}^{n} X_j$, wobei die X_j $b(1,p)$-verteilt und unabhängig sind. Nach Satz 4.19. konvergiert also die Verteilung von $Z_n = (X - np)/\sqrt{np(1-p)}$ gegen die $N(0,1)$-Verteilung.

Tabelle 4.1 bis Tabelle 4.4 geben für $n = 10$ und $n = 100$ sowie für $p = 0.05$ und $p = 0.5$ Aufschluss über die Approximationsgüte. □

Man sieht, dass die für binomialverteilte Zufallsvariable natürliche Bedingung $0 \leq X$ die Approximation durch die Normalverteilung nur verschlechtert.

Der zentrale Grenzwertsatz gilt für jede Verteilung von X_j. Man kann bessere Approximationen erhalten, wenn man ausnutzt, dass die X_j binomialverteilt sind. Der

Tabelle 4.3: $n = 10 \quad p = 0.05$

i	binomial $P(X \leq i)$	normal $P(0 \leq X \leq i)$	normal $P(X \leq i)$	*Satz* 4.23.
0	0.5987	0.0000	0.2341	0.4266
1	0.9139	0.5318	0.7659	0.8532
2	0.9885	0.7512	0.9852	0.9247
3	0.9990	0.7658	0.9999	0.9266

Tabelle 4.4: $n = 100 \quad p = 0.05$

i	binomial $P(X \leq i)$	normal $P(0 \leq X \leq i)$	normal $P(X \leq i)$	*Satz* 4.23.
0	0.0059	0.0000	0.0109	0.0137
2	0.1183	0.0734	0.0843	0.1199
4	0.4360	0.3123	0.3232	0.4035
6	0.7660	0.6659	0.6768	0.7485
8	0.9369	0.9048	0.9157	0.9400
10	0.9885	0.9782	0.9891	0.9884

folgende Satz wurde von **de Moivre** 1733 für den Fall $p = \frac{1}{2}$ und von **Laplace** 1812 für beliebiges p bewiesen.

Satz 4.23. (de Moivre, Laplace)
Es sei S_n $b(n,p)$-verteilt. Dann gilt für alle $a, b \in \mathbb{N}_0$, $a \leq b$

$$\lim_{n\to\infty} \left\{ P(a \leq S_n \leq b) - \Phi\left(\frac{b - np + 0.5}{\sqrt{np(1-p)}}\right) + \Phi\left(\frac{a - np - 0.5}{\sqrt{np(1-p)}}\right) \right\} = 0 \ ,$$

Bemerkung 4.24. Bis auf die "0.5-Terme" (die "Stetigkeitskorrekturen") folgt Satz 4.23. sofort aus dem zentralen Grenzwertsatz. Den Beweis findet man in **Parzen** [23, p. 239]. Wir haben in Tabelle 4.1 bis Tabelle 4.4 die Approximationen durch den Satz von de Moivre, Laplace bereits in der letzten Spalte angegeben. □

Wir kehren noch einmal zu der Frage zurück, wie groß die Anzahl n der Versuche bei einem Bernoulliexperiment gewählt werden muss, um die Wahrscheinlichkeit $p = P(A)$ mit einer vorgegebenen Genauigkeit zu schätzen. Wir bestimmen also bei gegebenen $\varepsilon > 0$ und $\delta > 0$ ein möglichst kleines $n \in \mathbb{N}$, so dass

$$P\left(\left|\frac{1}{n}S_n - p\right| \leq \varepsilon\right) \geq 1 - \delta \ ,$$

wobei S_n $b(n,p)$-verteilt ist.

Es gilt nach Satz 4.19.

$$\begin{aligned}
P\left(\left|\frac{1}{n}S_n - p\right| \leq \varepsilon\right) &= P\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq \frac{n\cdot\varepsilon}{\sqrt{np(1-p)}}\right) \\
&= P\left(-\frac{n\varepsilon}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{n\varepsilon}{\sqrt{np(1-p)}}\right) \\
&\approx \Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\right) \\
&= 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\right) - 1 \ .
\end{aligned}$$

Wegen der Monotonie von Φ und wegen $p(1-p) \leq 0.25$ erhalten wir das gewünschte n durch

$$\min\left\{n \in \mathbb{N} \,:\, 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{0.25}}\right) - 1 \geq 1-\delta\right\} = \min\{n \in \mathbb{N} \,:\, \Phi(2\varepsilon\sqrt{n}) \geq 1 - \frac{\delta}{2}\} \ .$$

Speziell für $\varepsilon = 0.01$, $\delta = 0.05$ erhält man $n \geq 9604$ (zum Vergleich: Die Tschebyscheffsche Ungleichung lieferte $n \geq 50\,000$). Nutzt man die Vorinformation $p \approx 0.1$ aus, so erhalten wir das gewünschte n durch

$$\min\left\{n \in \mathbb{N} \,:\, \Phi\left(\frac{\sqrt{n}\cdot\varepsilon}{\sqrt{0.1\cdot 0.9}}\right) \geq 1 - \frac{\delta}{2}\right\},$$

d.h. für $\varepsilon = 0.01$,$\delta = 0.05$ ergibt sich $n \geq 3458$ (mit der Tschebyscheffschen Ungleichung $n \geq 9000$).

Wir bemerken abschließend, dass man auf die identische Verteilung der X_i in dem zentralen Grenzwertsatz verzichten kann. Allgemeiner als Satz 4.19. gilt (vgl. **Moran** [22, p. 389]).

Satz 4.25. (Zentraler Grenzwertsatz von Ljapunov)
Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit $E(X_i) = \mu_i$ und $V(X_i) = \sigma_i^2$.

Es gelte mit $s_n^2 := \sum_{i=1}^n \sigma_i^2$ für ein $\delta > 0$

$$\sum_{i=1}^n \frac{E|X_i|^{2+\delta}}{s_n^{2+\delta}} \to 0 \quad \text{für } n \to \infty \ .$$

Dann gilt für alle $x < y$

$$\lim_{n\to\infty} P\left(x \leq \frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i) \leq x\right) = \Phi(y) - \Phi(x) \ .$$

Diesen Satz kann man so interpretieren: Jede Zufallsvariable, die als Summe einer großen Zahl von unabhängigen Zufallsvariablen dargestellt werden kann, ist normalverteilt.

4.2 Monte Carlo Methode

Eine wichtige Anwendung des zentralen Grenzwertsatzes ist die Begründung der Monte-Carlo-Methoden. Erste systematische Forschungen auf diesem Gebiet wurden während des zweiten Weltkrieges im Zusammenhang mit dem Bau der Atombombe ausgeführt und bezogen sich auf die Simulation von Zufallsprozessen bei der Diffusion von Neutronen in spaltbarem Material. Anschließend wurde von Fermi, von Neumann und Ulam gezeigt, dass die Methode auch auf deterministische Probleme anwendbar ist.

Wir beschränken uns in dieser Vorlesung auf die Berechnung von Integralen. Weitere Anwendungen (lineare Gleichungssysteme, Randwertprobleme, Extremalprobleme) findet man in **Hengartner & Theodorescu** [15].

Wir betrachten das Problem, das r-dimensionale Integral

$$I := \int_0^1 \int_0^1 \dots \int_0^1 g(x_1, x_2, \dots, x_r)\, dx_1 \dots dx_r$$

zu berechnen.

Sind $X_{11}, X_{12}, \dots, X_{1r}, X_{21}, X_{22}, \dots, X_{n-1,r}, X_{n1}, X_{n2}, \dots$ unabhängige, auf $[0,1]$ gleichverteilte Zufallsvariable, so sind die Zufallsvariablen

$$Y_1 := g(X_{11}, \dots, X_{1r}), Y_2 := g(X_{21}, \dots, X_{2r}), \dots$$

unabhängige und auf $[0,1]^r$ identisch verteilt mit dem Erwartungswert

$$E(Y_i) = \int_0^1 \dots \int_0^1 g(x_1, \dots, x_r)\, dx_1 \dots dx_r = I \ ,$$

und nach dem schwachen Gesetz der großen Zahlen gilt

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i - I\right| > \varepsilon\right) = 0 \quad \text{für alle} \quad \varepsilon > 0 \ .$$

Dies motiviert den folgenden Algorithmus zur approximativen Berechnung von I.

(i) Erzeuge eine Folge von auf $[0,1]$ gleichverteilten Zufallszahlen

$$x_{11}, x_{12}, \ldots, x_{1r}, x_{21}, \ldots, x_{n-1,r}, x_{n1}, x_{n2}, \ldots .$$

Algorithmen hierfür findet man in **Hengartner & Theodorescu** [15] oder in **Weiß** [28]. In MATLAB erhält man mit dem Aufruf rand ein Element einer Folge von in $[0,1]$ gleichverteilten Pseudo-Zufallszahlen.

(ii) Berechne für $i = 1, 2, \ldots, n$

$$y_i := g(x_{i1}, \ldots, x_{ir})$$

(iii) Berechne als Näherung für I

$$\bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i \ .$$

Die erforderliche Größe von n, um eine vorgegebene Genauigkeit zu erreichen, kann man wieder mit Hilfe des zentralen Grenzwertsatzes schätzen.

Beispiel 4.26.

$$I = \int_0^1 \frac{x}{\sqrt[4]{1-x^2}}\, dx \ .$$

Es ist

$$\begin{aligned} E(Y_i) &= I = \int_0^1 \frac{x}{\sqrt[4]{1-x^2}}\, dx = \left[-\frac{2}{3} \left(\sqrt[4]{1-x^2} \right)^3 \right]_0^1 = \frac{2}{3} \ , \\ V(Y_i) &= \int_0^1 \frac{x^2}{\sqrt{1-x^2}}\, dx - I^2 \le 0.36 \end{aligned}$$

(eine obere Schranke für das zweite Integral wurde mit der Monte Carlo Methode ermittelt).

Um zu sichern, dass

$$P\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i - \int_0^1 \frac{x}{\sqrt[4]{1-x^2}}\, dx \right| \le 0.01 \right) \ge 0.95$$

gilt, muss nach dem zentralen Grenzwertsatz wie oben

$$\Phi\left(\frac{0.01\sqrt{n}}{\sqrt{0.36}} \right) \ge 0.975$$

Tabelle 4.5: Beispiel 4.26.

0.0051	0.0039	0.0041	0.0047	0.0060	0.0027	0.0052	0.0017	0.0006	0.0091
0.0099	0.0033	0.0065	0.0008	0.0037	0.0030	0.0047	0.0132	0.0051	0.0032
0.0020	0.0040	0.0032	0.0025	0.0018	0.0096	0.0008	0.0017	0.0005	0.0025
0.0006	0.0110	0.0061	0.0025	0.0056	0.0038	0.0027	0.0049	0.0089	0.0065
0.0027	0.0019	0.0019	0.0046	0.0005	0.0119	0.0005	0.0001	0.0020	0.0008
0.0004	0.0081	0.0032	0.0027	0.0063	0.0021	0.0054	0.0037	0.0047	0.0028
0.0018	0.0005	0.0008	0.0023	0.0060	0.0012	0.0019	0.0050	0.0005	0.0115
0.0031	0.0081	0.0003	0.0081	0.0005	0.0016	0.0020	0.0097	0.0062	0.0066
0.0044	0.0082	0.0047	0.0080	0.0032	0.0007	0.0008	0.0105	0.0013	0.0006
0.0041	0.0001	0.0021	0.0068	0.0007	0.0023	0.0055	0.0038	0.0071	0.0020

Tabelle 4.6: Beispiel 4.26.

0.0066	0.0032	0.0012	0.0010	0.0033	0.0005	0.0025	0.0015	0.0001	0.0018
0.0012	0.0033	0.0041	0.0007	0.0045	0.0045	0.0096	0.0044	0.0120	0.0042
0.0016	0.0026	0.0015	0.0021	0.0049	0.0008	0.0047	0.0021	0.0046	0.0056
0.0005	0.0044	0.0039	0.0085	0.0028	0.0021	0.0064	0.0020	0.0025	0.0065
0.0010	0.0003	0.0041	0.0036	0.0018	0.0016	0.0034	0.0017	0.0047	0.0024
0.0016	0.0046	0.0014	0.0000	0.0021	0.0034	0.0059	0.0046	0.0019	0.0105
0.0035	0.0060	0.0018	0.0037	0.0017	0.0066	0.0039	0.0012	0.0071	0.0018
0.0030	0.0029	0.0029	0.0014	0.0059	0.0035	0.0028	0.0131	0.0006	0.0002
0.0022	0.0026	0.0068	0.0018	0.0056	0.0050	0.0010	0.0012	0.0111	0.0103
0.0004	0.0015	0.0030	0.0035	0.0008	0.0030	0.0062	0.0017	0.0000	0.0031

gelten, und dies ist für $n \geq 13830$ erfüllt.

Tabelle 4.5 enthält die Fehler der Monte Carlo Methode mit jeweils 13830 Funktionsauswertungen für 100 Testläufe. In 5 von den 100 Fällen wurde die vorgegebene Fehlerschranke 0.01 überschritten ($\delta = 0.05!$). Diese Versuche sind in Tabelle 4.5 unterstrichen. $\square$

Einfache Integrale werden in der Regel nicht mit Monte-Carlo-Methoden behandelt, da hierfür wesentliche bessere Quadraturformeln zur Verfügung stehen (vgl. Grundlagen der Numerischen Mathematik). Bei mehrfachen Integralen werden die deterministischen numerischen Verfahren mit wachsender Dimension sehr kompliziert, während die Monte-Carlo-Methoden sehr einfach bleiben.

Beispiel 4.27.

$$I = \int_0^1 \int_0^1 \int_0^1 f(x,y,z)\,dx\,dy\,dz := \int_0^1 \int_0^1 \int_0^1 \sin(\frac{1}{2}(x^2+y^2+z))\,dx\,dy\,dz \ .$$

Mit jeweils 100 Funktionsauswertungen wurde die untere Schranke 0.52 für I und die obere Schranke 0.34 für

$$\int_0^1 \int_0^1 \int_0^1 f^2(x,y,z)\,dx\,dy\,dz \ ,$$

und damit $V(Y_i) \leq 0.07$ berechnet.

Hiermit erhält man wie eben, dass

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - I\right| \leq 0.01\right) \geq 0.95$$

für $n \geq 2690$ erfüllt wird.

Tabelle 4.6 enthält die Abweichungen der Monte-Carlo-Methode in 100 Testläufen mit jeweils 2690 Funktionsauswertungen von dem Mittelwert 0.532781 mit 269000 Funktionsauswertungen. □

Kapitel 5

Stochastische Prozesse

5.1 Einleitende Beispiele

Wir haben bisher (bis auf wenige Ausnahmen) Zufallsvariable oder Zufallsvektoren betrachtet und hiermit Zufallsexperimente beschrieben. Wir untersuchen nun das Verhalten von Zufallsvariablen in Abhängigkeit von einem oder mehreren Parametern. Auf diese Weise kann man die zeitliche Entwicklung oder die räumliche Struktur eines zufälligen Systems beschreiben.

Definition 5.1. *Ein* **stochastischer Prozess** *ist eine Abbildung* $X(\omega, t)$ *von* $\Omega \times I$, $I \subset \mathbb{R}$, *in die reellen Zahlen, wobei für jeden Parameterwert* $t \in I$ $X(\cdot, t) : \Omega \to \mathbb{R}$ *eine Zufallsvariable auf einem festen Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *ist.* I *heißt die* **Parametermenge**, Ω *der* **Zustandsraum**.

Die Abbildungen $X(\omega, \cdot) : I \to \mathbb{R}$ *heißen* **Pfade** *oder Trajektorien des stochastischen Prozesses.*

Wir werden von nun an einen stochastischen Prozess beschreiben durch $\{X(t) : t \in I\}$, wobei jedes $X(t) : \Omega \to \mathbb{R}$ eine Zufallsvariable ist.

Beispiel 5.2. Die Entwicklung einer Population in einer Nährlösung (vgl. Beispiel 3.81.). Für $n \in \mathbb{N}$ bezeichnet N_n die Anzahl der Individuen der n-ten Generation. Die Parametermenge $I = \mathbb{N}_0$ ist also abzählbar, der Zustandsraum $\Omega = \mathbb{N}_0$ ebenfalls. $\square$

Beispiel 5.3. Brownsche Bewegung: 1826 beobachtete der englische Botaniker Brown die Bewegung eines kleinen Teilchens in einer Flüssigkeit. Das Teilchen führt eine ungeordnete Zickzack-Bewegung aus, die durch Zusammenstöße des Teilchens mit den Molekülen der Flüssigkeit hervorgerufen wird.

