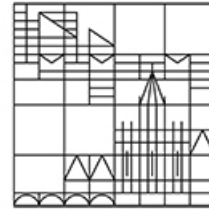


Fachbereich Informatik und  
Informationswissenschaft

Universität  
Konstanz



**Skriptum**  
**zur Vorlesung**  
**Mathematische Grundlagen 2**

*gehalten im Sommersemester 2010*

*von*

*Sven Kosub*

**20. Juli 2010**

*Version v1.12*

---



---

# Inhaltsverzeichnis

---

<b>5</b>	<b>Kombinatorik</b>	<b>1</b>
5.1	Grundregeln des Abzählens . . . . .	1
5.2	Urnenmodelle . . . . .	3
5.3	Binomialkoeffizienten . . . . .	5
5.4	Permutationen . . . . .	10
5.5	Weitere Abzählprinzipien . . . . .	15
<b>6</b>	<b>Wahrscheinlichkeitstheorie</b>	<b>21</b>
6.1	Diskrete Wahrscheinlichkeitsräume . . . . .	21
6.2	Kombinatorische Prinzipien . . . . .	24
6.3	Bedingte Wahrscheinlichkeiten . . . . .	27
6.4	Unabhängigkeit . . . . .	31
6.5	Zufallsvariablen . . . . .	33
6.6	Wichtige diskrete Verteilungen . . . . .	39
<b>7</b>	<b>Analysis</b>	<b>43</b>
7.1	Konvergenz von Folgen . . . . .	43
7.2	Konvergenz von Reihen . . . . .	47
7.3	Potenzreihen . . . . .	50
<b>8</b>	<b>Lineare Algebra</b>	<b>55</b>
8.1	Lineare Räume . . . . .	55
8.2	Lineare Abbildungen . . . . .	60
8.3	Eigenwerte und Eigenvektoren . . . . .	64
	<b>Literaturverzeichnis</b>	<b>70</b>



Der Schwerpunkt in diesem einführenden Kapitel über Kombinatorik liegt auf dem Abzählen endlicher Mengen.

## 5.1 Grundregeln des Abzählens

**Lemma 5.1 (Summenregel)** *Es seien  $A_1, \dots, A_n$  endliche, paarweise disjunkte Mengen. Dann gilt:*

$$\|A_1 \cup \dots \cup A_n\| = \sum_{j=1}^n \|A_j\|$$

**Beweis:** Wegen der paarweisen Disjunktheit der Mengen kommt jedes Element aus  $A_1 \cup \dots \cup A_n$  in genau einer Menge  $A_j$  vor. ■

**Lemma 5.2** *Es seien  $A$  und  $B$  endliche Mengen. Es gibt genau dann eine Bijektion  $f : A \rightarrow B$ , wenn  $\|A\| = \|B\|$  gilt.*

**Beweis:** Siehe Satz 3.19 (aus dem Kapitel über Funktionen und Abbildungen). ■

**Lemma 5.3 (Produktregel)** *Es seien  $A_1, \dots, A_n$  endliche Mengen. Dann gilt:*

$$\|A_1 \times \dots \times A_n\| = \prod_{j=1}^n \|A_j\|$$

**Beweis:** Wir beweisen die Aussage mittels Induktion über die Anzahl  $n$  der Mengen.

- *Induktionsanfang:* Für  $n = 1$  ist die Aussage offensichtlich.
- *Induktionsschritt:* Es sei  $n > 1$ . Weiterhin seien  $A_1, \dots, A_n$  endliche Mengen. Wir setzen

$$\begin{aligned} A^* &=_{\text{def}} A_1 \times \dots \times A_{n-1} \\ B_y &=_{\text{def}} \{ (x_1, \dots, x_{n-1}, y) \mid (x_1, \dots, x_{n-1}) \in A^* \} \quad \text{für } y \in A_n \end{aligned}$$

Für die so definierten Mengen gelten folgende Eigenschaften:

- (i) Die Mengenfamilie  $\{ B_y \mid y \in A_n \}$  ist eine Partition von  $A_1 \times \dots \times A_n$ .

(ii) Für jedes  $y \in A_n$  ist die Funktion

$$f_y : B_y \rightarrow A^* : (x_1, \dots, x_{n-1}, y) \mapsto (x_1, \dots, x_{n-1})$$

eine Bijektion, d.h.  $\|B_y\| = \|A^*\|$  für alle  $y \in A_n$  (nach Lemma 5.2).

Damit erhalten wir:

$$\begin{aligned} \|A_1 \times \dots \times A_n\| &= \left\| \bigcup_{y \in A_n} B_y \right\| && \text{(nach Eigenschaft (i))} \\ &= \sum_{y \in A_n} \|B_y\| && \text{(nach Lemma 5.1 und Eigenschaft (i))} \\ &= \sum_{y \in A_n} \|A^*\| && \text{(nach Lemma 5.2 und Eigenschaft (ii))} \\ &= \|A^*\| \cdot \|A_n\| \\ &= \left( \prod_{j=1}^{n-1} \|A_j\| \right) \cdot \|A_n\| && \text{(nach Induktionsvoraussetzung)} \\ &= \prod_{j=1}^n \|A_j\| \end{aligned}$$

Damit ist das Lemma bewiesen. ■

**Lemma 5.4 (Potenzregel)** *Es seien  $A$  und  $B$  endliche Mengen mit  $\|A\| = m$  und  $\|B\| = n$ . Dann existieren genau  $n^m$  Funktionen  $f : A \rightarrow B$ .*

**Beweis:** Nach Lemma 5.2 dürfen wir  $A = \{1, \dots, m\}$  ohne Beeinträchtigung der Allgemeinheit annehmen. Jeder Funktion  $f : A \rightarrow B$  kann nun eindeutig (injektiv) ein Tupel  $(f(1), \dots, f(m)) \in B^m$  zugeordnet werden. Außerdem entspricht jedes Tupel (die Wertetabelle)  $(y_1, \dots, y_m) \in B^m$  einer Funktion  $f : A \rightarrow B : j \mapsto y_j$ . Damit ist die Zuordnung sowohl injektiv als auch surjektiv, also eine Bijektion. Aus Lemma 5.2 und Produktregel (Lemma 5.3) folgt somit

$$\|\{ f \mid f : A \rightarrow B \}\| = \|B^m\| = \|B\|^m = n^m.$$

Damit ist das Lemma bewiesen. ■

**Beispiel:** Wie viele boolesche Funktionen mit  $n$  Variablen gibt es? Die Antwort lautet  $\|\{ f \mid f : \{0, 1\}^n \rightarrow \{0, 1\} \}\| = 2^{2^n}$ .

**Korollar 5.5** Für endliche Mengen  $A$  mit  $\|A\| = n$  gilt  $\|\mathcal{P}(A)\| = 2^n$ .

**Beweis:** Wir konstruieren eine Bijektion zwischen  $\mathcal{P}(A)$  und der Menge der Funktionen  $f : A \rightarrow \{0, 1\}$ . Dazu definieren wir für eine Menge  $B \in \mathcal{P}(A)$  die Funktion:

$$c_B : A \rightarrow \{0, 1\} : x \mapsto \begin{cases} 1 & \text{falls } x \in B \\ 0 & \text{falls } x \notin B \end{cases}$$

Diese Zuordnung ist offensichtlich eine Bijektion zwischen  $\mathcal{P}(A)$  und der Menge der Funktionen  $f : A \rightarrow \{0, 1\}$ . Nach der Potenzregel (Lemma 5.4) und Lemma 5.2 gilt folglich

$$\|\mathcal{P}(A)\| = \|\{ f \mid f : A \rightarrow \{0, 1\} \}\| = 2^n.$$

Damit ist das Korollar bewiesen. ■

Die im Beweis von Korollar 5.5 angegebenen Funktionen haben einen Namen: Für eine Menge  $B \subseteq A$  heißt  $c_B$  die *charakteristische Funktion* von  $B$ .

## 5.2 Urnenmodelle

Urnenmodelle stellen ein typisches Szenario für kombinatorische Problemstellungen dar. Die einfachste Situation ist die folgende: In einer Urne liegen  $n$  unterscheidbare Kugeln, von den  $k$  Kugel gezogen werden dürfen. Die zu beantwortende Frage ist dann: Wie viele Möglichkeiten gibt es diese  $k$  Kugeln zu ziehen? Zur Präzisierung des Szenarios werden Unterschiede danach gemacht, ob

- die Reihenfolge, in der die Kugeln gezogen werden, eine Rolle spielt,
- gezogene Kugeln wieder zurückgelegt werden oder nicht.

Damit ergeben sich vier verschiedene Szenarios.

**Theorem 5.6** Die Anzahl der Möglichkeiten, aus einer Urne mit  $n$  Kugeln  $k$  Kugeln auszuwählen, ist durch folgende Tabelle gegeben:

	mit Zurücklegen	ohne Zurücklegen
mit Reihenfolge	$n^k$	$n^{\underline{k}} =_{\text{def}} \frac{n!}{(n-k)!}$
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k} =_{\text{def}} \frac{n!}{k!(n-k)!}$

Die im Theorem mitdefinierten Größen  $n^k$  und  $\binom{n}{k}$  heißen *fallende Faktorielle von  $n$  der Länge  $k$*  sowie *Binomialkoeffizient („ $n$  über  $k$ “).*

**Beispiel:** Wir geben für jedes der vier Szenarien Beispiele an:

	mit Zurücklegen	ohne Zurücklegen
mit Reihenfolge (geordnet)	PIN-Codes: <ul style="list-style-type: none"> <li>• <math>n = 10</math> Ziffern</li> <li>• <math>k = 4</math> Stellen</li> <li>• 10.000 Codes</li> </ul>	Wettkämpfe: <ul style="list-style-type: none"> <li>• <math>n = 10</math> Starter</li> <li>• <math>k = 3</math> Medaillen</li> <li>• 720 Siegerehrungen</li> </ul>
ohne Reihenfolge (ungeordnet)	Wahlen: <ul style="list-style-type: none"> <li>• <math>n = 3</math> Kandidaten</li> <li>• <math>k = 100</math> Wähler</li> <li>• 5.151 Wahlausgänge</li> </ul>	Lotto: <ul style="list-style-type: none"> <li>• <math>n = 49</math> Kugeln</li> <li>• <math>k = 6</math> Kugeln</li> <li>• 13.983.816 Ziehungen</li> </ul>

**Beweis:** Wir beweisen alle Fälle einzeln, aber aufeinander aufbauend:

- *Ziehen mit Zurücklegen, mit Reihenfolge:* Jede Auswahlmöglichkeit entspricht einer Funktion  $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ , wobei  $f(i)$  genau der Kugel entspricht, die als  $i$ -te Kugel gezogen wurde. Nach der Potenzregel (Lemma 5.4) gibt es somit  $n^k$  Möglichkeiten.
- *Ziehen ohne Zurücklegen, mit Reihenfolge:* Für die erste gezogene Kugel gibt es  $n$  Möglichkeiten, für die zweite gezogene Kugel gibt es  $n - 1$  Möglichkeiten. Für die  $k$ -te gezogene Kugel ( $k \leq n$ ) gibt es mithin noch  $n - k + 1$  Möglichkeiten. Insgesamt gibt es damit

$$n \cdot (n - 1) \cdot \dots \cdot (n - k + 1) = \frac{n!}{(n - k)!} = n^{\underline{k}}$$

Möglichkeiten.

- *Ziehen ohne Zurücklegen, ohne Reihenfolge:* Mit Berücksichtigung der Reihenfolge gibt es  $\frac{n!}{(n - k)!}$  Auswahlmöglichkeiten. Wenn die Reihenfolge keine Rolle mehr spielt, zählen alle Auswahlfolgen, bei denen die gleichen  $k$  Kugeln gezogen wurden, nur noch als eine Auswahlmöglichkeit. Dies sind gerade  $k!$  viele. Damit gibt es insgesamt

$$\frac{n!}{(n - k)!} \cdot \frac{1}{k!} = \frac{n!}{k!(n - k)!} = \binom{n}{k}$$

Möglichkeiten.

- *Ziehen mit Zurücklegen, ohne Reihenfolge:* Da jede Kugel mehrmals gezogen werden kann, die Reihenfolge jedoch keine Rolle spielt, ist nur wichtig, wie oft eine Kugel gezogen wird. Es sei also  $(a_1, \dots, a_n)$  ein Tupel mit den entsprechenden Anzahlen, wobei  $a_j$  gerade angibt, wie oft die Kugel  $j$  gezogen wird. Für ein Anzahltuplel  $(a_1, \dots, a_n)$  muss nun gelten:

$$(i) \ a_j \in \{0, \dots, k\} \text{ für alle } j \in \{1, \dots, n\}$$



$$(ii) \quad a_1 + \dots + a_n = k$$

Wir müssen nun zählen, wie viele derartige Tupel es geben kann. Dazu repräsentieren wir die Tupel in einer anderen Weise, die es uns ermöglicht, das Szenario zu wechseln. Wir verwenden  $k$ -mal das Symbol  $*$  und  $(n-1)$ -mal das Symbol  $|$ . Ein Anzahltupel  $(a_1, \dots, a_n)$  wird nun (injektiv) durch die Symbolfolge

$$\underbrace{**\dots*}_{a_1} | \underbrace{**\dots*}_{a_2} | \dots | \underbrace{**\dots*}_{a_n}$$

repräsentiert. Umgekehrt entspricht auch jede Symbolfolge, die  $k$ -mal das Symbol  $*$  und  $(n-1)$ -mal das Symbol  $|$  enthält, einem Anzahltupel mit obigen Eigenschaften. Demzufolge ist die Repräsentierung eine Bijektion zwischen der Menge der Anzahltupel und der Menge der Symbolfolgen. Die Anzahl möglicher Symbolfolgen zu bestimmen, entspricht aber gerade dem Ziehen von  $k$  Positionen für das Symbol  $*$  aus  $n+k-1$  möglichen Positionen ohne Zurücklegen und ohne Reihenfolge. Mithin gibt es insgesamt

$$\binom{n+k-1}{k}$$

Möglichkeiten.

Damit ist das Theorem bewiesen. ■

## 5.3 Binomialkoeffizienten

Aus dem letzten Abschnitt (Theorem 5.6) wissen wir, dass die Anzahl der Möglichkeiten aus  $n$  Kugeln  $k$  Kugeln ungeordnet ohne Zurücklegen zu ziehen gerade dem Binomialkoeffizienten  $\binom{n}{k}$  entspricht. Da Binomialkoeffizienten auch über die reine Kombinatorik hinaus wichtig sind, wollen in diesem Abschnitt die wichtigsten Eigenschaften von Binomialkoeffizienten festhalten. Dazu definieren wir den Binomialkoeffizienten noch einmal explizit: Für  $n, k \in \mathbb{N}$  definieren wir

$$\binom{n}{k} \stackrel{\text{def}}{=} \frac{n!}{k!(n-k)!}, \quad \text{mit } \binom{n}{k} = 0 \quad \text{für } k > n.$$

Ein einfache, sofort einsichtige Beobachtung ist:

$$\binom{n}{k} = \binom{n}{n-k}$$

Damit lässt sich der Binomialkoeffizient konsistent auch für negative Werte für  $k$  definieren:

$$\binom{n}{k} \stackrel{\text{def}}{=} 0 \quad \text{für } k \in \mathbb{Z} \setminus \mathbb{N}.$$

**Theorem 5.7 (PASCALSches Dreieck)** Für  $n \in \mathbb{N}_+$  und  $k \in \mathbb{N}$  gilt

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Wir geben zwei Beweise für das Theorem an.

**Beweis:** (*rechnerisch*) Wir führen eine Fallunterscheidung bezüglich der Werte von  $k$  durch:

- Für  $k = 0$  und  $n > 1$  gilt  $\binom{n}{0} = 1 = \binom{n-1}{-1} + \binom{n-1}{0}$ .

- Für  $0 < k < n$  rechnen wir aus:

$$\begin{aligned} \binom{n-1}{k-1} + \binom{n-1}{k} &= \frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!(n-1-k)!} \\ &= \frac{(n-1)!}{(k-1)!(n-k)!} \cdot \frac{k}{k} + \frac{(n-1)!}{k!(n-1-k)!} \cdot \frac{n-k}{n-k} \\ &= \frac{(n-1)! \cdot k}{k!(n-k)!} + \frac{(n-1)!(n-k)}{k!(n-k)!} \\ &= \frac{(n-1)!(k+n-k)}{k!(n-k)!} \\ &= \frac{(n-1)! \cdot n}{k!(n-k)!} \\ &= \binom{n}{k} \end{aligned}$$

- Für  $k = n$  und  $n > 1$  gilt  $\binom{n}{n} = 1 = \binom{n-1}{n-1} + \binom{n-1}{n}$ .

- Für  $k > n > 1$  gilt  $\binom{n}{k} = 0 = \binom{n-1}{k-1} + \binom{n-1}{k}$ .

Damit ist das Theorem durch Nachrechnen bewiesen. ■

**Beweis:** (*kombinatorisch*) Wir interpretieren die Gleichung als Bestimmung der Kardinalität von Mengen auf zwei verschiedene Weisen. Es seien  $n \in \mathbb{N}_+$  und  $k \in \mathbb{N}$ . Wir definieren folgende Mengenfamilien:

$$\begin{aligned} \mathcal{F} &=_{\text{def}} \{ \{a_1, \dots, a_k\} \mid a_i \in \{1, \dots, n\} \text{ und } a_i \neq a_j \text{ für } i \neq j \} \\ \mathcal{F}_+ &=_{\text{def}} \{ A \mid A \in \mathcal{F} \text{ und } 1 \in A \} \\ \mathcal{F}_- &=_{\text{def}} \{ A \mid A \in \mathcal{F} \text{ und } 1 \notin A \} \end{aligned}$$



**Beweis:** (*kombinatorisch*) Es seien  $x, y \in \mathbb{R}$  und  $n \in \mathbb{N}$  beliebig. Durch Ausmultiplizieren von  $(x + y)^n$  erhalten wir:

$$\begin{aligned}
 (x + y)^n &= \underbrace{x \cdot \cdots \cdot x \cdot x}_{n \text{ Faktoren}} + \\
 &+ x \cdot \cdots \cdot x \cdot y + \\
 &+ x \cdot \cdots \cdot y \cdot x + \\
 &+ x \cdot \cdots \cdot y \cdot y + \\
 &\vdots \\
 &+ \underbrace{y \cdot \cdots \cdot y \cdot y}_{n \text{ Faktoren}}
 \end{aligned}$$

Die Summanden können zusammengefasst werden zu Produkten von jeweils  $n$  Faktoren, von denen  $k$  Faktoren gerade  $y$  und  $n - k$  Faktoren gerade  $x$  sind. Die Summanden sind also von der Form  $x^{n-k}y^k$ , da die Reihenfolge bei der Multiplikation keine Rolle spielt. Die Anzahl der Produkte  $x^{n-k}y^k$  entspricht somit gerade dem Ziehen von  $k$  Kugeln (die Positionen für  $y$  im Produkt) aus  $n$  Kugeln (die Gesamtheit aller Positionen für Faktoren), d.h.  $\binom{n}{k}$ . Folglich gilt insgesamt:

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

Damit ist das Theorem bewiesen. ■

**Korollar 5.9** Für alle  $n \in \mathbb{N}_+$  gilt

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

**Beweis:** Nach dem Binomialtheorem gilt

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

Damit ist das Korollar bewiesen. ■

**Korollar 5.10** Für alle  $n \in \mathbb{N}$  gilt

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

**Beweis:** Nach dem Binomialtheorem gilt

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

Damit ist das Korollar bewiesen. ■

**Theorem 5.11 (VANDERMONDESche Identität)** Für  $k, m, n \in \mathbb{N}$  gilt

$$\binom{n+m}{k} = \sum_{j=0}^k \binom{n}{j} \binom{m}{k-j}.$$

**Beweis:** (*kombinatorisch*) Es seien  $A$  und  $B$  disjunkte Mengen mit  $\|A\| = n$  und  $\|B\| = m$ . Für jedes  $j \in \{0, \dots, k\}$  definieren wir die Mengenfamilie

$$\mathcal{X}_j =_{\text{def}} \{ X \mid X \subseteq A \cup B, \|X \cap A\| = j \text{ und } \|X \cap B\| = k - j \}$$

Es gibt  $\binom{n}{j}$  viele  $j$ -elementige Teilmengen von  $A$  und  $\binom{m}{k-j}$  viele  $(k-j)$ -elementige Teilmengen von  $B$ . Damit gilt

$$\|\mathcal{X}_j\| = \binom{n}{j} \binom{m}{k-j}.$$

Wegen  $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$  für  $i \neq j$  folgt nun

$$\binom{n+m}{k} = \sum_{j=0}^k \|\mathcal{X}_j\| = \sum_{j=0}^k \binom{n}{j} \binom{m}{k-j}.$$

Damit ist das Theorem bewiesen. ■

**Beispiel:** Die Beweistechnik für Theorem 5.11 heißt *Doppeltes Abzählen*. Wenn zum Beispiel in einer Vorlesung  $n + m$  Studenten sitzen,  $n$  weibliche und  $m$  männliche, wie viele verschiedene Gruppen mit genau  $k$  Studenten gibt es dann? Dies lässt sich auf zwei Arten bestimmen:

- Ohne Berücksichtigung des Geschlechts erhalten wir  $\binom{n+m}{k}$  Gruppen.
- Mit Berücksichtigung des Geschlechts zählen wir für jedes  $j \in \{0, 1, \dots, k\}$  alle Gruppen mit jeweils genau  $j$  weiblichen und genau  $k - j$  männlichen Studenten, damit also insgesamt  $\sum_{j=0}^k \binom{n}{j} \binom{m}{k-j}$  Gruppen.

Da wir über dieselbe Menge von Studenten argumentieren, sind bei Anzahlen gleich.

## 5.4 Permutationen

Es sei  $A$  eine endliche Menge mit  $\|A\| = n$ . Eine *Permutation* von  $A$  ist eine bijektive Funktion  $\pi : A \rightarrow A$ . Ohne Beeinträchtigung der Allgemeinheit setzen wir stets  $A = \{1, \dots, n\}$  voraus. Die Menge  $\{1, \dots, n\}$  notieren wir auch als  $[n]$ . Weiterhin definieren wir die Menge

$$\mathcal{S}_n =_{\text{def}} \{ \pi \mid \pi : [n] \rightarrow [n] \text{ ist eine Permutation} \},$$

die sogenannte *symmetrische Gruppe* von  $n$  Elementen.

**Theorem 5.12** Für alle  $n \in \mathbb{N}_+$  gilt  $\|\mathcal{S}_n\| = n!$ .

**Beweis:**  $\|\mathcal{S}_n\|$  entspricht dem Ziehen von  $n$  Kugeln aus einer Urne mit  $n$  Kugeln ohne Zurücklegen mit Berücksichtigung der Reihenfolge. Nach Theorem 5.6 gilt

$$\|\mathcal{S}_n\| = \frac{n!}{(n-n)!} = n!.$$

Damit ist das Theorem bewiesen. ■

Ohne Beweis geben wir folgendes Resultat über das Verhalten der Fakultätsfunktion an:

**Theorem 5.13 (STIRLINGSche Formel)** Für alle  $n \in \mathbb{N}_+$  gilt

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}},$$

wobei  $e = e^1 = 2,718281828459\dots$  die EULERSche Konstante ist.

Permutationen können auf verschiedene Arten geschrieben werden. Im Folgenden behandeln wir drei Schreibweisen:

**Matrixschreibweise:** Dazu schreiben die Permutation  $\pi : [n] \rightarrow [n]$  als  $2 \times n$ -Matrix der Form

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \pi(1) & \pi(2) & \pi(3) & \dots & \pi(n) \end{pmatrix}$$

Da  $\pi$  bijektiv ist, kommen alle Werte  $1, \dots, n$  in der zweiten Zeile vor.

