

**Mathematik III für Ingenieure  
im Bachelor-Studiengang Maschinenbau**

Vorlesung Wintersemester 2010/2011

B. Schuster

*aktualisiert am 27. Januar 2011*



# INHALT

I. Eigenwerte und Eigenvektoren	1
1. Komplexe Matrizen	1
2. Determinanten	2
3. Eigenwerte und Diagonalisierbarkeit	7
4. Hermitesche und unitäre Matrizen	16
II. Kurven	24
1. Kurven im Raum	24
2. Bogenlänge	27
3. Krümmung	31
4. Evolute und Evolvente	35
5. Vektorfelder	37
6. Kurvenintegrale	39
III. Grundlagen der Statistik	45
1. Zufallsexperimente	46
2. Wahrscheinlichkeitsmaße	48
3. Ein wenig Kombinatorik	49
4. Bedingte Wahrscheinlichkeit	54
5. Zufallsvariable und Verteilungsfunktion	59
6. Diskrete Zufallsvariablen	60
7. Erwartungswerte und Varianz diskreter Zufallsvariablen	67
8. Stetige Zufallsvariablen	72
9. Zusammengesetzte Zufallsvariablen	80

# I Eigenwerte und Eigenvektoren

Die Wiederholung des Kapitels zur Linearen Algebra des ersten Semesters wird dringend empfohlen.

## 1. Komplexe Matrizen

Auch wenn bisher nur Matrizen mit reellen Einträgen betrachtet wurden, bleiben die Konzepte der Linearen Algebra sinnvoll, wenn man komplexe Zahlen zulässt. Solche Matrizen treten in ingenieurwissenschaftlichen Anwendungen häufig auf (etwa bei Schaltkreisen). Wir betrachten in der Folge also Matrizen

$$A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad \text{mit } a_{ij} \in \mathbb{C}, i = 1, \dots, m, j = 1, \dots, n.$$

Mit komplexen Matrizen rechnet man so wie mit reellen: Addition, Subtraktion, skalare Multiplikation und Multiplikation sind genauso erklärt. Die Menge der komplexen  $m \times n$ -Matrizen bezeichnen wir mit  $\text{Mat}(m \times n, \mathbb{C})$  (oder auch  $\mathbb{C}^{m \times n}$ ).

**Definition 1.1.** Sei  $A$  eine (komplexe)  $m \times n$ -Matrix.

- (a)  $\bar{A} = (\bar{a}_{ij})$  heißt die zu  $A$  konjugierte Matrix.
- (b)  $A^* = (\bar{A})^\top = \bar{a}_{ji}$  heißt die zu  $A$  konjugiert transponierte Matrix.

In der konjugierten Matrix wird also jeder Eintrag durch sein komplex Konjugiertes ersetzt. Bei der konjugiert Transponierten wird die Matrix zusätzlich noch „gestürzt“, d.h. an der Hauptdiagonalen gespiegelt (so dass die  $i$ -te Zeile zur  $i$ -ten Spalte wird und umgekehrt).

**Achtung:** In der Literatur wird gelegentlich eine andere Bezeichnung gebraucht: der Stern für die komplex Konjugierte und der Querstrich für die komplex Transponierte, gerade umgekehrt, wie es hier eingeführt wurde!

**Beispiel 1.2.**

$$A = \begin{pmatrix} 1+j & 3-j \\ 3-2j & 12 \\ 4j & 1+j \end{pmatrix}$$

$$\bar{A} = \begin{pmatrix} 1-j & 3+j \\ 3+2j & 12 \\ -4j & 1-j \end{pmatrix}$$

$$A^* = \begin{pmatrix} 1-j & 3+2j & -4j \\ 4+j & 12 & 1-j \end{pmatrix}$$

Es gelten die Rechengesetze

$$\begin{aligned} \overline{\bar{A}} &= A \\ \overline{A+B} &= \bar{A} + \bar{B} \\ \overline{A \cdot B} &= \bar{A} \cdot \bar{B} \\ A^* &= \overline{A^T} \\ (A^*)^* &= A \\ (A+B)^* &= A^* + B^* \\ (A \cdot B)^* &= B^* \cdot A^* \end{aligned}$$

**2. Determinanten**

Determinanten von  $2 \times 2$ - und  $3 \times 3$ -Matrizen sollten aus Mathematik 1 und 2 bekannt sein; hier geht es darum, sie auch für höhere Dimensionen zu erklären.

**2.1. Determinanten für beliebige Dimensionen.** Zunächst sei an  $2 \times 2$ - und  $3 \times 3$ -Determinanten erinnert.