Bezeichnet $(X(t), Y(t), Z(t))$ die Koordinaten des Teilchens, so lassen sich (wegen der Vielzahl der Flüssigkeitsmolküle) diese am leichtesten als Zufallsvariablen auffassen. In diesem Fall sind der Parameterraum $I = \mathbb{R}_+$ und auch der Zustandsraum $\Omega = \mathbb{R}^3$ überabzählbar. □

Beispiel 5.4. Beobachtet man die Lage des Teilchens in Beispiel 5.3. nur in festen Zeitabständen Δt, so erhält man einen stochastischen Prozess

$$(X(n \cdot \Delta t), Y(n \cdot \Delta t), Z(n \cdot \Delta t))$$

mit abzählbarer Parametermenge $I = \mathbb{N}_0$ und überabzählbarem Zustandsraum $\Omega = \mathbb{R}^3$. □

Beispiel 5.5. Es sei $X(t)$ die Anzahl der in einer Telefonzentrale bis zum Zeitpunkt $t > 0$ eintreffenden Anrufe. Dann ist die Parametermenge $I = \mathbb{R}_+$ überabzählbar und der Zustandsraum $\Omega = \mathbb{N}_0$ abzählbar. □

Entsprechend den obigen Beispielen nimmt man bei stochastischen Prozessen die folgende Klassifikation vor.

Definition 5.6. *Ein stochastischer Prozess heißt* **diskret** , *wenn der Zustandsraum diskret ist (Beispiel 5.2. und Beispiel 5.5.), sonst nennt man ihn* **kontinuierlich**. *Ist die Parametermenge I abzählbar, so spricht man von einer (zufälligen)* **Kette**. *(Beispiel 5.2. und Beispiel 5.4.).*

Um das zufällige Verhalten einer Zufallsvariable vollständig zu beschreiben, benötigt man die Verteilungsfunktion, für einen Zufallsvektor die gemeinsame Verteilungsfunktion. Im Falle eines stochastischen Prozesses genügt die Menge aller endlichdimensionalen gemeinsamen Verteilungsfunktionen.

Definition 5.7. *Sei $\{X(t) : t \in I\}$ ein stochastischer Prozess. Dann heißt die Gesamtheit aller endlichdimensionalen Verteilungsfunktionen*

$$F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) := P(X(t_1) \le x_1, \ldots, X(t_n) \le x_n)$$

mit beliebigen, paarweise verschiedenen $t_1, \ldots, t_n \in I$ und beliebigem endlichen n die dem Prozess $\{X(t) : t \in I\}$ **zugeordnete Verteilung** .

Stochastische Prozesse mit gleichen zugeordneten Verteilungen stimmen zwar nicht überein, unterscheiden sich aber so wenig, dass man die Unterschiede unberücksichtigt lassen kann. Man nennt sie äquivalente stochastische Prozesse.

Beispiel 5.8. Es sei $\Omega := \mathbb{R}$, $I = \mathbb{R}_+$ und P stetiges Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B})$. Dann haben die stochastischen Prozesse

$$\{X(t), t \in \mathbb{R}_+\},\ X(\omega, t) = 0 \quad \text{für alle } t \in \mathbb{R}_+ \text{ und } \omega \in \mathbb{R} =: \Omega,$$

und

$$\{Y(t) \,:\, t \in \mathbb{R}_+\}, \quad Y(\omega, t) = \begin{cases} 1 & \text{für} \quad t = \omega \\ 0 & \text{sonst} \end{cases},$$

beide die Verteilung

$$F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \begin{cases} 1 & \text{falls} \quad x_j \geq 0, j = 1, \ldots, n \\ 0 & \text{sonst} \end{cases},$$

sind also äquivalent.

Für $X(t)$ ist dies klar, für $Y(t)$ sieht man das so ein. Es ist für jedes $x_j \in \mathbb{R}$ und $t_j \in \mathbb{R}_+$

$$\{\omega \,:\, Y(\omega, t_j) \leq x_j\} = \begin{cases} \emptyset & \text{falls} \quad x_j < 0 \\ \mathbb{R} \setminus \{t_j\} & \text{falls} \quad 0 \leq x_j < 1 \\ \mathbb{R} & \text{falls} \quad 1 \leq x_j, \end{cases}$$

und daher

$$\{\omega \,:\, Y(\omega, t_j) \leq x_j, j = 1, \ldots, n\} = \begin{cases} \supset \mathbb{R} \setminus \bigcup_{j=1}^{n} \{t_j\} & \text{falls} \quad 0 \leq x_j\,,\ j = 1, \ldots, n \\ \emptyset & \text{sonst} \end{cases}.$$

Die Trajektorien

$$X(\omega, t) = 0 \quad \text{für alle } t$$

und

$$Y(\omega, t) = \begin{cases} 0 & \text{für} \quad t \neq \omega \\ 1 & \text{für} \quad t = \omega \end{cases}$$

unterscheiden sich für jedes feste ω nur in einem Punkt. □

Im allgemeinen Fall (wenn also keine zusätzliche Struktur bekannt ist) ist es unmöglich, alle endlichdimensionalen Verteilungen eines Prozesses anzugeben. Man wird sich dann auf die Angabe von Parametern beschränken, ähnlich wie wir zur Charakterisierung von Zufallsvektoren den Erwartungswert, die Varianz und die Kovarianzmatrix benutzt haben. Diese Begriffe lassen sich sofort auf stochastische Prozesse übertragen.

Definition 5.9. *Ist $\{X(t) : t \in I\}$ ein stochastischer Prozess und existiert für jedes $t \in I$ der Erwartungswert $\mu(t)$ von $X(t)$, so heißt die Funktion*

$$\mu : I \to \mathbb{R}, \quad \mu(t) := E[X(t)]$$

die **Mittelwertfunktion** *des Prozesses.*

Existiert für jedes $t \in I$ die Varianz $v(t)$ von $X(t)$, so heißt die Funktion

$$v : I \to \mathbb{R}, \quad v(t) := V[X(t)] = E[(X(t) - \mu(t))^2],$$

die **Varianzfunktion** *des Prozesses.*

Auch wenn $\mu(t)$ und $v(t)$ existieren, reichen sie nicht einmal für eine grobe Beschreibung des Prozesses aus, da die Abhängigkeiten der $X(t_j)$, $t_j \in I$, nicht erfasst werden. Ein Maß für die Abhängigkeit von Zufallsvariablen war in Kapitel 3 die Kovarianz bzw. Korrelation. Die folgende Definition ist nicht ganz konsequent in der Namensgebung.

Definition 5.10. *Es sei $\{X(t) : t \in I\}$ ein stochastischer Prozess. Dann heißt die Funktion*

$$k : I \times I \to \mathbb{R}, \quad k(t,s) := \mathrm{Cov}(X(t), X(s)) = E[(X(t) - \mu(t))(X(s) - \mu(s))],$$

die **Autokorrelationsfunktion** *(oder* **Korrelationsfunktion***) und*

$$r : I \times I \to \mathbb{R}, \quad r(t,s) := \frac{k(t,s)}{\sqrt{v(t)v(s)}}$$

die **normierte Korrelationsfunktion** *des Prozesses.*

Es ist klar, dass die Autokorrelation symmetrisch ist ($k(s,t) = k(t,s)$ für alle $s, t \in I$) und dass $k(t,t) = v(t)$ für alle $t \in I$ gilt.

Beispiel 5.11. Viele stochastische Prozesse ergeben sich als Funktionen von endlich vielen Zufallsvariablen. Es sei $X(t)$ die Lage eines Teilchens, das sich mit konstanter Geschwindigkeit bewegt. Dann kann man $X(t)$ in der Gestalt annehmen:

$$X(t) = X_0 + V \cdot t,$$

wobei X_0 und V Zufallsvariable sind, die die Anfangslage X_0 und die Geschwindigkeit V repräsentieren. Dann gilt für die Mittelwertfunktion und für die Autokorrelationsfunktion

$$\begin{aligned}\mu(t) &= E(X_0) + tE(V), \\ k(s,t) &= E[(X_0 - E(X_0) + t(V - E(V)))(X_0 - E(X_0) + s(V - E(V)))] \\ &= V(X_0) + (s+t)\mathrm{Cov}(X_0, V) + stV(V).\end{aligned}$$

Man benötigt also nicht die gemeinsame Verteilung von X_0 und V, sondern nur ihre Mittelwerte, ihre Varianzen und ihre Kovarianz. □

Sind $\{X(t) : t \in I\}$ und $\{Y(t) : t \in I\}$ zwei stochastische Prozesse, die sich nur um eine additive Funktion unterscheiden, d.h. $Y(t) = X(t) + f(t)$, so besitzen sie dieselbe Autokorrelationsfunktion, denn für die Mittelwertfunktion gilt $\mu_Y(t) = \mu_X(t) + f(t)$, und daher

$$\begin{aligned}k_Y(t,s) &= E[(Y(t) - \mu_Y(t))(Y(s) - \mu_Y(s))] \\ &= E[(X(t) + f(t) - (\mu_X(t) + f(t)))(X(s) + f(s) - (\mu_X(s) + f(s)))] \\ &= E[(X(t) - \mu_X(t))(X(s) - \mu_X(s))] = k_X(t,s).\end{aligned}$$

Wegen dieses Ergebnisses kann man ohne Beschränkung der Allgemeinheit beim Operieren mit stochastischen Prozessen annehmen, dass die Mittelwertfunktion identisch gleich null ist.

5.2 Poisson Prozesse und Erneuerungsprozesse

Wir wollen nun zwei wichtige stochastische Prozesse, die wir bereits in Beispiel 5.3. und Beispiel 5.5. am Anfang dieses Abschnitts angesprochen haben, genauer beschreiben.

Sei $X(t)$ die Anzahl der in dem Intervall $(0, t]$, $t > 0$, eintreffenden Anrufe in einer Telefonzentrale (bzw. bei der Anode einer Elektronenröhre eintreffenden α-Teilchen). Dann ist die Annahme, dass die Anzahl der in einem Intervall $(t, s]$ $(s > t)$ eingehenden Anrufe nicht durch die Anzahl $X(t)$ der in $(0, t]$ eingehenden Anrufe beeinflusst wird, plausibel. Allgemeiner können wir annehmen, dass $X(t)$ die folgende Eigenschaft hat:

Definition 5.12. *Der stochastische Prozess* $\{X(t) : t \in I\}$ ($I = \mathbb{R}$ *oder* $I = \mathbb{N}$) *heißt* **Prozess mit unabhängigen Zuwächsen** *, wenn für alle* $n \in \mathbb{N}$ *und für jede Wahl von* $t_0, \ldots, t_n \in I$ *mit* $0 \leq t_0 < t_1 < \ldots < t_n$ *die Zuwächse* $\Delta X_i := X(t_i) - X(t_{i-1})$, $i = 1, \ldots, n$, *stochastisch unabhängig sind.*

Ferner ist die Annahme plausibel, dass die Anzahl der in einem Intervall $(s, t]$ eingehenden Anrufe nicht von der Lage des Intervalls abhängt, sondern nur von seiner Länge $t - s$. $X(t)$ hat also die Eigenschaft

Definition 5.13. *Der stochastische Prozess* $\{X(t) : t \in I\}$ *hat* **stationäre Zuwächse** *, wenn für alle* $t, s, t+h, s+h \in I$, $s < t$, *die Zufallsvariablen* $X(t) - X(s)$ *und* $X(t+h) - X(s+h)$ *identisch verteilt sind.*

Wir haben bereits gesehen, dass die Wartezeit zwischen zwei Anrufen exponentialverteilt ist und dass die Wartezeit S_n bis zum n-ten Anruf Erlang-verteilt ist.

Daher folgt aus

$$\begin{aligned} P(S_n \leq t) &= P([S_n \leq t, S_{n+1} > t] \cup [S_{n+1} \leq t]) \\ &= P(X(t) = n) + P(S_{n+1} \leq t) \end{aligned}$$

für die Zähldichte von $X(t)$, $t > 0$,

$$\begin{aligned} P(X(t) = n) &= P(S_n \leq t) - P(S_{n+1} \leq t) \\ &= 1 - \exp(-\alpha t) \sum_{k=0}^{n-1} \frac{(\alpha t)^k}{k!} - \left(1 - \exp(-\alpha t) \sum_{k=0}^{n} \frac{(\alpha t)^k}{k!}\right) \\ &= \exp(-\alpha t) \cdot \frac{(\alpha t)^n}{n!}, \end{aligned}$$

d.h. $X(t)$ ist Poisson-verteilt mit Parameter αt.

Schließlich ist klar, dass $P(X(0) = 0) = 1$ gilt.

Definition 5.14. *Ein diskreter stochastischer Prozess* $\{X(t) : t \in \mathbb{R}_+\}$ *mit Werten in* $\mathbb{N}_0$ *heißt* **Poisson Prozess** *mit der Intensität* α, *falls gilt*

(i) $P(X(0) = 0) = 1$

(ii) $\{X(t) : t \in \mathbb{R}_+\}$ *hat unabhängige Zuwächse*

(iii) $\{X(t) : t \in \mathbb{R}_+\}$ *hat stationäre Zuwächse*

(iv) $X(t)$ *ist Poisson-verteilt mit Parameter* αt *für alle* $t > 0$.

Die Intensität α gibt die durchschnittliche Anzahl der eintretenden Ereignisse (Anrufe) pro Zeiteinheit an.

Für die Parameter eines Poisson Prozesses gilt wegen (i) und (iv)

$$\begin{aligned} \mu(t) &= E(X(t)) = \alpha t \\ v(t) &= V(X(t)) = \alpha t. \end{aligned}$$

Zur Berechnung der Autokorrelation betrachten wir $Y(t) := X(t) - \mu(t)$. Dann gilt für $t \geq s$

$$\begin{aligned} k_X(t,s) &= k_Y(t,s) = E[Y(t)Y(s)] \\ &= E[Y(s)(Y(s) + (Y(t) - Y(s)))] \\ &= E[Y^2(s)] + E[Y(s)(Y(t) - Y(s))]. \end{aligned}$$

Wegen $Y(0) = 0$, der Unabhängigkeit der Zuwächse und wegen $E[Y(s)] = 0$ gilt

$$\begin{aligned} E[Y(s)(Y(t) - Y(s))] &= E[(Y(s) - Y(0))(Y(t) - Y(s))] \\ &= E[Y(s) - Y(0)] \cdot E[Y(t) - Y(s)] = 0, \end{aligned}$$

und daher

$$k_X(t,s) = V(X(s)) = \lambda s.$$

Genauso gilt $k_X(t,s) = \lambda t$ für $s \geq t$, zusammengenommen also

$$k_X(t,s) = \lambda \min(t,s).$$

Für einen Poisson Prozess kann man die endlichdimensionalen Verteilungen leicht angeben. Für $t \geq s$ und $m_s, m_t \in \mathbb{N}_0$ gilt

$$P(X(t) = m_t, X(s) = m_s) = 0 \quad \text{für} \quad m_t < m_s$$

und wegen $X(0) = 0$, der Unabhängigkeit und der Stationarität der Zuwächse für $m_t \geq m_s$

$$\begin{aligned} P(X(t) = m_t, X(s) = m_s) &= P(X(s) - X(0) = m_s, X(t) - X(s) = m_t - m_s) \\ &= P(X(s) = m_s) \cdot P(X(t) - X(s) = m_t - m_s) \\ &= \exp(-\alpha s)\frac{(\alpha s)^{m_s}}{m_s!} \exp(-\alpha(t-s))\frac{(\alpha(t-s))^{m_t - m_s}}{(m_t - m_s)!}. \end{aligned}$$

Entsprechend erhält man die drei- und höher dimensionalen Verteilungen:

Es seien $0 < t_1 < t_2 < \ldots < t_n$ und $m_j \in \mathbb{N}_0$. Dann gilt

$$P(X(t_j) = m_j,\ j = 1, \ldots, n) = 0 \quad \text{falls nicht } m_j \leq m_{j+1},\ j = 1, \ldots, n-1,$$

und sonst (mit $t_0 = 0$ und $m_0 = 0$)

$$
\begin{aligned}
& P(X(t_j) = m_j,\ j = 1, \ldots, n) \\
&= P(X(t_1) - X(t_0) = m_1 - m_0, \ldots, X(t_n) - X(t_{n-1}) = m_n - m_{n-1}) \\
&= \prod_{j=1}^{n} P(X(t_j) - X(t_{j-1}) = m_j - m_{j-1}) \\
&= \prod_{j+1}^{n} \exp(-\alpha(t_j - t_{j-1})) \frac{(\alpha(t_j - t_{j-1}))^{m_j - m_{j-1}}}{(m_j - m_{j-1})!} \\
&= \exp(-\alpha t_n)\alpha^{m_n} \prod_{j=1}^{n} \frac{(t_j - t_{j-1})^{m_j - m_{j-1}}}{(m_j - m_{j-1})!}.
\end{aligned}
$$

Wir betrachten nun die Brownsche Bewegung (Beispiel 5.3.). Wir beschränken uns auf eindimensionale Bewegungen, die Verallgemeinerung auf den dreidimensionalen Fall ist evident.