**Beispiel:** Folgende Permutation ist Matrixschreibweise gegeben:

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 6 & 2 & 5 & 3 \end{pmatrix}$$

**Tupelschreibweise:** Im Prinzip genügt es, von der Matrixschreibweise lediglich die zweite Zeile zu übernehmen, d.h. Permutationen können angegeben werden in der Form

$$\pi = (\pi(1), \pi(2), \pi(3), \dots, \pi(n)).$$

**Beispiel:**  $\pi = (4, 1, 6, 2, 5, 3)$  ist obige Permutation in Tupelschreibweise.

**Zyklenschreibweise:** Die Zyklenschreibweise entsteht, wenn wir für  $x \in [n]$  die iterierte Hintereinanderausführung von  $\pi$  auf  $x$  betrachten. Dadurch entsteht eine Folge:

$$\begin{aligned} \pi^0(x) &=_{\text{def}} x, \\ \pi^1(x) &=_{\text{def}} \pi(x), \\ \pi^2(x) &=_{\text{def}} \pi(\pi(x)), \\ &\vdots \\ \pi^k(x) &=_{\text{def}} \pi(\pi^{k-1}(x)), \\ &\vdots \end{aligned}$$

Für jedes  $x \in [n]$  gibt es dann ein minimales  $0 < k < n$  mit  $\pi^k(x) = x$ .

**Beispiel:** Für die Permutation  $\pi = (4, 1, 6, 2, 5, 3)$  gilt

$$\begin{array}{llll} \pi^0(1) = 1, & \pi^1(1) = 4, & \pi^2(1) = 2, & \pi^3(1) = 1; \\ \pi^0(2) = 2, & \pi^1(2) = 1, & \pi^2(2) = 4, & \pi^3(2) = 2; \\ \pi^0(3) = 3, & \pi^1(3) = 6, & \pi^2(3) = 3; & \\ \pi^0(4) = 4, & \pi^1(4) = 2, & \pi^2(4) = 1, & \pi^3(4) = 4; \\ \pi^0(5) = 5, & \pi^1(5) = 5; & & \\ \pi^0(6) = 6, & \pi^1(6) = 3, & \pi^2(6) = 6. & \end{array}$$

Eine Folge  $x, \pi(x), \pi^2(x), \dots, \pi^{k-1}(x)$  mit minimalem  $k > 0$ , so dass  $\pi^k(x) = x$ , heißt *Zyklus* der Länge  $k$  und wird als  $(x \ \pi(x) \ \pi^2(x) \ \dots \ \pi^{k-1}(x))$  geschrieben.

**Beispiel:**  $\pi = (4, 1, 6, 2, 5, 3)$  enthält die Zyklen  $(1 \ 4 \ 2)$ ,  $(3 \ 6)$  und  $(5)$ .

Jede Permutation kann als *Produkt von Zyklen* geschrieben werden, indem die Zyklen einfach hintereinander gesetzt werden. Die Schreibweise ist jedoch nicht eindeutig. Insbesondere kann jeder Zyklus der Länge  $k$  auf genau  $k$  Arten geschrieben werden.

**Beispiel:** Die Permutation  $\pi = (4, 1, 6, 2, 5, 3)$  können wir als Produkt von Zyklen wie folgt schreiben:

$$(4, 1, 6, 2, 5, 3) = (1 \ 4 \ 2)(3 \ 6)(5)$$

$$\begin{aligned}
 &= (4\ 2\ 1)(6\ 3)(5) \\
 &= (5)(2\ 1\ 4)(6\ 3)
 \end{aligned}$$

Insbesondere gilt  $(1\ 4\ 2) = (4\ 2\ 1) = (2\ 1\ 4)$ .

Wie wir bereits aus dem letzten Semester wissen, ist die Hintereinanderausführung bijektiver Funktionen wiederum eine bijektive Funktion. Zum Abschluss dieses Abschnitts über Permutationen interessieren wir uns deshalb für die Struktur der Hintereinanderausführung von Permutationen. Diese kann durch gruppentheoretische Begriffe beschrieben werden:

**Definition 5.14** *Es seien  $G$  eine Menge und  $*$  :  $G \times G \rightarrow G : (x, y) \mapsto x * y$  eine Funktion (Operation). Das Paar  $(G, *)$  heißt Gruppe, falls folgende Eigenschaften erfüllt sind:*

1. Für alle  $x, y, z \in G$  gilt  $(x * y) * z = x * (y * z)$  (Assoziativität)
2. Es gibt ein  $e \in G$ , so dass  $e * x = x * e = x$  für alle  $x \in G$  gilt (neutrales Element)
3. Für alle  $x \in G$  gibt es ein  $x^{-1} \in G$ , so dass  $x^{-1} * x = x * x^{-1} = e$  gilt (inverses Element)

Gilt zusätzlich die Bedingung

4. Für alle  $x, y \in G$  gilt  $x * y = y * x$  (Kommutativität),

so heißt  $(G, *)$  kommutative (oder abelsche) Gruppe.

Bei der Definition einer Gruppe ist zu beachten, dass wir durch die Forderung an die Operation  $*$  eine Funktion zu sein, implizit vorausgesetzt haben, dass die Menge  $G$  abgeschlossen ist unter der Operation  $*$ . Mengen, die nicht abgeschlossen sind unter einer Operation  $*$ , können keine Gruppe sein. Insbesondere haben die drei geforderten Eigenschaften dann keinen Sinn.

**Proposition 5.15** *Es sei  $(G, *)$  eine Gruppe. Dann sind sowohl das neutrale Element als auch alle inversen Elemente eindeutig bestimmt.*

**Beweis:** Es seien  $e, e'$  zwei neutrale Elemente von  $(G, *)$ . Dann gilt:

$$\begin{aligned}
 e &= e * e' && \text{(da } e' \text{ neutrales Element ist)} \\
 &= e' && \text{(da } e \text{ neutrales Element ist)}
 \end{aligned}$$

Damit ist die Eindeutigkeit von  $e$  bewiesen.

Zum Nachweis der Eindeutigkeit der inversen Elemente seien nun für ein beliebiges  $x \in G$  zwei inverse Elemente  $y, z \in G$  von  $x$  gegeben. Dann gilt:

$$y = y * e \quad \text{(da } e \text{ neutrales Element ist)}$$



$$\begin{aligned}
&= y * (x * z) && \text{(da } z \text{ invers zu } x \text{ ist)} \\
&= (y * x) * z && \text{(wegen Assoziativität)} \\
&= e * z && \text{(da } y \text{ inverse zu } x \text{ ist)} \\
&= z && \text{(da } e \text{ neutrales Element ist)}
\end{aligned}$$

Somit ist das inverse Element von  $x$  eindeutig und die Proposition bewiesen. ■

**Beispiel:** Wir diskutieren einige Beispiele zur Verdeutlichung der Gruppeneigenschaften:

- $(\mathbb{Z}, +)$  mit der üblichen Addition ist eine abelsche Gruppe mit dem neutralen Element 0.
- $(\mathbb{N}, +)$  ist keine Gruppe, da z.B. für 1 kein inverses Element existiert.
- $(\mathbb{R} \setminus \{0\}, \cdot)$  mit der üblichen Multiplikation ist eine abelsche Gruppe mit neutralem Element 1.
- $(\mathbb{Q}, \cdot)$  ist keine Gruppe, da für 0 kein inverses Element existiert.
- $(\mathbb{Z}_n, +)$  mit  $\mathbb{Z}_n =_{\text{def}} \{0, 1, \dots, n-1\}$  und der Addition

$$+ : \mathbb{Z}_n \times \mathbb{Z}_n \rightarrow \mathbb{Z}_n : (x, y) \mapsto \text{mod}(x + y, n)$$

ist abelsche Gruppe mit neutralem Element 0.

- $(\mathbb{Z}_n^*, \cdot)$  mit  $\mathbb{Z}_n^* =_{\text{def}} \{x \in \mathbb{Z}_n \setminus \{0\} \mid \text{ggT}(x, n) = 1\}$  und der Multiplikation

$$\cdot : \mathbb{Z}_n^* \times \mathbb{Z}_n^* \rightarrow \mathbb{Z}_n^* : (x, y) \mapsto \text{mod}(x \cdot y, n)$$

ist abelsche Gruppe mit neutralem Element 1.

Für Permutationen  $f, g : [n] \rightarrow [n]$  sei die Hintereinanderausführung  $f \circ g$  wie folgt definiert:

$$f \circ g : [n] \rightarrow [n] : x \mapsto f(g(x)).$$

**Theorem 5.16** Für  $n \in \mathbb{N}_+$  ist  $(\mathcal{S}_n, \circ)$  eine Gruppe, die für  $n \geq 3$  nicht abelsch ist.

**Beweis:** Wir überprüfen zunächst die Gruppeneigenschaften:

1. *Assoziativität:* Für  $f, g, h \in \mathcal{S}_n$  müssen wir  $(f \circ g) \circ h = f \circ (g \circ h)$  zeigen, d.h. es muss  $((f \circ g) \circ h)(x) = (f \circ (g \circ h))(x)$  für alle  $x \in [n]$  gelten. Es sei also  $x \in [n]$ . Unter mehrmaliger Anwendung der Definition der Hintereinanderausführung erhalten wir:

$$\begin{aligned}
((f \circ g) \circ h)(x) &= (f \circ g)(h(x)) \\
&= f(g(h(x))) \\
&= f((g \circ h)(x)) \\
&= (f \circ (g \circ h))(x)
\end{aligned}$$

2. *neutrales Element:* Wir wählen  $e = \text{id}_n =_{\text{def}} (1, 2, \dots, n)$ .
3. *inverses Element:* Wir wählen für eine beliebige Permutation  $\pi$  die Umkehrfunktion  $\pi^{-1}$  als inverses Element, da Umkehrfunktionen von Bijektionen stets existieren und selbst wieder Bijektionen sind.

Damit ist der Nachweis erbracht, dass  $(\mathcal{S}_n, \circ)$  eine Gruppe ist. Um zu zeigen, dass  $(\mathcal{S}_3, \circ)$  (und somit auch  $(\mathcal{S}_n, \circ)$  für  $n \geq 3$ ) nicht abelsch ist, betrachten wir die zwei Permutationen  $\pi_1 =_{\text{def}} (1, 3, 2)$  und  $\pi_2 =_{\text{def}} (2, 1, 3)$ . Wir rechnen aus:

$$\pi_1 \circ \pi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

$$\pi_2 \circ \pi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

Mithin gilt  $\pi_1 \circ \pi_2 \neq \pi_2 \circ \pi_1$ . Damit ist das Theorem bewiesen. ■

Mit dem Nachweis der Gruppeneigenschaft für die symmetrische Gruppe  $(\mathcal{S}_n, \circ)$  gelten die allgemeinen Rechenregeln für Gruppen. Einige davon sind in folgendem Theorem zusammengefasst.

**Theorem 5.17** *Es sei  $(G, *)$  eine Gruppe.*

1. Für alle  $x, y \in G$  gilt  $(x * y)^{-1} = y^{-1} * x^{-1}$ .
2. Für alle  $x \in G$  gilt  $x = (x^{-1})^{-1}$ .
3. Für alle  $a, b \in G$  und für alle  $x \in G$  gilt:

$$(a) \quad a * x = b \iff x = a^{-1} * b$$

$$(b) \quad x * a = b \iff x = b * a^{-1}$$

Insbesondere die dritte Aussage des Theorems ist bemerkenswert. Sie garantiert stets die Existenz und Eindeutigkeit einer Lösung für lineare Gleichungen in einer Gruppe.

**Beweis:** Wir beweisen die Aussagen einzeln:

1. Es seien  $x, y \in G$ . Dann gilt:

$$\begin{aligned} y^{-1} * x^{-1} &= (y^{-1} * x^{-1}) * e \\ &= (y^{-1} * x^{-1}) * (x * y) * (x * y)^{-1} \\ &= y^{-1} * \underbrace{(x^{-1} * x)}_e * y * (x * y)^{-1} \end{aligned}$$

$$\begin{aligned}
&= \underbrace{(y^{-1} * y)}_e * (x * y)^{-1} \\
&= (x * y)^{-1}
\end{aligned}$$

2. Es sei  $x \in G$ . Dann gilt:

$$(x^{-1})^{-1} = (x^{-1})^{-1} * (x^{-1} * x) = \underbrace{((x^{-1})^{-1} * x^{-1})}_e * x = x$$

3. Übungsaufgabe.

Damit ist das Theorem bewiesen. ■

## 5.5 Weitere Abzählprinzipien

Im letzten Abschnitt dieses Kapitels über Kombinatorik diskutieren wir noch zwei weitere Abzählprinzipien.

**Inklusion-Exklusion.** Zunächst wollen wir eine Verallgemeinerung der Summenregel (siehe Lemma 5.1) auf beliebige, nicht notwendig paarweise disjunkte Mengen angeben.

**Theorem 5.18 (Inklusions-Exklusions-Prinzip)** *Es seien  $A_1, \dots, A_n$  endliche Mengen. Dann gilt:*

$$\left\| \bigcup_{j=1}^n A_j \right\| = \sum_{\emptyset \neq K \subseteq \{1, \dots, n\}} (-1)^{1+\|K\|} \left\| \bigcap_{k \in K} A_k \right\|$$

**Beispiel:**

- Für  $n = 2$  reduzieren sich die Ausdrücke in Theorem 5.18 zu folgender Identität:

$$\|A_1 \cup A_2\| = \|A_1\| + \|A_2\| - \|A_1 \cap A_2\|$$

- Für  $n = 3$  reduzieren sich die Ausdrücke in Theorem 5.18 zu folgender Identität:

$$\begin{aligned}
\|A_1 \cup A_2 \cup A_3\| &= \|A_1\| + \|A_2\| + \|A_3\| \\
&\quad - \|A_1 \cap A_2\| - \|A_1 \cap A_3\| - \|A_2 \cap A_3\| \\
&\quad + \|A_1 \cap A_2 \cap A_3\|
\end{aligned}$$

**Beweis:** Wir bestimmen, wie oft jedes Element auf beiden Seiten der Gleichung gezählt wird. Es sei  $x \in \bigcup_{j=1}^n A_j$ .

- *Linke Seite:* Das Element  $x$  wird genau einmal gezählt.
- *Rechte Seite:* Wir müssen zeigen, dass  $x$  auch hier genau einmal gezählt wird. Dazu sei  $\ell =_{\text{def}} \|\{j \mid x \in A_j\}\|$ . Ohne Beeinträchtigung der Allgemeinheit komme  $x$  genau in den Mengen  $A_1, \dots, A_\ell$  vor. Dann gilt:
  - Für  $\emptyset \neq K \subseteq \{1, \dots, \ell\}$  wird  $x$  genau  $(-1)^{1+\|K\|}$ -mal gezählt.
  - Für alle anderen Menge  $K$  trägt  $x$  wird  $x$  gar nicht gezählt.

Somit folgt für den Beitrag von  $x$  zur rechten Seite der Gleichung insgesamt:

$$\begin{aligned} \sum_{\emptyset \neq K \subseteq \{1, \dots, \ell\}} (-1)^{1+\|K\|} &= \sum_{k=1}^{\ell} \binom{\ell}{k} (-1)^{1+k} = - \sum_{k=1}^{\ell} \binom{\ell}{k} (-1)^k \\ &= 1 - \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^k \\ &= 1 \end{aligned} \quad (\text{nach Korollar 5.10})$$

Damit ist das Theorem bewiesen. ■

Wir wollen an einem Beispiel verdeutlichen, wie der doch recht kompliziert wirkende Ausdruck auf der rechten Seite gewinnbringend angewendet werden kann.

**Beispiel:** Wie viele Primzahlen gibt es zwischen 2 und 100? Um diese Frage zu beantworten, bestimmen wir die zusammengesetzten Zahlen zwischen 2 und 100 mit Hilfe des Inklusions-Exklusions-Prinzip. Es sei  $A =_{\text{def}} \{2, \dots, 100\}$ . Eine Zahl  $x \in A$  ist zusammengesetzt, falls  $x = p \cdot n$  für geeignete Zahlen  $p, n \in A$  gilt, wobei  $p$  eine Primzahl mit  $p \leq \sqrt{100} = 10$  ist. Damit kommen als Primzahlen nur  $p_1 = 2$ ,  $p_2 = 3$ ,  $p_3 = 5$  und  $p_4 = 7$  in Frage. Für  $i \in \{1, 2, 3, 4\}$  betrachten wir die Menge der Vielfachen von  $p_i$ , d.h. die Menge

$$A_i =_{\text{def}} \{x \in A \mid (\exists n \in A)[x = p_i \cdot n]\}.$$

Damit gilt:

- $A_1 \cup A_2 \cup A_3 \cup A_4$  ist die Menge der zusammengesetzten Zahlen aus  $A$
- Die Kardinalitäten der möglichen Schnittmengen sind

$$\begin{aligned} \|A_i\| &= \left\lfloor \frac{100}{p_i} \right\rfloor - 1 \quad (\text{da } p_i \notin A_i) \\ \left\| \bigcap_{j=1}^k A_{i_j} \right\| &= \left\lfloor \frac{100}{\prod_{j=1}^k p_{i_j}} \right\rfloor \quad \text{für } k \in \{2, 3, 4\} \text{ und } 1 \leq i_1 < \dots < i_k \leq 4 \end{aligned}$$

Nach Theorem 5.18 erhalten wir:

$$\begin{aligned}
 & \|A_1 \cup A_2 \cup A_3 \cup A_4\| \\
 &= \left( \left\lfloor \frac{100}{2} \right\rfloor - 1 + \left\lfloor \frac{100}{3} \right\rfloor - 1 + \left\lfloor \frac{100}{5} \right\rfloor - 1 + \left\lfloor \frac{100}{7} \right\rfloor - 1 \right) \\
 &\quad - \left( \left\lfloor \frac{100}{6} \right\rfloor + \left\lfloor \frac{100}{10} \right\rfloor + \left\lfloor \frac{100}{14} \right\rfloor + \left\lfloor \frac{100}{15} \right\rfloor + \left\lfloor \frac{100}{21} \right\rfloor + \left\lfloor \frac{100}{35} \right\rfloor \right) \\
 &\quad + \left( \left\lfloor \frac{100}{30} \right\rfloor + \left\lfloor \frac{100}{42} \right\rfloor + \left\lfloor \frac{100}{70} \right\rfloor + \left\lfloor \frac{100}{105} \right\rfloor \right) \\
 &\quad - \left\lfloor \frac{100}{210} \right\rfloor \\
 &= 49 + 32 + 19 + 13 - 16 - 10 - 7 - 6 - 4 - 2 + 3 + 2 + 1 + 0 - 0 \\
 &= 74
 \end{aligned}$$

Damit gibt es  $99 - 74 = 25$  Primzahlen zwischen 2 und 100.

**Schubfachschluss.** Ein weiteres wichtiges Abzählprinzip, um die Existenz von Objekten zu beweisen, ist der Schubfachschluss (engl. *pigeonhole principle*).

**Theorem 5.19 (Schubfachschluss)** *Es seien  $A$  und  $B$  endliche Mengen mit  $\|A\| > \|B\| > 0$  und  $f : A \rightarrow B$  eine Funktion. Dann gibt es ein  $y \in B$  mit  $\|f^{-1}(y)\| > 1$ .*

**Beweis:** (*Widerspruch*) Angenommen es gilt  $\|f^{-1}(y)\| \leq 1$  für alle  $y \in B$ . Dann wissen wir aus dem letzten Semester, dass  $f$  eine injektive Funktion ist. Daraus folgt  $\|A\| \leq \|B\|$ . Dies ist ein Widerspruch zu  $\|A\| > \|B\|$ . Mithin war die Annahme falsch, und das Theorem ist bewiesen. ■

Mit anderen Worten: Um  $\|A\|$  Objekte in  $\|B\|$  Schubfächer zu stecken, müssen sich in mindestens einem Schubfach 2 Objekte befinden (falls  $\|A\| > \|B\|$  ist).

**Beispiele:** An folgenden Fällen wollen wir die Anwendung des Schubfachschlusses demonstrieren:

- Von 13 Personen feiern mindestens zwei Personen im gleichen Monat ihren Geburtstag.
- In jeder Menge  $P$  von mindestens zwei Personen gibt es immer mindestens zwei Personen, die die gleiche Anzahl von Personen in  $P$  kennen. (Hierbei sei angenommen, dass „kennen“ eine symmetrische Relation ist.)

Zur Begründung: Es seien  $P = \{p_1, \dots, p_n\}$  die Personenmenge mit  $n \geq 2$  Personen sowie  $f : P \rightarrow \{0, \dots, n-1\}$  eine Funktion, die der Person  $p_i$  die Anzahl ihrer Bekannten in  $P$  zuordnet. Wegen  $\|P\| = \|\{0, \dots, n-1\}\| = n$  kann Theorem 5.19 nicht direkt angewendet werden. Eine genauere Analyse ermöglicht jedoch die folgende Fallunterscheidung:

- Es gibt ein  $p \in P$  mit  $f(p) = 0$ . Wegen der Symmetrie der Bekanntschaftsrelation gibt es auch keine Person, die alle Personen in  $P$  kennt. Also gilt  $f(q) \neq n-1$  für alle  $q \in P$  und folglich  $f(P) \subseteq \{0, \dots, n-2\}$ .
- Für alle  $p \in P$  gilt  $f(p) \neq 0$ . Damit gilt  $f(P) \subseteq \{1, \dots, n-1\}$ .

In beiden Fällen gilt also  $\|f(P)\| < \|P\|$ . Nach Theorem 5.19 folgt die Aussage.

**Theorem 5.20 (Verallgemeinerter Schubfachschluss)** *Es seien  $A$  und  $B$  endliche, nichtleere Mengen und  $f : A \rightarrow B$  eine Funktion. Dann existiert ein  $y \in B$  mit  $\|f^{-1}(y)\| \geq \left\lceil \frac{\|A\|}{\|B\|} \right\rceil$ .*

**Beweis:** (*Widerspruch*) Wir nehmen wiederum an, dass  $\|f^{-1}(y)\| \leq \left\lceil \frac{\|A\|}{\|B\|} \right\rceil - 1$  für alle  $y \in B$  gilt. Dann folgt:

$$\begin{aligned} \|A\| &= \sum_{y \in B} \|f^{-1}(y)\| \\ &\leq \|B\| \cdot \left( \left\lceil \frac{\|A\|}{\|B\|} \right\rceil - 1 \right) \\ &\leq \|B\| \cdot \left( \frac{\|A\| + \|B\| - 1}{\|B\|} - 1 \right) \\ &= \|B\| \cdot \frac{\|A\| - 1}{\|B\|} \\ &= \|A\| - 1 \end{aligned}$$

Dies ist jedoch ein Widerspruch. Mithin war die Annahme falsch, und das Theorem ist bewiesen. ■

**Beispiel:** Wir wollen wieder an zwei Beispielen den verallgemeinerten Schubfachschluss verdeutlichen.

- Von 100 Personen feiern mindestens 9 Personen im gleichen Monat ihren Geburtstag.
- In jeder Menge von 6 Personen gibt es 3 Personen, die sich alle untereinander kennen, oder 3, die sich alle nicht kennen. (Hierbei nehmen wir wiederum an, dass „kennen“ eine symmetrische Relation ist.)

Zur Begründung: Es sei  $P = \{p_1, \dots, p_6\}$  eine beliebige Personenmenge. Wir betrachten für die Person  $p_1$  die Funktion

$$f : \{p_2, \dots, p_5\} \rightarrow \{\text{„bekannt“}, \text{„nicht bekannt“}\},$$

die jeder Person  $p_2, \dots, p_5$  zuordnet, ob  $p_1$  diese Person kennt. Nach Theorem 5.20 sind  $\lceil \frac{5}{2} \rceil = 3$  Personen mit  $p_1$  „bekannt“ oder 3 Personen mit  $p_1$  „nicht bekannt“. Ohne Beeinträchtigung der Allgemeinheit seien 3 Personen mit  $p_1$  bekannt (sonst vertauschen wir in nachfolgender Argumentation einfach „kennen“ und „nicht kennen“) und zwar  $p_2, p_3$  und  $p_4$ . Nun gibt es zwei Möglichkeiten für die Personen  $p_2, p_3$  und  $p_4$ :

- Es gibt zwei Personen in  $\{p_2, p_3, p_4\}$ , die sich kennen. Diese beiden Personen kennen aber auch  $p_1$ . Somit gibt es also 3 Personen, die sich alle untereinander kennen.
- Die Personen  $p_2, p_3$  und  $p_4$  kennen sich nicht. Also gibt es 3 Personen, die sich untereinander nicht kennen.