Im Fall  $n = 2$  nimmt die Differenz der Produkte der Diagonaleinträge:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

Für  $n = 3$  hat man die „Regel von Sarrus“,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Diese Rechenvorschrift merkt man sich am leichtesten durch das wohlbekanntes graphische Schema:

$$(2.1)$$

Hierbei werden die Produkte der Elemente auf den durchgezogenen Diagonalen mit positiven, die auf den durchbrochenen Diagonalen mit negativen Vorzeichen versehen und dann addiert. Man kann diesen Ausdruck aber auch noch anders schreiben, nämlich als eine Linearkombination von Unterdeterminanten:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Das Rezept lautet dabei: man versehe die Elemente der ersten Zeile mit wechselnden Vorzeichen und multipliziere den  $j$ -ten Eintrag mit der Determinante der Matrix, die aus der ursprünglichen Matrix durch Streichen der ersten Zeile und der  $j$ -ten Spalte hervorgeht. Mit dieser Regel kann man Determinanten auf quadratische Matrizen beliebiger Zeilenzahl verallgemeinern.

Wir illustrieren dies erst noch einmal am Beispiel  $n = 4$ , bevor wir die allgemeine Definition angeben.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} \\ + a_{13} \begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14} \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{32} \\ a_{41} & a_{42} & a_{43} \end{vmatrix}$$

Bezeichnet man mit  $A_{1j}$  die Matrix, die aus  $A$  durch Streichen der ersten Zeile und der  $j$ -ten Spalte hervorgeht, so schreibt sich die obige Gleichung für eine  $4 \times 4$ -Matrix  $A$  kürzer als

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13}) - a_{14} \det(A_{14}) = \sum_{j=1}^4 (-1)^{j+1} \det(A_{1j})$$

(Den scheinbar überflüssigen ersten unteren Index 1 werden wir bald rechtfertigen.) Damit sind  $4 \times 4$ -Determinanten definiert, und wir können das Rezept induktiv fortschreiben: Sei

$$A = (a_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

eine  $n \times n$ -Matrix. Dann ist

$$\det(A) = \sum_{j=1}^n (-1)^{j+1} \cdot a_{1j} \cdot \det(A_{1j}) \quad (2.2)$$

wobei wieder die Matrix  $A_{1j}$  aus  $A$  durch Streichen der ersten Zeile und der  $j$ -ten Spalte hervorgeht.

**Beispiel 2.1.** Sei  $A = \begin{pmatrix} 1 & 2 & 4 & -1 \\ 2 & 3 & 0 & 1 \\ -1 & 0 & 2 & 1 \\ 5 & 1 & 1 & 2 \end{pmatrix}$ .

$$\det(A) = 1 \cdot \underbrace{\begin{vmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix}}_{=7} - 2 \cdot \underbrace{\begin{vmatrix} 2 & 0 & 1 \\ -1 & 2 & 1 \\ 5 & 1 & 2 \end{vmatrix}}_{=-5} + 4 \cdot \underbrace{\begin{vmatrix} 2 & 3 & 1 \\ -1 & 0 & 1 \\ 5 & 1 & 2 \end{vmatrix}}_{=18} - (-1) \cdot \underbrace{\begin{vmatrix} 2 & 3 & 0 \\ -1 & 0 & 2 \\ 5 & 1 & 1 \end{vmatrix}}_{=29} \\ = 7 + 10 + 72 + 29 = 118$$

**Warnung:** Es gibt keine „Regel von Sarrus“ für höhere Dimensionen!

**2.2. Eigenschaften der Determinante.** Sei im Folgenden  $A$  eine  $n \times n$ -Matrix.

**2.2.1.** Die Einheitsmatrix  $E_n$  hat die Determinante 1.

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

**2.2.2.** Ist  $A^\top$  die Transponierte von  $A$ , so gilt

$$\det(A^\top) = \det(A).$$

Für  $3 \times 3$ -Matrizen ist dies rechnerisch leicht nachzuvollziehen, denn es treten die gleichen Produkte mit den gleichen Vorzeichen auf.

**2.2.3.** Entsteht  $B$  aus  $A$  durch Vertauschen zweier Zeilen oder Spalten, so gilt

$$\det(B) = -\det(A).$$

So ist etwa

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = -\det \begin{pmatrix} d & e & f \\ a & b & c \\ g & h & i \end{pmatrix}$$

wie man schnell nachrechnet.

**2.2.4.** Hat  $A$  eine Zeile oder eine Spalte, die nur aus Nullen besteht, so ist  $\det(A) = 0$ .

Ist die erste Zeile eine Nullzeile, folgt dies direkt aus der definierenden Formel (2.2):

$$\sum_{j=1}^n (-1)^{j+1} \cdot a_{1j} \cdot \det(A_{1j}) = \sum_{j=1}^n (-1)^{j+1} \cdot 0 \cdot \det(A_{1j}) = 0$$

Ansonsten tausche man die Nullzeile nach oben und benutze 2.2.3; für Spalten folgt die Aussage dann aus 2.2.2.