Es sei $X(t)$ die Verschiebung eines Teilchens zu einem Zeitpunkt $t \geq 0$ aus seiner Ruhelage $X(0) = 0$. Da man die Verschiebung $X(t) - X(s)$ in einem Zeitintervall $(s, t]$ (das lang ist verglichen mit der Zeit zwischen zwei Zusammenstößen des Teilchens mit Flüssigkeitsmolekülen) als Summe einer sehr großen Zahl von kleinen Einzelverschiebungen betrachten kann, folgt aus dem zentralen Grenzwertsatz, dass die Zuwächse $X(t) - X(s)$ als normalverteilt angenommen werden können.

Da man die Zuwächse auch als stationär und unabhängig annehmen kann, lässt sich die Brownsche Bewegung mathematisch modellieren durch die folgende Klasse von Prozessen, die von N. Wiener (1923) untersucht und nach ihm benannt wurden:

Definition 5.15. *Ein kontinuierlicher stochastischer Prozess* $\{X(t) : t \in \mathbb{R}_+\}$ *mit Werten in* $\mathbb{R}$ *heißt (eindimensionaler)* **Wiener Prozess** *, wenn gilt*

(i) $P(X(0) = 0) = 1$

(ii) $\{X(t) : t \in \mathbb{R}_+\}$ *hat unabhängige, stationäre Zuwächse*

(iii) $X(t), t > 0$, *ist normalverteilt mit Erwartungswert* $E(X(t)) = 0$ *und Varianz* $V(X(t)) = \sigma^2 t$.

Ist $\sigma = 1$, *so heißt* $\{X(t) : t \in \mathbb{R}_+\}$ **standardisierter Wiener Prozess** .

Völlig analog den Betrachtungen für den Poisson Prozess erhält man für die Autokorrelation

$$k(t, s) = \sigma^2 \min(t, s).$$

Den Poisson Prozess kann man als mathematisches Modell des folgenden Vorgangs betrachten: Gegeben sei eine Folge gleichartiger Geräte, deren Lebensdauer Y_j exponentialverteilt sei mit Parameter α (keine Alterung!). Zum Zeitpunkt $t = 0$ beginnt das erste Gerät zu arbeiten. Bei seinem Ausfall übernimmt sofort das zweite Geräte die Arbeit, bei dessen Ausfall das dritte usw. Dann ist $S_n := \sum\limits_{j=1}^{n} Y_j$ die Lebensdauer bis zur n-ten Erneuerung, und $X(t)$, die Anzahl der Erneuerungen im Intervall $(0, t]$, ist ein Poisson Prozess mit der Intensität α.

Wir verzichten nun darauf, dass die Y_j exponentialverteilt sind.

Definition 5.16. *Sei $\{Y_j\}$ eine Folge von nichtnegativen, unabhängigen, identisch verteilten Zufallsvariablen. Dann heißt $\{Y_j\}$ ein* **Erneuerungsprozess**. *Die Verteilungsfunktion F der Y_j heißt* **Lebensdauerverteilung** *des Erneuerungsprozesses, und der diskrete Prozess*

$$N(t) := \max\left\{n \, : \, \sum_{j=1}^{n} Y_j \le t\right\}$$

heißt der zugehörige **Zählprozess**.

Es sei $S_k := \sum\limits_{j=1}^{k} Y_j$ die Wartezeit bis zur k-ten Erneuerung. Dann besitzt S_k wegen der Unabhängigkeit der Y_j als Verteilungsfunktion die k-fache Faltung von F, d.h. ist f die Dichte von F, so gilt für die Verteilungsfunktion F_k von S_k wegen $f(x) = 0$ für $x < 0$ und $F_{k-1}(t - x) = 0$ für $x > t$

$$F_k(t) = \int\limits_0^t F_{k-1}(t - x) f(x) \, dx, k = 1, 2, \dots, \tag{5.1}$$

wobei

$$F_0(t) := \begin{cases} 1 & \text{für} \quad t \ge 0 \\ 0 & \text{für} \quad t < 0 \end{cases}$$

gesetzt ist.

Als Zähldichte von $N(t)$ erhält man

$$P(N(t) = n) = P(S_n \le t < S_{n+1}) = P(S_n \le t) - P(S_{n+1} \le t) = F_n(t) - F_{n+1}(t).$$

Wir nennen $M(t) := E(N(t))$, die durchschnittliche Anzahl der Erneuerungen im Intervall $(0,t]$, die **Erneuerungsfunktion** des Prozesses. Hierfür gilt

$$\begin{aligned} M(t) &= \sum_{n=0}^{\infty} nP(N(t)=n) = \sum_{n=0}^{\infty} n(F_n(t) - F_{n+1}(t)) \\ &= \sum_{n=1}^{\infty} F_n(t) = F(t) + \sum_{n=1}^{\infty} F_{n+1}(t), \end{aligned} \tag{5.2}$$

und wegen (5.1)

$$\begin{aligned} M(t) &= F(t) + \sum_{n=1}^{\infty} \int_0^t F_n(t-x) f(x)\, dx \\ &= F(t) + \int_0^t \left(\sum_{n=1}^{\infty} F_n(t-x) \right) f(x)\, dx \\ &= F(t) + \int_0^t M(t-x) f(x)\, dx, \end{aligned} \tag{5.3}$$

(5.3) heißt die **Erneuerungsgleichung** des Prozesses.

Auch die **Erneuerungsdichte**

$$m(t) := \frac{d}{dt} M(t)$$

genügt einer Erneuerungsgleichung. Durch Differentiation von (5.3) erhält man wegen $M(0) = 0$

$$m(t) = f(t) + \int_0^t m(t-x) f(x)\, dx. \tag{5.4}$$

Die Erneuerungsgleichung (5.4) kann man mit Hilfe der Laplace Transformation lösen. Da die Laplace Transformierte der Faltung zweier Funktionen gleich dem Produkt der Laplace Transformierten ist, folgt aus (5.4)

$$\mathcal{L}m(s) = \mathcal{L}f(s) + \mathcal{L}m(s)\mathcal{L}f(s),$$

d.h.

$$\mathcal{L}m(s) = \frac{\mathcal{L}f(s)}{1 - \mathcal{L}f(s)} \tag{5.5}$$

und

$$\mathcal{L}f(s) = \frac{\mathcal{L}m(s)}{1 + \mathcal{L}m(s)}.$$

Beispiel 5.17. Es seien die Zufallsvariablen Y_i $E(2,\alpha)$-verteilt (Man zählt in einem Poisson Prozess also nur jedes zweite Ereignis). Dann besitzt die Dichte

$$f(t) = \alpha^2 t \exp(-\alpha t)$$

die Laplace Transformierte (vgl. Mathematik für Elektrotechniker)

$$\mathcal{L}f(s) = \frac{\alpha^2}{(s+\alpha)^2}.$$

Gleichung (5.5) liefert

$$\mathcal{L}m(s) = \frac{\alpha^2}{s^2 + 2\alpha s} = \frac{\alpha}{2}\left(\frac{1}{s} - \frac{1}{s+2\alpha}\right)$$

mit der Originalfunktion

$$m(t) = \frac{\alpha}{2}(1 - \exp(-2\alpha t)), \quad t > 0.$$

Durch Integration erhält man

$$M(t) = \int_0^t m(x)\,dx = \frac{1}{4}(-1 + 2\alpha t + \exp(-2\alpha t)).$$

In diesem Beispiel sind die n-fachen Faltungen F_n von F für alle n bekannt. S_n ist $E(2n, \alpha)$-verteilt (vgl. Beispiel 3.50.):

$$f_n(t) := \frac{d}{dt}F_n(t) = \alpha\exp(-\alpha t)\frac{(\alpha t)^{2n-1}}{(2n-1)!}.$$

Daher kann man hier m ohne Benutzung der Laplace Transformation berechnen. Es gilt

$$\begin{aligned} m(t) &= \sum_{n=1}^{\infty} f_n(t) = \alpha\exp(-\alpha t)\sum_{n=1}^{\infty}\frac{(\alpha t)^{2n-1}}{(2n-1)!} \\ &= \alpha\exp(-\alpha t)\sinh(\alpha t) = \frac{\alpha}{2}\exp(-\alpha t)(\exp(\alpha t) - \exp(-\alpha t)) \\ &= \frac{\alpha}{2}(1 - \exp(-2\alpha t)). \qquad \square \end{aligned}$$

5.3 Markov Prozesse

Wir haben bereits erwähnt, dass es unmöglich ist, die Verteilung eines stochastischen Prozesses zu bestimmen, wenn nicht spezielle Strukturen vorhanden sind.

Der einfachste Fall liegt vor, wenn die Zufallsvariablen $X(t)$, $t \in I$, unabhängig sind, denn dann gilt für alle $t_1, \ldots, t_n \in I$, $t_i \neq t_j$, und für alle $x_1, \ldots, x_n \in \mathbb{R}$

$$F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \prod_{j=1}^{n} F_{X(t_j)}(x_j).$$

Dieser Fall bietet nichts Neues gegenüber Kapitel 3. Zudem ist die Annahme der Unabhängigkeit viel zu einschränkend für stochastische Prozesse der Praxis.

Eine allgemeinere Eigenschaft, die eine Vielzahl von wichtigen Prozessen besitzt, ist die Markov Eigenschaft. Man kann sie als das stochastische Analogon des folgenden Sachverhalts auffassen: Das zukünftige Verhalten eines isolierten dynamischen Systems von n Teilchen ist vollständig bestimmt durch die gegenwärtige Lage seiner Teilchen. Alle Informationen über die Vergangenheit sind überflüssig für die Vorhersage der Zukunft.

Definition 5.18. *Der stochastische Prozess* $\{X(t) : t \in I\}$, $I \subset \mathbb{R}$, *heißt* **Markov Prozess**, *falls für jede Folge von Parametern* $t_1 < t_2 < \ldots < t_n$ *und für alle* $x_1, \ldots, x_n \in \mathbb{R}$

$$P(X(t) \leq x \,|\, X(t_j) = x_j,\ j = 1, \ldots, n) = P(X(t) \leq x \,|\, X(t_n) = x_n)$$

für alle $t > t_n$ *und alle* $x \in \mathbb{R}$ *gilt.*

Ist der Parameterraum diskret, so heißt $\{X(t) : t \in I\}$ *eine* **Markov Kette**.

Es ist anschaulich klar, dass der Wiener Prozess, der Poisson Prozess (und allgemeiner der Zählprozess eines Erneuerungsprozesses) die Markov Eigenschaft besitzt. Für den Poisson Prozess folgt dies aus

Satz 5.19. *Jeder diskrete stochastische Prozess* $\{X(t) : t \in I\}$ *mit unabhängigen Zuwächsen und* $P(X(0) = 0) = 1$ *ist ein Markov Prozess.*

Beweis: Es gilt für $t_1 < t_2 < \ldots < t_n < t$ und für alle $x, x_j \in \mathbb{R}$ und $t_0 = 0$, $x_0 = 0$ (beachte $P(X(0) = 0) = 1$)

$$\begin{aligned}
&P(X(t) \leq x \,|\, X(t_j) = x_j, j = 1, \ldots, n)\\
&= \frac{P(X(t) \leq x, X(t_j) = x_j, j = 1, \ldots, n)}{P(X(t_j) = x_j, j = 1, \ldots, n)}\\
&= \frac{P(X(t) - X(t_n) \leq x - x_n, X(t_j) - X(t_{j-1}) = x_j - x_{j-1}, j = 1, \ldots, n)}{P(X(t_j) - X(t_{j-1}) = x_j - x_{j-1}, j = 1, \ldots, n)}\\
&= P(X(t) - X(t_n) \leq x - x_n),
\end{aligned}$$

und man erhält weiter

$$\begin{aligned}
&P(X(t) \leq x \mid X(t_j) = x_j, j = 1, \ldots, n) \\
=\ & P(X(t) - X(t_n) \leq x - x_n) \cdot \frac{P(X(t_n) - X(0) = x_n)}{P(X(t_n) - X(0) = x_n)} \\
=\ & \frac{P(X(t) - X(t_n) \leq x - x_n, X(t_n) - X(0) = x_n)}{P(X(t_n) - X(0) = x_n)} \\
=\ & \frac{P(X(t) \leq x, X(t_n) = x_n)}{P(X(t_n) = x_n)} \\
=\ & P(X(t) \leq x \mid X(t_n) = x_n). \qquad \blacksquare
\end{aligned}$$

Der Beweis lässt sich nicht auf den Fall eines kontinuierlichen stochastischen Prozesses übertragen, da dann die bedingten Wahrscheinlichkeiten sich nicht mehr so einfach schreiben lassen. Für den Wiener Prozess kann man aber (ähnlich wie beim Poisson Prozess) alle endlichdimensionalen gemeinsamen Dichten berechnen (dies sind mehrdimensionale Normalverteilungen), hiermit die bedingten Dichten ausrechnen und so die Gültigkeit der Markov Eigenschaft nachweisen (ein langwieriger, aber elementarer Weg).

Wir betrachten von nun an nur noch diskrete stochastische Prozesse. Als Zustandsraum nehmen wir dabei ohne Beschränkung der Allgemeinheit stets $\mathbb{N}_0$ (oder im endlichen Fall $\{1, \ldots, N\}$) an.

Es seien $t_1 < t_2 < \ldots < t_n$, $t_i \in I$. Dann gilt für die gemeinsame Verteilung von $(X(t_1), \ldots, X(t_n))$ für alle $i_1, \ldots, i_n \in \mathbb{N}_0$ (genauer durch vollständige Induktion)

$$\begin{aligned}
&P(X(t_n) = i_n, \ldots, X(t_1) = i_1) \\
=\ & P(X(t_n) = i_n \mid X(t_j) = i_j,\ j = 1, \ldots, n-1) \cdot P(X(t_j) = i_j,\ j = 1, \ldots, n-1) \\
=\ & P(X(t_n) = i_n \mid X(t_{n-1}) = i_{n-1}) P(X(t_j) = i_j,\ j = 1, \ldots, n-1) \\
=\ & \ldots \\
=\ & \prod_{j=2}^{n} P(X(t_j) = i_j \mid X(t_{j-1}) = i_{j-1}) \cdot P(X(t_1) = i_1).
\end{aligned}$$

Die Verteilung ist also bekannt, wenn für alle $i, j \in \mathbb{N}_0$ und alle $t, s \in I$, $s < t$, die bedingten Wahrscheinlichkeiten $P(X(t) = j \mid X(s) = i)$ bekannt sind. Man kann diese deuten als die Wahrscheinlichkeit dafür, dass ein Teilchen, das sich zur Zeit s im Punkte i befindet, sich im Intervall $(s, t]$ nach j bewegt.

Definition 5.20. *Die bedingte Wahrscheinlichkeit*

$$p_{ij}(s,t) := P(X(t) = j \mid X(s) = i)$$

($i, j \in \mathbb{N}_0$, $s,t \in I, s < t$) heißt **Übergangswahrscheinlichkeit** *des Zustandes i zur Zeit s in den Zustand j zur Zeit t.*

Satz 5.21. *Für die Übergangswahrscheinlichkeiten $p_{ij}(s,t)$ eines diskreten Markov Prozesses gilt im Falle $P(X(s) = i) \neq 0$*

(i) $0 \leq p_{ij}(s,t) \leq 1$

(ii) $\sum_j p_{ij}(s,t) \ = 1$

(iii) $p_{ij}(s,t) = \sum_k p_{ik}(s,\tau)p_{kj}(\tau,t), s < \tau < t.$

Beweis: Für festes i, s, t ist $p_{ij}(s,t)$ ein Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$. Daher gelten (i) und (ii). (iii) ist gerade die Formel von der totalen Wahrscheinlichkeit. ∎

Bemerkung 5.22. Gleichung (iii) heißt **Chapman Kolmogorow Gleichung** . □

Bemerkung 5.23. Für die Übergangswahrscheinlichkeiten eines Poisson Prozesses gilt

$$\begin{aligned} p_{ij}(s,t) &= P(X(t) = j \mid X(s) = i) \\ &= \frac{P(X(t) = j, X(s) = i)}{P(X(s) = i)} \\ &= \begin{cases} \exp(-\alpha(t-s))\frac{(\alpha(t-s))^{j-i}}{(j-i)!} & , \ i < j \\ 0 & , \ i \geq j. \end{cases} \end{aligned} \qquad \square$$

Definition 5.24. *Ein diskreter Markov Prozess $\{X(t) \ : \ t \in I\}$ heißt* **homogen** *, wenn für beliebige $i, j \in \mathbb{N}_0$ und $t, s \in I$ die Übergangswahrscheinlichkeit $p_{ij}(s,t)$ nicht von s und t, sondern nur von der Differenz $\tau := t - s$ abhängt. Man schreibt*

$$p_{ij}(\tau) := p_{ij}(s,t).$$

Beispiel 5.25. Der Poisson Prozess ist ein diskreter, homogener Markov Prozess mit der Übergangswahrscheinlichkeit

$$p_{ij}(\tau) = \begin{cases} \exp(-\alpha\tau)\dfrac{(\alpha\tau)^{j-i}}{(j-i)!} & , \ j > i \\ 0 & , \ j \leq i. \end{cases} \qquad \square$$

5.4 Diskrete Markov Ketten

Wir betrachten nun diskrete Markov Ketten, wobei wir ohne Einschränkung $I = \mathbb{N}_0$ annehmen. Da dann nur ganzzahlige "zeitliche Übergänge" möglich sind und da nach der Chapman Kolmogorow Gleichung

$$\begin{aligned} p_{ij}(n,m) &= \sum_{k_1} p_{ik_1}(n,n+1)p_{k_1j}(n+1,m) = \dots \\ &= \sum_{k_1} \dots \sum_{k_{m-n}} p_{ik_1}(n,n+1)p_{k_1k_2}(n+1,n+2)\cdot\ldots\cdot p_{k_{m-n}j}(m-1,m), \end{aligned}$$

gilt, ist das zufällige Verhalten der Kette vollbeständig bestimmt, wenn die Übergangswahrscheinlichkeiten $p_{ij}(n,n+1)$ für alle Schritte der Länge 1 bekannt sind.