In diesem Kapitel legen wir die Grundlagen der diskreten Wahrscheinlichkeitstheorie, wie sie in den informatischen Disziplinen häufig zur Anwendung kommen.

## 6.1 Diskrete Wahrscheinlichkeitsräume

**Definition 6.1** Ein diskreter Wahrscheinlichkeitsraum ist ein Paar  $(\Omega, \mathbf{P})$ , wobei

- $\Omega = \{\omega_1, \omega_2, \dots\}$  eine abzählbare Menge von Elementarereignissen  $\omega_j$  und
- $\mathbf{P} : \Omega \rightarrow [0, 1]$  mit  $\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$  ein Wahrscheinlichkeitsmaß sind.

Eine Menge  $E \subseteq \Omega$  heißt Ereignis. Die Wahrscheinlichkeit  $\mathbf{P}(E)$  eines Ereignisses ist definiert als

$$\mathbf{P}(E) =_{\text{def}} \sum_{\omega \in E} \mathbf{P}(\omega).$$

*Bemerkungen:* Wir führen einige zusätzliche Anmerkungen zu obiger Definition an.

1. Das Wahrscheinlichkeitsmaß  $\mathbf{P}$  ist eigentlich eine Funktion  $\mathcal{P}(\Omega) \rightarrow [0, 1]$ ; statt  $\mathbf{P}(\{\omega_j\})$  schreiben wir jedoch weiterhin  $\mathbf{P}(\omega_j)$ .
2. Die leere Menge  $\emptyset$  heißt *unmögliches Ereignis*, da  $\mathbf{P}(\emptyset) = 0$  gilt; die volle Menge  $\Omega$  heißt *sicheres Ereignis*, da  $\mathbf{P}(\Omega) = 1$  gilt.
3. Statt  $\mathbf{P}(A)$  schreiben wir auch  $\mathbf{P}[x \in A]$ .
4. Gilt  $\mathbf{P}(\omega) = \mathbf{P}(\omega')$  für alle  $\omega, \omega' \in \Omega$ , so heißt  $\mathbf{P}$  gleichverteilt.

Wir wollen die Begriffsbildungen an Hand einiger Beispiele verdeutlichen.

**Beispiel:** Wir modellieren zunächst den Wahrscheinlichkeitsraum  $(\Omega, \mathbf{P})$  für den *idealen Würfel*. Die Menge der Elementarereignisse ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Das Wahrscheinlichkeitsmaß  $\mathbf{P}$  ist die Gleichverteilung, d.h.  $\mathbf{P}(\omega) = \frac{1}{6}$  für alle  $\omega \in \Omega$ . Betrachten wir das Ereignis  $E = \{1, 3, 5\}$ , so ergibt sich:

$$\begin{aligned} \mathbf{P}(E) &= \mathbf{P}[ \text{„Würfel zeigt eine ungerade Augenzahl“} ] \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

Als zweites Beispiel modellieren wir nun einen *abgenutzten Würfel*, bei dem durch äußere Einwirkungen die 5 nicht von der 1 zu unterscheiden ist. Die Menge der Elementarereignisse ist damit  $\Omega = \{1, 2, 3, 4, 6\}$ . Das Wahrscheinlichkeitsmaß ergibt sich zu

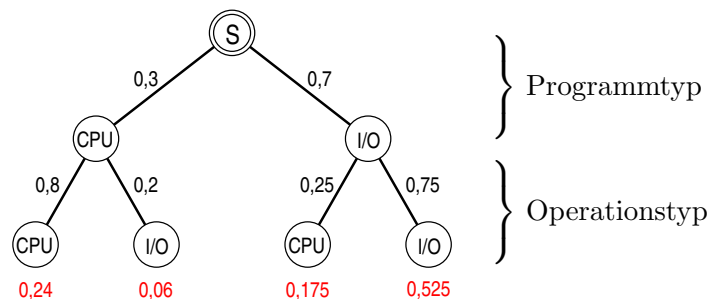
$$\mathbf{P}(\omega) = \begin{cases} \frac{1}{6} & \text{falls } \omega \in \Omega \setminus \{1\} \\ \frac{1}{3} & \text{falls } \omega = 1 \end{cases}$$

Zur Überprüfung der Normierungseigenschaft für  $\mathbf{P}$  rechnet man einfach nach:  $\mathbf{P}(1) + \dots + \mathbf{P}(4) + \mathbf{P}(6) = \frac{1}{3} + 4 \cdot \frac{1}{6} = 1$ . Betrachten wir nun das Ereignis  $E = \{1, 2, 3\}$ , so ergibt sich:

$$\begin{aligned} \mathbf{P}(E) &= \mathbf{P}[\text{„Würfel zeigt höchstens eine 3“}] \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3} \end{aligned}$$

**Beispiel:** Das dritte Beispiel betrifft die Modellierung eines *Prozessors*. Dabei gehen wir von folgendem Szenario aus: Ein Prozessor führt in jedem Schritt entweder eine Ein-/Ausgabe-Operation (Operation vom Typ I/O) oder eine Rechenoperation (Operation vom Typ CPU) aus. Programme für den Prozessor können in Ein-/Ausgabe-lastige (Typ I/O) oder rechenintensive (Typ CPU) unterschieden werden. In einem gegebenem Kontext gelte für Programme vom Typ I/O, dass die Wahrscheinlichkeit, dass eine Rechenoperation ausgeführt wird, gerade 0,25 ist; für Programme vom Typ CPU sei diese Wahrscheinlichkeit dagegen 0,8. Das aktuell laufende Programm ist mit Wahrscheinlichkeit 0,7 vom Typ I/O. Wie groß ist die Wahrscheinlichkeit, dass im nächsten Schritt eine Rechenoperation ansteht?

Wir veranschaulichen das Szenario zunächst mit einem *Entscheidungsbaum*.



Der zugehörige Wahrscheinlichkeitsraum ist  $\Omega =_{\text{def}} \{\text{CPU}, \text{I/O}\}^2$  mit den Elementarwahrscheinlichkeiten

$$\mathbf{P} : \begin{aligned} (\text{CPU}, \text{CPU}) &\mapsto 0,24 = 0,3 \cdot 0,8 \\ (\text{CPU}, \text{I/O}) &\mapsto 0,06 = 0,3 \cdot 0,2 \\ (\text{I/O}, \text{CPU}) &\mapsto 0,175 = 0,7 \cdot 0,25 \\ (\text{I/O}, \text{I/O}) &\mapsto 0,525 = 0,7 \cdot 0,75 \end{aligned}$$

Die Wahrscheinlichkeit, dass im nächsten Schritt eine Rechenoperation ansteht, ergibt sich damit als

$$\mathbf{P}(\{(CPU, CPU), (I/O, CPU)\}) = 0,24 + 0,175 = 0,415$$

**Lemma 6.2** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A, B, A_1, \dots, A_n \subseteq \Omega$  Ereignisse. Dann gilt:*

1.  $\mathbf{P}(\emptyset) = 0 \leq \mathbf{P}(A) \leq 1 = \mathbf{P}(\Omega)$ .
2.  $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$ .
3. Ist  $A \subseteq B$ , so gilt  $\mathbf{P}(A) \leq \mathbf{P}(B)$ .
4. Sind  $A_1, \dots, A_n$  paarweise disjunkt, so gilt

$$\mathbf{P}\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n \mathbf{P}(A_j).$$

**Beweis:** Wir beweisen die Aussagen einzeln:

1. folgt direkt aus Definition 6.1 (siehe auch die nachfolgenden Bemerkungen).
2. Es gilt:

$$\begin{aligned} 1 &= \mathbf{P}(\Omega) && \text{(nach Aussage 1)} \\ &= \sum_{\omega \in \Omega} \mathbf{P}(\omega) && \text{(nach Definition 6.1)} \\ &= \sum_{\omega \in A} \mathbf{P}(\omega) + \sum_{\omega \notin A} \mathbf{P}(\omega) && \text{(wegen } A \cup \bar{A} = \Omega) \\ &= \mathbf{P}(A) + \mathbf{P}(\bar{A}) && \text{(nach Definition 6.1)} \end{aligned}$$

Somit folgt  $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$ .

3. Es gilt:

$$\begin{aligned} \mathbf{P}(B) &= \sum_{\omega \in B} \mathbf{P}(\omega) && \text{(nach Definition 6.1)} \\ &= \sum_{\omega \in A} \mathbf{P}(\omega) + \sum_{\omega \in B \setminus A} \mathbf{P}(\omega) && \text{(wegen } A \subseteq B \text{ bzw. } A \cup (B \setminus A) = B) \\ &= \mathbf{P}(A) + \mathbf{P}(B \setminus A) && \text{(nach Definition 6.1)} \end{aligned}$$

Daraus folgt  $\mathbf{P}(B) - \mathbf{P}(A) = \mathbf{P}(B \setminus A) \geq 0$ , d.h.  $\mathbf{P}(A) \leq \mathbf{P}(B)$ .

4. Es gilt:

$$\begin{aligned}
 \mathbf{P}\left(\bigcup_{j=1}^n A_j\right) &= \sum_{\omega \in \bigcup_{j=1}^n A_j} \mathbf{P}(\omega) && \text{(nach Definition 6.1)} \\
 &= \sum_{j=1}^n \sum_{\omega \in A_j} \mathbf{P}(\omega) \quad \text{(wegen paarweiser Disjunktheit von } A_1, \dots, A_n) \\
 &= \sum_{j=1}^n \mathbf{P}(A_j) && \text{(nach Definition 6.1)}
 \end{aligned}$$

Damit ist das Lemma bewiesen. ■

## 6.2 Kombinatorische Prinzipien

Wir übertragen einige Resultate und Techniken aus der Kombinatorik in die diskrete Wahrscheinlichkeitstheorie. Anwendungen der Kombinatorik sind insbesondere in gleichverteilten Wahrscheinlichkeitsräumen von Bedeutung, da hier die Elementarwahrscheinlichkeiten meistens ausgeklammert werden können.

### Theorem 6.3 (Inklusions-Exklusions-Prinzip für Wahrscheinlichkeitsräume)

Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \subseteq \Omega$  Ereignisse. Dann gilt

$$\mathbf{P}\left(\bigcup_{j=1}^n A_j\right) = \sum_{\emptyset \neq K \subseteq \{1, \dots, n\}} (-1)^{1+\|K\|} \cdot \mathbf{P}\left(\bigcap_{k \in K} A_k\right)$$

**Beweis:** Wir beweisen die Aussage mittels vollständiger Induktion über  $n$ :

- *Induktionsanfang:* Für  $n = 1$  ist die Aussage trivialerweise wahr. Für den Nachweis der Aussage für  $n = 2$  siehe Übungsblatt 5.
- *Induktionsschritt:* Für  $n > 2$  seien  $A_1, \dots, A_n \subseteq \Omega$  beliebige Ereignisse. Dann gilt:

$$\begin{aligned}
 \mathbf{P}\left(\bigcup_{j=1}^n A_j\right) &= \mathbf{P}\left(A_n \cup \bigcup_{j=1}^{n-1} A_j\right) \\
 &= \mathbf{P}(A_n) + \mathbf{P}\left(\bigcup_{j=1}^{n-1} A_j\right) - \mathbf{P}\left(A_n \cap \bigcup_{j=1}^{n-1} A_j\right) \\
 &\hspace{15em} \text{(nach Induktionsvoraussetzung)}
 \end{aligned}$$

$$\begin{aligned}
&= \mathbf{P}(A_n) + \mathbf{P}\left(\bigcup_{j=1}^{n-1} A_j\right) - \mathbf{P}\left(\bigcup_{j=1}^{n-1} (A_n \cap A_j)\right) \\
&= \mathbf{P}(A_n) + \sum_{\emptyset \neq K \subseteq \{1, \dots, n-1\}} (-1)^{1+\|K\|} \cdot \mathbf{P}\left(\bigcap_{k \in K} A_k\right) - \\
&\quad \sum_{\emptyset \neq K \subseteq \{1, \dots, n-1\}} (-1)^{1+\|K\|} \cdot \mathbf{P}\left(\bigcap_{k \in K} (A_n \cap A_k)\right) \\
&\quad \text{(nach Induktionsvoraussetzung)} \\
&= \mathbf{P}(A_n) + \sum_{\emptyset \neq K \subseteq \{1, \dots, n-1\}} (-1)^{1+\|K\|} \cdot \mathbf{P}\left(\bigcap_{k \in K} A_k\right) - \\
&\quad \sum_{\emptyset \neq K \subseteq \{1, \dots, n-1\}} (-1)^{1+\|K \cup \{n\}\|} \cdot \mathbf{P}\left(\bigcap_{k \in K \cup \{n\}} A_k\right) \\
&= \sum_{\emptyset \neq K \subseteq \{1, \dots, n\}} (-1)^{1+\|K\|} \cdot \mathbf{P}\left(\bigcap_{k \in K} A_k\right)
\end{aligned}$$

Damit ist das Theorem bewiesen. ■

Wir wollen auch hier das Theorem an einem Beispiel verdeutlichen.

		<b>0</b>				
<b>PASSE</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>MANQUE</b>	<b>IMPAIR</b>	
	<b>4</b>	<b>5</b>	<b>6</b>			
	<b>7</b>	<b>8</b>	<b>9</b>			
	<b>10</b>	<b>11</b>	<b>12</b>			
<b>PAIR</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>IMPAIR</b>	<b>PAIR</b>	
	<b>16</b>	<b>17</b>	<b>18</b>			
	<b>19</b>	<b>20</b>	<b>21</b>			
	<b>22</b>	<b>23</b>	<b>24</b>			
<b>PAIR</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>IMPAIR</b>	<b>PAIR</b>	
	<b>28</b>	<b>29</b>	<b>30</b>			
	<b>31</b>	<b>32</b>	<b>33</b>			
	<b>34</b>	<b>35</b>	<b>36</b>			
<div style="display: flex; justify-content: space-around; font-size: small;"> <span><math>12^P</math></span> <span><math>12^M</math></span> <span><math>12^D</math></span> </div>					<div style="display: flex; justify-content: space-around; font-size: small;"> <span><math>12^D</math></span> <span><math>12^M</math></span> <span><math>12^P</math></span> </div>	

**Beispiel:** Der Wahrscheinlichkeitsraum beim (französischen) *Roulette* besteht aus  $\Omega = \{0, 1, \dots, 36\}$  mit dem Wahrscheinlichkeitsmaß  $\mathbf{P}(\omega) = \frac{1}{37}$  für alle  $\omega \in \Omega$ . Unser Plan für ein Spiel (*coup*) ist nun, einen Chip auf „schwarz“ (*noir*) und einen Chip auf „ungerade“ (*impair*) zu setzen. Wie hoch ist dann die Wahrscheinlichkeit, dass wir nichts verlieren? Nach Inspektion des Einsatzfeldes (siehe Abbildung links) halten wir zunächst fest, dass es 8 schwarze ungerade Zahlen gibt. Wir definieren folgende Ereignisse (ohne die zugehörigen Elementarereignisse aufzuzählen):

$E_s$  =<sub>def</sub> Menge aller schwarzen Zahlen  
 $E_u$  =<sub>def</sub> Menge aller ungeraden Zahlen

Das uns interessierende Ereignis ist  $E_s \cup E_u$ . Seine Wahrscheinlichkeit lässt sich mit Hilfe des Inklusions-Exklusions-Prinzip wie folgt bestimmen:

$$\begin{aligned} \mathbf{P}(E_s \cup E_u) &= \mathbf{P}(E_s) + \mathbf{P}(E_u) - \mathbf{P}(E_s \cap E_u) \\ &= \frac{18}{37} + \frac{18}{37} - \frac{8}{37} \\ &= \frac{28}{37} = 0,7568\dots \end{aligned}$$

Die Wahrscheinlichkeit, dass wir mit unserem Spielplan nichts verlieren, ist also ca. 76%. Später werden wir sehen, dass der zu erwartende Gewinn jedoch negativ ist!

**Theorem 6.4** *Es sei  $N = \{1, \dots, n\}$ . Die Wahrscheinlichkeitsräume  $\Omega$  beim Ziehen von  $k$  Elementen aus  $n$  Elementen können wie folgt gewählt werden:*

- mit Reihenfolge, mit Zurücklegen:  $\{ (x_1, \dots, x_k) \mid x_i \in N \}$
- mit Reihenfolge, ohne Zurücklegen:  $\{ (x_1, \dots, x_k) \mid x_i \in N, x_i \neq x_j \text{ für } i \neq j \}$
- ohne Reihenfolge, mit Zurücklegen:  $\{ (a_1, \dots, a_n) \mid a_i \in \mathbb{N}, a_1 + \dots + a_n = k \}$
- ohne Reihenfolge, ohne Zurücklegen:  $\{ \{x_1, \dots, x_k\} \mid x_i \in N, x_i \neq x_j \text{ für } i \neq j \}$

Im Fall „ohne Reihenfolge, mit Zurücklegen“ gibt  $a_i \in \mathbb{N}$  an, wie oft das  $i$ -te Element gezogen wurde. Sind alle Elementarereignisse gleichwahrscheinlich, so gilt für die Wahrscheinlichkeit der Elementarereignisse

	mit Zurücklegen	ohne Zurücklegen
mit Reihenfolge	$n^{-k}$	$\frac{(n-k)!}{n!}$
ohne Reihenfolge	$\binom{n+k-1}{k}^{-1}$	$\binom{n}{k}^{-1}$

**Beweis:** Wegen der vorausgesetzten Wahrscheinlichkeitsverteilung gilt für ein spezielles  $\omega^* \in \Omega$

$$1 = \sum_{\omega \in \Omega} \mathbf{P}(\omega) = \|\Omega\| \cdot \mathbf{P}(\omega^*),$$

d.h.  $\mathbf{P}(\omega^*) = \|\Omega\|^{-1}$ . Damit folgen die Wahrscheinlichkeiten aus Theorem 5.6. ■

**Beispiel:** Beim Pokerspiel (*Texas hold'em*) erhält jeder Spieler verdeckt zwei Spielkarten ausgeteilt. Die Spielkarten stammen aus der Menge

$$M = \{\diamond, \heartsuit, \clubsuit, \spadesuit\} \times \{2, 3, 4, 5, 6, 7, 8, 9, 10, \text{Bube, Dame, König, As}\}.$$

$M$  enthält also 52 Spielkarten. Die Menge der Elementarereignisse ist damit  $\Omega = \{ \{x, y\} \mid x, y \in M, x \neq y \}$ , d.h.  $|\Omega| = \binom{52}{2} = 1326$ . Wir interessieren uns für das Ereignis  $E_p =$  „pocket pair“, d.h., dass wir sofort zwei Karten des gleichen Wertes auf der Hand haben:

$$\mathbf{P}(E_p) = \sum_{\omega \in E_p} \mathbf{P}(\omega) = 13 \cdot \binom{4}{2} \cdot \frac{1}{1326} = \frac{78}{1326} = \frac{3}{51} = 0,0588\dots$$

*Bemerkungen:* In Theorem 6.4 entspricht die Gleichverteilung auf der Menge der Elementarereignisse für den Fall „ohne Reihenfolge, mit Zurücklegen“ *nicht* der Wahrscheinlichkeitsverteilung beim ungeordneten Ziehen mit Zurücklegen. Wenn zum Beispiel zweimal gezogen wird, dann ist in letzterem Szenario die Wahrscheinlichkeit, dass das erste und das zweite Element gezogen wird, doppelt so groß, wie zweimal das erste Element zu ziehen. Damit ist die Wahrscheinlichkeit für die Elementarereignisse nicht gleichverteilt. Die Wahrscheinlichkeit für ein Elementarereignis  $\omega = (a_1, \dots, a_n) \in \Omega$  ergibt sich wie folgt:

$$\mathbf{P}(\omega) = \frac{1}{n^k} \cdot \frac{k!}{a_1! a_2! \dots a_n!} = \frac{1}{n^k} \cdot \binom{k}{a_1} \binom{k-a_1}{a_2} \dots \binom{k-(a_1+a_2+\dots+a_{n-1})}{a_n}$$

## 6.3 Bedingte Wahrscheinlichkeiten

Im Folgenden interessieren wir uns für die Wahrscheinlichkeit, dass ein Ereignis  $A$  eintritt unter der Bedingung, dass ein bestimmtes Ereignis  $B$  ebenfalls eintritt.

**Definition 6.5** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A, B \subseteq \Omega$  Ereignisse mit  $\mathbf{P}(B) > 0$ . Die bedingte Wahrscheinlichkeit  $\mathbf{P}(A|B)$  von  $A$  unter der Bedingung  $B$  ist definiert als*

$$\mathbf{P}(A|B) \stackrel{\text{def}}{=} \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Es gilt ganz offensichtlich stets  $0 \leq \mathbf{P}(A|B) \leq 1$  (letzteres wegen  $\mathbf{P}(A \cap B) \leq \mathbf{P}(B)$ ).

**Beispiel:** Wir betrachten den Wahrscheinlichkeitsraum zu „zweimal Würfeln“, d.h.  $\Omega = \{1, 2, 3, 4, 5, 6\}^2$  mit Gleichverteilung  $\mathbf{P}$ . Auf dem Wahrscheinlichkeitsraum seien folgende Ereignisse definiert:

$$\begin{aligned} A & \stackrel{\text{def}}{=} \{ (j, k) \mid j + k \geq 9 \} \subseteq \Omega \\ B_i & \stackrel{\text{def}}{=} \{ (j, k) \mid j = i \} \quad \text{für } i \in \{1, \dots, 6\} \end{aligned}$$

Extensional lassen sich die Ereignisse wie folgt beschreiben:

$$\begin{aligned} A & = \{ (3, 6), (4, 5), (4, 6), (5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6) \} \\ B_i & = \{ (i, 1), (i, 2), (i, 3), (i, 4), (i, 5), (i, 6) \} \end{aligned}$$

Damit gilt  $\mathbf{P}(A) = \frac{10}{36}$  sowie  $\mathbf{P}(B_i) = \frac{1}{6}$  für alle  $i \in \{1, \dots, 6\}$ . Es ergeben sich folgende bedingte Wahrscheinlichkeiten:

$$\begin{aligned} \mathbf{P}(A|B_1) &= \mathbf{P}(\emptyset) \cdot \mathbf{P}(B_1)^{-1} &&= 0 \\ \mathbf{P}(A|B_2) &= \mathbf{P}(\emptyset) \cdot \mathbf{P}(B_2)^{-1} &&= 0 \\ \mathbf{P}(A|B_3) &= \mathbf{P}(\{(3, 6)\}) \cdot \mathbf{P}(B_3)^{-1} &&= \frac{1}{36} \cdot \left(\frac{1}{6}\right)^{-1} = \frac{1}{6} \\ \mathbf{P}(A|B_4) &= \mathbf{P}(\{(4, 5), (4, 6)\}) \cdot \mathbf{P}(B_4)^{-1} &&= \frac{2}{36} \cdot \left(\frac{1}{6}\right)^{-1} = \frac{1}{3} \\ \mathbf{P}(A|B_5) &= \mathbf{P}(\{(5, 4), (5, 5), (5, 6)\}) \cdot \mathbf{P}(B_5)^{-1} &&= \frac{3}{36} \cdot \left(\frac{1}{6}\right)^{-1} = \frac{1}{2} \\ \mathbf{P}(A|B_6) &= \mathbf{P}(\{(6, 3), (6, 4), (6, 5), (6, 6)\}) \cdot \mathbf{P}(B_6)^{-1} &&= \frac{4}{36} \cdot \left(\frac{1}{6}\right)^{-1} = \frac{2}{3} \end{aligned}$$

**Beispiel:** Wir betrachten wieder den Wahrscheinlichkeitsraum zum Pokerspiel (*Texas hold'em*). Nach wie vor seien die Spielkartenmenge

$$M = \{\diamond, \heartsuit, \clubsuit, \spadesuit\} \times \{2, 3, 4, 5, 6, 7, 8, 9, 10, \text{Bube, Dame, König, As}\}$$

und die Menge der Elementarereignisse  $\Omega = \{ \{x, y\} \mid x, y \in M, x \neq y \}$ . Neben den beiden verdeckten Karten, die an jeden Spieler ausgegeben werden, liegen noch fünf Spielkarten offen aufgedeckt. In unserem Beispiel sei dies die Spielkartenmenge

$$O =_{\text{def}} \{ (\diamond, 4), (\heartsuit, 7), (\clubsuit, 10), (\spadesuit, 4), (\spadesuit, \text{Dame}) \}$$

Wir interessieren uns diesmal für die Wahrscheinlichkeit des Ereignisses

$$A =_{\text{def}} \text{„Gegenspieler hat ein pocket pair“.}$$

Dazu können wir die offenen Karten sowie unsere eigenen Karten einbeziehen. Unsere eigenen Karten seien durch die Menge

$$E =_{\text{def}} \{ (\diamond, 7), (\heartsuit, \text{Dame}) \}$$

repräsentiert. Damit können wir das Ereignis

$$B =_{\text{def}} \text{„Gegenspieler hat keine der offenen oder von meinen Karten“}$$

betrachten. Es gilt  $B = (M \setminus (O \cup E))^2$  und folglich  $\|B\| = \binom{45}{2} = 990$  bzw.  $\mathbf{P}(B) = \frac{990}{1326}$ . Weiterhin erhalten wir

$$\mathbf{P}(A \cap B) = \left( 1 + 1 + 1 + \binom{3}{2} + 9 \cdot \binom{4}{2} \right) \cdot \frac{1}{1326} = \frac{60}{1326}.$$

Somit ergibt sich

$$\mathbf{P}(A|B) = \mathbf{P}(A \cap B) \cdot \mathbf{P}(B)^{-1} = \frac{60}{1326} \cdot \left(\frac{990}{1326}\right)^{-1} = \frac{60}{990} = 0,0606\dots$$

Die Wahrscheinlichkeit eines *pocket pair* für einen Gegenspieler hat sich in dieser Situation erhöht, allerdings im Vergleich zu  $0,0588\dots$  nur geringfügig.