**2.2.5.** Sei  $\lambda \neq 0$  eine Konstante. Entsteht die Matrix  $B$  aus  $A$  durch Multiplikation einer Zeile (oder einer Spalte) mit  $\lambda$ , so ist  $\det(B) = \lambda \det(A)$ .

Multipliziert man etwa die erste Zeile mit  $\lambda \neq 0$ , so wird jeder Summand in Formel (2.2) mit dem Faktor  $\lambda$  multipliziert. Für die anderen Zeilen sieht man dies wieder durch Vertauschen und 2.2.3. Die entsprechende Aussage für Spalten folgt aus 2.2.2

**2.2.6.** Sind zwei Spalten oder zwei Zeilen von  $A$  gleich, so ist  $\det(A) = 0$ .

Für  $3 \times 3$ -Matrizen ist das wieder leicht nachzurechnen

**2.2.7.** Sei  $B$  eine Matrix, die aus  $A$  durch Addition des Vielfachen einer Zeile zu einer anderen Zeile hervorgeht. Dann ist  $\det(B) = \det(A)$ . Das gleiche gilt für Addition von Spalten.

**Beispiel.** Wir überlegen uns, dass sich die Determinante nicht verändert, wenn man die zweite Zeile zur ersten addiert. Sei also  $B$  die Matrix, die aus  $A$  durch Ersetzen der ersten Zeile durch die Summe der ersten beiden Zeilen entsteht. Dadurch verändern sich die Streichmatrizen  $A_{1j}$  nicht, d.h. es gilt  $B_{1j} = A_{1j}$  (für  $j = 1, \dots, n$ ), und laut unserer Formel (2.2) ist dann

$$\begin{aligned} \det(B) &= \sum_{j=1}^n (-1)^{j+1} \cdot (a_{1j} + a_{2j}) \cdot \det(A_{1j}) \\ &= \sum_{j=1}^n (-1)^{j+1} \cdot a_{1j} \cdot \det(A_{1j}) + \sum_{j=1}^n (-1)^{j+1} \cdot a_{2j} \cdot \det(A_{1j}) = \det(A) + \det(A'), \end{aligned}$$

wobei die erste und die zweite Zeile der Matrix  $A'$  gleich sind (denn beide sind gleich der zweiten Zeile von  $A$ ). Mit 2.2.6 folgt  $\det(A') = 0$  und daher  $\det(B) = \det(A)$ .

Den allgemeinen Fall führt man mit Vertauschungen und Multiplikation einzelner Zeilen mit einer Konstanten auf dieses Beispiel zurück.

**2.2.8.** Ist  $A$  eine obere oder untere Dreiecksmatrix, so ist  $\det(A)$  das Produkt der Diagonalelemente.

Diese letzten beiden Eigenschaften erleichtern die Berechnung von Determinanten ganz erheblich, denn man kann zunächst die Matrix auf Dreiecksform bringen und dann das Produkt der Diagonalelemente nehmen. Der Rechenaufwand für dieses Verfahren liegt in der Größenordnung von  $n^3$  Multiplikationen, wenn  $n$  die Zeilenzahl ist. Benutzte man statt dessen Formel (2.2), wäre der Aufwand  $n!$ .

**Beispiel.** Wir illustrieren das Verfahren anhand der Matrix

$$A = \begin{pmatrix} 1 & 2 & -1 & 3 & 3 \\ 1 & 0 & -1 & 1 & 2 \\ 2 & 4 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 1 \\ 2 & 2 & -2 & 5 & 6 \end{pmatrix}$$

In einem ersten Schritt ziehen wir die erste Zeile von der zweiten und dritten Zeile und jeweils zweimal von der dritten und fünften Zeile ab mit dem Resultat

$$\begin{pmatrix} 1 & 2 & -1 & 3 & 3 \\ 0 & -2 & 0 & -2 & -1 \\ 0 & 0 & 4 & -2 & -3 \\ 0 & -2 & 2 & -2 & -2 \\ 0 & -2 & 0 & -1 & 0 \end{pmatrix}$$

Dann ziehen wir die zweite Zeile von der vierten und fünften ab und erhalten

$$\begin{pmatrix} 1 & 2 & -1 & 3 & 3 \\ 0 & -2 & 0 & -2 & -1 \\ 0 & 0 & 4 & -2 & -3 \\ 0 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$



Nun ziehen wir  $\frac{1}{2}$  mal die dritte Zeile von der vierten ab und anschließend die vierte von der fünften:

$$\begin{pmatrix} 1 & 2 & -1 & 3 & 3 \\ 0 & -2 & 0 & -2 & -1 \\ 0 & 0 & 4 & -2 & -3 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & -1 & 3 & 3 \\ 0 & -2 & 0 & -2 & -1 \\ 0 & 0 & 4 & -2 & -3 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Das Produkt der Diagonalelemente dieser letzten Matrix ist  $-4$ , also gilt  $\det(A) = -4$ .