Sind die Übergangswahrscheinlichkeiten $p_{ij}(n,n+1)$ für alle Zustände i und j unabhängig von n, so ist die Kette **homogen** . Wir schreiben dann kurz

$$p_{ij}(n,n+1) =: p_{ij}.$$

Es sei von nun an $\{X_n : n \in \mathbb{N}_0\}$ eine homogene Markov Kette mit dem Zustandsraum S, und es seien p_{ij} die Übergangswahrscheinlichkeiten. Wir bezeichnen mit $p_{ij}(m) := P(X_{n+m} = j \mid X_n = i)$ die Wahrscheinlichkeit dafür, dass man in genau m Schritten vom Zustand i zum Zustand j gelangt.

Dann gilt nach der Chapman Kolmogorow Gleichung

$$p_{ij}(2) = \sum_k p_{ik}p_{kj}$$

oder allgemeiner für $m \geq 2$

$$p_{ij}(m) = \sum_k p_{ik}p_{kj}(m-1).$$

Ordnet man die p_{ij} in einer Matrix $\boldsymbol{P}$ und die $p_{ij}(m)$ in einer Matrix $\boldsymbol{P}(m)$ an (die im Falle eines abzählbaren Zustandsraumes unendlich viele Zeilen und Spalten besitzen), so hat man offensichtlich $\boldsymbol{P}(m) = \boldsymbol{P} \cdot \boldsymbol{P}(m-1)$, und durch Induktion

$$\boldsymbol{P}(m) = \boldsymbol{P}^m, \quad m \geq 1.$$

Definition 5.26. *$\boldsymbol{P}$ heißt die* **Übergangsmatrix** *der homogenen Markov Kette, $\boldsymbol{P}(m)$ die m-Schritt Übergangsmatrix.*

Bevor wir uns weiter mit homogenen Markov Ketten beschäftigen, betrachten wir einige zusätzliche Beispiele.

Beispiel 5.27. Eindimensionale Irrfahrt
Ein Teilchen bewege sich nach folgenden Regeln auf der Menge der ganzen Zahlen: Zur Zeit $t = 0$ befinde es sich im Ursprung, d.h. $P(X_0 = 0) = 1$. Befindet es sich zur Zeit $t = n$ im Punkte i, so gehe es im nächsten Zeitintervall der Länge 1 mit Wahrscheinlichkeit $p \in (0, 1)$ in den Zustand $i + 1$ und mit Wahrscheinlichkeit $q := 1 - p$ in den Zustand $i - 1$ über.

Die eindimensionale Irrfahrt ist offensichtlich eine homogene Markov Kette. Die Übergangsmatrix hat die Gestalt

$$\boldsymbol{P} = \begin{pmatrix} \multicolumn{7}{c}{\dotfill} \\ \dots & q & 0 & p & 0 & 0 & \dots \\ \dots & 0 & q & 0 & p & 0 & \dots \\ \dots & 0 & 0 & q & 0 & p & \dots \\ \multicolumn{7}{c}{\dotfill} \end{pmatrix}$$

Die folgenden Fragen liegen auf der Hand:

(i) Wie groß ist die Wahrscheinlichkeit, dass ein Teilchen wieder in den Nullpunkt zurückkehrt?

(ii) Wie lange braucht ein Teilchen im Mittel, um in den Ursprung zurückzukehren?

(iii) Wie lange dauert es im Mittel, dass das Teilchen erstmals einen Punkt i erreicht? □

Beispiel 5.28. Eindimensionale Irrfahrt mit reflektierenden Barrieren
Ein Teilchen vollführe eine eindimensionale Irrfahrt, wobei es nun aber in den Punkten $a < 0$ und $b > 0$ reflektiert werde, d.h.

$$P(X_n = a + 1 \,|\, X_{n-1} = a) = 1, \quad P(X_n = b - 1 \,|\, X_{n-1} = b) = 1.$$

Auch dies ist eine homogene Markov Kette mit dem Zustandsraum $\{a, a+1, \dots, b-1, b\}$ und der Übergangsmatrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \multicolumn{8}{c}{\dotfill} \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

Wir stellen hier die weiteren Fragen.

(iv) Konvergieren die Verteilungen $(P(X_n = i))_{i=a,\dots,b}$ mit wachsendem n gegen eine Grenzverteilung?

(v) Hängt die Grenzverteilung von der Anfangsverteilung $(P(X_0 = i))_{i=a,\dots,b}$ ab (wenn wir nicht wie bisher $P(X_0 = 0) = 1$ voraussetzen)?

(vi) Gibt es eine "stationäre Verteilung" $\pi = (\pi_i)_{i=a,\dots,b}$, für die also

$$\sum_i \pi_i p_{ij} = \pi_j, \quad \sum_i \pi_i = 1,$$

gilt?

Ist π eine stationäre Verteilung und gilt $P(X_n = i) = \pi_i$, so folgt

$$P(X_{n+1} = i) = \sum_j P(X_{n+1} = i \,|\, X_n = j) P(X_n = j) = \sum_j p_{ji}\pi_j = \pi_i,$$

und genauso $P(X_m = i) = \pi_i$ für alle $m \geq n$. Das System bleibt also in diesem Zustand. $\square$

Beispiel 5.29. Eindimensionale Irrfahrt mit Absorption
Ein Teilchen vollführe eine eindimensionale Irrfahrt, werde aber in den Zuständen $a < 0$ und $b > 0$ absorbiert, d.h.

$$P(X_{n+1} = i \,|\, X_n = a) = \delta_{ia}, \quad P(X_{n+1} = i \,|\, X_n = b) = \delta_{ib}.$$

Die Übergangsmatrix dieser homogenen Markov Kette ist

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Diese Markov Kette ist auch unter dem Namen "Gambler's Ruin Chain" bekannt. Ein Spieler mit einem Anfangskapital von k DM führt eine Folge von Spielen aus, die er mit Wahrscheinlichkeit p gewinnt und mit Wahrscheinlichkeit q verliert. Beim Gewinn erhöht sich sein Kapital um 1 DM, beim Verlust erniedrigt es sich um 1 DM. Das Spiel endet, wenn das Kapital bei 0 DM oder bei einem $K > k$ angelangt ist.

Lässt man bei dem Spiel auch einen unentschiedenen Ausgang mit der Wahrscheinlichkeit r zu ($p, q, r \in (0,1),\ p+q+r = 1$), so hat man eine homogene Markov Kette mit der Übergangsmatrix

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & r & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & r & p & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & q & r & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Wir stellen hier die Fragen:

(vii) Mit welcher Wahrscheinlichkeit wird das Teilchen in a absorbiert (endet das Spiel mit der Pleite des Spielers)?

(viii) Wie lange dauert es im Mittel, bis ein Teilchen absorbiert wird (das Spiel verloren geht oder mit einem Gewinn von $K - k$ DM endet? □

Beispiel 5.30. Ehrenfestsches Urnenmodell
Es seien zwei Urnen gegeben. Am Anfang befinden sich k Kugeln in Urne 1 und $N-k$ Kugeln in Urne 2. Es wird nun laufend eine der N Kugeln zufällig ausgewählt, ihrer Urne entnommen und in die andere Urne gelegt.

Dieses Modell wurde von P. und T. Ehrenfest 1907 vorgeschlagen, um die Diffusion zwischen zwei Zellen oder den Wärmeaustausch zwischen zwei isolierten Körpern zu modellieren.

Das Ehrenfestsche Urnenmodell ist eine Markov Kette mit dem Zustandsraum $\{0, \dots, N\}$ ($[X_n = i]$ bedeutet, dass in Urne 1 sich genau i Kugeln befinden). Die Übergangsmatrix ist

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 1/N & 0 & (N-1)/N & 0 & \dots \\ 0 & 2/N & 0 & (N-2)/N & \dots \\ 0 & 0 & 3/N & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Wir fragen hier:

(ix) Gibt es einen Gleichgewichtszustand? Wie lautet der?

(x) Kann der Anfangszustand wieder erreicht werden? □

Definition 5.31. *Es sei $\{X_n\}$ eine Markov Kette. Dann heißt*

$$T_i := \inf\{n > 0 \,:\, X_n = i\}$$

die **Trefferzeit** *von i, wobei $\inf \emptyset := \infty$ gesetzt wird.*

Bemerkung 5.32. T_i ist die Wartezeit bis zum ersten Besuch von $\{X_n\}$ in i. □

Lemma 5.33. *Es sei $\{X_n\}$ eine homogene Markov Kette mit Übergangsmatrix $\boldsymbol{P}$. Dann gilt für alle $n \geq 1$*

(i) $p_{ij}(n) = \sum\limits_{m=1}^{n} P(T_j = m \,|\, X_0 = i) p_{jj}(n-m)$

(ii) $P(T_j = n+1 \,|\, X_0 = i) = \sum\limits_{k \neq j} p_{ik} P(T_j = n \,|\, X_0 = k)$.

Beweis: (i) Die Ereignisse $[T_j = m, X_n = j]$ sind für $1 \leq m \leq n$ disjunkt und $[X_n = j] = \bigcup\limits_{m=1}^{n} [T_j = m, X_n = j]$.

Daher gilt

$$\begin{aligned} p_{ij}(n) &= P(X_n = j \,|\, X_0 = i) \\ &= \sum_{m=1}^{n} P(T_j = m, X_n = j \,|\, X_0 = i) \\ &= \sum_{m=1}^{n} P(T_j = m \,|\, X_0 = i) P(X_n = j \,|\, X_0 = i, T_j = m) \end{aligned}$$

und man erhält weiter

$$\begin{aligned} p_{ij}(n) &= P(X_n = j \,|\, X_0 = i) \\ &= \sum_{m=1}^{n} P(T_j = m | X_0 = i) P(X_n = j | X_0 = i, X_1 \neq j, \ldots, X_{m-1} \neq j, X_m = j) \\ &= \sum_{m=1}^{n} P(T_j = m \,|\, X_0 = i) P(X_n = j \,|\, X_m = j) \\ &= \sum_{m=1}^{n} P(T_j = m \,|\, X_0 = i) p_{jj}(n-m). \end{aligned}$$

(ii) ist klar, da man von i nach j in genau $n+1$ Schritten gelangt, indem man von i im ersten Schritt nach $k \neq j$ geht und dann in genau n Schritten von k nach j. ∎

Wir bezeichnen nun für die Markov Kette $\{X_n\}$ mit

$$\rho_{ij} := P(T_j < \infty \,|\, X_0 = i)$$

die Wahrscheinlichkeit, dass ein in i startendes Teilchen in endlich vielen Schritten den Zustand j erreicht. Insbesondere ist ρ_{ii} die Wahrscheinlichkeit, dass ein in i startendes Teilchen nach i zurückkehrt.

Definition 5.34. *Gilt $\rho_{ii} = 1$ für einen Zustand i, so nennen wir i einen* **rekurrenten** *Zustand. Gilt $\rho_{ii} < 1$, so heißt i* **transient**.

Bemerkung 5.35. Ist i ein rekurrenter Zustand, so kehrt eine in i startende Markov Kette mit Wahrscheinlichkeit 1 nach i zurück. Ist i ein transienter Zustand und startet die Markov Kette in i, so ist die Wahrscheinlichkeit $1 - \rho_{ii}$, dass die Kette nie nach i zurückkehrt, positiv. □

Bemerkung 5.36. Ist i ein absorbierender Zustand, d.h. $p_{ii} = 1$, so gilt $P(T_i = 1 \,|\, X_0 = i) = p_{ii} = 1$ und daher $\rho_{ii} = 1$. Ein absorbierender Zustand ist also immer rekurrent. □

Beispiel 5.37. Wir betrachten die Irrfahrt mit Absorption.

Dann gilt

$$P(X_1 = 1, X_2 = 2, \ldots, X_b = b, X_{b+1} = b, \ldots, X_n = b \,|\, X_0 = 0)$$

$$= \prod_{j=1}^{b} P(X_j = j \,|\, X_{j-1} = j-1) \cdot \prod_{j=b+1}^{n} P(X_j = b \,|\, X_{j-1} = b) = p^b$$

für alle $n \in \mathbb{N}$, d.h.

$$\rho_{00} = P(T_0 < \infty \,|\, X_0 = 0) \le 1 - p^b < 1$$

und 0 ist ein transienter Zustand.

Genauso sind die Zustände $a+1, a+2, \ldots, b-1$ transient, und nur die beiden absorbierenden Zustände a und b sind rekurrent. □

Zur weiteren Untersuchung rekurrenter und transienter Zustände betrachten wir nun für jeden Zustand i die Anzahl $N(i)$ der Besuche in i.

Bezeichnet

$$1_i(j) := \begin{cases} 1 & \text{für} \quad i = j \\ 0 & \text{für} \quad i \neq j \end{cases}$$

die charakteristische Funktion der Menge $\{i\}$, so kann man $N(i)$ schreiben als

$$N(i) = \sum_{n=1}^{\infty} 1_i(X_n).$$

Offenbar gilt $[N(j) \geq 1] = [T_j < \infty]$. Daher folgt

$$P(N(j) \geq 1 \,|\, X_0 = i) = P(T_j < \infty \,|\, X_0 = i) = \rho_{ij}.$$

Es seien $n, m \in \mathbb{N}$. Dann ist die Wahrscheinlichkeit, dass ein Teilchen in i startet, im m-ten Schritt den Zustand j erstmals besucht und dann nach genau n Schritten den Zustand zum zweiten Mal besucht

$$P(T_j = m \,|\, X_0 = i) P(T_j = n \,|\, X_0 = j),$$

und daher gilt

$$\begin{aligned} P(N(j) \geq 2 \,|\, X_0 = i) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P(T_j = m \,|\, X_0 = i) P(T_j = n \,|\, X_0 = j) \\ &= \sum_{m=1}^{\infty} P(T_j = m \,|\, X_0 = i) \sum_{n=1}^{\infty} P(T_j = n \,|\, X_0 = j) = \rho_{ij}\rho_{jj}. \end{aligned}$$

Genauso erhält man

$$P(N(j) \geq m \,|\, X_0 = i) = \rho_{ij}\rho_{jj}^{m-1} \tag{5.6}$$

und wegen

$$P(N(j) = m \,|\, X_0 = i) = P(N(j) \geq m \,|\, X_0 = i) - P(N(j) \geq m+1 \,|\, X_0 = i)$$

folgt

$$P(N(j) = m \,|\, X_0 = i) = \rho_{ij}\rho_{jj}^{m-1}(1 - \rho_{jj}), \tag{5.7}$$

sowie

$$P(N(j) = 0 \,|\, X_0 = i) = 1 - P(N(j) \geq 1 \,|\, X_0 = i) = 1 - \rho_{ij}. \tag{5.8}$$

Die Formeln (5.6), (5.7) und (5.8) sind anschaulich klar. Die rechte Seite von (5.7) gibt z.B. die Wahrscheinlichkeit dafür an, dass ein Teilchen in i startet, den Zustand j erstmals erreicht, $m-1$ mal nach j zurückkehrt und dann nie wieder den Zustand j besucht.