**Theorem 6.6 (Multiplikationsregel)** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \subseteq \Omega$  Ereignisse mit  $\mathbf{P}(A_1 \cap \dots \cap A_n) > 0$ . Dann gilt*

$$\mathbf{P}(A_1 \cap \dots \cap A_n) = \prod_{i=1}^n \mathbf{P}(A_i | A_1 \cap \dots \cap A_{i-1})$$

**Beweis:** Wegen  $\mathbf{P}(A_1) \geq \mathbf{P}(A_1 \cap A_2) \geq \dots \geq \mathbf{P}(A_1 \cap \dots \cap A_n) > 0$  sind alle bedingten Wahrscheinlichkeit wohldefiniert. Wir beweisen das Theorem mittels vollständiger Induktion über die Anzahl  $n$  der Mengen:

- *Induktionsanfang  $n = 1$ :* Es gilt  $\mathbf{P}(A_1) = \mathbf{P}(A_1 | \Omega)$ .
- *Induktionsschritt  $n > 1$ :* Wegen  $\mathbf{P}(A|B) \cdot \mathbf{P}(B) = \mathbf{P}(A \cap B)$  gilt:

$$\begin{aligned} \mathbf{P}(A_1 \cap \dots \cap A_{n-1} \cap A_n) &= \mathbf{P}(A_n | A_1 \cap \dots \cap A_{n-1}) \cdot \mathbf{P}(A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbf{P}(A_n | A_1 \cap \dots \cap A_{n-1}) \cdot \prod_{i=1}^{n-1} \mathbf{P}(A_i | A_1 \cap \dots \cap A_{i-1}) \\ &\quad \text{(nach Induktionsvoraussetzung für } n-1) \\ &= \prod_{i=1}^n \mathbf{P}(A_i | A_1 \cap \dots \cap A_{i-1}) \end{aligned}$$

Damit ist das Theorem bewiesen. ■

**Beispiel:** Das *Geburtstagsproblem* besteht darin zu bestimmen, wie groß die Wahrscheinlichkeit dafür ist, dass unter  $n$  Personen zwei am gleichen Tag Geburtstag feiern. Zur Lösung dieses Problems nehmen wir an, dass Geburtstage gleich wahrscheinlich für jeden Tag des Jahres sind und es keine Schaltjahre gibt. Ist  $n > 365$ , so ist die Wahrscheinlichkeit  $P_1 = 1$  nach dem Schubfachprinzip. Für  $1 \leq n \leq 365$  sei die Menge der Elementarereignisse  $\Omega_n = \{1, \dots, 365\}^n$ . Wir betrachten folgende Ereignisse

$A_i \stackrel{\text{def}}{=} \text{„}i\text{-te Person feiert nicht am gleichen Tag wie eine der Personen } 1, \dots, i-1 \text{ Geburtstag“}$

Damit gilt für die gesuchte Wahrscheinlichkeit  $P_n$ :

$$\begin{aligned} P_n &= 1 - \mathbf{P}(A_1 \cap \dots \cap A_n) \\ &= 1 - \mathbf{P}(A_1) \cdot \mathbf{P}(A_2 | A_1) \cdot \dots \cdot \mathbf{P}(A_n | A_1 \cap \dots \cap A_{n-1}) \\ &= 1 - \prod_{i=1}^n \frac{365 - i + 1}{365} = 1 - \frac{365!}{365^n \cdot (365 - n)!} \end{aligned}$$

Die Wahrscheinlichkeit  $P_n$  steigt sehr schnell. So gilt zum Beispiel  $P_{23} > \frac{1}{2}$ , d.h., bereits unter 21 Personen ist die Wahrscheinlichkeit, dass zwei Personen am gleichen Tag Geburtstag feiern größer als 50%.

**Theorem 6.7 (Satz von der totalen Wahrscheinlichkeit)** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum,  $A_1, \dots, A_n \subseteq \Omega$  paarweise disjunkte Ereignisse mit  $\mathbf{P}(A_i) > 0$  für alle  $i \in \{1, \dots, n\}$  und  $B \subseteq A_1 \cup \dots \cup A_n$ . Dann gilt*

$$\mathbf{P}(B) = \sum_{i=1}^n \mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i).$$

**Beweis:** Es gilt  $B = (B \cap A_1) \cup \dots \cup (B \cap A_n)$ . Nach Lemma 6.2.4 folgt

$$\mathbf{P}(B) = \sum_{i=1}^n \mathbf{P}(B \cap A_i) = \sum_{i=1}^n \mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i).$$

Damit ist das Theorem bewiesen. ■

**Theorem 6.8 (Satz von BAYES)** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum,  $A_1, \dots, A_n \subseteq \Omega$  paarweise disjunkte Ereignisse mit  $\mathbf{P}(A_i) > 0$  für alle  $i \in \{1, \dots, n\}$  und  $B \subseteq A_1 \cup \dots \cup A_n$  mit  $\mathbf{P}(B) > 0$ . Dann gilt für alle  $i \in \{1, \dots, n\}$*

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i)}{\sum_{i=1}^n \mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i)}.$$

**Beweis:** Anwendung von Theorem 6.7 auf  $\mathbf{P}(A_i|B) = \mathbf{P}(A_i \cap B) \cdot \mathbf{P}(B)^{-1}$ . ■

**Beispiel:** Ein in einem Gebäude installierter Rauchmelder löst bei einem Brand zu 98% Feueralarm aus. Ein Fehlalarm wird dagegen pro Tag mit der Wahrscheinlichkeit 0,05% ausgelöst. Die Wahrscheinlichkeit pro Tag für einen Brand liegt bei 0,01%. Wie groß ist die Wahrscheinlichkeit, dass es wirklich brennt, falls Feueralarm ausgelöst wird? Wir modellieren das Szenario mittels folgender Zustände:

- $\mathbf{b}$  repräsentiere „Brand“
- $\bar{\mathbf{b}}$  repräsentiere „kein Brand“
- $\mathbf{a}$  repräsentiere „Alarm“
- $\bar{\mathbf{a}}$  repräsentiere „kein Alarm“

Die Menge der Elementarereignisse ist somit  $\Omega = \{ (\mathbf{b}, \mathbf{a}), (\bar{\mathbf{b}}, \mathbf{a}), (\mathbf{b}, \bar{\mathbf{a}}), (\bar{\mathbf{b}}, \bar{\mathbf{a}}) \}$ . Für die Analyse des Szenarios sind darüber hinaus folgende Ereignisse relevant:

$$\begin{aligned} A &=_{\text{def}} \{ (\mathbf{b}, \mathbf{a}), (\bar{\mathbf{b}}, \mathbf{a}) \} && (= \text{„Alarm!“}) \\ B &=_{\text{def}} \{ (\mathbf{b}, \mathbf{a}), (\mathbf{b}, \bar{\mathbf{a}}) \} && (= \text{„Es brennt!“}) \end{aligned}$$

Aus der Beschreibung sind die Wahrscheinlichkeiten

$$\mathbf{P}(A|B) = 0,98, \quad \mathbf{P}(B) = 0,0001, \quad \mathbf{P}(A|\bar{B}) = 0,0005$$

bekannt. Mit Hilfe des Satzes von Bayes errechnen wir damit die gesuchte Wahrscheinlichkeit  $\mathbf{P}(B|A)$  zu

$$\begin{aligned} \mathbf{P}(B|A) &= \frac{\mathbf{P}(A|B) \cdot \mathbf{P}(B)}{\mathbf{P}(A|B) \cdot \mathbf{P}(B) + \mathbf{P}(A|\bar{B}) \cdot \mathbf{P}(\bar{B})} \\ &= \frac{0,98 \cdot 0,0001}{0,98 \cdot 0,0001 + 0,0005 \cdot (1 - 0,0001)} = 0,1639\dots \end{aligned}$$

Die Wahrscheinlichkeit, dass es bei ausgelöstem Feueralarm tatsächlich brennt, ist angesichts der Ausgangsparameter mit  $\approx 16,39\%$  überraschend niedrig.

## 6.4 Unabhängigkeit

Ein fundamentales Konzept in der Wahrscheinlichkeitstheorie ist die Unabhängigkeit von Ereignissen.

**Definition 6.9** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A, B \subseteq \Omega$  Ereignisse.  $A$  und  $B$  heißen unabhängig, falls  $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$  gilt.*

Die folgende Proposition gibt eine intuitive Interpretation der Definition der Unabhängigkeit zweier Ereignisse. Es wird aber deutlich, dass Definition 6.9 allgemeiner ist, da sie auch bei verschwindenden Wahrscheinlichkeit angewendet werden kann.

**Proposition 6.10** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A, B \subseteq \Omega$  Ereignisse mit  $\mathbf{P}(A) > 0$  und  $\mathbf{P}(B) > 0$ . Dann sind folgende Aussagen äquivalent:*

1.  $A$  und  $B$  sind unabhängig
2.  $\mathbf{P}(A|B) = \mathbf{P}(A)$
3.  $\mathbf{P}(B|A) = \mathbf{P}(B)$

**Beweis:** Wegen  $\mathbf{P}(A) > 0$  und  $\mathbf{P}(B) > 0$  sind alle bedingten Wahrscheinlichkeit wohldefiniert. Wir zeigen lediglich die Äquivalenz (1)  $\Leftrightarrow$  (2). Durch Vertauschung von  $A$  und  $B$  ergibt sich sofort auch die Äquivalenz (1)  $\Leftrightarrow$  (3).

- (1)  $\Rightarrow$  (2):  $A$  und  $B$  seien unabhängig, d.h., es gilt  $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$ . Es folgt

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A) \cdot \mathbf{P}(B)}{\mathbf{P}(B)} = \mathbf{P}(A).$$

- (2)  $\Rightarrow$  (1): Es gelte  $\mathbf{P}(A|B) = \mathbf{P}(A)$ . Dann gilt

$$\mathbf{P}(A \cap B) = \mathbf{P}(A|B) \cdot \mathbf{P}(B) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Mithin sind  $A$  und  $B$  unabhängig.

Damit ist die Proposition bewiesen. ■

**Beispiel:** Wir betrachten wiederum den Wahrscheinlichkeitsraum „zweimal Würfeln“ mit der Menge der Elementarereignisse  $\Omega = \{1, \dots, 6\}^2$  und der Gleichverteilung. Für die beiden Ereignisse

$$\begin{aligned} A &=_{\text{def}} \{ (j, k) \mid j + k \equiv 0 \pmod{2} \} \subseteq \Omega \\ B &=_{\text{def}} \{ (j, k) \mid j \equiv 0 \pmod{2} \} \subseteq \Omega \end{aligned}$$

gilt dann

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{P}(\{ (1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), \\ &\quad (4, 2), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (6, 2), (6, 4), (6, 6) \}) \\ &= \frac{18}{36} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \mathbf{P}(B) &= \mathbf{P}(\{ (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (4, 1), (4, 2), (4, 3), \\ &\quad (4, 4), (4, 5), (4, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \}) \\ &= \frac{18}{36} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \mathbf{P}(A \cap B) &= \mathbf{P}(\{ (2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6) \}) \\ &= \frac{9}{36} = \frac{1}{4} \end{aligned}$$

Damit sind  $A$  und  $B$  unabhängig. Wenn wir das Beispiel um die beiden Mengen

$$\begin{aligned} C_1 &=_{\text{def}} \{ (j, k) \mid k \equiv 0 \pmod{2} \} \subseteq \Omega \\ C_2 &=_{\text{def}} \{ (j, k) \mid k \geq 3 \} \subseteq \Omega \end{aligned}$$

mit den zugehörigen Wahrscheinlichkeiten  $\mathbf{P}(C_1) = \frac{1}{2}$  und  $\mathbf{P}(C_2) = \frac{2}{3}$  erweitern, so ergeben sich folgende Wahrscheinlichkeiten:

$$\begin{aligned} \mathbf{P}(A \cap C_1) &= \frac{1}{4} = \mathbf{P}(A) \cdot \mathbf{P}(C_1) \\ \mathbf{P}(A \cap C_2) &= \frac{1}{3} = \mathbf{P}(A) \cdot \mathbf{P}(C_2) \\ \mathbf{P}(B \cap C_1) &= \frac{1}{4} = \mathbf{P}(B) \cdot \mathbf{P}(C_1) \\ \mathbf{P}(B \cap C_2) &= \frac{1}{3} = \mathbf{P}(B) \cdot \mathbf{P}(C_2) \\ \mathbf{P}(A \cap B \cap C_1) &= \frac{1}{4} \neq \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C_1) \\ \mathbf{P}(A \cap B \cap C_2) &= \frac{1}{6} = \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C_2) \end{aligned}$$

Damit sind die Ereignisfamilien  $\{A, B, C_1\}$  sowie  $\{A, B, C_2\}$  jeweils paarweise unabhängig. Zusätzlich nennen wir die Familie  $\{A, B, C_2\}$  unabhängig (und nicht nur paarweise),  $\{A, B, C_1\}$  dagegen ist nicht unabhängig.

**Definition 6.11** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \subseteq \Omega$  Ereignisse.  $A_1, \dots, A_n$  heißen unabhängig, falls für alle  $K \subseteq \{1, \dots, n\}$  gilt*

$$\mathbf{P}\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} \mathbf{P}(A_k).$$

Das obige Beispiel zeigt, dass paarweise Unabhängigkeit (Definition 6.9) nicht Unabhängigkeit im Sinne von Definition 6.11 impliziert.

## 6.5 Zufallsvariablen

Zufallsvariablen bieten eine Möglichkeit komplexere zufällige Ereignisse und Situationen zu modellieren.

**Definition 6.12** *Es sei  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum. Eine Funktion  $X : \Omega \rightarrow \mathbb{R}$  heißt Zufallsvariable.*

**Beispiel:** Wir wollen den Einsatz von Zufallsvariablen bei der Modellierung des Wahrscheinlichkeitsraumes zu „Summe bei zweimal Würfeln“ verdeutlichen. Mit dem Vokabular aus den letzten Abschnitten könnten wir wie folgt vorgehen: Die Menge der Elementarereignisse ist  $\Omega_1 = \{2, \dots, 12\}$  und das zugehörige Wahrscheinlichkeitsmaß ermittelt man leicht zu

$$\mathbf{P}_1(\omega) = \begin{cases} \frac{\omega-1}{36} & \text{falls } \omega \leq 7 \\ \frac{13-\omega}{36} & \text{falls } \omega \geq 8 \end{cases}$$

Eine natürliche Modellierung mit einer Zufallsvariable ist wie folgt: Der Wahrscheinlichkeitsraum  $(\Omega_2, \mathbf{P}_2)$  ist derselbe wie bei „zweimal Würfeln“, d.h.  $\Omega_2 = \{1, \dots, 6\}^2$  und  $\mathbf{P}_2(\omega) = \frac{1}{36}$ . Als Zufallsvariable betrachten wir

$$X : \Omega_2 \rightarrow \mathbb{R} : (j, k) \mapsto j + k.$$

Damit können wir zum Beispiel die Wahrscheinlichkeit „gewürfelte Augenzahl ist 10“ wie folgt ausdrücken:

$$\mathbf{P}_2[X = 10] =_{\text{def}} \mathbf{P}_2(X^{-1}(\{10\})) = \mathbf{P}_2(\{(4, 6), (5, 5), (6, 4)\}) = \frac{3}{36} = \frac{1}{12}$$

Aus einer grundsätzlicheren Perspektive vermittelt die Zufallsvariable  $X$  einen Zusammenhang zwischen den beiden Wahrscheinlichkeitsräumen  $(\Omega_1, \mathbf{P}_1)$  und  $(\Omega_2, \mathbf{P}_2)$ . Mit der Definition

$$W_X =_{\text{def}} \{ x \mid (\exists \omega \in \Omega_2)[X(\omega) = x] \}$$

gilt nämlich  $\Omega_1 = W_X$  sowie  $\mathbf{P}_1(x) = \mathbf{P}_2(X = x) = \mathbf{P}_2(X^{-1}(\{x\}))$  für  $x \in \Omega_1$ .

**Beispiel:** Wir kommen zurück auf unser Beispiel zum *Roulette*. Der Wahrscheinlichkeitsraum  $(\Omega, \mathbf{P})$  ist gegeben durch  $\Omega = \{0, 1, \dots, 36\}$  und  $\mathbf{P}(\omega) = \frac{1}{37}$ . Unser Spielplan war jeweils 1 Chip auf „schwarz“ und 1 Chip auf „ungerade“ zu setzen. Die zugehörigen Ereignisse sind deshalb

$$\begin{aligned} E_s &= \text{„Kugel fällt auf schwarze Zahl“} \\ E_u &= \text{„Kugel fällt auf ungerade Zahl“} \end{aligned}$$

In beiden Fällen bekommen wir das Doppelte des Einsatzes als Auszahlung zurück, falls das entsprechende Ereignis eintritt. Unsere möglichen Gewinn können wir deshalb als Zufallsvariable modellieren:

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto \begin{cases} -2 & \text{falls } \omega \notin E_s \cup E_u \\ 0 & \text{falls } \omega \in E_s \triangle E_u \\ 2 & \text{falls } \omega \in E_s \cap E_u \end{cases}$$

Durch  $X$  wird folgender diskreter Wahrscheinlichkeitsraum  $(W_X, \mathbf{P}_X)$  beschrieben:  $W_X = \{-2, 0, 2\}$  mit den Wahrscheinlichkeiten

$$\begin{aligned} \mathbf{P}_X(-2) &= \mathbf{P}[X = -2] = \frac{9}{37}, \\ \mathbf{P}_X(0) &= \mathbf{P}[X = 0] = \frac{20}{37}, \\ \mathbf{P}_X(2) &= \mathbf{P}[X = 2] = \frac{8}{37}. \end{aligned}$$

(Die Wahrscheinlichkeiten erhält man durch Inspektion des Tableaus.)

Bevor wir mit der Analyse des Beispiels fortfahren, führen wir die wichtige Definition des Erwartungswertes ein.

**Definition 6.13** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable. Der Erwartungswert von  $X$  ist definiert als*

$$\mathbf{E}(X) =_{\text{def}} \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\omega) = \sum_{x \in W_X} x \cdot \mathbf{P}[X = x].$$

*Hierbei setzen wir voraus, dass die Summe existiert.*

**Beispiel (Fortsetzung):** Der für unseren Spielplan zu erwartende Gewinn beim Roulette ist somit

$$\mathbf{E}(X) = \sum_{x \in W_X} x \cdot \mathbf{P}[X = x] = (-2) \cdot \frac{9}{37} + 0 \cdot \frac{20}{37} + 2 \cdot \frac{8}{37} = -\frac{2}{37} < 0.$$

Trotz einer mehr als 75%-Wahrscheinlichkeit nichts zu verlieren, ist der zu erwartende Gewinn also negativ. (Angesichts der Wahrscheinlichkeiten  $\mathbf{P}_X$  ist dies nicht überraschend, da auch die Wahrscheinlichkeit nichts zu gewinnen bei mehr als 75% liegt.)

**Lemma 6.14** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum,  $X, Y : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen sowie  $a \in \mathbb{R}$ . Dann gelten die folgende Aussagen:*

1. Gilt  $X(\omega) \leq Y(\omega)$  für alle  $\omega \in \Omega$ , so gilt  $\mathbf{E}(X) \leq \mathbf{E}(Y)$ .
2. Ist  $X(\omega) = 1$  für alle  $\omega \in \Omega$ , so gilt  $\mathbf{E}(X) = 1$ .
3.  $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ .
4.  $\mathbf{E}(a \cdot X) = a \cdot \mathbf{E}(X)$ .
5. Ist  $W_X \subseteq \mathbb{N}$ , so gilt  $\mathbf{E}(X) = \sum_{i=1}^{\infty} \mathbf{P}[X \geq i]$ .

Nach der ersten Eigenschaft ist der Erwartungswert monoton und nach der dritten und vierten Eigenschaft linear. Insbesondere gilt  $\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$  für beliebige reelle Zahlen  $a, b$  und Zufallsvariablen  $X$ .

**Beweis:** Wir zeigen die Aussagen des Lemmas einzeln.

$$1. \mathbf{E}(X) = \sum_{\omega \in \Omega} \underbrace{X(\omega)}_{\leq Y(\omega)} \cdot \mathbf{P}(\omega) \leq \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbf{P}(\omega) = \mathbf{E}(Y).$$

$$2. \mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\omega) = \sum_{\omega \in \Omega} 1 \cdot \mathbf{P}(\omega) = \sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1.$$

3. Mit  $(X + Y) : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) + Y(\omega)$  ergibt sich:

$$\begin{aligned} \mathbf{E}(X + Y) &= \sum_{\omega \in \Omega} (X + Y)(\omega) \cdot \mathbf{P}(\omega) \\ &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot \mathbf{P}(\omega) \\ &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\omega) + \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbf{P}(\omega) \\ &= \mathbf{E}(X) + \mathbf{E}(Y) \end{aligned}$$

4. Mit  $(a \cdot X) : \Omega \rightarrow \mathbb{R} : \omega \mapsto a \cdot X(\omega)$  ergibt sich:

$$\begin{aligned} \mathbf{E}(a \cdot X) &= \sum_{\omega \in \Omega} (a \cdot X)(\omega) \cdot \mathbf{P}(\omega) \\ &= \sum_{\omega \in \Omega} a \cdot X(\omega) \cdot \mathbf{P}(\omega) \\ &= a \cdot \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\omega) \\ &= a \cdot \mathbf{E}(X) \end{aligned}$$

5. Wegen  $W_X \subseteq \mathbb{N}$  und durch Vertauschung der Summationsreihenfolge erhalten wir:

$$\begin{aligned} \mathbf{E}(X) &= \sum_{x \in W_X} x \cdot \mathbf{P}[X = x] \\ &= \sum_{j=0}^{\infty} j \cdot \mathbf{P}[X = j] \\ &= \sum_{j=0}^{\infty} \sum_{i=1}^j \mathbf{P}[X = j] \\ &= \sum_{i=1}^{\infty} \underbrace{\sum_{j=i}^{\infty} \mathbf{P}[X = j]}_{=\mathbf{P}[X \geq i]} \\ &= \sum_{i=1}^{\infty} \mathbf{P}[X \geq i] \end{aligned}$$

Damit ist das Lemma bewiesen. ■

**Definition 6.15** Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable mit  $\mu = \mathbf{E}(X)$ . Die Varianz von  $X$  ist definiert als

$$\text{Var}(X) =_{\text{def}} \mathbf{E}((X - \mathbf{E}(X))^2) = \sum_{x \in W_X} (x - \mu)^2 \cdot \mathbf{P}[X = x].$$

Die Standardabweichung von  $X$  ist definiert als  $\sigma(X) =_{\text{def}} \sqrt{\text{Var}(X)}$ .