Wir betrachten nun den bedingten Erwartungswert von $N(j)$ unter der Bedingung $X_0 = i$. Es gilt

$$E(1_j(X_n) \,|\, X_0 = i) = P(X_n = j \,|\, X_0 = i) = p_{ij}(n),$$

und daher (die Grenzprozesse dürfen wegen der gleichmäßigen Konvergenz vertauscht werden!)

$$\begin{aligned} E(N(j) \mid X_0 = i) &= E\left(\sum_{n=1}^{\infty} 1_j(X_n) \mid X_0 = i\right) \\ &= \sum_{n=1}^{\infty} E(1_j(X_n) \mid X_0 = i) = \sum_{n=1}^{\infty} p_{ij}(n). \end{aligned}$$

Wir setzen

$$g_{ij} := E(N(j) \mid X_0 = i) = \sum_{n=1}^{\infty} p_{ij}(n).$$

Dies ist die mittlere Anzahl der Besuche einer homogenen Markov Kette im Zustand j, die in i startet. Hiermit gilt

Satz 5.38. *(i) Sei j ein transienter Zustand. Dann gilt für alle i*

$$P(N(j) < \infty \mid X_0 = i) = 1$$

und

$$g_{ij} = \frac{\rho_{ij}}{1 - \rho_{jj}} < \infty \ .$$

(ii) Sei j ein rekurrenter Zustand. Dann gilt

$$P(N(j) = \infty \mid X_0 = j) = 1 \text{ und } g_{jj} = \infty \ .$$

Ferner ist

$$P(N(j) = \infty \mid X_0 = i) = P(T_j < \infty \mid X_0 = i) = \rho_{ij} \ .$$

Ist $\rho_{ij} = 0$, so gilt $g_{ij} = 0$, ist $\rho_{ij} > 0$, so gilt $g_{ij} = \infty$.

Insbesondere folgt aus Satz 5.38.(i)

Korollar 5.39. *Ist j ein transienter Zustand, so gilt*

$$\lim_{n\to\infty} p_{ij}(n) = 0 \quad \text{für alle } i \in S.$$

Bemerkung 5.40. Satz 5.38. macht den fundamentalen Unterschied zwischen transienten und rekurrenten Zuständen deutlich. Ist j transient und startet die Markov Kette an einem beliebigen Zustand i, so wird j höchstens endlich oft besucht und der Erwartungswert der Anzahl der Besuche in j ist endlich. Ist j rekurrent und startet die Kette in j, so kehrt sie unendlich oft in j zurück. Startet sie in einem anderen Zustand, so kann es sein, dass sie j niemals erreicht. Wenn sie aber j erreichen kann, so kehrt sie unendlich oft nach j zurück. □

Beispiel 5.41. Wir betrachten die eindimensionale Irrfahrt. Dann gilt

$$p_{00}(2n-1) = 0 \quad \text{und} \quad p_{00}(2n) = \binom{2n}{n} p^n q^n = \frac{(2n)!}{(n!)^2} p^n q^n$$

für alle n (die erste Formel ist klar, da ein Teilchen nur nach einer geraden Anzahl von Schritten nach 0 zurückkehren kann).

Nach der Stirlingschen Formel gilt asymptotisch $n! \sim n^{n+1/2} e^{-n} \sqrt{2\pi}$, und daher

$$p_{00}(2n) \sim \frac{p^n q^n 2^{2n}}{\sqrt{\pi n}} = \frac{(4pq)^n}{\sqrt{\pi n}}.$$

Für $p = q = \frac{1}{2}$ ist $pq = \frac{1}{4}$, und es gilt dann $g_{00} = \infty$. Anderenfalls ist $pq < \frac{1}{4}$, und es folgt $g_{00} < \infty$. Der Zustand 0 (und genauso jeder andere Zustand) ist also genau dann rekurrent, wenn die Irrfahrt symmetrisch ist ($p = q = \frac{1}{2}$).

Es ist anschaulich klar, dass ein in 0 startendes Teilchen im Falle $p > \frac{1}{2}$ gegen ∞ driften wird und im Falle $p < \frac{1}{2}$ gegen $-\infty$ gehen wird. □

Beweis: (von Satz 5.38.)
(i) Sei j ein transienter Zustand. Dann folgt aus $\rho_{jj} < 1$ wegen (5.6)

$$P(N(j) = \infty \,|\, X_0 = i) = \lim_{m\to\infty} P(N(j) \geq m \,|\, X_0 = i) = \lim_{m\to\infty} \rho_{ij} \rho_{jj}^{m-1} = 0.$$

Wegen (5.7) gilt

$$\begin{aligned} g_{ij} &= E(N(j) \,|\, X_0 = i) = \sum_{m=1}^{\infty} m P(N(j) = m \,|\, X_0 = i) \\ &= \sum_{m=1}^{\infty} m \rho_{ij} \rho_{jj}^{m-1}(1 - \rho_{jj}) = \frac{\rho_{ij}}{1 - \rho_{jj}} < \infty \ . \end{aligned}$$

(Beachte: Für $|t| < 1$ konvergiert die Potenzreihe $\sum\limits_{m=0}^{\infty} t^m = 1/(1-t)$, und durch gliedweise Differentiation erhält man $\sum\limits_{m=1}^{\infty} m t^{m-1} = 1/(1-t)^2$).
(ii) Ist j rekurrent, so gilt $\rho_{jj} = 1$, und aus (5.6) folgt

$$P(N(j) = \infty \,|\, X_0 = i) = \lim_{m\to\infty} P(N(j) \geq m \,|\, X_0 = i) = \rho_{ij} \ .$$

Insbesondere gilt $P(N(j) = \infty \,|\, X_0 = j) = 1$, sowie $g_{jj} = E(N(j) \,|\, X_0 = j) = \infty$, da der Erwartungswert einer nichtnegativen Zufallsvariable X mit $P(X = \infty) > 0$ stets unendlich ist.

Ist $\rho_{ij} = 0$, so gilt $P(T(j) = m \,|\, X_0 = i) = 0$ für alle $m > 0$, und aus Lemma 5.33. folgt $p_{ij}(n) = 0$ für alle $n \geq 1$. Daher ist $g_{ij} = 0$ in diesem Fall. Ist $\rho_{ij} > 0$, so gilt $P(N(j) = \infty \,|\, X_0 = i) = \rho_{ij} > 0$, und daher $E(N(j) \,|\, X_0 = i) = \infty$. ∎

Definition 5.42. *Eine Markov Kette heißt* **rekurrente Kette** *, wenn alle Zustände rekurrent sind, sie heißt* **transiente Kette** *, wenn alle Zustände transient sind.*

Ist der Zustandsraum S einer Markov Kette endlich, so existiert wenigstens ein rekurrenter Zustand, denn wenn alle Zustände transient sind, so folgt

$$0 = \sum_j \lim_{n\to\infty} p_{ij}(n) = \lim_{n\to\infty} \sum_j p_{ij}(n) = \lim_{n\to\infty} P(X_n \in S \mid X_0 = i) = 1.$$

Die Vertauschung der Summe mit dem Grenzwert ist nur für endliches S gerechtfertigt. Eine Markov Kette mit abzählbarem Zustandsraum kann also transient sein (z.B. die nichtsymmetrische Irrfahrt).

Definition 5.43. *Es seien nun i und j nicht notwendig verschiedene Zustände. i* **führt nach** *j, wenn $p_{ij}(n) > 0$ für ein $n > 0$ gilt.*

Wir schreiben dann $i \to j$.

Es ist klar, dass i genau dann nach j führt, wenn $\rho_{ij} > 0$ gilt und dass $\to$ eine transitive Relation ist, d.h.

$$i \to j,\ j \to k \ \Rightarrow\ i \to k.$$

Satz 5.44. *Sei i ein rekurrenter Zustand, der nach j führe. Dann gilt $\rho_{ij} = \rho_{ji} = 1$, und j ist ein rekurrenter Zustand.*

Beweis: Es sei $i \neq j$, denn sonst ist nichts zu zeigen. Wegen $P(T_j < \infty \mid X_0 = i) = \rho_{ij} > 0$ gilt $P(T_j = n \mid X_0 = i) > 0$ für ein $n \in \mathbb{N}$. Es sei $n_0 := \min\{n \in \mathbb{N} : P(T_j = n \mid X_0 = i) > 0\}$. Dann folgt aus Lemma 5.33.

$$\begin{aligned} p_{ij}(n) &= \sum_{m=1}^{n} P(T_j = m \mid X_0 = i) p_{jj}(n-m) = 0, \quad n = 1, \ldots, n_0 - 1, \\ p_{ij}(n_0) &= P(T_j = n_0 \mid X_0 = i) p_{jj}(0) > 0. \end{aligned}$$

Es gibt daher Zustände $k_1, \ldots, k_{n_0-1} =: k'$ mit

$$P(X_1 = k_1, \ldots, X_{n_0-1} = k', X_{n_0} = j \mid X_0 = i) = p_{ik_1} p_{k_1k_2} \ldots p_{k'j} > 0,$$

und alle $k_1, \ldots, k'$ sind von i und j verschieden, da sonst n_0 nicht minimal gewählt wäre.

Wir nehmen nun an, dass $\rho_{ji} < 1$ gilt. Dann besitzt eine Markov Kette, die in j gestartet wird, die positive Wahrscheinlichkeit $1 - \rho_{ji}$, den Zustand i niemals zu erreichen. Daher besitzt eine in i gestartete Markov Kette (wenigstens) die positive Wahrscheinlichkeit $p_{ik_1}p_{k_1k_2}\dots p_{k'j}(1-\rho_{ji})$, nicht nach i zurückzukehren im Widerspruch zur Rekurrenz von i.

Also folgt $\rho_{ji} = 1$.

Wegen $\rho_{ji} = 1$ gibt es ein n_1 mit $p_{ji}(n_1) > 0$. Daher ist

$$\begin{aligned} p_{jj}(n_1+n+n_0) &= P(X_{n_1+n+n_0} = j \mid X_0 = j) \\ &\geq P(X_{n_1} = i, X_{n_1+n} = i, X_{n_1+n+n_0} = j \mid X_0 = j) \\ &= p_{ji}(n_1)p_{ii}(n)p_{ij}(n_0), \end{aligned}$$

und es folgt

$$\begin{aligned} g_{jj} &\geq \sum_{n=n_1+1+n_0}^{\infty} p_{jj}(n) = \sum_{n=1}^{\infty} p_{jj}(n_1+n+n_0) \\ &\geq p_{ji}(n_1)p_{ij}(n_0)\sum_{n=1}^{\infty} p_{ii}(n) \\ &= p_{ji}(n_1)p_{ij}(n_0)g_{ii} = \infty. \end{aligned}$$

Also ist j rekurrent. Da j nach i führt, folgt wie oben

$$\rho_{ij} = 1. \qquad \blacksquare$$

Beispiel 5.45. Bei der Irrfahrt mit reflektierenden Barrieren führt offenbar jeder Zustand zu jedem anderen Zustand. Da der Zustandsraum endlich ist, existiert wenigstens ein rekurrenter Zustand, und Satz 5.44. sagt, dass dann alle Zustände rekurrent sind. □

Definition 5.46. *Es sei $\{X_n\}$ eine Markov Kette mit dem Zustandsraum S. Eine nichtleere Menge $C \subset S$ heißt* **abgeschlossen** *, wenn gilt*

$$p_{ij} = 0 \quad \textit{für alle} \quad i \in C,\ j \notin C.$$

Bemerkung 5.47. Ist C abgeschlossen, $i \in C$, $j \notin C$, so gilt

$$p_{ij}(2) = \sum_{k\in S} p_{ik}p_{kj} = \sum_{k\in C} p_{ik}p_{kj} = 0,$$

und durch Induktion

$$p_{ij}(n) = 0 \quad \text{für alle} \quad i \in C,\ j \notin C \text{ und alle } n \in \mathbb{N}.$$

Hieraus folgt

$$\rho_{ij} = 0 \quad \text{für alle} \quad i \in C,\ j \notin C. \qquad \square$$

Bemerkung 5.48. Ist a ein absorbierender Zustand, so ist die Menge $\{a\}$ natürlich abgeschlossen. $\square$

Bemerkung 5.49. Ist C abgeschlossen und startet eine Markov Kette in C, so bleibt sie mit Wahrscheinlichkeit 1 in C. $\square$

Definition 5.50. *Eine abgeschlossene Menge C heißt* **irreduzibel**, *wenn jeder Zustand aus C zu jedem Zustand aus C führt.*

Aus Satz 5.44. folgt, dass eine irreduzible, abgeschlossene Menge nur aus rekurrenten oder nur aus transienten Zuständen besteht.

Entsprechend der vorhergehenden Argumentation besteht eine endliche, abgeschlossene, irrduzible Menge C nur aus rekurrenten Zuständen.

Beispiel 5.51. Wir betrachten die Markov Kette mit dem Zustandsraum $S := \{1, \ldots, 8\}$ und der Übergangsmatrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.25 & 0 & 0.25 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Offenbar ist 8 ein absorbierender Zustand, also $C_1 = \{8\}$ eine abgeschlossene, irreduzible, rekurrente Menge. Ferner sieht man an dem zu $\boldsymbol{P}$ gehörenden Graphen, dass $C_2 = \{2, 3, 4\}$ eine abgeschlossene irreduzible Menge von rekurrenten Zuständen ist ("Es führen keine Pfeile aus C_2 heraus"). Die Zustände $1, 5, 6$ und 7 führen in C_2 hinein, sind also transient. $\square$

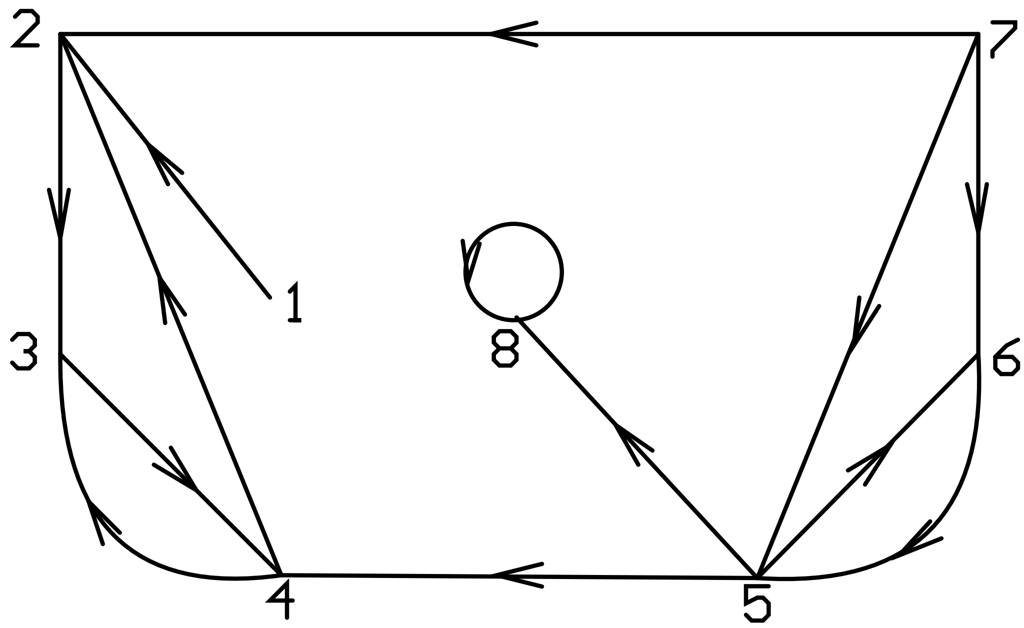

Abbildung 5.1: Graph zu Beispiel 5.51.

Wir haben eben den Zustandsraum $S := \{1, 2, \ldots, 8\}$ disjunkt zerlegt in $S = C_1 \cup C_2 \cup T$, $T := \{1, 5, 6, 7\}$, wobei T die Menge aller transienten Zustände ist und C_1, C_2 disjunkte, abgeschlossene, irreduzible Mengen von rekurrenten Zuständen.

Eine solche Zerlegung kann man stets vornehmen.

Satz 5.52. *Es sei $S = S_R \cup S_T$, wobei S_R bzw. S_T die Menge aller rekurrenten bzw. transienten Zustände bezeichnet. Ist $S_R \neq \emptyset$, so ist S_R die Vereinigung von (endlich oder abzählbar unendlich vielen) disjunkten, irreduziblen, abgeschlossenen Mengen $C_1, C_2, \ldots$.*

Diese Zerlegung von S hilft, das Verhalten von Markov Ketten zu verstehen. Startet eine Markov Kette in einem C_i, so bleibt sie mit Wahrscheinlichkeit 1 in C_i und besucht alle Zustände in C_i unendlich oft. Startet sie in S_T, so bleibt sie entweder immer in S_T, oder sie gelangt in ein C_i und verhält sich dann wie oben beschrieben.

Definition 5.53. *Es sei $\{X_n\}$ eine Markov Kette mit dem Zustandsraum S, es sei C eine irreduzible, abgeschlossene Menge von rekurrenten Zuständen und $i \in S$. Dann heißt*

$$\rho_C(i) := P(T_C < \infty \mid X_0 = i)$$

die **Absorptionswahrscheinlichkeit** *von i durch C, wobei $T_C := \inf\{n > 0 : X_n \in C\}$ entsprechend T_i definiert ist.*

Bemerkung 5.54. $\rho_C(i)$ ist die Wahrscheinlichkeit dafür, dass eine Markov Kette, die in i startet, die Menge C in endlicher Zeit trifft. Da C rekurrent ist, verbleibt die Kette danach mit Wahrscheinlichkeit 1 in C. Dies erklärt den Namen Absorptionswahrscheinlichkeit von i durch C. □

Bemerkung 5.55. Natürlich gilt $\rho_C(i) = 1$ für jeden Zustand $i \in C$ und $\rho_C(i) = 0$ für jeden rekurrenten Zustand $i \notin C$. □

Wir geben nun einen Weg an, wie man für einen transienten Zustand i die Absorptionswahrscheinlichkeit $\rho_C(i)$ berechnen kann, falls S_T eine endliche Menge ist.

Eine Kette, die in $i \in S_T$ startet, kann nach C gelangen, indem sie im ersten Schritt in C eintritt oder indem sie im ersten Schritt in einen transienten Zustand übergeht und später nach C gelangt. Sie kann nicht im ersten Schritt in einen nicht in C enthaltenen rekurrenten Zustand j übergehen, da sie sonst von der rekurrenten, irreduziblen, abgeschlossenen Menge, in der j liegt, absorbiert würde. Es gilt also für $i \in S_T$

$$\rho_C(i) = \sum_{j\in C} p_{ij} + \sum_{j\in S_T} p_{ij}\rho_C(j).$$

Die Absorptionswahrscheinlichkeiten erfüllen also ein lineares Gleichungssystem.

Satz 5.56. *Die Menge S_T der transienten Zustände sei endlich und es sei C eine irreduzible, abgeschlossene Menge rekurrenter Zustände. Dann besitzt das lineare Gleichungssystem*

$$f(i) = \sum_{j\in C} p_{ij} + \sum_{j\in S_T} p_{ij} f(j), \quad i \in S_T \tag{5.9}$$

die eindeutige Lösung $f(i) = \rho_C(i), i \in S_T$.

Beweis: Es sei f eine Lösung von (5.9). Setzt man f auf der rechten Seite von (5.9) ein, so folgt

$$\begin{aligned} f(i) &= \sum_{j\in C} p_{ij} + \sum_{j\in S_T}\sum_{k\in C} p_{ij}p_{jk} + \sum_{j\in S_T}\sum_{k\in S_T} p_{ij}p_{jk}f(k) \\ &= P(T_C \le 2 \mid X_0 = i) + \sum_{j\in S_T} p_{ij}(2) f(j). \end{aligned}$$

Durch Induktion erhält man auf dieselbe Weise

$$f(i) = P(T_C \le n \mid X_0 = i) + \sum_{j\in S_T} p_{ij}(n) f(j) \quad \text{für alle } n \in \mathbb{N}.$$

Nach Korollar 5.39. gilt $\lim\limits_{n\to\infty} p_{ij}(n) = 0$ für $j \in S_T$, und da S_T eine endliche Menge ist, konvergiert die Summe auf der rechten Seite gegen 0 für n gegen ∞. Daher folgt

$$f(i) = \lim_{n\to\infty} P(T_C \leq n \,|\, X_0 = i) = P(T_C < \infty \,|\, X_0 = i) = \rho_C(i). \qquad \blacksquare$$

Beispiel 5.57. Wir betrachten Beispiel 5.51. mit $C = \{8\}$. Dann erfüllen die $\rho(i) := \rho_{\{8\}}(i)$, $i = 1, 5, 6, 7$ das Gleichungssystem

$$\begin{pmatrix}\rho(1)\\ \rho(5)\\ \rho(6)\\ \rho(7)\end{pmatrix} = \begin{pmatrix}0\\ 1/4\\ 0\\ 0\end{pmatrix} + \begin{pmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 1/4 & 0\\ 0 & 1 & 0 & 0\\ 0 & 1/4 & 1/4 & 0\end{pmatrix} \begin{pmatrix}\rho(1)\\ \rho(5)\\ \rho(6)\\ \rho(7)\end{pmatrix}$$

mit der Lösung $\rho(1) = 0$, $\rho(5) = 1/3$, $\rho(6) = 1/3$,$\rho(7) = 1/6$.

Ist $C = \{2, 3, 4\}$, so erfüllen die $\rho(i) := \rho_C(i)$, $i = 1, 5, 6, 7$ das System

$$\begin{pmatrix}\rho(1)\\ \rho(5)\\ \rho(6)\\ \rho(7)\end{pmatrix} = \begin{pmatrix}1\\ 1/2\\ 0\\ 1/2\end{pmatrix} + \begin{pmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 1/4 & 0\\ 0 & 1 & 0 & 0\\ 0 & 1/4 & 1/4 & 0\end{pmatrix} \begin{pmatrix}\rho(1)\\ \rho(5)\\ \rho(6)\\ \rho(7)\end{pmatrix}$$

mit der Lösung $\rho(1) = 1$, $\rho(5) = 2/3$, $\rho(6) = 2/3$, $\rho(7) = 5/6$.

Das letzte Ergebnis hätte man auch erhalten können, ohne das zugehörige Gleichungssystem zu lösen, denn ist S gemäß Satz 5.52. zerlegt und S_T endlich, so gilt für jedes $i \in S_T$

$$\sum_k \rho_{C_k}(i) = \sum_k P(T_{C_k} < \infty \,|\, X_0 = i) = P(T_{S_R} < \infty \,|\, X_0 = i) = 1,$$

denn, da S_T endlich ist und jeder transiente Zustand nur endlich oft besucht wird, muss $P(T_{S_R} < \infty \,|\, X_0 = i) = 1$ gelten. $\square$

Bemerkung 5.58. Trifft eine Markov Kette, die in einem transienten Zustand i startet, eine irreduzible, abgeschlossene Menge C von rekurrenten Zuständen, so besucht sie jeden Zustand $j \in C$. Es gilt also $\rho_{ij} = \rho_C(i)$ für alle $j \in C$.

In Beispiel 5.51. erhalten wir daher

$$\boldsymbol{\rho} = (\rho_{ij}) = \begin{pmatrix}0 & 1 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 2/3 & 2/3 & 2/3 & 1/4 & 1/4 & 0 & 1/3\\ 0 & 2/3 & 2/3 & 2/3 & 1 & 1/4 & 0 & 1/3\\ 0 & 5/6 & 5/6 & 5/6 & 1/2 & 5/16 & 0 & 1/6\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

ρ_{ij} für $j \in \{2, 3, 4, 8\}$ liefert die Theorie, den Rest erhält man aus einfachen Überlegungen. $\square$

5.5 Stationäre Verteilungen

Wir untersuchen nun die Existenz von stationären Verteilungen einer Markov Kette, also Verteilungen, die sich im Laufe des Prozesses nicht verändern. Für diese Untersuchung benötigen wir das folgende Hilfsmittel aus der Matrizentheorie.

Satz 5.59. (Perron, Frobenius)
Sei $\boldsymbol{A}$ eine nichtnegative, nichtzerfallende Matrix und ρ der Spektralradius von $\boldsymbol{A}$, d.h. $\rho = \max_i |\lambda_i|$, wenn λ_i die Eigenwerte von A bezeichnet. Dann gilt

(i) ρ ist algebraisch einfacher Eigenwert von $\boldsymbol{A}$.

(ii) Zu ρ existiert ein positiver Eigenvektor, d.h. ein $\boldsymbol{z} \in \mathbb{R}^n$ mit $\boldsymbol{Az} = \rho\boldsymbol{z}$ und $z_i > 0$ für $i = 1, \ldots, n$.

(iii) Ist $\boldsymbol{v}$ irgendein nichtnegativer Eigenvektor von $\boldsymbol{A}$, so ist $\boldsymbol{v}$ ein Vielfaches von $\boldsymbol{z}$.

Beweis s. **R.S. Varga**, Matrix Iterative Analysis, S. 28 ff.

Dabei heißt eine Matrix nicht zerfallend, wenn durch Umordnung der Zeilen und gleichlautende Umordnung der Spalten nicht die Gestalt

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{O} \\ \boldsymbol{A}_2 & \boldsymbol{A}_3 \end{pmatrix}$$

mit quadratischen Matrizen $\boldsymbol{A}_1$ und $\boldsymbol{A}_3$ hergestellt werden kann.

Eine beliebige nichtnegative Matrix kann als Grenzwert nichtnegativer, nichtzerfallender (sogar positiver) Matrizen dargestellt werden. Deshalb gilt für sie noch

Satz 5.60. *Ist $\boldsymbol{A}$ eine nichtnegative Matrix, so ist der Spektralradius ρ ein Eigenwert von $\boldsymbol{A}$, und es existiert $\boldsymbol{z} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ mit $\boldsymbol{Az} = \rho\boldsymbol{z}$ und $z_i \geq 0$ für $i = 1, \ldots, n$.*

Es sei nun $\boldsymbol{P}$ die Übergangsmatrix einer homogenen Markov Kette. Dann ist 1 ein Eigenwert von $\boldsymbol{P}$, denn mit $\boldsymbol{w} = (1, \ldots, 1)^T$ gilt

$$\boldsymbol{Pw} = (\sum_j p_{ij} w_j)_i = (\sum_j p_{ij})_i = \boldsymbol{w}.$$

Ist λ ein beliebiger Eigenwert mit Eigenvektor $\boldsymbol{z}$, so folgt

$$\begin{aligned} |\lambda| \cdot \|\boldsymbol{z}\|_\infty &= \|\lambda \boldsymbol{z}\|_\infty = \|\boldsymbol{Pz}\|_\infty = \max_i |\sum p_{ij} z_j| \\ &\leq \max_i \sum_j |p_{ij} z_j| \leq \max_i \sum_j |p_{ij}| \cdot \max_j |z_j| = \|\boldsymbol{z}\|_\infty \ , \end{aligned}$$

d.h. $|\lambda| \leq 1$, und daher ist $\rho = 1$ der Spektralradius von $\boldsymbol{P}$.

Da $\boldsymbol{P}$ und $\boldsymbol{P}^T$ identische Eigenwerte haben, ist $\rho = 1$ auch Spektralradius von $\boldsymbol{P}^T$, und nach Satz 5.60. existiert zu ρ ein nichtnegativer Eigenvektor $\boldsymbol{\pi} = (\pi_i)$, den wir normieren können zu $\sum\limits_i \pi_i = 1$.

Für das so gewonnene $\boldsymbol{\pi}$ gilt offenbar $\pi_j = \sum\limits_i p_{ij}\pi_i$ für alle j, d.h. $\boldsymbol{\pi}$ ist eine stationäre Verteilung der Markov Kette.

Ist umgekehrt $\boldsymbol{\pi}$ eine stationäre Verteilung einer Markov Kette, so ist $\boldsymbol{\pi}$ ein nichtnegativer Eigenvektor der Transponierten der Übergangsmatrix.

Um uns einen Überblick über alle stationären Verteilungen zu verschaffen, numerieren wir nun die Zustände so, dass entsprechend Satz 5.52. zunächst die Elemente von C_1, dann von $C_2, \ldots,$ bis C_r und dann die transienten Zustände aufgeführt sind (C_j sind die disjunkten, abgeschlossenen, irreduziblen Mengen rekurrenter Zustände). Dann hat die Übergangsmatrix die Gestalt

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_1 & \boldsymbol{O} & \ldots & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{P}_2 & \ldots & \boldsymbol{O} & \boldsymbol{O} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \boldsymbol{O} & \boldsymbol{O} & \ldots & \boldsymbol{P}_r & \boldsymbol{O} \\ \boldsymbol{P}_{T1} & \boldsymbol{P}_{T2} & \ldots & \boldsymbol{P}_{Tr} & \boldsymbol{P}_T \end{pmatrix}, \tag{5.10}$$

wobei $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_r$ stochastische Matrizen sind (d.h. nichtnegativ mit der Zeilensumme 1), die zu den Mengen $C_1, \ldots, C_r$ gehören.

Sei $\boldsymbol{\pi}$ eine stationäre Verteilung, d.h. $\boldsymbol{\pi}^T\boldsymbol{P} = \boldsymbol{\pi}^T$.

Zerlegt man $\boldsymbol{\pi}$ entsprechend der Zerlegung von $\boldsymbol{P}$ in $\boldsymbol{\pi}^T = (\boldsymbol{\pi}_1^T, \ldots, \boldsymbol{\pi}_r^T, \boldsymbol{\pi}_T^T)$, so gilt

$$\begin{aligned} \boldsymbol{\pi}_j^T\boldsymbol{P}_j + \boldsymbol{\pi}_T^T\boldsymbol{P}_{Tj} &= \boldsymbol{\pi}_j^T, \quad j = 1, \ldots, r, \\ \boldsymbol{\pi}_T^T\boldsymbol{P}_T &= \boldsymbol{\pi}_T^T. \end{aligned} \tag{5.11}$$

Nun ist die Matrix $\boldsymbol{E} - \boldsymbol{P}_T$ regulär, denn wegen $\boldsymbol{P}_T^n \to \boldsymbol{O}$ ist $\boldsymbol{E} - \boldsymbol{P}_T^n$ für genügend großes n regulär und dann auch die beiden Faktoren in der Zerlegung

$$\boldsymbol{E} - \boldsymbol{P}_T^n = (\boldsymbol{E} - \boldsymbol{P}_T)(\boldsymbol{E} + \boldsymbol{P}_T + \boldsymbol{P}_T^2 + \ldots + \boldsymbol{P}_T^{n-1}) \ .$$

Daher folgt aus der letzten Gleichung von (5.11) $\boldsymbol{\pi}_T = \boldsymbol{0}$, und (5.11) geht über in

$$\boldsymbol{\pi}_j^T\boldsymbol{P}_j = \boldsymbol{\pi}_j^T \ .$$

Die Matrizen $\boldsymbol{P}_j^T$ sind nichtzerfallend, denn sonst könnte man durch Umordnung der Zeilen und Spalten erreichen, dass $\boldsymbol{P}_j$ die Gestalt hat

$$\boldsymbol{P}_j = \begin{pmatrix} \tilde{\boldsymbol{P}} & \tilde{\boldsymbol{Q}} \\ \boldsymbol{O} & \tilde{\boldsymbol{R}} \end{pmatrix}$$

mit quadratischen Matrizen $\tilde{\boldsymbol{P}}$ und $\tilde{\boldsymbol{R}}$. Damit bildeten bereits die zu $\tilde{\boldsymbol{R}}$ gehörenden Zustände eine abgeschlossene Menge von Zuständen, und die Menge C_j wäre nicht irreduzibel.

Nach dem Satz von Perron und Frobenius besitzt $\boldsymbol{P}_j^T$ genau einen nichtnegativen, normierten Eigenvektor $\hat{\boldsymbol{\pi}}_j$ zum Eigenwert 1.

Ist $r = 1$, gibt es also genau eine abgeschlossene, irreduzible, rekurrente Menge, so besitzt die Markov Kette eine eindeutige stationäre Verteilung $\hat{\boldsymbol{\pi}}^T = (\hat{\boldsymbol{\pi}}_1^T, \mathbf{0}^T)$.

Sonst ist jeder Vektor

$$\boldsymbol{\pi}^T = (\lambda_1 \hat{\boldsymbol{\pi}}_1^T, \ldots, \lambda_r \hat{\boldsymbol{\pi}}_r^T, \mathbf{0}^T)$$

mit $\lambda_j \geq 0$, $\sum_{j=1}^r \lambda_j = 1$, eine stationäre Verteilung.

Wir fassen unsere Ergebnisse zusammen.

Satz 5.61. *Es sei $\{X_n : n \in \mathbb{N}_0\}$ eine homogene Markov Kette mit endlichem Zustandsraum, und es sei die Übergangsmatrix $\boldsymbol{P}$ gemäß (5.10)) zerlegt. Dann gilt*

(i) Es gibt stets eine stationäre Verteilung.

(ii) Existiert genau eine rekurrente, abgeschlossene, irreduzible Menge von Zuständen, so ist die stationäre Verteilung eindeutig.

(iii) Jede stationäre Verteilung besitzt die Darstellung

$$\boldsymbol{\pi}^T = (\lambda_1 \hat{\boldsymbol{\pi}}_1^T, \ldots, \lambda_r \hat{\boldsymbol{\pi}}_r^T, \mathbf{0}^T), \quad \textit{wobei } \lambda_j \geq 0 \textit{ und } \sum_{j=1}^n \lambda_j = 1$$

gilt und $\hat{\boldsymbol{\pi}}_j$ der eindeutige positive Eigenvektor von $\boldsymbol{P}_j^T$ mit $\|\hat{\boldsymbol{\pi}}_j\|_1 = 1$ ist.

Beispiel 5.62. Wir kehren zu Beispiel 5.51. zurück.