**Beispiel:** Wir betrachten den Wahrscheinlichkeitsraum  $(\Omega, \mathbf{P})$  zu „zweimal Würfeln“, d.h.  $\Omega = \{1, \dots, 6\}^2$  und  $\mathbf{P}(\omega) = \frac{1}{36}$  für alle  $\omega \in \Omega$ . Die Summe von zwei Würfeln ist durch die Zufallsvariable  $X : \Omega \rightarrow \mathbb{R} : (j, k) \mapsto j + k$  gegeben. Dann gilt für den Erwartungswert von  $X$

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



$$\begin{aligned}
&= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + \\
&\quad + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\
&= 7.
\end{aligned}$$

Für die Varianz von  $X$  erhalten wir mit  $\mu = 7$

$$\begin{aligned}
\text{Var}(X) &= \sum_{x \in W_X} (x - 7)^2 \cdot \mathbf{P}[X = x] \\
&= (-5)^2 \cdot \frac{1}{36} + (-4)^2 \cdot \frac{2}{36} + (-3)^2 \cdot \frac{3}{36} + (-2)^2 \cdot \frac{4}{36} + (-1)^2 \cdot \frac{5}{36} \\
&\quad + 0^2 \cdot \frac{6}{36} + 1^2 \cdot \frac{5}{36} + 2^2 \cdot \frac{4}{36} + 3^2 \cdot \frac{3}{36} + 4^2 \cdot \frac{2}{36} + 5^2 \cdot \frac{1}{36} \\
&= \frac{210}{36}.
\end{aligned}$$

Somit beträgt die Standardabweichung  $\sigma(X) = \sqrt{210/36} = 2,415\dots$

**Lemma 6.16** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum,  $X, Y : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen und  $a, b \in \mathbb{R}$ . Dann gelten die folgenden Aussagen:*

1.  $\text{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$
2.  $\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$
3.  $\sigma(aX + b) = |a| \cdot \sigma(X)$

**Beweis:** Wir zeigen die Aussagen einzeln:

1. Mit Hilfe der Rechenregeln für den Erwartungswert (Lemma 6.14) erhalten wir:

$$\begin{aligned}
\text{Var}(X) &= \mathbf{E}((X - \mathbf{E}(X))^2) \\
&= \mathbf{E}(X^2 - 2X \cdot \mathbf{E}(X) + \mathbf{E}(X)^2) \\
&= \mathbf{E}(X^2) - 2 \cdot \mathbf{E}(X \cdot \mathbf{E}(X)) + \mathbf{E}(\mathbf{E}(X)^2) \\
&= \mathbf{E}(X^2) - 2 \cdot \mathbf{E}(X)^2 + \mathbf{E}(X)^2 \quad (\mathbf{E}(X), \mathbf{E}(X)^2 \text{ sind Konstanten}) \\
&= \mathbf{E}(X^2) - \mathbf{E}(X)^2
\end{aligned}$$

2. Mit Hilfe der ersten Aussage und der Linearität des Erwartungswertes ergibt sich:

$$\begin{aligned}
\text{Var}(aX + b) &= \mathbf{E}((aX + b)^2) - \mathbf{E}(aX + b)^2 \\
&= \mathbf{E}(a^2X^2 + 2abX + b^2) - (a\mathbf{E}(X) + b)^2 \\
&= a^2\mathbf{E}(X^2) + 2ab\mathbf{E}(X) + b^2 - (a^2\mathbf{E}(X)^2 + 2ab\mathbf{E}(X) + b^2) \\
&= a^2(\mathbf{E}(X^2) - \mathbf{E}(X)^2) \\
&= a^2 \cdot \text{Var}(X)
\end{aligned}$$

3. folgt direkt aus der zweiten Aussage.

Damit ist das Lemma bewiesen. ■

**Definition 6.17** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen.  $X_1, \dots, X_n$  heißen unabhängig, falls für alle  $x_1, \dots, x_n \in \mathbb{R}$  gilt*

$$\mathbf{P}\left(\bigcap_{i=1}^n X_i^{-1}(\{x_i\})\right) = \prod_{i=1}^n \mathbf{P}(X_i^{-1}(\{x_i\}))$$

oder, mit Prädikaten ausgedrückt,

$$\mathbf{P}[X_1 = x_1 \wedge \dots \wedge X_n = x_n] = \prod_{i=1}^n \mathbf{P}[X_i = x_i].$$

**Lemma 6.18** *Es seien  $(\Omega, \mathbf{P})$  ein diskreter Wahrscheinlichkeitsraum und  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariablen. Dann gelten die folgenden Aussagen:*

1.  $\mathbf{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbf{E}(X_i)$
2.  $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$

**Beweis:** Wir beweisen die Aussagen lediglich für den Fall  $n = 2$ . Der allgemeine Fall kann mit vollständiger Induktion über die Anzahl der Variablen behandelt werden.

1. Es seien  $X$  und  $Y$  unabhängige Zufallsvariablen. Dann gilt:

$$\begin{aligned} \mathbf{E}(X \cdot Y) &= \sum_{z \in W_{X \cdot Y}} z \cdot \mathbf{P}[X \cdot Y = z] \\ &= \sum_{x \in W_X} \sum_{y \in W_Y} x \cdot y \cdot \mathbf{P}[X = x \wedge Y = y] \\ &= \sum_{x \in W_X} \sum_{y \in W_Y} x \cdot y \cdot \mathbf{P}[X = x] \cdot \mathbf{P}[Y = y] \\ &\hspace{15em} (X \text{ und } Y \text{ sind unabhängig}) \\ &= \sum_{x \in W_X} x \cdot \mathbf{P}[X = x] \cdot \sum_{y \in W_Y} y \cdot \mathbf{P}[Y = y] \\ &= \left( \sum_{x \in W_X} x \cdot \mathbf{P}[X = x] \right) \cdot \left( \sum_{y \in W_Y} y \cdot \mathbf{P}[Y = y] \right) \\ &= \mathbf{E}(X) \cdot \mathbf{E}(Y) \end{aligned}$$

2. Es seien  $X$  und  $Y$  unabhängige Zufallsvariablen. Mit Hilfe von Lemma 6.16 und der erste Aussage gilt dann:

$$\begin{aligned}
 \text{Var}(X + Y) &= \mathbf{E}((X + Y)^2) - (\mathbf{E}(X + Y))^2 \\
 &= \mathbf{E}(X^2 + 2 \cdot X \cdot Y + Y^2) - (\mathbf{E}(X) + \mathbf{E}(Y))^2 \\
 &= \mathbf{E}(X^2) + 2 \cdot \underbrace{\mathbf{E}(X \cdot Y)}_{=\mathbf{E}(X) \cdot \mathbf{E}(Y)} + \mathbf{E}(Y^2) - \mathbf{E}(X)^2 - 2 \cdot \mathbf{E}(X) \cdot \mathbf{E}(Y) - \mathbf{E}(Y)^2 \\
 &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 + \mathbf{E}(Y^2) - \mathbf{E}(Y)^2 \\
 &= \text{Var}(X) + \text{Var}(Y)
 \end{aligned}$$

Damit ist das Lemma bewiesen. ■

## 6.6 Wichtige diskrete Verteilungen

Es seien  $(\Omega, \mathbf{P})$  ein diskrete Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable. Dann heißt die Funktion

$$f_X(x) =_{\text{def}} \mathbf{P}[X = x]$$

*Dichtefunktion* (oder kurz *Dichte*) von  $X$ .

Im Folgenden werden einige wichtige Verteilungen von Zufallsvariablen über ihre Dichtefunktionen spezifiziert.

**Bernoulli-Verteilung.** Eine Zufallsvariable  $X$  mit  $W_X = \{0, 1\}$  und der Dichte

$$f_X(x) = \begin{cases} p & \text{falls } x = 1 \\ 1 - p & \text{falls } x = 0 \end{cases}$$

heißt *Bernoulli-verteilt* mit dem Parameter  $0 \leq p \leq 1$ . Der Parameter  $p$  heißt auch *Erfolgswahrscheinlichkeit*. Die Komplementärwahrscheinlichkeit  $1 - p$  wird üblicherweise mit  $q$  bezeichnet.

**Beispiel:** Für das Werfen einer Münze gibt  $p$  die Wahrscheinlichkeit an, dass Kopf oben liegt.

Für Bernoulli-verteilte Zufallsvariablen  $X$  gilt:

1.  $\mathbf{E}(X) = p$
2.  $\text{Var}(X) = p - p^2 = p(1 - p) = pq$

**Binomialverteilung.** Eine Zufallsvariable  $X$  mit  $W_X = \{0, 1, \dots, n\}$  und der Dichte

$$f_X(x) =_{\text{def}} \binom{n}{x} p^x (1-p)^{n-x}$$

heißt *binomialverteilt* mit den Parametern  $n \in \mathbb{N}_+$  und  $0 \leq p \leq 1$ ; symbolisch geschrieben:  $X \sim \text{Bin}(n, p)$ .

**Beispiel:** Binomialverteilte Zufallsvariablen entstehen nach dem folgenden Erzeugungsprinzip: Sind  $X_1, \dots, X_n$  unabhängige, mit der Erfolgswahrscheinlichkeit  $p$  Bernoulli-verteilte Zufallsvariablen, so ist  $X =_{\text{def}} X_1 + \dots + X_n$  binomialverteilt mit den Parametern  $n, p$ . Zur Begründung rechnet man nach:

$$f_X(x) = \mathbf{P}(X^{-1}(\{x\})) = \sum_{\substack{\kappa \subseteq \{1, \dots, n\} \\ \|\kappa\| = x}} p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$$

Umgekehrt kann jede binomialverteilte Zufallsvariable so aufgefasst werden.

Für binomialverteilte Zufallsvariablen  $X$  gilt mit dem beschriebenen Erzeugungsprinzip:

1.  $\mathbf{E}(X) = \mathbf{E}(X_1 + \dots + X_n) = n \cdot p$
2.  $\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = n \cdot p \cdot q$

**Geometrische Verteilung.** Eine Zufallsvariable  $X$  mit  $W_X = \mathbb{N}_+$  und der Dichte

$$f_X(x) =_{\text{def}} (1-p)^{x-1} p$$

heißt *geometrisch verteilt* mit dem Parameter  $0 < p < 1$ .

**Beispiel:** Wenn wir eine Münze so lange werfen, bis Kopf oben liegt, so entspricht  $f_X(x)$  der Wahrscheinlichkeit, dass dies genau im  $x$ -ten Versuch geschieht.

Wir müssen uns zunächst davon überzeugen, dass die Dichte tatsächlich ein Wahrscheinlichkeitsmaß definiert:

$$\sum_{x=1}^{\infty} (1-p)^{x-1} p = p \cdot \sum_{x=0}^{\infty} (1-p)^x = p \cdot \frac{1}{1-(1-p)} = 1.$$

Hierbei habe wir folgende Gleichung für die unendliche geometrische Reihe (wodurch gleichzeitig die Namensgebung für die geometrische Verteilung klar wird) verwendet:

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \quad \text{für } 0 \leq q < 1$$

Die Gleichheit lässt sich wie folgt herleiten:

$$A(q) =_{\text{def}} \sum_{n=0}^{\infty} q^n = 1 + \sum_{n=1}^{\infty} q^n = 1 + q \cdot \sum_{n=1}^{\infty} q^{n-1} = 1 + q \cdot \sum_{n=0}^{\infty} q^n = 1 + q \cdot A(q)$$

Mithin gilt  $(1-q) \cdot A(q) = 1$ , also  $A(q) = (1-q)^{-1}$ . Warum wir alle Schritte so durchführen dürfen, wie wir es eben getan haben, werden wir im nächsten Kapitel über Analysis sehen.

Für geometrisch verteilte Zufallsvariablen  $X$  gilt:

$$1. \mathbf{E}(X) = \frac{1}{p}$$

*Begründung:* Wenn wir die erste Ableitung  $A(q)'$  der Funktion  $A(q) =_{\text{def}} \sum_{x=0}^{\infty} q^x$  für  $0 \leq q < 1$  aus zwei Perspektiven betrachten, so erhalten wir einerseits

$$A(q)' = \left( \sum_{x=0}^{\infty} q^x \right)' = \sum_{x=0}^{\infty} (q^x)' = \sum_{x=1}^{\infty} x \cdot q^{x-1}$$

und andererseits

$$A(q)' = \left( \frac{1}{1-q} \right)' = \frac{1}{(1-q)^2}.$$

Damit ergibt sich für den Erwartungswert von  $X$ :

$$\mathbf{E}(X) = \sum_{x=1}^{\infty} x \cdot q^{x-1} \cdot p = p \cdot \sum_{x=1}^{\infty} x \cdot q^{x-1} = p \cdot A(q)' = \frac{p}{(1-q)^2} = \frac{1}{p}$$

$$2. \text{Var}(X) = \frac{1-p}{p^2}$$

*Begründung:* Wir benutzen Lemma 6.16.1. Weiterhin verwenden wir wieder die Funktion  $A(q) =_{\text{def}} \sum_{x=0}^{\infty} q^x$ , die diesmal zweimal differenziert wird:

$$\begin{aligned} A(q)'' &= \left( \sum_{x=0}^{\infty} q^x \right)'' = \sum_{x=0}^{\infty} x(x-1) \cdot q^{x-2} = \sum_{x=0}^{\infty} x^2 \cdot q^{x-2} - \sum_{x=0}^{\infty} x \cdot q^{x-2} \\ &= \frac{1}{q} \cdot \sum_{x=1}^{\infty} x^2 \cdot q^{x-1} - \frac{1}{q} \cdot \sum_{x=1}^{\infty} x \cdot q^{x-1} \\ &= \frac{1}{q} \cdot \sum_{x=1}^{\infty} x^2 \cdot q^{x-1} - \frac{1}{q} \cdot A(q)' \end{aligned}$$

Durch Umstellen und Ausdifferenzieren ergibt sich:

$$\sum_{x=1}^{\infty} x^2 \cdot q^{x-1} = q \cdot A(q)'' + A(q)' = q \cdot \frac{2}{(1-q)^3} + \frac{1}{(1-q)^2} = \frac{1+q}{(1-q)^3} = \frac{2-p}{p^3}$$

Mithin erhalten wir:

$$\text{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \sum_{x=1}^{\infty} x^2 \cdot q^{x-1} \cdot p - \frac{1}{p^2} = p \cdot \frac{2-p}{p^3} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Wir schließen das Kapitel mit einem komplexeren Szenario und seiner wahrscheinlichkeitstheoretischen Modellierung ab.

**Beispiel:** Das Coupon-Collector-Problem besteht in Folgendem: Ein Unternehmen möchte den Absatz seiner Produkt dadurch erhöhen, dass jedes seiner Produkt mit einem Abziehbild versehen wird. Es gäbe  $n$  verschiedene Typen von Abziehbildern. Wie viele Produkte müssen wir im Mittel kaufen, um von jedem Typ ein Abziehbild zu besitzen? Zur Analyse nehmen wir an, dass die Abziehbilder bei jedem Kauf unabhängig voneinander mit gleicher Wahrscheinlichkeit auftreten. Es sei nun  $X$  die Anzahl der Käufe bis zur Vervollständigung der Kollektion von Abziehbildern. Bei der Analyse gehen wir in  $n$  Phasen vor. In der  $i$ -ten Phase betrachten wir die Zufallsvariable

$$X_i \stackrel{\text{def}}{=} \begin{array}{l} \text{Anzahl der Käufe vom Erwerb des } (i-1)\text{-ten Bildes} \\ \text{(ausschließlich) bis zum Erwerb des } i\text{-ten Bildes} \\ \text{(einschließlich)} \end{array}$$

Zum Beispiel könnte für  $n = 4$  folgende Kauffolge auftreten:

$$\underbrace{3}_1, \underbrace{3, 2}_2, \underbrace{2, 3, 3, 1}_3, \underbrace{2, 1, 3, 1, 1, 2, 4}_4$$

Die zugehörigen Zufallsvariablen nehmen damit die Werte  $X_1 = 1$ ,  $X_2 = 2$ ,  $X_3 = 4$ ,  $X_4 = 7$  sowie insgesamt  $X = 14$  an. Im Allgemeinen gilt:

- $X = X_1 + \dots + X_n$
- $X_i$  ist geometrisch verteilt mit  $p = \frac{n-i+1}{n}$
- $\mathbf{E}(X_i) = \frac{n}{n-i+1}$

Damit erhalten wir insgesamt

$$\mathbf{E}(X) = \sum_{i=1}^n \mathbf{E}(X_i) = \sum_{i=1}^n \frac{n}{n-i+1} = n \cdot \sum_{i=1}^n \frac{1}{i} = n \cdot H_n \approx n \cdot \ln n,$$

wobei  $H_n \stackrel{\text{def}}{=} \sum_{i=1}^n \frac{1}{i}$  die  $n$ -te *harmonische Zahl* ist.  $H_n$  verhält sich ungefähr wie  $\ln n$ . Etwas genauer gilt  $H_n \approx \ln n + 0,5772 \dots$ . Im besten Fall benötigen wir  $n$  Käufe, um alle Abziehbilder zu erwerben. Im Mittel benötigen wir dagegen nur um den Faktor  $\ln n$  mehr Käufe.

In diesem Kapitel wollen wir eine Auswahl grundlegender Begriffe der reellen Analysis entwickeln mit dem Ziel, das Konzept der Reihenentwicklung von Funktionen in Potenzreihen zu entwickeln.

## 7.1 Konvergenz von Folgen

Eine (reelle) Folge ist eine Abbildung  $a : \mathbb{N} \rightarrow \mathbb{R}$ . Wenn wir die natürliche Anordnung von  $\mathbb{N}$  zugrunde legen, so kann  $a$  durch die Folge der Funktionswerte  $a(0), a(1), a(2), \dots$  beschrieben werden. Allgemein hat sich dafür die Schreibweise  $(a_0, a_1, a_2, \dots)$  oder kompakt  $(a_n)_{n \in \mathbb{N}}$  etabliert.

Der zentrale Begriff der Analysis ist die Konvergenz.

**Definition 7.1** *Es seien  $(a_n)_{n \in \mathbb{N}}$  eine Folge reeller Zahlen und  $c \in \mathbb{R}$ .*

1. *Die Folge  $(a_n)_{n \in \mathbb{N}}$  konvergiert gegen  $c$ , falls folgende Aussage wahr ist:*

$$(\forall \varepsilon > 0) (\exists n_0 \in \mathbb{N}) (\forall n \geq n_0) [ |a_n - c| < \varepsilon ]$$

*Falls  $(a_n)_{n \in \mathbb{N}}$  gegen  $c$  konvergiert, so heißt  $c$  Grenzwert von  $(a_n)_{n \in \mathbb{N}}$  und wir schreiben  $\lim_{n \rightarrow \infty} a_n = c$ .*

2. *Die Folge  $(a_n)_{n \in \mathbb{N}}$  heißt konvergent, falls ein Grenzwert für  $(a_n)_{n \in \mathbb{N}}$  existiert.*

Die Annäherung einer Folge an einen Grenzwert wird gemäß dieser Definition so verstanden, dass in *jeder* beliebig kleinen Umgebung des Grenzwertes alle Folgenglieder bis auf eine endliche Menge von Ausnahmen (nämlich höchstens diejenigen Folgenglieder mit einem Index kleiner als  $n_0$ ) zu finden sind.

Die logische Struktur der Definitionsbildung ist einigermaßen komplex und bedarf Trainings in der Handhabung. Zur Übung negiere man die formale Definition einer konvergenten Folge mit allen in der vollständigen Definition enthaltenen Quantoren.

Bevor wir Beispiele geben, überzeugen wir uns von der Wohldefiniertheit des Grenzwertes.

**Proposition 7.2** *Der Grenzwert einer konvergenten Folge ist eindeutig.*

**Beweis:** Wir führen einen Widerspruchsbeweis. Es sei  $(a_n)_{n \in \mathbb{N}}$  eine beliebige konvergente Folge. Es seien  $c_1, c_2 \in \mathbb{R}$  Grenzwerte von  $(a_n)_{n \in \mathbb{N}}$ . Angenommen  $c_1 \neq c_2$ . Dann gibt es für  $\varepsilon =_{\text{def}} \frac{1}{2} \cdot |c_1 - c_2| > 0$  natürliche Zahlen  $n_0^{(1)}$  und  $n_0^{(2)}$  mit

- $|a_n - c_1| < \varepsilon$  für alle  $n \geq n_0^{(1)}$  und
- $|a_n - c_2| < \varepsilon$  für alle  $n \geq n_0^{(2)}$ .

Somit gilt für  $N =_{\text{def}} \max \{n_0^{(1)}, n_0^{(2)}\}$

$$|c_1 - c_2| = |c_1 - a_N + a_N - c_2| \leq |c_1 - a_N| + |a_N - c_2| < 2\varepsilon = |c_1 - c_2|.$$

Dies ist jedoch ein Widerspruch. Damit ist die Annahme  $c_1 \neq c_2$  falsch und die Proposition ist bewiesen. ■

**Beispiele:** Wir führen einige Beispiele für die Definition 7.1 an:

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} \frac{1}{n+1}$  konvergiert gegen 0. Dies ist wie folgt einzusehen: Für  $\varepsilon > 0$  definiere  $n_0 =_{\text{def}} \lfloor \varepsilon^{-1} \rfloor$ . Dann gilt für alle  $n \geq n_0$ :

$$|a_n - 0| = \left| \frac{1}{n+1} - 0 \right| = \frac{1}{n+1} \leq \frac{1}{n_0+1} = \frac{1}{\lfloor \varepsilon^{-1} \rfloor + 1} \leq \frac{1}{\varepsilon^{-1}} = \varepsilon$$

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} 2$  konvergiert gegen 2. Dies ist wie folgt einzusehen: Für  $\varepsilon > 0$  definiere  $n_0 =_{\text{def}} 0$ . Dann gilt  $|a_n - 2| = 0 < \varepsilon$  für  $n \geq n_0$ .
- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} \lfloor \sqrt{2} \cdot 10^n \rfloor \cdot 10^{-n}$  konvergiert gegen  $\sqrt{2}$ . Dies ist wie folgt einzusehen: Für  $\varepsilon > 0$  definiere  $n_0 =_{\text{def}} -\lfloor \log_{10} \varepsilon \rfloor$ . Dann gilt für alle  $n \geq n_0$ :

$$\begin{aligned} |a_n - \sqrt{2}| &= \left| \frac{\lfloor \sqrt{2} \cdot 10^n \rfloor}{10^n} - \sqrt{2} \right| = \frac{\sqrt{2} \cdot 10^n - \lfloor \sqrt{2} \cdot 10^n \rfloor}{10^n} \\ &< 10^{-n} \leq 10^{\lfloor \log_{10} \varepsilon \rfloor} \leq \varepsilon \end{aligned}$$

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} (-1)^n$  ist nicht konvergent. Angenommen  $\lim_{n \rightarrow \infty} a_n = c$ . Dann gibt es für  $\varepsilon = \frac{1}{3}$  ein  $n_0 \in \mathbb{N}$ , sodass  $|a_n - c| < \frac{1}{3}$  für alle  $n \geq n_0$  gilt. Dann gilt aber auch für  $n \geq n_0$ :

$$2 = |a_n - a_{n+1}| = |a_n - c - a_{n+1} + c| \leq |a_n - c| + |a_{n+1} - c| < \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

Dies ist jedoch ein Widerspruch. Somit existiert kein Grenzwert.