Mit $C_1 = \{8\}$, $C_2 = \{2, 3, 4\}$, $T = \{1, 5, 6, 7\}$ geht $\boldsymbol{P}$ über in

$$\left(\begin{array}{cccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.25 & 0 & 0 & 0.5 & 0 & 0 & 0.25 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0.5 & 0 & 0 & 0 & 0.25 & 0.25 & 0
\end{array}\right)
\begin{array}{c}
C_1 \\ C_2 \\ C_2 \\ C_2 \\ T \\ T \\ T \\ T
\end{array} .$$

$\boldsymbol{P}_1^T = (1)$ besitzt den positiven, normierten Eigenvektor $\hat{\boldsymbol{\pi}}_1 = (1)$ und

$$\boldsymbol{P}_2^T = \begin{pmatrix} 0 & 0 & 1/2 \\ 1 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

besitzt den positiven, normierten Eigenvektor $\hat{\boldsymbol{\pi}}_2 = (0.2, 0.4, 0.4)^T$.

Daher ist die Menge aller stationären Verteilungen (in der ursprünglichen Reihenfolge der Zustände)

$$\boldsymbol{\pi} = (0, 0.2\alpha, 0.4\alpha, 0.4\alpha, 0, 0, 0, 1-\alpha)^T \ , \ \alpha \in [0,1] \ . \qquad \square$$

Beispiel 5.63. Ist der Zustandsraum abzählbar, so braucht keine stationäre Verteilung zu existieren.

Ist $\{X_n : n \in \mathbb{N}_0\}$ die Markov Kette im Zustandsraum $\mathbb{N}_0$ mit $P(X_{n+1} = i+1 | X_n = i) = 1$ für alle i, so existiert keine stationäre Verteilung, denn aus $P(X_n = j) = P(X_{n+1} = j)$ für alle $j \in \mathbb{N}_0$ würde folgen

$$P(X_n = j) = P(X_{n+1} = j) = \sum_k P(X_{n+1} = j | X_n = k) P(X_n = k) = P(X_n = j-1) \ ,$$

d.h. $P(X_n = j) = \text{const}$ (bzgl. j), und dies ist bei nicht endlichem Zustandsraum unmöglich. $\square$

Beispiel 5.64. In dem letzten Beispiel waren alle Zustände transient. Auch wenn alle Zustände rekurrent sind, braucht keine stationäre Verteilung zu existieren.

Ist $\{X_n : n \in \mathbb{N}_0\}$ die symmetrische Irrfahrt, d.h. $p_{i,i-1} = p_{i,i+1} = 0.5$ und ist $\boldsymbol{\pi} = (\pi_j)$ eine stationäre Verteilung, so folgt

$$\begin{aligned} \pi_j &= P(X_{n+1} = j) = \sum_k P(X_{n+1} = j | X_n = k) P(X_n = k) \\ &= \frac{1}{2}(P(X_n = j-1) + P(X_n = j+1)) = \frac{1}{2}(\pi_{j-1} + \pi_{j+1}). \end{aligned}$$

Hieraus erhält man $\pi_j \geq \pi_{j-1}$ für alle j oder $\pi_j \leq \pi_{j-1}$ für alle j, und beides ist nur möglich für $\pi_j = 0$ für alle j. $\square$

Wir kommen zu der Frage nach stationären Verteilungen bei Markov Ketten mit abzählbarem Zustandsraum zurück.

Wir untersuchen das Langzeitverhalten von Markov Ketten, fragen also, welche Aussagen über die Verteilung von X_n für große n gemacht werden können.

Beispiel 5.65. Als Beispiel betrachten wir ein einfaches Modell der Wettervorhersage. Es beschreibe X_n das Wetter am n-ten Tag, und zwar

$$
\begin{aligned}
[X_n = 0] &\mathrel{\hat{=}} \text{es regnet am } n\text{-ten Tag,} \\
[X_n = 1] &\mathrel{\hat{=}} \text{es regnet am } n\text{-ten Tag nicht.}
\end{aligned}
$$

Wir nehmen an, dass das Wetter am n-ten Tag nur vom Wetter am $(n-1)$-ten Tag abhängt gemäß

$$
P(X_{n+1} = 0 \,|\, X_n = 1) = p, \quad P(X_{n+1} = 1 \,|\, X_n = 0) = q \quad \text{mit } 0 < p, q < 1 \; .
$$

Es ist anschaulich klar, dass die Verteilung von X_n für große n nicht von der Verteilung von X_0 abhängig ist, sondern dass $P(X_n = i) = \pi_i$, $i = 0, 1$ gilt, wobei $365 \cdot \pi_0$ die mittlere Anzahl der Regentage im Jahr ist.

Die Übergangsmatrix dieses Prozesses ist

$$
\boldsymbol{P} = \begin{pmatrix} 1-q & q \\ p & 1-p \end{pmatrix} \; ,
$$

und man rechnet leicht nach (Induktion!), dass gilt

$$
\boldsymbol{P}^n = \begin{pmatrix} \dfrac{p}{p+q} + q \cdot \dfrac{(1-p-q)^n}{p+q} & \dfrac{q}{p+q} - q \cdot \dfrac{(1-p-q)^n}{p+q} \\ \dfrac{p}{p+q} - p \cdot \dfrac{(1-p-q)^n}{p+q} & \dfrac{q}{p+q} + p \cdot \dfrac{(1-p-q)^n}{p+q} \end{pmatrix} .
$$

Wegen $|1-p-q| < 1$ konvergiert $\boldsymbol{P}^n$ gegen die Matrix

$$
\hat{\boldsymbol{P}} = \begin{pmatrix} \dfrac{p}{p+q} & \dfrac{q}{p+q} \\ \dfrac{p}{p+q} & \dfrac{q}{p+q} \end{pmatrix} \; .
$$

$\hat{\boldsymbol{P}}$ besitzt identische Zeilen

$$
\boldsymbol{\pi}^T = (\pi_0 \quad \pi_1) = \begin{pmatrix} \dfrac{p}{p+q} & \dfrac{q}{p+q} \end{pmatrix} .
$$

Man beachte, dass $\boldsymbol{\pi}$ eine stationäre Verteilung des Prozesses ist.

Ist $\boldsymbol{\pi}^0 = \begin{pmatrix} \pi_0^0 & \pi_1^0 \end{pmatrix}^T$ eine beliebige Startverteilung, so gilt für große n

$$
P(X_n = 0) = (\boldsymbol{\pi}^0)^T \boldsymbol{P}^n \approx \pi_0^0 \frac{p}{p+q} + \pi_1^0 \frac{p}{p+q} = \frac{p}{p+q} = \pi_0
$$

und genauso $P(X_n = 1) \approx \dfrac{q}{p+q} = \pi_1$.

Die stationäre Verteilung π ist also stabil, d.h. für jede Startverteilung konvergiert die Verteilung von X_n gegen π. □

Ein solches Konvergenzresultat ist i.a. nicht zu erwarten. Gibt es z.B. für den Prozess zwei rekurrente, abgeschlossene Mengen und ist die Startverteilung in einer der beiden Mengen konzentriert, so muss auch die Grenzverteilung in dieser Menge konzentriert sein.

Ein anderes Beispiel dafür, dass $\boldsymbol{P}^n$ nicht konvergiert, erhält man, wenn man in dem Wettervorhersagemodell $p = q = 1$ wählt. Dann ist

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \boldsymbol{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; \ldots, \; \boldsymbol{P}^{2n-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \boldsymbol{P}^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

Die Besonderheit in diesem Beispiel (die die Konvergenz verhindert) ist, dass $P(X_{n+1} = i \,|\, X_n = i) = 0$ für $i = 0, 1$ gilt, dass ein Teilchen also immer mindestens zwei Schritte benötigt, um an seinen Ort zurückzukehren.

Definition 5.66. *Es sei $\{X_n \,:\, n \in \mathbb{N}_0\}$ eine homogene Markov Kette mit Übergangsmatrix $\boldsymbol{P}$. Dann heißt d_i, der größte gemeinsame Teiler der Menge*

$$\{n \in \mathbb{N} \,:\, p_{ii}(n) > 0\},$$

die **Periode** *des Zustands i. Gilt $d_i = 1$, so heißt der Zustand* **aperiodisch**.

Bemerkung 5.67. Offensichtlich gilt $1 \leq d_i \leq \min\{n \,:\, p_{ii}(n) > 0\}$ und daher $d_i = 1$, falls $p_{ii} > 0$. Ferner gilt sicher $p_{ii}(nd_i + k) = 0$ für alle $n \in \mathbb{N}_0$ und alle $k = 1, \ldots, d_i - 1$. Ein Teilchen, das sich zu irgendeinem Zeitpunkt in dem Ort i befindet, kann also frühestens nach d_i Zeitschritten nach i zurückkehren. □

Wir zeigen nun, dass die Periode auf abgeschlossenen, irreduziblen Mengen konstant ist.

Lemma 5.68. *Sind i und j Zustände mit $i \to j$ und $j \to i$, so gilt $d_i = d_j$.*

Beweis: Wegen $i \to j$ und $j \to i$ existieren $n_1, n_2 \in \mathbb{N}$ mit $p_{ij}(n_1) > 0, p_{ji}(n_2) > 0$, und daher gilt

$$p_{ii}(n_1 + n_2) = \sum_k p_{ik}(n_1)p_{ki}(n_2) \geq p_{ij}(n_1)p_{ji}(n_2) > 0 \;,$$

d.h. d_i ist ein Teiler von $n_1 + n_2$.

Ist $p_{jj}(n) > 0$, so folgt genauso

$$p_{ii}(n_1 + n + n_2) \geq p_{ij}(n_1)p_{jj}(n)p_{ji}(n_2) > 0 \;,$$

d.h. d_i ist ein Teiler von $n_1 + n + n_2$, und da d_i Teiler von $n_1 + n_2$ ist, muss d_i auch Teiler von n sein.

Es ist also d_i ein gemeinsamer Teiler der Menge $\{n : p_{jj}(n) > 0\}$, und da d_j der größte gemeinsame Teiler dieser Menge ist, ist d_i ein Teiler von d_j.

Aus Symmetriegründen ist auch d_j ein Teiler von d_i, und daher gilt $d_i = d_j$. ∎

Als Hilfsmittel für unsere weiteren Untersuchungen benötigen wir den folgenden Satz.

Satz 5.69. (diskretes Erneuerungstheorem)
Es seien $\{a_n\}, \{b_n\}, \{x_n\}$ reelle Zahlenfolgen mit

$$a_n \geq 0,\ \sum_{n=0}^{\infty} a_n = 1,\ \sum_{n=0}^{\infty} n a_n > 0,\ b_n \geq 0,\ \sum_{n=0}^{\infty} b_n < \infty\ ,$$

es sei der größte gemeinsame Teiler von $\{n : a_n > 0\}$ gleich 1, und es gelte die Erneuerungsgleichung

$$x_n - \sum_{k=0}^{n} a_{n-k} x_k = b_n,\ n = 0, 1, \dots\ .$$

Ist die Folge $\{x_n\}$ beschränkt, so existiert $\lim_{n\to\infty} x_n$ und

$$\lim_{n\to\infty} x_n = \sum_{n=0}^{\infty} b_n \Big/ \sum_{n=0}^{\infty} n a_n\ , \tag{5.12}$$

wobei die rechte Seite von (5.12) gleich 0 zu setzen ist, falls $\sum_{n=1}^{\infty} n a_n$ nicht konvergiert.

Beweis: **Karlin** [19, p. 67 ff.]

Satz 5.70. *Es sei $\{X_n : n \in \mathbb{N}_0\}$ eine homogene, rekurrente, irreduzible, aperiodische Markov Kette (d.h. alle $n \in \mathbb{N}_0$ seien rekurrent und aperiodisch, und $\mathbb{N}_0$ sei irreduzibel), und es sei*

$$m_i := E(T_i \,|\, X_0 = i) = \sum_{n=1}^{\infty} n P(T_i = n \,|\, X_0 = i)$$

die mittlere Rückkehrzeit für den Zustand i. Dann gilt

$$\lim_{n\to\infty} p_{ji}(n) = \frac{1}{m_i}$$

für alle Zustände i, j.

Beweis: Für $j = i$ folgt die Aussage sofort aus Satz 5.69., denn nach Lemma 5.33. gilt für $n \geq 1$

$$p_{ii}(n) = \sum_{k=0}^{n} P(T_i = k \mid X_0 = i) p_{ii}(n-k).$$

Wegen $p_{ii}(0) = 1$ ist also die Erneuerungsgleichung mit $x_n = p_{ii}(n)$, $a_n = P(T_i = n \mid X_0 = i)$ und $b_n = \delta_{n0}$ erfüllt.

Um die Behauptung auch für $j \neq i$ nachzuweisen, gehen wir von (vgl. Lemma 5.33.)

$$p_{ji}(n) = \sum_{k=0}^{n} P(T_i = k \mid X_0 = j) p_{ii}(n-k)\ ,\ n \geq 0\ ,$$

aus, d.h. von

$$y_n = \sum_{k=0}^{n} a_{n-k} x_k$$

mit $a_k \geq 0$, $\sum_{k=0}^{\infty} a_k = 1$, $\lim\limits_{k\to\infty} x_k = c$.

Wir haben zu zeigen, dass $\lim\limits_{k\to\infty} y_k = c$ gilt.

$$y_n - c = \sum_{k=0}^{n} a_{n-k} x_k - c \sum_{m=0}^{\infty} a_m = \sum_{k=0}^{n} a_{n-k}(x_k - c) - c \sum_{m=n+1}^{\infty} a_m\ .$$

Es sei $\varepsilon > 0$ vorgegeben und $K = K(\varepsilon)$ so groß, dass $|x_k - c| < \varepsilon/3$ für alle $k \geq K$. Dann folgt mit $M := \max\limits_k |x_k - c|$ für $n > K$

$$\begin{aligned}
|y_n - c| &\leq \sum_{k=0}^{K} a_{n-k}|x_k - c| + \sum_{k=K+1}^{n} a_{n-k}|x_k - c| + |c| \sum_{m=n+1}^{\infty} a_m \\
&\leq M \sum_{k=0}^{K} a_{n-k} + \frac{\varepsilon}{3} \sum_{k=K+1}^{n} a_{n-k} + |c| \sum_{m=n+1}^{\infty} a_m\ .
\end{aligned}$$

Wir wählen nun $N = N(\varepsilon) > K$ so, dass $|c| \sum_{m=n+1}^{\infty} a_m < \varepsilon/3$ und $\sum_{k=0}^{K} a_{n-k} = \sum_{m=n-K}^{n} a_m < \varepsilon/(3M)$ sowie $\sum_{m=n+1}^{\infty} a_m < \varepsilon/(3c)$ für alle $n \geq N$ gilt. Dann folgt $|y_n - c| < \varepsilon$ für alle $a \geq N$. ∎

Bemerkung 5.71. Es ist klar, dass Satz 5.70. sofort auf den Fall übertragen werden kann, dass der Zustandsraum in rekurrente, irreduzible, aperiodische Mengen $C_1, \ldots, C_r$ zerlegt werden kann. Dann konvergiert $\boldsymbol{P}^n$ gegen eine Matrix, in der alle Zeilen, die zu einer Menge C_i gehören, übereinstimmen. □

Beispiel 5.72. Wir betrachten die homogene Markov Kette mit der Übergangsmatrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{pmatrix}\ .$$

Dann gilt $1 \to 2 \to 3 \to 4 \to 1$, d.h. die Kette ist irreduzibel und wegen $p_{44} > 0$ aperiodisch.

(Dass die Zustände 1, 2 und 3 aperiodisch sind, sieht man so direkt: Es gilt $p_{11} = p_{11}(2) = p_{11}(3) = 0$, $p_{11}(4) = 0.1$, $p_{11}(5) = 0.04$, also $d_1 = 1$. Aus $p_{22}(1) = p_{22}(2) = 0$, $p_{22}(3) = 0.2$, $p_{22}(4) = 0.18$ folgt $d_2 = 1$, und aus $p_{33}(1) = 0$, $p_{33}(2) = 0.3$, $p_{33}(3) = 0.32$ erhält man $d_3 = 1$).

Die Potenzen $\boldsymbol{P}^n$ konvergieren gegen

$$\begin{pmatrix} 0.05 & , & 0.15 & , & 0.30 & , & 0.50 \\ 0.05 & , & 0.15 & , & 0.30 & , & 0.50 \\ 0.05 & , & 0.15 & , & 0.30 & , & 0.50 \\ 0.05 & , & 0.15 & , & 0.30 & , & 0.50 \end{pmatrix}$$

(in $\boldsymbol{P}^{30}$ ist dieses Ergebnis auf 9 Stellen genau erreicht), also eine Matrix mit den identischen Zeilen $\boldsymbol{\pi}^T = (0.05, 0.15, 0.3, 0.5)$.

Der Vektor $\boldsymbol{\pi}$ ist eine stationäre Verteilung der betrachteten Markov Kette. □

Definition 5.73. *Es sei $\{X_n : n \in \mathbb{N}_0\}$ eine Markov Kette und für einen rekurrenten Zustand i sei $m_i := E(T_i \mid X_0 = i)$. i heißt* **positiv rekurrent** *, falls $m_i < \infty$ gilt, und* **null rekurrent** *, falls $m_i = \infty$.*

Bemerkung 5.74. Ist i ein aperiodischer, positiv rekurrenter Zustand und gilt $i \to j$ und $j \to i$, so ist auch j positiv rekurrent, denn zunächst existieren n_1, n_2 mit $p_{ij}(n_1) > 0, p_{ji}(n_2) > 0$, und daher folgt (wie in Lemma 5.68.)

$$p_{jj}(n_2 + n + n_1) \geq p_{ji}(n_2) p_{ii}(n) p_{ij}(n_1) \ ,$$

für $n \to \infty$ also

$$\frac{1}{m_j} \geq p_{ji}(n_2) \frac{1}{m_i} p_{ij}(n_1) > 0.$$

Die Existenz der Grenzwerte $\lim\limits_{n\to\infty} p_{ii}(n)$ und $\lim\limits_{n\to\infty} p_{jj}(n)$ folgt so: Es sei C diejenige irreduzible, abgeschlossene Menge von Zuständen, die i und j enthält. Dann ist die Restriktion der Markov Kette auf C eine homogene, rekurrente, irreduzible, aperiodische Markov Kette, und hierauf kann man Satz 5.70. anwenden. □

Satz 5.75. *Es sei $\{X_n : n \in \mathbb{N}_0\}$ eine homogene, positiv rekurrente, irreduzible, aperiodische Markov Kette mit dem Zustandsraum $\mathbb{N}_0$, und es sei $\pi_j = \lim\limits_{n\to\infty} p_{jj}(n)$. Dann existiert genau eine stationäre Verteilung, nämlich $\boldsymbol{\pi} = (\pi_j)$.*

Beweis: Für alle n und m gilt $1 = \sum_{j=0}^{\infty} p_{ij}(n) \geq \sum_{j=0}^{m} p_{ij}(n)$, für $n \to \infty$ erhält man $\sum_{j=0}^{m} \pi_j \leq 1$, und da dies für alle m richtig ist, folgt auch $\sum\limits_{j=0}^{\infty} \pi_j \leq 1$.

Es gilt

$$p_{ij}(n+1) = \sum_{k=0}^{\infty} p_{ik}(n)p_{kj} \geq \sum_{k=0}^{m} p_{ik}(n)p_{kj} \ .$$

Für $n \to \infty$ folgt $\pi_j \geq \sum_{k=0}^{m} \pi_k p_{kj}$, und da die linke Seite von m unabhängig ist, liefert $m \to \infty$

$$\pi_j \geq \sum_{k=0}^{\infty} \pi_k p_{kj}.$$

Durch zweimalige Anwendung dieser Ungleichung folgt

$$\pi_i \geq \sum_{j=0}^{\infty} \pi_j p_{ji} \geq \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj}p_{ji} = \sum_{k=0}^{\infty} \pi_k p_{ki}(2) \ , \tag{5.13}$$

und nach n-maliger Anwendung entsprechend

$$\pi_j \geq \sum_{k=0}^{\infty} \pi_k p_{kj}(n) \quad \text{für alle} \quad n \in \mathbb{N} \ .$$

Wir nehmen nun an, dass für ein j und ein n eine strikte Ungleichung gilt. Dann erhält man durch Summation über j den Widerspruch

$$\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{kj}(n) = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj}(n) = \sum_{k=0}^{\infty} \pi_k \ .$$

Also ist $\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}(n)$ für alle j und alle n. Lässt man hierin n gegen ∞ gehen, so erhält man (da $\sum_{k=0}^{\infty} \pi_k$ konvergiert und da $p_{kj}(n)$ gleichmäßig beschränkt ist)

$$\pi_j = \sum_{k=0}^{\infty} \pi_k \lim_{n\to\infty} p_{kj}(n) = \pi_j \sum_{k=0}^{\infty} \pi_k \quad \text{für alle } j.$$

Wegen $\pi_j > 0$ (j positiv rekurrent!) folgt $\sum_{k=0}^{\infty} \pi_k = 1$, d.h. wegen $\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}$ ist $\boldsymbol{\pi}$ eine stationäre Verteilung.

Die Eindeutigkeit erhält man so: Es sei $\{x_j\}$ eine stationäre Verteilung der Kette. Dann gilt (wie oben)

$$x_k = \sum_{j=0}^{\infty} x_j p_{jk} = \sum_{j=0}^{\infty} x_j p_{jk}(n), \tag{5.14}$$

und für $n \to \infty$ folgt

$$x_k = \sum_{j=0}^{\infty} x_j \lim_{n\to\infty} p_{jk}(n) = \pi_k \sum_{j=0}^{\infty} x_j = \pi_k \ . \qquad \blacksquare$$

Bemerkung 5.76. Die letzte Schlussweise zeigt auch, dass eine homogene Markov Kette, die nur transiente oder null rekurrente Zustände besitzt, keine stationären Verteilungen haben kann. Denn aus (5.14) würde ja wegen Korollar 5.39. oder Satz 5.70. folgen

$$x_k = \sum_{j=0}^{\infty} x_j \lim_{n\to\infty} p_{jk}(n) = 0 \quad \text{für alle } k \ . \qquad \square$$

Bemerkung 5.77. Besitzt eine homogene Markov Kette mehr als eine irreduzible, positiv rekurrente Menge von Zuständen $C_1, \ldots, C_r$ und sind alle positiv rekurrenten Zustände aperiodisch, so erhält man zu jeder Menge C_m eine stationäre Verteilung $(\boldsymbol{\pi}^{(m)})^T$ (=Zeile in $\lim_{n\to\infty}(p_{jk}(n))$ zu einem Zustand aus C_m), und die Menge aller stationären Verteilungen ist die Menge aller Konvexkombinationen

$$\left\{ \sum_{m=1}^{r} \alpha_m \boldsymbol{\pi}^{(m)} \ : \ \alpha_m \geq 0, \ \sum_{m=1}^{r} \alpha_m = 1) \right\}$$

der $\boldsymbol{\pi}^{(m)}$. $\square$

Abschließend betrachten wir den periodischen Fall. Es sei $\{X_n \ : \ n \in \mathbb{N}_0\}$ eine homogene, irreduzible, rekurrente Markov Kette der Periode $d > 1$. Dann gilt wegen $p_{ii}(nd + k) = 0$ für $k = 1, \ldots, d-1$ nach Lemma 5.33. für $n > 1$

$$\begin{aligned} p_{ii}(nd) &= \sum_{k=0}^{nd} P(T_i = nd - k \,|\, X_0 = i) p_{ii}(k) \\ &= \sum_{j=0}^{n} P(T_i = d(n-j) \,|\, X_0 = i) p_{ii}(jd) \ , \end{aligned}$$

und aus Satz 5.70. folgt wegen $P(T_i = nd + k \,|\, X_0 = i) = 0$ für $k = 1, \ldots, d-1$:

$$\begin{aligned} \lim_{n\to\infty} p_{ii}(nd) &= 1 \Big/ \sum_{n=0}^{\infty} n P(T_i = nd \,|\, X_0 = i) \\ &= 1 \Big/ \left(\frac{1}{d} \sum_{n=0}^{\infty} (nd) P(T_i = nd \,|\, X_0 = i) \right) \\ &= d \Big/ \sum_{k=0}^{\infty} k P(T_i = k \,|\, X_0 = i) = \frac{d}{m_i} \ . \end{aligned}$$

Ist i positiv rekurrent, so ist $\lim\limits_{n\to\infty} p_{ii}(nd) > 0$ und man kann wie im aperiodischen Fall $m_j > 0$ für jeden Zustand mit $i \to j$ und $j \to i$ schließen. Damit sind die Zustände einer irreduziblen, abgeschlossenen, rekurrenten Menge entweder alle positiv rekurrent oder alle null rekurrent.

Es seien nun i und j beliebige Zustände, und es sei $r_1 := \min\{n : p_{ij}(n) > 0\}$. Sei $n_1 \in \mathbb{N}$ mit $p_{ji}(n_1) > 0$. Dann folgt $p_{jj}(r_1 + n_1) \geq p_{ji}(n_1)p_{ij}(r_1) > 0$, und daher ist $r_1 + n_1$ ein Vielfaches von d. Ist $p_{ij}(n) > 0$, so folgt genauso aus $p_{ii}(n + n_1) \geq p_{ij}(n)p_{ji}(n_1) > 0$, dass $n+n_1$ Vielfaches von d ist, und daher ist $(n+n_1)-(n_1+r_1) = n - r_1$ Vielfaches von d. Da schließlich r_1 dargestellt werden kann als $r_1 = m_1 d + r$ mit $m_1 \in \mathbb{N}_0$ und $0 \leq r < d$, haben wir gezeigt:

Für alle Zustände i, j einer homogenen, irreduziblen, rekurrenten Markov Kette der Periode d gibt es ein $r \in \mathbb{N}_0$, $0 \leq r < d$, so dass

$$p_{ij}(n) = 0 \quad \text{für alle} \quad n \notin \{md + r \, : \, m \in \mathbb{N}_0\} \; .$$

Hiermit kann man nun ähnlich wie im Fall $j = i$ unter Benutzung von Lemma 5.33. und Satz 5.70. zeigen:

$$\lim_{n\to\infty} p_{ij}(nd + r) = \frac{d}{m_j} \; .$$

Da schließlich für jede konvergente Zahlenfolge $\{a_k\}$ mit $\lim\limits_{k\to\infty} a_k = a$ auch $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} a_k = a$ gilt, erhalten wir mit $m(n) := \max\{m \in \mathbb{N}_0 \, : \, n \geq md + r\}$

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}(k) &= \lim_{n\to\infty} \frac{m(n)}{n} \frac{1}{m(n)} \sum_{k=0}^{m(n)} p_{ij}(kd + r) \\
&= \lim_{n\to\infty} \frac{m(n)}{n} \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m} p_{ij}(kd + r) = \frac{1}{d} \cdot \frac{d}{m_j} = \frac{1}{m_j}.
\end{aligned}$$

Die obigen Überlegungen kann man natürlich für jede rekurrente, abgeschlossene Menge einer homogenen Markov Kette anstellen. Wir haben also insgesamt gezeigt, dass stets der Grenzwert

$$\mathbf{\Pi} = \lim \left(\frac{1}{n} \sum_{k=1}^{n} p_{ij}(k) \right)$$

existiert, wobei genau die Spalten, die zu positiv rekurrenten Zuständen gehören, von Null verschiedene Elemente enthalten.

Man kann wieder (wie im Falle $d = 1$) zeigen, dass

(i) die von Null verschiedenen Zeilen von $\mathbf{\Pi}$ stationäre Verteilungen sind

(ii) es genau eine stationäre Verteilung gibt, wenn genau eine positiv rekurrente, abgeschlossene Menge existiert

(iii) die stationären Verteilungen genau die Konvexkombinationen der zu den verschiedenen positiv rekurrenten Klassen gehörenden Zeilen von $\mathbf{\Pi}$ sind.

Beispiel 5.78. Wir betrachten die homogene Markov Kette mit der Übergangsmatrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} .$$

Dann gilt ($*$ bezeichnet von Null verschiedene Elemente)

$$\boldsymbol{P}^{2n} = \begin{pmatrix} * & 0 & * & 0 \\ 0 & * & 0 & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \end{pmatrix} , \quad \boldsymbol{P}^{2n+1} = \begin{pmatrix} 0 & * & 0 & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \\ * & 0 & * & 0 \end{pmatrix} .$$

Es gilt also $d = 2$ (denn $p_{33}(2) > 0$), und es existiert nicht $\lim\limits_{n\to\infty} \boldsymbol{P}^n$.

$\dfrac{1}{n}\sum\limits_{k=1}^{n} \boldsymbol{P}^k$ ist für

$$n = 100 \quad : \quad \begin{pmatrix} 0.164 & , & 0.171 & , & 0.336 & , & 0.329 \\ 0.164 & , & 0.164 & , & 0.336 & , & 0.336 \\ 0.168 & , & 0.164 & , & 0.332 & , & 0.336 \\ 0.168 & , & 0.168 & , & 0.332 & , & 0.332 \end{pmatrix}$$

$$n = 1000 \quad : \quad \begin{pmatrix} 0.1664 & , & 0.1671 & , & 0.3336 & , & 0.3329 \\ 0.1664 & , & 0.1664 & , & 0.3336 & , & 0.3336 \\ 0.1668 & , & 0.1664 & , & 0.3332 & , & 0.3336 \\ 0.1668 & , & 0.1668 & , & 0.3332 & , & 0.3332 \end{pmatrix} .$$

Man erhält also (wenn auch sehr langsame) Konvergenz gegen eine Matrix mit identischen Zeilen $\boldsymbol{\pi}^T = (1/6, 1/6, 1/3, 1/3)$ und dies ist die eindeutige stationäre Verteilung der Markov Kette. $\square$

Literaturverzeichnis

[1] **A. O. Allen:** *Probability, Statistics, and Queuing Theory with Computer Applications*. Academic Press, New York 1978

[2] **K. Behnen**, G. Neuhaus**:** *Grundkurs Stochastik*. Teubner, Stuttgart 1984

[3] **F. Beichelt:** *Stochastik für Ingenieure*. Teubner Verlag, Stuttbart 1995

[4] **O. Beyer**, H. J. Girlich, H.-U. Zschiesche**:** *Stochastische Prozesse und Modelle*. Harri Deutsch, Frankfurt a. Main 1978

[5] **O. Beyer**, H. Hacke, V. Pieper, J. Tiedge**:** *Wahrscheinlichkeitsrechnung und mathematische Statistik*. Harri Deutsch, Frankfurt a. Main 1980

[6] **A. T. Bharucha-Reid:** *Elements of the Theory of Markov Processes and Their Applications*. McGraw-Hill, New York 1960

[7] **U. N. Bhat:** *Elements of Applied Stochastic Processes*. Wiley, New York 1972

[8] **L. Breiman:** *Probability and Stochastic Processes. With a View Toward Applications*. Houghton Mifflin, Boston 1969

[9] **A. B. Clarke**, R. L. Disney**:** *Probability and Random Processes: A First Course with Applications*. Wiley, New York 1985

[10] **H. Cramer:** *Mathematical Methods of Statistics*. Princeton 1946

[11] **J. L. Doob:** *Stochastic Processes*. Wiley, New York 1953

[12] **W. Feller:** *An Introduction to Probability Theory and its Applications*. Wiley, New York 1968

[13] **D. Freedman:** *Markov Chains*. Holden Day, San Francisco 1971

[14] **K. W. Gaede:** *Zuverlässigkeit. Mathematische Modelle*. Hanser, München 1977

[15] **W. Hengartner**, R. Theodorescu**:** *Einführung in die Monte Carlo Methode*. Hanser, München 1978

[16] **K. Hinderer:** *Grundbegriffe der Wahrscheinlichkeitstheorie*. Springer, Berlin 1972

[17] **P. G. Hoel**, S. C. Port, C. J. Stone**:** *Introduction to Stochastic Processes*. Houghton Mifflin, Boston 1972

[18] **G. Hübner:** *Stochastik. Eine anwendungsorientierte Einführung für Informatiker, Ingenieure und Mathematiker*. Vieweg Verlag, Braunschweig 1996

[19] **S. Karlin:** *A First Course in Stochastic Processes*. Academic Press, New York 1966

[20] **J. E. C. Kingman**, S. J. Taylor**:** *Introduction to Measure and Probability*. Cambridge University Press, Cambridge 1966

[21] **I. Miller**, J. E. Freund**:** *Probability and Statistics for Engineers*. Prentice Hall, Englewood Cliffs 1985

[22] **P. A. P. Moran:** *An Introduction to Probability Theory*. Clarendon Press, Oxford 1984

[23] **A. Parzen:** *Modern Probability Theory and its Applications*. Wiley, New York 1980

[24] **A. Parzen:** *Stochastic Processes*. Holden Day, San Francisco 1962

[25] **V. S. Pugachev:** *Probability Theory and Mathematical Statistics for Engineers*. Pergamon Press, Oxford 1984

[26] **D. Stoyan:** *Stochastik für Ingenieure und Naturwissenschaftler*. Akademie Verlag, Berlin 1993

[27] **K. S. Trivedi:** *Probability & Statistics with Reliability, Queuing and Computer Science Applications*. Prentice Hall, Englewood Cliffs 1982

[28] **P. Weiß:** *Stochastische Modelle für Anwender*. Teubner, Stuttgart 1987

Index