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_0 =_{\text{def}} 1$  und  $a_n =_{\text{def}} \frac{n^{(-1)^n}}{n}$  ist nicht konvergent. Die Folge beginnt mit  $1, -1, 1, \frac{1}{9}, 1, \frac{1}{25}, 1, \frac{1}{49}, \dots$ . Die Begründung für die Nichtkonvergenz folgt analog zum vorangegangenen Beispiel aus  $|a_n - a_{n+1}| \geq \frac{8}{9}$  für alle  $n \geq 2$ .



Das dritte Beispiel ist bemerkenswert: Obwohl alle Folgenglieder rationale Zahlen sind, liegt der Grenzwert  $\sqrt{2}$  der Folge nicht im Bereich der rationalen Zahlen. Nicht jede Menge ist also abgeschlossen unter Grenzwertbildung. Die Menge der reellen Zahlen kann gerade als die Menge aller Grenzwerte konvergenter rationaler Folgen konstruiert werden. Damit kann gezeigt werden, dass die reellen Zahlen im Gegensatz zu den rationalen Zahlen auch abgeschlossen unter Grenzwertbildung sind.

Das folgende Lemma gibt Regeln zur Bestimmung von Grenzwerten an.

**Lemma 7.3** *Es seien  $(a_n)_{n \in \mathbb{N}}$ ,  $(b_n)_{n \in \mathbb{N}}$  und  $(c_n)_{n \in \mathbb{N}}$  reelle konvergente Folgen. Dann gilt:*

1. *Gilt  $a_n \leq b_n$  für alle  $n \geq n_0$  und geeignetes  $n_0 \in \mathbb{N}$ , so gilt  $\lim_{n \rightarrow \infty} a_n \leq \lim_{n \rightarrow \infty} b_n$*
2. *Gilt  $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = x$  und gilt  $a_n \leq b_n \leq c_n$  für alle  $n \geq n_0$  und geeignetes  $n_0 \in \mathbb{N}$ , so gilt  $\lim_{n \rightarrow \infty} b_n = x$*
3.  $\lim_{n \rightarrow \infty} (a_n + b_n) = \left( \lim_{n \rightarrow \infty} a_n \right) + \left( \lim_{n \rightarrow \infty} b_n \right)$
4.  $\lim_{n \rightarrow \infty} (a_n \cdot b_n) = \left( \lim_{n \rightarrow \infty} a_n \right) \cdot \left( \lim_{n \rightarrow \infty} b_n \right)$
5.  $\lim_{n \rightarrow \infty} \left( \frac{a_n}{b_n} \right) = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}$ , falls  $\lim_{n \rightarrow \infty} b_n \neq 0$  und  $b_n \neq 0$  für alle  $n \in \mathbb{N}$

**Beweis:** Wir beweisen die Aussagen einzeln:

1. Es gelte  $\lim_{n \rightarrow \infty} a_n = x$  und  $\lim_{n \rightarrow \infty} b_n = y$  sowie  $a_n \leq b_n$  für alle  $n \geq n_0$  und ein geeignetes  $n_0 \in \mathbb{N}$ . Angenommen es gilt  $x > y$ . Dann gibt es für  $\varepsilon = \frac{1}{2}(x - y) > 0$  natürliche Zahlen  $n_0^{(1)}$  und  $n_0^{(2)}$  mit

- $|a_n - x| < \varepsilon$  und folglich  $a_n > x - \varepsilon$  für alle  $n \geq n_0^{(1)}$
- $|b_n - y| < \varepsilon$  und folglich  $b_n < y + \varepsilon$  für alle  $n \geq n_0^{(2)}$

Damit gilt für alle  $n \geq \max \{n_0, n_0^{(1)}, n_0^{(2)}\}$

$$a_n - b_n > (x - \varepsilon) - (y + \varepsilon) = x - y - 2\varepsilon = 0,$$

d.h.  $a_n > b_n$ . Dies ist ein Widerspruch. Mithin gilt  $x \leq y$ .

2. folgt aus der ersten Aussage.
3. Es gelte  $\lim_{n \rightarrow \infty} a_n = x$  und  $\lim_{n \rightarrow \infty} b_n = y$ . Dann gibt es für jedes  $\varepsilon > 0$  natürliche Zahlen  $n_0^{(1)}$  und  $n_0^{(2)}$  mit

- $|a_n - x| < \frac{\varepsilon}{2}$  für alle  $n \geq n_0^{(1)}$

- $|b_n - y| < \frac{\varepsilon}{2}$  für alle  $n \geq n_0^{(2)}$

Für  $n \geq \max\{n_0^{(1)}, n_0^{(2)}\}$  gilt somit auch

$$a_n + b_n - (x + y) = |a_n - x + b_n - y| \leq |a_n - x| + |b_n - y| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Folglich konvergiert die Folge  $(a_n + b_n)_{n \in \mathbb{N}}$  gegen  $x + y$ .

4. Es gelte  $\lim_{n \rightarrow \infty} a_n = x$  und  $\lim_{n \rightarrow \infty} b_n = y$ . Zunächst halten wir fest, dass  $(a_n)_{n \in \mathbb{N}}$  beschränkt ist (siehe Übungsblatt 9). Es gibt also ein  $s > 0$  mit  $|a_n| \leq s$  für alle  $n \in \mathbb{N}$ . Weiterhin existieren zu jedem  $\varepsilon > 0$  geeignete natürliche Zahlen  $n_0^{(1)}$  und  $n_0^{(2)}$  mit

- $|a_n - x| < \frac{\varepsilon}{2} \cdot (\max\{|y|, 1\})^{-1}$  für alle  $n \geq n_0^{(1)}$
- $|b_n - y| < \frac{\varepsilon}{2} \cdot s^{-1}$  für alle  $n \geq n_0^{(2)}$

Für  $n \geq \max\{n_0^{(1)}, n_0^{(2)}\}$  gilt somit

$$\begin{aligned} |a_n b_n - xy| &= |a_n b_n - a_n y + a_n y - xy| \\ &= |a_n(b_n - y) + (a_n - x)y| \\ &\leq |a_n| \cdot |b_n - y| + |y| \cdot |a_n - x| \\ &< s \cdot \left(\frac{\varepsilon}{2} \cdot s^{-1}\right) + |y| \cdot \left(\frac{\varepsilon}{2} \cdot (\max\{|y|, 1\})^{-1}\right) \\ &\leq \varepsilon \end{aligned}$$

Mithin konvergiert die Folge  $(a_n b_n)_{n \in \mathbb{N}}$  gegen  $xy$ .

5. Wegen der vierten Aussage genügt es,  $\lim_{n \rightarrow \infty} (b_n^{-1}) = \left(\lim_{n \rightarrow \infty} b_n\right)^{-1}$  unter den angegebenen Voraussetzungen zu zeigen. Es gelte also  $\lim_{n \rightarrow \infty} b_n = y \neq 0$  sowie  $b_n \neq 0$  für alle  $n \in \mathbb{N}$ . Wegen  $b_n \neq 0$  ist  $b_n^{-1}$  stets definiert. Wegen  $y \neq 0$  und da  $(b_n)_{n \in \mathbb{N}}$  gegen  $y$  konvergiert, gibt es eine natürliche Zahl  $n_0$ , sodass für alle  $n \geq n_0$

$$|y| = |y - b_n + b_n| \leq |y - b_n| + |b_n| \leq \frac{|y|}{2} + |b_n|$$

also  $|b_n| \geq \frac{|y|}{2}$  gilt. Weiterhin gibt es für alle  $\varepsilon > 0$  ein  $n'_0 \in \mathbb{N}$  mit  $|b_n - y| < \frac{\varepsilon}{2} \cdot |y|^2$ . Für  $n \geq \max\{n_0, n'_0\}$  folgt somit

$$\left| \frac{1}{b_n} - \frac{1}{y} \right| = \frac{|y - b_n|}{|b_n| \cdot |y|} \leq \frac{2 \cdot |b_n - y|}{|y|^2} < \varepsilon$$

Mithin konvergiert  $(b_n^{-1})_{n \in \mathbb{N}}$  gegen  $y^{-1}$ .

Damit ist das Lemma bewiesen. ■

**Beispiele:** Folgende Grenzwerte werden mit Hilfe von Lemma 7.3 bestimmt:

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} \frac{2n^2+17n}{3n^2+5}$  konvergiert gegen  $\frac{2}{3}$ . Denn es gilt:

$$\lim_{n \rightarrow \infty} \frac{2n^2 + 17}{3n^2 + 5} = \lim_{n \rightarrow \infty} \frac{n^2 \cdot (2 + \frac{17}{n})}{n^2 \cdot (3 + \frac{5}{n^2})} = \frac{\lim_{n \rightarrow \infty} 2 + \frac{17}{n}}{\lim_{n \rightarrow \infty} 3 + \frac{5}{n^2}} = \frac{2}{3}$$

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} \frac{n^2}{2^n}$  konvergiert gegen 0. Wegen  $2^n \geq n^3$  für  $n \geq 10$  gilt nämlich

$$0 \leq \lim_{n \rightarrow \infty} \frac{n^2}{2^n} \leq \lim_{n \rightarrow \infty} \frac{n^2}{n^3} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

- Die Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} \sqrt[n]{n}$  konvergiert gegen 1. Zum Nachweis betrachten wir zunächst die Hilfsfolge  $(d_n)_{n \in \mathbb{N}}$  mit  $d_n =_{\text{def}} (1 + \sqrt{2/n})^n$ . Dann gilt für  $n \geq 2$  nach dem Binomialtheorem

$$d_n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} (\sqrt{2/n})^k \geq 1 + \binom{n}{1} \sqrt{2/n} + \binom{n}{2} \frac{2}{n} = n + \sqrt{2n} \geq n.$$

Wir definieren Folgen  $(b_n)_{n \in \mathbb{N}}$  mit  $b_n =_{\text{def}} 1$  und  $(c_n)_{n \in \mathbb{N}}$  mit  $c_n =_{\text{def}} \sqrt[n]{d_n}$ . Für  $n \geq 2$  gilt

$$b_n \leq a_n \leq c_n = \sqrt[n]{d_n} = \sqrt[n]{(1 + \sqrt{2/n})^n} = 1 + \sqrt{2/n}.$$

Wir erhalten  $1 = \lim_{n \rightarrow \infty} b_n = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = 1$ .

## 7.2 Konvergenz von Reihen

Zu einer Folge  $(a_n)_{n \in \mathbb{N}}$  definieren wir die *Reihe*  $(s_n)_{n \in \mathbb{N}}$  wie folgt für  $n \in \mathbb{N}$ :

$$s_n =_{\text{def}} \sum_{k=0}^n a_k$$

Die Folgenglieder  $a_k$  heißen *Koeffizienten*. Falls  $(s_n)_{n \in \mathbb{N}}$  konvergiert, schreiben wir

$$\sum_{k=0}^{\infty} a_k =_{\text{def}} \lim_{n \rightarrow \infty} s_n.$$

Üblicherweise wird die unendliche Summe  $\sum_{k=0}^{\infty} a_k$  (also ein Grenzwert) mit der Reihe  $(s_n)_{n \in \mathbb{N}}$  identifiziert.

**Beispiele:** Wir geben einige Reihen und ihre Grenzwerte an, falls sie existieren:

- Zur Folge  $(a_n)_{n \in \mathbb{N}}$  mit  $a_n =_{\text{def}} q^n$  für  $0 \leq q < 1$  gehört die geometrische Reihe  $\sum_{n=0}^{\infty} q^n$ . Den Grenzwert haben wir bereits im letzten Kapitel durch einen algebraischen Trick mit  $(1-q)^{-1}$  bestimmt. Wir wollen überprüfen, ob die Rechnung korrekt war. Es sei  $s_n =_{\text{def}} \sum_{k=0}^n q^k$ . Wir wissen bereits, dass mittels vollständiger Induktion gezeigt werden kann:

$$s_n = \frac{1 - q^{n+1}}{1 - q} \quad \text{für } n \in \mathbb{N}$$

Wegen  $\lim_{n \rightarrow \infty} (1 - q^{n+1}) = 1$  und  $\lim_{n \rightarrow \infty} (1 - q) = 1 - q$  gilt

$$\sum_{n=0}^{\infty} q^n = \lim_{n \rightarrow \infty} s_n = \frac{\lim_{n \rightarrow \infty} (1 - q^{n+1})}{\lim_{n \rightarrow \infty} (1 - q)} = \frac{1}{1 - q}.$$

- $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$ , denn es gilt:

$$\begin{aligned} s_n &=_{\text{def}} \sum_{k=1}^{n+1} \frac{1}{k(k+1)} = \sum_{k=1}^{n+1} \frac{(k+1) - k}{k(k+1)} = \sum_{k=1}^{n+1} \left( \frac{1}{k} - \frac{1}{k+1} \right) \\ &= \left( \sum_{k=1}^{n+1} \frac{1}{k} \right) - \left( \sum_{k=2}^{n+2} \frac{1}{k} \right) = 1 - \frac{1}{n+2} \end{aligned}$$

Also folgt  $\lim_{n \rightarrow \infty} s_n = 1$ .

- $\sum_{n=1}^{\infty} \frac{1}{n}$  existiert nicht. Dies ist wie folgt einzusehen: Es sei  $s_n =_{\text{def}} \sum_{k=1}^{n+1} \frac{1}{k}$ .

Angenommen  $s_n$  konvergiert gegen  $c$ . Dann gibt es für  $\varepsilon = \frac{1}{2}$  eine natürliche Zahl  $n_0$  mit  $|s_n - c| = c - s_n < \frac{1}{2}$ . Hierbei ist zu beachten, dass  $s_n$  monoton wachsend ist, d.h.  $s_n \leq s_m$  für  $n \leq m$ . Mithin gilt für  $m \geq n \geq n_0$  auch  $s_m - s_n < \frac{1}{2}$ . Insbesondere erhalten wir für  $m \geq 2n + 1 \geq n \geq n_0$

$$\begin{aligned} \frac{1}{2} > s_m - s_n &= \sum_{k=n+2}^{m+1} \frac{1}{k} \geq \frac{1}{m+1} \cdot (m+1 - (n+2) + 1) \\ &= 1 - \frac{n+1}{m+1} \geq 1 - \frac{n+1}{2n+2} = \frac{1}{2} \end{aligned}$$

Dies ist ein Widerspruch. Somit gibt es keinen Grenzwert für die Reihe.

Die zentrale Frage für Reihen ist: Welche Kriterien müssen die Koeffizienten erfüllen, damit die Reihe konvergiert? Um dieser Frage nachzugehen, führen wir einen verschärften Konvergenzbegriff für Reihen ein.

**Definition 7.4** Die Reihe  $\sum_{n=0}^{\infty} a_n$  heißt absolut konvergent, falls  $\sum_{n=0}^{\infty} |a_n|$  konvergent ist.

Klarerweise ist jede absolut konvergente Reihe auch konvergent. Die Umkehrung gilt nicht.

**Beispiel:**  $\sum_{n=0}^{\infty} (-1)^n \frac{1}{n}$  ist konvergent aber nicht absolut konvergent.

Folgende hinreichende Kriterien können für die absolute Konvergenz von Reihen aufgestellt werden.

**Lemma 7.5** Es seien  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  Folgen.

1. Ist  $\sum_{n=0}^{\infty} b_n$  absolut konvergent und gilt  $|a_n| \leq |b_n|$  für alle  $n \geq n_0$  und ein geeignetes  $n_0$ , so ist  $\sum_{n=0}^{\infty} a_n$  absolut konvergent (Majoranten-Kriterium).
2. Gibt es  $0 \leq q < 1$  und  $n_0 \in \mathbb{N}$  mit  $\sqrt[n]{|a_n|} \leq q$  für alle  $n \geq n_0$ , so ist  $\sum_{n=0}^{\infty} a_n$  absolut konvergent (Wurzel-Kriterium).
3. Gibt es  $0 \leq q < 1$  und  $n_0 \in \mathbb{N}$  mit  $|a_{n+1}| \leq q \cdot |a_n|$  für alle  $n \geq n_0$ , so ist  $\sum_{n=0}^{\infty} a_n$  absolut konvergent (Quotienten-Kriterium).

**Beweis:** Wir zeigen die Aussagen einzeln (und nehmen zur Vereinfachung  $n_0 = 0$  an):

1. Es seien  $s_n =_{\text{def}} \sum_{k=0}^n |a_k|$  und  $x =_{\text{def}} \sum_{k=0}^{\infty} |b_k|$ . Dann gilt  $s_n \leq x$ . Weiterhin gilt  $s_n \leq s_m$  für alle  $n \leq m$ , d.h.  $(s_n)_{n \in \mathbb{N}}$  ist monoton wachsend. Wir definieren

$$s =_{\text{def}} \sup \{ s_n \mid n \in \mathbb{N} \}.$$

Wgen  $s_n \leq x$  für alle  $n \in \mathbb{N}$  existiert  $s$  als reelle Zahl und es gilt  $\lim_{n \rightarrow \infty} s_n = s$ . Letzteres ist wie folgt einzusehen: Für  $\varepsilon > 0$  gibt es ein  $n_0$  mit  $s_{n_0} > s - \varepsilon$  nach Definition von  $s$  als Supremum für die Menge aller Folgenglieder  $s_n$ . Da  $s_n$  monoton wachsend ist, folgt  $s \geq s_n > s - \varepsilon$  für alle  $n \geq n_0$ . Mithin gilt  $|s_n - s| < \varepsilon$  für alle  $n \geq n_0$ . Somit konvergiert  $s_n$  gegen  $s$ .

2. Mit  $\sqrt[n]{|a_n|} \leq q$  gilt  $|a_n| \leq q^n$  für alle  $n$ . Folglich ist  $\sum_{n=0}^{\infty} |a_n|$  konvergent nach dem Majorantenkriterium, denn es gilt

$$\sum_{n=0}^{\infty} |a_n| \leq \sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \quad \text{für } 0 \leq q < 1$$

3. Mit  $|a_{n+1}| \leq q \cdot |a_n|$  gilt  $|a_n| \leq q^n \cdot |a_0|$  für alle  $n$ . Folglich ist  $\sum_{n=0}^{\infty} |a_n|$  konvergent nach dem Majorantenkriterium, denn es gilt

$$\sum_{n=0}^{\infty} |a_n| \leq \sum_{n=0}^{\infty} q^n \cdot |a_0| = \frac{|a_0|}{1-q} \quad \text{für } 0 \leq q < 1$$

Damit ist das Lemma bewiesen. ■

### 7.3 Potenzreihen

**Definition 7.6** Es sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge. Eine Reihe der Form

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n$$

mit  $x, x_0 \in \mathbb{R}$  heißt Potenzreihe (um den Entwicklungspunkt  $x_0$ ).

**Theorem 7.7** Zu jeder Potenzreihe  $\sum_{n=0}^{\infty} a_n (x - x_0)^n$  existiert eine Konvergenzradius  $R$  mit  $0 \leq R \leq \infty$  und

1.  $\sum_{n=0}^{\infty} a_n (x - x_0)^n$  ist absolut konvergent, falls  $|x - x_0| < R$ ,
2.  $\sum_{n=0}^{\infty} a_n (x - x_0)^n$  ist divergent, falls  $|x - x_0| > R$ .

Eine allgemeine Aussage für den Konvergenzradius, d.h.  $|x - x_0| = R$ , ist nicht möglich.

**Lemma 7.8** Es sei  $\sum_{n=0}^{\infty} a_n(x - x_0)^n$  eine Potenzreihe mit  $a_n \neq 0$  für alle  $n \in \mathbb{N}$ .

1. Konvergiert (oder divergiert bestimmt)  $\frac{|a_n|}{|a_{n+1}|}$ , so gilt  $\lim_{n \rightarrow \infty} \frac{|a_n|}{|a_{n+1}|} = R$   
(Quotienten-Kriterium).

2. Konvergiert (oder divergiert bestimmt)  $\frac{1}{\sqrt[n]{|a_n|}}$ , so gilt  $\lim_{n \rightarrow \infty} \frac{1}{\sqrt[n]{|a_n|}} = R$   
(Wurzel-Kriterium).

Hierbei ist die bestimmte Divergenz wie folgt definiert: Eine Folge  $(a_n)_{n \in \mathbb{N}}$  divergiert bestimmt gegen  $+\infty$ , falls gilt

$$(\forall c > 0)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)[a_n > c].$$

Gilt die Aussage mit  $c < 0$  statt  $c > 0$ , so divergiert die Folge bestimmt gegen  $-\infty$ .

**Beispiel:** Die Konvergenzradien für folgende Potenzreihen (um den Entwicklungspunkt  $x_0 = 0$ ) können mittels des Wurzel-Kriteriums aus Lemma 7.8 bestimmt werden:

- $\sum_{n=0}^{\infty} x^n$  besitzt den Konvergenzradius  $R = 1$  mit Divergenz für  $|x| = 1$
- $\sum_{n=0}^{\infty} \frac{x^n}{n}$  besitzt den Konvergenzradius  $R = 1$  mit Divergenz für  $x = 1$  und Konvergenz für  $x = -1$
- $\sum_{n=0}^{\infty} \frac{x^n}{n^2}$  besitzt den Konvergenzradius  $R = 1$  mit Konvergenz für  $|x| = 1$

**Lemma 7.9** Es sei  $f(x) =_{\text{def}} \sum_{n=0}^{\infty} a_n \cdot x^n$  eine Potenzreihe (um den Entwicklungspunkt  $x_0 = 0$ ) mit Konvergenzradius  $R > 0$ , wobei  $a_n \neq 0$  für mindestens ein  $n \in \mathbb{N}$ . Dann gibt es ein  $r \in \mathbb{R}$  mit  $0 < r < R$ , so dass  $f(x)$  in der Menge  $U_r =_{\text{def}} \{ x \mid x \in \mathbb{R}, |x| < r \}$  nur endlich viele Nullstellen besitzt.

**Beweis:** Bevor wir die Aussage beweisen, leiten wir eine Ungleichung her, auf der der eigentliche Beweis basiert. Es sei  $N =_{\text{def}} \min \{ n \mid a_n \neq 0 \} < \infty$ . Dann gilt für jedes  $r$  mit  $0 < r < R$  und jedes  $x \in U_r$ :

$$|f(x) - a_N \cdot x^N| = \left| \left( a_N \cdot x^N + \sum_{n=N+1}^{\infty} a_n \cdot x^n \right) - a_N \cdot x^N \right| = \left| \sum_{n=N+1}^{\infty} a_n \cdot x^n \right|$$

$$\begin{aligned}
&\leq \sum_{n=N+1}^{\infty} |a_n| \cdot |x|^n = \sum_{n=N+1}^{\infty} |a_n| \cdot \underbrace{|x| \cdots |x|}_{n-(N+1)} \cdot \underbrace{|x| \cdots |x|}_{N+1} \\
&\leq \sum_{n=N+1}^{\infty} |a_n| \cdot \underbrace{r \cdots r}_{n-(N+1)} \cdot \underbrace{|x| \cdots |x|}_{N+1} \\
&= |x|^{N+1} \cdot \underbrace{\sum_{n=N+1}^{\infty} |a_n| \cdot r^{n-(N+1)}}_{s \stackrel{\text{def}}{=} } \\
&= s \cdot |x|^{N+1}
\end{aligned}$$

Hierbei ist zweierlei zu beachten: Einmal hängt  $s$  nicht mehr von  $x$  ab und zweitens ist  $s$  tatsächlich definiert, da wegen  $r < R$  die Reihe in jedem Fall konvergiert.

Wir führen nun einen Widerspruchsbeweis, um die Aussage zu beweisen. Dazu fixieren wir zunächst ein  $r \in \mathbb{R}$  mit  $0 < r < R$ . Wir nehmen an, dass  $f$  in jeder Umgebung  $U_{r/k}$  für alle  $k \in \mathbb{N}_+$  unendlich viele Nullstellen besitzt. Wir betrachten eine beliebige Folge  $(x_k)_{k \in \mathbb{N}_+}$ , wobei  $x_k$  gerade eine Nullstelle von  $f$  in  $U_{r/k} \setminus \{0\}$  ist. Mit obiger Abschätzung gilt damit für alle  $k \in \mathbb{N}_+$ :

$$s \cdot |x_k|^{N+1} \geq |f(x_k) - a_N \cdot x_k^N| = |a_N| \cdot |x_k|^N$$

Folglich ist  $|a_N| \leq s \cdot |x_k|$  für alle  $k \in \mathbb{N}_+$ . Nun gilt aber:

$$0 \leq |a_N| \leq \lim_{k \rightarrow \infty} s \cdot |x_k| \leq s \cdot \lim_{k \rightarrow \infty} \frac{r}{k} = 0$$

Dies ist jedoch ein Widerspruch zu  $a_N \neq 0$ . Folglich gibt es ein  $k \in \mathbb{N}_+$ , sodass  $f$  in  $U_{r/k}$  nur endlich viele Nullstellen besitzt. Damit ist das Lemma bewiesen. ■

**Theorem 7.10** *Es seien  $f_a(x) =_{\text{def}} \sum_{n=0}^{\infty} a_n \cdot x^n$  und  $f_b(x) =_{\text{def}} \sum_{n=0}^{\infty} b_n \cdot x^n$  zwei Potenzreihen (um  $x_0 = 0$ ) mit den Konvergenzradien  $R_a \geq R_b > 0$ . Gibt es eine reelle Zahl  $r$  mit  $0 < r < R_b$  und  $f_a(x) = f_b(x)$  für alle  $x$  mit  $|x| < r$ , so gilt  $a_n = b_n$  für alle  $n \in \mathbb{N}$ .*

**Beweis:** Wir betrachten die Funktion  $f =_{\text{def}} f_a - f_b$  mit der zugehörigen Potenzreihe  $\sum_{n=0}^{\infty} (a_n - b_n)x^n$ . Dann gilt  $f(x) = 0$  für alle  $x$  mit  $|x| < r$ . Für alle  $0 < r^* < r$  existieren also unendlich viele Nullstellen von  $f$  in den Mengen  $U_{r^*}$ . Nach Lemma 7.9 ist dies für eine Potenzreihe mit wenigstens einem von 0 verschiedenen Koeffizienten nicht möglich. Mithin muss  $a_n - b_n = 0$  bzw.  $a_n = b_n$  für alle  $n \in \mathbb{N}$  gelten. Damit ist das Theorem bewiesen. ■



Wie können wir nun Potenzreihen für gegebenen Funktionen bestimmen? Eine Antwort darauf gibt die folgende Definition: Für eine beliebig oft differenzierbare Funktion  $f$  heißt

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} \cdot (x - x_0)^n$$

TAYLOR-Reihe von  $f$  am Punkt  $x_0$ . Hierbei bezeichnet wie üblich  $f^{(n)}(x_0)$  den Wert der  $n$ -ten Ableitung von  $f$  nach  $x$  an der Stelle  $x_0$ .

**Beispiele:** Einige wichtige TAYLOR-Reihen für Funktionen (ohne Begründung der Konvergenzradien) sind wie folgt:

- $e^x = \sum_{n=0}^{\infty} \frac{1}{n!} \cdot x^n$  für alle  $x \in \mathbb{R}$  wegen  $(e^x)' = e^x$
- $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$  für alle  $-1 < x < 1$  wegen  $\left(\frac{1}{1-x}\right)^{(n)} = \frac{n!}{(1-x)^{n+1}}$
- $\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \cdot x^n$  für  $-1 < x \leq 1$  wegen  $\ln(1+0) = 0$   
und  $(\ln(1+x))^{(n)} = \frac{(-1)^{n+1}(n-1)!}{(1+x)^n}$ . Ein interessanter Spezialfall dieser Potenzreihe ergibt sich für  $x = 1$ :

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = - \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \cdot 1^n = -\ln(1+1) = -\ln 2 = -0,6931 \dots$$

**Theorem 7.11** Jede Potenzreihe  $f(x) =_{\text{def}} \sum_{n=0}^{\infty} a_n(x - x_0)^n$  mit dem Konvergenzradius  $R > 0$  ist für alle  $x$  mit  $|x - x_0| < R$  gleich ihrer TAYLOR-Reihe.

**Beweis:** Wir beweisen die Aussage lediglich für  $x_0 = 0$ . Zunächst halten wir fest, dass die Ableitung der Potenzreihe von  $f(x)$

$$\left( \sum_{n=0}^{\infty} a_n \cdot x^n \right)' = \sum_{n=0}^{\infty} (a_n \cdot x^n)' = \sum_{n=1}^{\infty} n a_n \cdot x^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} \cdot x^n$$

den gleichen Konvergenzradius  $R$  wie die Potenzreihe von  $f(x)$  besitzt (siehe Übungsblatt 10). Damit können wir  $f(x)$  innerhalb des Konvergenzradius der Potenzreihe beliebig oft

differenzieren. Wir erhalten im Einzelnen folgende Ableitungen:

$$\begin{array}{rcl}
 f(x) & = & a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots + a_n x^n + \dots \\
 f'(x) & = & a_1 + 2a_2 x + 3a_3 x^2 + \dots + na_n x^{n-1} + \dots \\
 f''(x) & = & 2a_2 + 6a_3 x + \dots + n(n-1)a_n x^{n-2} + \dots \\
 f'''(x) & = & 6a_3 + \dots + n(n-1)(n-2)a_n x^{n-3} + \dots \\
 & \vdots & \\
 f^{(n)}(x) & = & n! a_n + \dots \\
 & \vdots &
 \end{array}$$

Insbesondere gilt  $f^{(n)}(0) = n! \cdot a_n$  für alle  $n \in \mathbb{N}$ . Durch Umstellung nach  $a_n$  ergibt sich das Theorem. ■

In diesem Kapitel geben wir eine kurze Einführung in die Grundbegriffe der linearen Algebra.

## 8.1 Lineare Räume

**Definition 8.1** Eine linearer Raum ist eine Menge  $V$  mit zwei Funktionen (Operationen)  $+$  :  $V \times V \rightarrow V$  (Vektoraddition) und  $\cdot$  :  $\mathbb{R} \times V \rightarrow V$  (skalare Multiplikation), sodass folgende Eigenschaften erfüllt sind:

1.  $(V, +)$  ist eine abelsche Gruppe
2. Für alle  $a, b \in \mathbb{R}$  und  $v \in V$  gilt  $(a + b) \cdot v = a \cdot v + b \cdot v$
3. Für alle  $a \in \mathbb{R}$  und  $v, w \in V$  gilt  $a \cdot (v + w) = a \cdot v + a \cdot w$
4. Für alle  $a, b \in \mathbb{R}$  und  $v \in V$  gilt  $(a \cdot b) \cdot v = a \cdot (b \cdot v)$
5. Für alle  $v \in V$  gilt  $1 \cdot v = v$

Die Elemente von  $V$  heißen Vektoren.

**Beispiele:** Die folgenden Menge mit den geeigneten zugehörigen Operationen bilden lineare Räume:

- $\mathbb{R}^n$  für  $n \in \mathbb{N}_+$  mit komponentenweiser Addition und Multiplikation mit Konstanten:

$$v + w = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} \quad \text{sowie} \quad a \cdot v = \begin{pmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{pmatrix}$$

- Die Menge aller Polynome  $\sum_{i=0}^n a_i \cdot x^i$  vom Grad  $n$ ,  $a_i \in \mathbb{R}$  mit der üblichen Addition und Multiplikation von Konstanten:

$$\begin{aligned} \left( \sum_{i=0}^n a_i \cdot x^i \right) + \left( \sum_{i=0}^n b_i \cdot x^i \right) &= \sum_{i=0}^n (a_i + b_i) \cdot x^i && \text{sowie} \\ c \sum_{i=0}^n a_i \cdot x^i &= \sum_{i=0}^n (c \cdot a_i) \cdot x^i \end{aligned}$$

- Die Menge aller konvergenten Folgen mit komponentenweise Addition und Multiplikation mit Konstanten
- Die Menge aller (formalen) Potenzreihen

**Definition 8.2** Es seien  $V$  ein linearer Raum und  $\emptyset \neq W \subseteq V$ . Dann heißt  $W$  linearer Unterraum (von  $V$ ), falls für alle  $v, w \in W$  und  $a \in \mathbb{R}$  die folgenden Bedingungen erfüllt sind:

1. Sind  $v, w \in W$ , so ist  $v + w \in W$  (Abgeschlossenheit unter  $+$ )
2. Ist  $v \in W$ , so ist  $a \cdot v \in W$  (Abgeschlossenheit unter  $\cdot$ )

Die Benennung einer Teilmenge  $W$  von  $V$ , die die beiden obigen Eigenschaften erfüllt, als Unterraum ist plausibel.

**Proposition 8.3** Jeder lineare Unterraum eines linearen Raumes ist ein linearer Raum.

**Beweis:** Offensichtlich durch Überprüfung der Eigenschaften eines linearen Raumes. ■

**Beispiele:** Folgende Mengen bilden Unterräume in den entsprechenden linearen Räumen:

- $\mathbb{R}^2 \times \{0\}$  ist ein Unterraum von  $\mathbb{R}^3$ , denn es gilt

$$\underbrace{\begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix}}_{\in \mathbb{R}^2 \times \{0\}} + \underbrace{\begin{pmatrix} w_1 \\ w_2 \\ 0 \end{pmatrix}}_{\in \mathbb{R}^2 \times \{0\}} = \underbrace{\begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ 0 \end{pmatrix}}_{\in \mathbb{R}^2 \times \{0\}} \quad \text{sowie} \quad a \cdot \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix}}_{\in \mathbb{R}^2 \times \{0\}} = \underbrace{\begin{pmatrix} av_1 \\ av_2 \\ 0 \end{pmatrix}}_{\in \mathbb{R}^2 \times \{0\}}$$

- Die Menge der Polynome vom Grad  $m$  ist ein Unterraum in der Menge der Polynome vom Grad  $n \geq m$ , denn

$$\begin{aligned} \left( \sum_{i=0}^m a_i \cdot x^i \right) + \left( \sum_{i=0}^m b_i \cdot x^i \right) &= \sum_{i=0}^m (a_i + b_i) \cdot x^i \quad \text{und} \\ c \sum_{i=0}^m a_i \cdot x^i &= \sum_{i=0}^m (c \cdot a_i) \cdot x^i \end{aligned}$$

sind wieder Polynome vom Grad  $m$ .

**Definition 8.4** Es seien  $V$  ein linearer Raum und  $v_1, \dots, v_n \in V$  Vektoren. Der Vektor  $w \in V$  heißt Linearkombination von  $v_1, \dots, v_n$ , falls  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  existieren mit

$$w = \lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n.$$

Die Menge aller Linearkombinationen von  $v_1, \dots, v_n$  heißt lineare Hülle (engl. span) von  $v_1, \dots, v_n$  und wird mit  $\text{span}\{v_1, \dots, v_n\}$  bezeichnet.

**Proposition 8.5** Es seien  $V$  ein linearer Raum und  $v_1, \dots, v_n \in V$  Vektoren. Dann ist die lineare Hülle von  $v_1, \dots, v_n$  ein linearer Unterraum von  $V$ .

**Beweis:** Wir müssen zeigen, dass  $\text{span}\{v_1, \dots, v_n\}$  abgeschlossen unter Addition und skalarer Multiplikation ist. Es seien  $w, w' \in V$  Vektoren mit  $w = \lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n$  und  $w' = \lambda'_1 \cdot v_1 + \dots + \lambda'_n \cdot v_n$ . Dann gilt

$$\begin{aligned} w + w' &= (\lambda_1 + \lambda'_1) \cdot v_1 + \dots + (\lambda_n + \lambda'_n) \cdot v_n && \in \text{span}\{v_1, \dots, v_n\} \\ c \cdot w &= (c\lambda_1) \cdot v_1 + \dots + (c\lambda_n) \cdot v_n && \in \text{span}\{v_1, \dots, v_n\} \end{aligned}$$

Damit ist die Proposition bewiesen. ■

**Beispiele:** Die folgenden Vektoren bilden Erzeugendensysteme in den jeweiligen linearen Räumen, d.h. ihre lineare Hülle spannt immer gerade den gesamten Raum auf:

- Für  $V = \mathbb{R}^2$  und die Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{und} \quad v_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

gilt  $\text{span}\{v_1, v_2, v_3\} = \mathbb{R}^2$ , denn für  $x_1, x_2 \in \mathbb{R}$  gilt:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

- Für  $V = \mathbb{R}^2$  und die Vektoren

$$v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \text{und} \quad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

gilt  $\text{span}\{v_1, v_2\} = \mathbb{R}^2$ , denn für  $x_1, x_2 \in \mathbb{R}$  gilt:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \left(\frac{x_2}{2} - \frac{x_1}{2}\right) \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \left(\frac{x_1}{2} + \frac{x_2}{2}\right) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

**Definition 8.6** Es seien  $V$  ein linearer Raum und  $v_1, \dots, v_n \in V$  Vektoren. Dann heißen  $v_1, \dots, v_n$  linear unabhängig, falls für alle  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  gilt:

$$\text{Ist } \lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n = 0, \text{ so sind } \lambda_1 = \dots = \lambda_n = 0$$

Hierbei steht 0 für das neutrale Element von  $V$  (Nullvektor).

Mit anderen Worten: Sind die Vektoren  $v_1, \dots, v_n$  linear abhängig, so gibt es einen Vektor  $v_i$ , der in der linearen Hülle von  $\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n\}$  liegt. Dies kann man sehr leicht einsehen: Angenommen die Gleichung  $\lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n = 0$  ist auch mit  $\lambda_1 \neq 0$  möglich, dann folgt

$$v_1 = \left(-\frac{\lambda_2}{\lambda_1}\right) \cdot v_2 + \dots + \left(-\frac{\lambda_n}{\lambda_1}\right) \cdot v_n,$$

d.h.  $v_1$  ist eine Linearkombination von  $v_2, \dots, v_n$ . Das Argument läßt sich natürlich auf jeden Vektor  $v_i$  mit  $\lambda_i \neq 0$  verallgemeinern.

**Beispiele:** Folgende Vektoren verdeutlichen die Definition der linearen Unabhängigkeit im Raum  $V = \mathbb{R}^2$ :

- Die Vektoren  $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ ,  $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  und  $v_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  sind linear abhängig, denn es gilt:

$$1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Die Vektoren  $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  und  $v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  sind linear unabhängig, denn aus

$$\lambda_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

folgt  $\lambda_1 + \lambda_2 = 0$  und  $\lambda_2 = 0$ . Damit gilt  $\lambda_1 = \lambda_2 = 0$ .

**Definition 8.7** Es seien  $V$  ein linearer Raum und  $\{v_1, \dots, v_n\}$  ein Erzeugendensystem von  $V$ . Dann heißt die Menge  $\{v_1, \dots, v_n\}$  Basis von  $V$ , falls  $v_1, \dots, v_n$  linear unabhängig sind. Die Anzahl der Vektoren einer Basis heißt Dimension von  $V$  und wird mit  $\dim(V)$  bezeichnet.

Zur Wohldefiniertheit der Dimension merken wir ohne Beweis an, dass jede (endliche) Basis von  $V$  die gleiche Anzahl von Vektoren besitzt.

**Beispiel:** Als Vektorraum betrachten wir den  $\mathbb{R}^2$ .

- Die Menge der Vektoren  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ ,  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  und  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  ist ein linear abhängiges Erzeugendensystem und damit keine Basis

- Die Menge der Vektoren  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  und  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  ist ein linear unabhängiges Erzeugendensystem und damit eine Basis. Mithin ist  $\dim(\mathbb{R}^2) = 2$ .
- Die Menge der Vektoren  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  und  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  sowie die Menge der Vektoren  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  und  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  sind ebenfalls Basen von  $\mathbb{R}^2$ .

Im Folgenden interessieren wir uns für eine spezielle Klasse von Basen gegebener Vektorräume. Dazu führen wir den Begriff des Skalarproduktes ein, welches eine Abstraktion des Winkelkonzeptes zwischen Vektoren darstellt. Da der Schwerpunkt in diesem Kapitel nicht auf der geometrischen Interpretation linearalgebraischer Konzepte liegt, gehen wir auf die Winkelinterpretation nicht näher ein.

**Definition 8.8** *Es seien  $V$  ein linearer Raum. Eine Abbildung  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$  heißt (euklidisches) Skalarprodukt, falls für alle  $u, v, w \in V$  und  $a \in \mathbb{R}$  gilt:*

1.  $\langle v, w \rangle = \langle w, v \rangle$  (Symmetrie)
2.  $\langle a \cdot v + w, u \rangle = a \cdot \langle v, u \rangle + \langle w, u \rangle$  (Linearität)
3.  $\langle v, v \rangle \geq 0$  und  $\langle v, v \rangle = 0 \iff v = 0$  (positive Definitheit)

*Ein linearer Raum mit einem Skalarprodukt heißt euklidischer Raum. Vektoren in einem euklidischen Raum mit  $\langle v, w \rangle = 0$  heißen orthogonal.*

**Proposition 8.9** *Es sei  $V$  ein euklidischer Raum. Sind zwei Vektoren  $v, w \in V \setminus \{0\}$  orthogonal zueinander, so sind sie linear unabhängig.*

**Beweis:** Wir beweisen die Kontraposition der Aussage. Es seien also  $v, w \neq 0$  linear abhängige Vektoren, d.h. es gibt  $\lambda_1, \lambda_2 \neq 0$  mit  $\lambda_1 \cdot v + \lambda_2 \cdot w = 0$ . (Eigentlich dürfen wir nur voraussetzen, dass nur eines der  $\lambda$ 's verschieden von 0 ist, da wir aber nur zwei Vektoren betrachten, ist das andere  $\lambda$  somit auch stets verschieden von 0.) Damit gilt also  $v = (-\lambda_2/\lambda_1) \cdot w$  und wir erhalten

$$\begin{aligned} \langle v, w \rangle &= \langle (-\lambda_2/\lambda_1) \cdot w, w \rangle \\ &= -\frac{\lambda_2}{\lambda_1} \cdot \langle w, w \rangle && \text{(Linearität von } \langle \cdot, \cdot \rangle \text{)} \\ &\neq 0 && (\lambda_1, \lambda_2 \neq 0 \text{ und } \langle w, w \rangle > 0 \text{ wegen } w \neq 0) \end{aligned}$$

Damit sind  $v$  und  $w$  nicht orthogonal und die Proposition ist bewiesen. ■

**Beispiel:** Im  $\mathbb{R}^3$  ist für Vektoren  $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$  und  $w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$  das Standardskalarprodukt definiert als:

$$\langle v, w \rangle =_{\text{def}} v_1 w_1 + v_2 w_2 + v_3 w_3$$

Die Überprüfung der Axiome für Skalarprodukte ist eine Übungsaufgabe.

**Definition 8.10** *Es seien  $V$  ein euklidischer Raum und  $B = \{v_1, \dots, v_n\}$  eine Basis von  $V$ . Dann heißt  $B$  Orthogonalbasis von  $V$ , falls  $\langle v_j, v_k \rangle = 0$  für alle  $j, k$  mit  $j \neq k$  gilt.*

**Beispiele:** Wir betrachten wieder den  $\mathbb{R}^3$  mit dem Standardskalarprodukt.

- $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$  ist eine Basis aber keine Orthogonalbasis
- $\left\{ \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$  ist eine Orthogonalbasis
- $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right\}$  ist eine Orthogonalbasis

## 8.2 Lineare Abbildungen

In diesem Abschnitt interessieren wir uns für bestimmte Abbildungen  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Eine solche Abbildung  $f$  nennen wir *linear*, falls wir die Funktion in der Form  $f(x) = A \cdot x$  schreiben können, wobei  $A \in \mathbb{R}^{m \times n}$  eine Matrix und  $x \in \mathbb{R}^n$  ein Vektor (eine einspaltige Matrix) ist. Das Produkt  $A \cdot B$  zweier Matrizen  $A \in \mathbb{R}^{\ell \times m}$  und  $B \in \mathbb{R}^{m \times n}$  ist definiert als die Matrix  $C \in \mathbb{R}^{\ell \times n}$  mit den Einträgen

$$c_{ij} =_{\text{def}} \sum_{k=1}^m a_{ik} b_{kj}.$$

Eine typische Anwendung linearer Abbildungen ist die *Koordinatentransformation*. Wir betrachten dabei vereinfachend den  $\mathbb{R}^n$  und die Vektoren  $w, v_1, \dots, v_n \in \mathbb{R}^n$ . Ist  $w$  eine Linearkombination von  $v_1, \dots, v_n$ , d.h. gilt  $w = \lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n$ , so lässt sich dies wie folgt ausdrücken:

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{1n} & v_{2n} & \dots & v_{nn} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$



Die Werte  $\lambda_1, \dots, \lambda_n$  heißen *Koordinaten* von  $w$  bezüglich der Vektoren  $v_1, \dots, v_n$ .

Damit stellt sich als prinzipielle Frage, wie zu gegebenen Vektoren  $a_1, \dots, a_n \in \mathbb{R}^n$  für beliebige  $w \in \mathbb{R}^n$  die Koordinaten bestimmt werden können. Wir formulieren dieses Problem wie folgt in ein Problem für Matrizen um. Ausgehend von der Gleichheit  $w = A \cdot u$ , wobei  $A$  die wie oben aus den Vektoren  $a_1, \dots, a_n$  gebildete Matrix und  $u$  ein Vektor der Koordinaten sind, bestimmen wir, falls dies möglich ist, eine Matrix  $B \in \mathbb{R}^{n \times n}$  mit der Eigenschaft  $B \cdot A = I$ , wobei  $I \in \mathbb{R}^{n \times n}$  die *Einheitsmatrix* im  $\mathbb{R}^n$  ist:

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Damit gilt dann unter Ausnutzung der Assoziativität der Matrizenmultiplikation:

$$B \cdot w = B \cdot (A \cdot u) = (B \cdot A) \cdot u = I \cdot u = u$$

Mit dem Kenntnis der Matrix  $B$ , die in einem gewissen Sinne die zu  $A$  *inverse* Matrix darstellt, hätten wir das Problem der Koordinationberechnung gelöst.

**Beispiel:** Wir betrachten den linearen Raum  $\mathbb{R}^2$ . Zunächst wollen wir die inverse Matrix von  $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$  bestimmen, d.h. wir wollen eine Gleichung

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot w = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot u$$

in eine Gleichung

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \cdot w = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot u$$

für geeignete  $b_{11}, b_{12}, b_{21}$  und  $b_{22}$  umwandeln. Dazu schreiben wir die Matrizen nebeneinander und versuchen durch GAUSS-Elimination die Einheitsmatrix von der linken auf die rechten Seite zu bringen:

$$\begin{array}{l} \left( \begin{array}{cc|cc} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \end{array} \right) \\ \left( \begin{array}{cc|cc} 1 & 0 & 1 & -1 \\ -1 & 1 & 0 & 2 \end{array} \right) \quad \text{(Ziehe Zeile (1) von Zeile (2) ab)} \\ \left( \begin{array}{cc|cc} 1 & 0 & 1 & -1 \\ -1/2 & 1/2 & 0 & 1 \end{array} \right) \quad \text{(Multipliziere Zeile (2) mit 1/2)} \\ \left( \begin{array}{cc|cc} 1/2 & 1/2 & 1 & 0 \\ -1/2 & 1/2 & 0 & 1 \end{array} \right) \quad \text{(Addiere Zeile (2) zu Zeile (1))} \end{array}$$

Damit gilt  $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ . Zur Überprüfung rechnen wir nach:

$$\frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Als zweites Beispiel wollen wir einsehen, dass eine inverse Matrix nicht immer existieren muss. Dazu betrachten wir die Matrix  $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ . Zur Berechnung der inversen Matrix müssen geeignete reelle Zahlen  $a, b, c, d \in \mathbb{R}$  existieren mit

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Insbesondere muss also  $a + 2b = 1$  sowie  $2a + 4b = 0$  gelten, was nicht möglich ist. Mithin gibt es keine inverse Matrix.

**Definition 8.11** Es sei  $A \in \mathbb{R}^{m \times n}$  für  $m, n \in \mathbb{N}_+$  eine Matrix.

1.  $A$  heißt quadratisch, falls  $m = n$  gilt.
2.  $A^T \in \mathbb{R}^{n \times m}$  heißt die zu  $A$  transponierte Matrix, falls  $a_{jk} = (a^T)_{kj}$  für alle  $j \in \{1, \dots, m\}$  und  $k \in \{1, \dots, n\}$  gilt.
3.  $A$  heißt invertierbar, falls  $A$  quadratisch ist und eine Matrix  $A^{-1}$  existiert mit  $A \cdot A^{-1} = A^{-1} \cdot A = I$ .
4.  $A$  heißt symmetrisch, falls  $A = A^T$  gilt.
5.  $A$  heißt orthogonal, falls  $A^{-1} = A^T$  gilt.

**Beispiel:** Die Matrix  $\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$  ist orthogonal, denn:

$$\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

**Definition 8.12** Die Determinante einer Matrix  $A \in \mathbb{R}^{n \times n}$  ist definiert durch

$$\det(A) =_{\text{def}} \sum_{\pi \in S_n} \text{sgn}(\pi) \cdot \prod_{i=1}^n a_{i, \pi(i)}.$$

Hierbei steht  $\text{sgn}(\pi) =_{\text{def}} (-1)^{\|F(\pi)\|}$  für das Vorzeichen von  $\pi$ , wobei  $F(\pi)$  gerade die Menge der Fehlstände der Permutation  $\pi$  ist, d.h.  $F(\pi) =_{\text{def}} \{ (j, k) \mid j < k, \pi(j) > \pi(k) \}$ .

**Beispiele:** Die Determinante einer  $2 \times 2$ -Matrix ergibt sich wie folgt:

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \underbrace{(+1) \cdot a_{11}a_{22}}_{\pi = \begin{pmatrix} 12 \\ 12 \end{pmatrix}} + \underbrace{(-1) \cdot a_{12}a_{21}}_{\pi = \begin{pmatrix} 12 \\ 21 \end{pmatrix}}$$

Für eine  $3 \times 3$ -Matrix erhalten wir als Determinante:

$$\begin{aligned} \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} &= \underbrace{(+1) \cdot a_{11}a_{22}a_{33}}_{\pi = \begin{pmatrix} 123 \\ 123 \end{pmatrix}} + \underbrace{(+1) \cdot a_{12}a_{23}a_{31}}_{\pi = \begin{pmatrix} 123 \\ 231 \end{pmatrix}} + \underbrace{(+1) \cdot a_{13}a_{21}a_{32}}_{\pi = \begin{pmatrix} 123 \\ 312 \end{pmatrix}} + \\ &+ \underbrace{(-1) \cdot a_{11}a_{23}a_{32}}_{\pi = \begin{pmatrix} 123 \\ 132 \end{pmatrix}} + \underbrace{(-1) \cdot a_{13}a_{22}a_{31}}_{\pi = \begin{pmatrix} 123 \\ 321 \end{pmatrix}} + \underbrace{(-1) \cdot a_{12}a_{21}a_{33}}_{\pi = \begin{pmatrix} 123 \\ 213 \end{pmatrix}} \end{aligned}$$

Im Allgemeinen sind  $n!$  Produkte von Matrixeinträgen zu bestimmen. Das folgende Theorem gibt einen wichtigen Spezialfall von Matrizen an, für die die Determinante sehr einfach zu berechnen ist.

**Theorem 8.13** *Es sei  $A \in \mathbb{R}^{n \times n}$  eine Matrix in oberer Dreiecksform, d.h.*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & \dots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \dots & 0 & a_{nn} \end{pmatrix}$$

bzw.  $a_{jk} = 0$  für alle  $j, k$  mit  $j > k$ . Dann gilt:

$$\det(A) = \prod_{i=1}^n a_{ii}$$

**Beweis:** Es sei  $\pi \in S_n$  eine Permutation mit  $\pi(j) < j$  für ein  $j \in \{1, \dots, n\}$ . Dann gilt  $a_{j,\pi(j)} = 0$  wegen der oberen Dreiecksform der Matrix  $A$ . Somit gilt

$$\prod_{i=1}^n a_{i,\pi(i)} = 0.$$

Die einzige Permutation, die obige Eigenschaft nicht besitzt, ist die Identität  $\pi = \text{id}_n$ . Insgesamt folgt damit

$$\det(A) = \sum_{\pi \in S_n} \text{sgn}(\pi) \cdot \prod_{i=1}^n a_{i,\pi(i)} = \prod_{i=1}^n a_{ii}$$

und das Theorem ist bewiesen. ■

Aus dem Beweis wird deutlich, dass das Theorem 8.13 auch für Matrizen in unterer Dreiecksform gilt. Weiterhin gibt uns Theorem 8.13 die GAUSS-Elimination als Verfahren an die Hand, um die Determinante einer Matrix schneller als gemäß der Definition berechnen zu können. Folgende Regeln sind dabei zu beachten:

- Entsteht eine Matrix  $A'$  aus  $A$  durch Addieren des  $x$ -fachen von Zeile ( $k$ ) zu Zeile ( $j$ ) mit  $j \neq k$ , so gilt:

$$\det(A') = \det(A)$$

- Entsteht eine Matrix  $A'$  aus  $A$  durch Vertauschen von Zeile ( $k$ ) und Zeile ( $j$ ) mit  $j \neq k$ , so gilt:

$$\det(A') = -\det(A)$$

- Entsteht eine Matrix  $A'$  aus  $A$  durch Multiplikation von Zeile ( $j$ ) mit  $x \neq 0$ , so gilt:

$$\det(A') = x \cdot \det(A)$$

**Theorem 8.14** *Es seien  $A, B \in \mathbb{R}^{n \times n}$ . Dann gilt:*

1.  $A$  ist invertierbar  $\iff \det(A) \neq 0$
2.  $\det(A \cdot B) = \det(A) \cdot \det(B)$
3.  $\det(A^{-1}) = \det(A)^{-1}$ , falls  $A$  invertierbar ist

**Beweis:** (Nur dritte Aussage) Es sei  $A$  eine invertierbare Matrix. Nach Theorem 8.13 und der zweiten Aussage gilt

$$1 = \det(I) = \det(A^{-1} \cdot A) = \det(A^{-1}) \cdot \det(A).$$

Nach der ersten Aussage ist  $\det(A) \neq 0$  und wir können die beiden äußeren Ausdrücke durch  $\det(A)$  teilen. Mithin gilt  $\det(A^{-1}) = \det(A)^{-1}$  und die dritte Aussage des Theorems ist bewiesen. ■

### 8.3 Eigenwerte und Eigenvektoren

Betrachten wir eine lineare Abbildung  $f : V \rightarrow V : x \mapsto Ax$  im linearen Raum  $V$ , so hängt die Matrix  $A$  von der Wahl der Basis in  $V$  ab.

**Beispiel:** Zunächst betrachten wir im linearen Raum  $V = \mathbb{R}^2$  die Basis

$$\mathcal{A}_{\text{def}} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

sowie die Abbildung  $f : V \rightarrow V : x \mapsto M_{\mathcal{A}} \cdot x$  mit der Matrix  $M_{\mathcal{A}} \in \mathbb{R}^{2 \times 2}$

$$M_{\mathcal{A}} =_{\text{def}} \frac{1}{2} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

Wählen wir nunmehr im  $\mathbb{R}^2$  die Basis

$$\mathcal{B} =_{\text{def}} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$$

Dann ist die lineare Abbildung  $f$  von oben jetzt gegeben durch die Zuordnung  $y \mapsto M_{\mathcal{B}} \cdot y$  mit der Matrix

$$M_{\mathcal{B}} =_{\text{def}} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Wieso ist das der Fall?

Wir wissen bereits aus dem letzten Abschnitt: Ist  $x \in V$  ein Vektor bezüglich der Basis  $\mathcal{A}$ , so entspricht  $x$  dem Vektor

$$y =_{\text{def}} \underbrace{\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1}}_{B =_{\text{def}}} \cdot x = \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \cdot x$$

bezüglich der Basis  $\mathcal{B}$ , wobei die Spalten der Matrix  $B$  gerade aus den Vektoren der Basis  $\mathcal{B}$  bestehen. Somit muss die Gleichung

$$M_{\mathcal{A}} \cdot x = B \cdot (M_{\mathcal{B}} \cdot (B^{-1} \cdot x)) = (B \cdot M_{\mathcal{B}} \cdot B^{-1}) \cdot x$$

für alle  $x \in V$  gelten. Dies ist jedoch äquivalent zu  $M_{\mathcal{A}} = B \cdot M_{\mathcal{B}} \cdot B^{-1}$  bzw.  $B^{-1} \cdot M_{\mathcal{A}} \cdot B = M_{\mathcal{B}}$ . Mithin ergibt sich:

$$\begin{aligned} M_{\mathcal{B}} &= \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \cdot \frac{1}{2} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\ &= \frac{1}{4} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & -2 \\ 4 & 2 \end{pmatrix} \\ &= \frac{1}{4} \cdot \begin{pmatrix} 8 & 0 \\ 0 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Die im Beispiel angegebene Matrix hat eine besonders einfache Struktur, da lediglich auf der Diagonalen der Matrix Werte, die verschieden von 0 sind, auftreten. Im Folgenden beschäftigen wir uns mit einem Verfahren–der *Hauptachsentransformation*–, mit dem wir

zu einer gegebenen Matrix eine solche Diagonalmatrix sowie die zugehörige Basis bestimmen können. Zur Vereinfachung werden wir uns auf den Fall symmetrischer Matrizen beschränken. Von grundlegender Bedeutung für dieses Verfahren sind die Begriffe *Eigenwert* und *Eigenvektor*.

**Definition 8.15** *Es sei  $V$  ein linearer Raum mit  $\dim(V) = n$  bezüglich einer beliebigen Basis. Weiterhin sei  $f : V \rightarrow V : x \mapsto A \cdot x$  eine lineare Abbildung. Ein Vektor  $v \in V \setminus \{0\}$  heißt Eigenvektor zum Eigenwert  $\lambda \in \mathbb{R}$ , falls gilt:*

$$A \cdot v = \lambda \cdot v$$

**Beispiel:** Wir betrachten wiederum den linearen Raum  $V = \mathbb{R}^2$  mit der Basis

$$\mathcal{A} =_{\text{def}} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

sowie die Abbildung  $f : V \rightarrow V : x \mapsto A \cdot x$  mit der Matrix  $A \in \mathbb{R}^{2 \times 2}$

$$A =_{\text{def}} \frac{1}{2} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

Dann ist

- $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  ist Eigenvektor zum Eigenwert 2. Zur Überprüfung:

$$\frac{1}{2} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

- $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$  ist Eigenvektor zum Eigenwert 1. Zur Überprüfung:

$$\frac{1}{2} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

**Definition 8.16** *Es sei  $V$  ein linearer Raum mit  $\dim(V) = n$  bezüglich einer beliebigen Basis. Weiterhin seien  $f : V \rightarrow V : x \mapsto A \cdot x$  eine lineare Abbildung sowie  $\lambda \in \mathbb{R}$  ein Eigenwert von  $A$ . Dann heißt*

$$E_\lambda(A) =_{\text{def}} \{ v \in V \mid A \cdot v = \lambda \cdot v \}$$

der Eigenraum von  $\lambda$ .

**Proposition 8.17** *Der Eigenraum  $E_\lambda(A)$  ist ein Unterraum von  $V$ .*

**Beweis:** Es seien  $v_1, v_2 \in E_\lambda(A)$  zwei Vektoren sowie  $a \in \mathbb{R}$  ein Skalar. Dann gilt

$$A \cdot (v_1 + v_2) = A \cdot v_1 + A \cdot v_2 = \lambda \cdot v_1 + \lambda \cdot v_2 = \lambda \cdot (v_1 + v_2)$$

Mithin gilt  $v_1 + v_2 \in E_\lambda(A)$ . Somit ist  $E_\lambda(A)$  abgeschlossen unter Addition. Weiterhin erhalten wir

$$A \cdot (a \cdot v_1) = a \cdot (A \cdot v_1) = a \cdot \lambda \cdot v_1 = \lambda \cdot (a \cdot v_1)$$

Mithin gilt  $a \cdot v_1 \in E_\lambda(A)$ . Somit ist  $E_\lambda(A)$  auch abgeschlossen unter Multiplikation mit Skalaren. Folglich ist  $E_\lambda(A)$  ein linearer Unterraum von  $V$ . ■

*Bemerkungen:* Wir führen einige ergänzende Anmerkungen an:

- Jeder Vektor  $v \in E_\lambda(A) \setminus \{0\}$  ist Eigenvektor zu  $\lambda$ .
- Der Nullvektor  $0$  erfüllt stets  $A \cdot 0 = \lambda \cdot 0$  und wird deshalb als Eigenvektor ausgeschlossen.
- $E_{\lambda_1}(A) \cap E_{\lambda_2}(A) = \{0\}$  für Eigenwerte  $\lambda_1, \lambda_2$  mit  $\lambda_1 \neq \lambda_2$ .

**Definition 8.18** *Es sei  $V$  ein linearer Raum mit  $\dim(V) = n$  bezüglich einer beliebigen Basis. Weiterhin seien  $f : V \rightarrow V : x \mapsto A \cdot x$  eine lineare Abbildung sowie  $\lambda \in \mathbb{R}$  ein Eigenwert von  $A$ . Dann heißt  $\dim(E_\lambda(A))$  die geometrische Vielfachheit von  $A$ .*

Die Frage ist nunmehr: *Wie bestimmen wir die Eigenwerte einer Matrix?* Dazu betrachten wir folgende Herleitung (in Form einer Folge von Äquivalenzen). Gegeben sei ein Matrix  $A \in \mathbb{R}^{n \times n}$ . Dann gilt:

$$\begin{aligned} \lambda \text{ ist Eigenwert von } A &\iff \text{es gibt ein } v \in V \setminus \{0\} \text{ mit } A \cdot v = \lambda \cdot v \\ &\iff \text{es gibt ein } v \in V \setminus \{0\} \text{ mit } A \cdot v - \lambda \cdot v = 0 \\ &\iff \text{es gibt ein } v \in V \setminus \{0\} \text{ mit } (A - \lambda I) \cdot v = 0 \\ &\iff \det(A - \lambda I) = 0 \end{aligned}$$

Die letzte Äquivalenz folgt aus der Charakterisierung der Existenz einer inversen Matrix zu  $A - \lambda I$ . Könnten wir nämlich die Matrix  $A - \lambda I$  invertieren, so wäre die einzige Lösung der Gleichung  $(A - \lambda I) \cdot v = 0$  der Nullvektor  $v = (A - \lambda I)^{-1} \cdot 0 = 0$ . Diesen hatten wir aber gerade ausgeschlossen.

**Definition 8.19** *Es sei  $A \in \mathbb{R}^{n \times n}$  eine Matrix. Dann ist*

$$p_A(x) =_{\text{def}} \det(A - xI) = \det \begin{pmatrix} a_{11} - x & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - x & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - x \end{pmatrix}$$

*ein Polynom vom Grad  $n$ , dessen Nullstellen  $\lambda \in \mathbb{R}$  die Eigenwerte von  $A$  sind. Das Polynom  $p_A$  heißt charakteristisches Polynom von  $A$ . Die Vielfachheit einer Nullstelle heißt algebraische Vielfachheit.*

**Beispiel:** Wir greifen wieder auf den linearen Raum  $V = \mathbb{R}^2$  sowie die Matrix

$$A =_{\text{def}} \frac{1}{2} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

zurück. Dann gilt:

$$\begin{aligned} \det(A - xI) &= \det \begin{pmatrix} 3/2 - x & 1/2 \\ 1/2 & 3/2 - x \end{pmatrix} \\ &= \left(\frac{3}{2} - x\right) \cdot \left(\frac{3}{2} - x\right) - \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{9}{4} - 3x + x^2 - \frac{1}{4} \\ &= x^2 - 3x + 2 \\ &= (x - 2)(x - 1) \end{aligned}$$

Die Matrix hat also das charakteristische Polynom  $p_A(x) = x^2 - 3x + 2$  (in expandierter Form) bzw.  $p_A(x) = (x - 2)(x - 1)$  (in Nullstellenform). Damit sind die Eigenwerte der Matrix die reellen Zahlen 1 und 2 jeweils mit algebraischer Vielfachheit 1.

Im Folgenden schränken wir uns auf den einfachen Fall symmetrischer Matrizen ein. Dafür gibt es vor allem zwei vereinfachende Gründe:

1. Für die Eigenwerte einer beliebigen Matrix gilt, dass die geometrische Vielfachheiten höchstens so groß wie die algebraische Vielfachheiten (aber auch kleiner) sein können. Bei symmetrischen Matrizen gilt stets die Gleichheit.
2. Ein Polynom vom Grad  $n$  hat maximal  $n$  reelle Nullstellen (Vielfachheiten mitgezählt) und in der Menge  $\mathbb{C}$  der komplexen Zahlen genau  $n$  Nullstellen. Im Allgemeinen kann ein charakteristisches Polynom somit auch komplexe Nullstellen und die Matrix damit komplexe Eigenwerte besitzen. Bei symmetrischen Matrizen treten dagegen keine komplexen Eigenwerte auf.

**Theorem 8.20 (Hauptachsentransformation für symmetrische Matrizen)** *Es seien  $V$  ein euklidischer Raum mit  $\dim(V) = n$  und  $A \in \mathbb{R}^{n \times n}$ . Dann gilt:*

1. *Alle Eigenwerte  $\lambda_1, \lambda_2, \dots, \lambda_n$  sind reelle Zahlen.*
2. *Es gibt eine Orthonormalbasis  $\mathcal{B} =_{\text{def}} \{b_1, b_2, \dots, b_n\}$  von  $v$ , die aus Eigenvektoren  $b_1, b_2, \dots, b_n$  (zu den Eigenwerten  $\lambda_1, \lambda_2, \dots, \lambda_n$ ) besteht.*
3.  *$A = Q \cdot D \cdot Q^T$  mit  $Q$  als Matrix mit den Spaltenvektoren  $b_1, b_2, \dots, b_n$  sowie*

$$D =_{\text{def}} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$



Wir geben das Theorem ohne Beweis an und führen stattdessen zur Verdeutlichung die Hauptachsentransformation für ein Beispiel an.

**Beispiel:** Gegeben sei die Matrix

$$A =_{\text{def}} \begin{pmatrix} 3/2 & 1/2 & 0 \\ 1/2 & 3/2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Wir führen die Hauptachsentransformation in Schritten durch.

1. *Bestimmung des charakteristischen Polynoms sowie der Eigenwerte  $\lambda_1, \lambda_2$  und  $\lambda_3$ :*

$$\begin{aligned} p_A(x) &= \det \begin{pmatrix} 3/2 - x & 1/2 & 0 \\ 1/2 & 3/2 - x & 0 \\ 0 & 0 & 2 - x \end{pmatrix} \\ &= \left(\frac{3}{2} - x\right) \cdot \left(\frac{3}{2} - x\right) \cdot (2 - x) - \frac{1}{2} \cdot \frac{1}{2} \cdot (2 - x) \\ &= (2 - x) \cdot \left[ \left(\frac{3}{2} - x\right)^2 - \frac{1}{4} \right] \\ &= -(x - 2)(x - 2)(x - 1) \end{aligned}$$

Damit sind die Eigenwerte  $\lambda_1 = 2, \lambda_2 = 2$  und  $\lambda_3 = 1$ . Der Eigenwert 2 tritt mit Vielfachheit 2 und der Eigenwert 1 mit Vielfachheit 1 auf.

2. *Bestimmung der Eigenräume  $E_1(A)$  und  $E_2(A)$ :  $E_2(A)$  besteht aus allen Vektoren  $v$  mit*

$$(A - 2 \cdot I) \cdot v = \begin{pmatrix} -1/2 & 1/2 & 0 \\ 1/2 & -1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot v = 0,$$

$$\text{d.h. } E_2(A) = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

$E_1(A)$  besteht aus allen Vektoren  $v$  mit

$$(A - 1 \cdot I) \cdot v = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot v = 0,$$

$$\text{d.h. } E_1(A) = \text{span} \left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

3. *Bestimmung der Orthonormalbasis:* Für die Eigenvektoren

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

gilt bereits  $\langle v_1, v_2 \rangle = 0$ ,  $\langle v_1, v_3 \rangle = 0$  und  $\langle v_2, v_3 \rangle = 0$ . Somit ist  $\{v_1, v_2, v_3\}$  eine Orthogonalbasis. Um eine Orthonormalbasis  $\{b_1, b_2, b_3\}$  zu bekommen, müssen die Vektoren so normiert werden, dass  $\langle b_i, b_i \rangle = 1$  gilt. Dazu wählen wir den Ansatz  $b_i = a_i \cdot v_i$ , wobei  $a_i \in \mathbb{R}$  gilt. Aus den Eigenschaften des Skalarproduktes erhalten wir:

$$1 = \langle a_i \cdot v_i, a_i \cdot v_i \rangle = a_i \cdot \langle v_i, a_i \cdot v_i \rangle = a_i \cdot \langle a_i \cdot v_i, v_i \rangle = a_i^2 \cdot \langle v_i, v_i \rangle$$

Damit ergibt sich  $a_i = \langle v_i, v_i \rangle^{-\frac{1}{2}}$  bzw. in unserem konkreten Falle

$$a_1 = \frac{1}{\sqrt{2}}, \quad a_2 = 1, \quad a_3 = \frac{1}{\sqrt{2}}$$

Die Orthonormalbasis  $\{b_1, b_2, b_3\}$  besteht daher aus den Eigenvektoren

$$b_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad b_3 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

4. *Bestimmung der Matrizen D und Q:*

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$$

Damit ist die Hauptachsentransformation abgeschlossen. Wir können nun zur Überprüfung der Korrektheit übergehen. Zunächst halten wir fest, dass  $Q$  eine Orthogonalmatrix ist, denn es gilt  $Q^{-1} = Q^T$ :

$$Q \cdot Q^T = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Ebenfalls leicht nachzurechnen ist die Gleichung  $Q^T \cdot Q = I$ . Wir überprüfen noch die Identität  $A = Q \cdot D \cdot Q^T$  rechnerisch:

$$\begin{aligned} & \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 2 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix} = \begin{pmatrix} 3/2 & 1/2 & 0 \\ 1/2 & 3/2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \end{aligned}$$

---

# Literaturverzeichnis

---

- [MM06] Christoph Meinel und Martin Mundhenk. *Mathematische Grundlagen der Informatik. Mathematisches Denken und Beweisen. Eine Einführung.* 3., überarbeitete und erweiterte Auflage. B. G. Teubner Verlag, Wiesbaden, 2006.
- [Ste07] Angelika Steger. *Diskrete Strukturen. Band 1: Kombinatorik-Graphentheorie-Algebra.* 2. Auflage. Springer-Verlag, Berlin, 2007.
- [SS02] Thomas Schickinger und Angelika Steger. *Diskrete Strukturen. Band 2: Wahrscheinlichkeitstheorie und Statistik.* Springer-Verlag, Berlin, 2002.
- [WHK04] Manfred Wolff, Peter Hauck und Wolfgang Küchlin. *Mathematik für Informatik und Bioinformatik.* Springer-Verlag, Berlin, 2004.