In diesem Beispiel kamen wir beim Gaussverfahren ohne Zeilenvertauschungen aus, aber das muss nicht immer so sein. In solchen Fällen muss man sich für jede Vertauschung zweier Zeilen ein Vorzeichen merken, weshalb man die Tauscherei am besten bis zum Schluss aufschiebt.

**Beispiel.** In der folgenden Rechnung soll  $Z3 - 2 \cdot Z1$  bedeuten, dass die dritte Zeile ersetzt wird durch „Zeile 3 minus 2mal Zeile 1“.

$$B = \begin{pmatrix} 1 & 3 & 2 & -3 & 8 \\ -1 & -3 & 1 & 5 & -15 \\ 2 & 4 & 5 & 6 & 11 \\ 6 & 12 & 12 & 17 & 41 \\ -1 & -1 & 4 & -1 & 2 \end{pmatrix} \xrightarrow{\substack{Z2+Z1 \\ Z3-2 \cdot Z1 \\ Z4+6 \cdot Z1 \\ Z5+Z1}} \begin{pmatrix} 1 & 3 & 2 & -3 & 8 \\ 0 & 0 & 3 & 2 & -7 \\ 0 & -2 & 1 & 12 & -5 \\ 0 & -6 & 0 & 35 & -7 \\ 0 & 2 & 5 & -4 & 10 \end{pmatrix} \xrightarrow{\substack{Z4-3 \cdot Z3 \\ Z5+Z3}}$$

$$\begin{pmatrix} 1 & 3 & 2 & -3 & 8 \\ 0 & 0 & 3 & 2 & -7 \\ 0 & -2 & 1 & 12 & -5 \\ 0 & 0 & -3 & -1 & 8 \\ 0 & 0 & 6 & 8 & 5 \end{pmatrix} \xrightarrow{\substack{Z4+Z2 \\ Z5-2 \cdot Z2}} \begin{pmatrix} 1 & 3 & 2 & -3 & 8 \\ 0 & 0 & 3 & 2 & -7 \\ 0 & -2 & 1 & 12 & -5 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 4 & 19 \end{pmatrix} \xrightarrow{Z5-4 \cdot Z4}$$

$$\begin{pmatrix} 1 & 3 & 2 & -3 & 8 \\ 0 & 0 & 3 & 2 & -7 \\ 0 & -2 & 1 & 12 & -5 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 15 \end{pmatrix} \xrightarrow{\substack{\text{vertausche} \\ Z2 \text{ mit } Z3}} \begin{pmatrix} 1 & 3 & 2 & -3 & 8 \\ 0 & -2 & 1 & 12 & -5 \\ 0 & 0 & 3 & 2 & -7 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 15 \end{pmatrix} = B'$$

Wegen der einen Vertauschung haben wir also

$$\det(B) = -\det(B') = -1 \cdot (-2) \cdot 3 \cdot 1 \cdot 15 = 90.$$

Als Folgerung aus all diesen Eigenschaften kann man eine neue Formel zur Determinantenberechnung angeben, die die Sonderrolle der ersten Zeile in Formel (2.2) überflüssig macht.

**2.2.9** (Entwicklungssatz). Für eine  $n \times n$ -Matrix  $A$  und ein Paar  $(i, j)$  von Indizes mit  $1 \leq i, j \leq n$  bezeichne  $A_{ij}$  die Matrix, die aus  $A$  durch Streichen der  $i$ -ten Zeile und der  $j$ -ten Spalte hervorgeht. Dann gilt für jedes  $i$  zwischen 1 und  $n$

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} \cdot a_{ij} \cdot \det(A_{ij}) \quad (2.3)$$

und für jedes  $j$  zwischen 1 und  $n$

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} \cdot a_{ij} \cdot \det(A_{ij}) \quad (2.4)$$

Die Formel (2.3) nennt man *Entwicklung nach der i-ten Zeile* und (2.4) *Entwicklung nach der j-ten Spalte*. Die Vorzeichen kann man sich nach einem einfachen Schachbrettmuster merken:

+	-	+	-	+	-	+	-
-	+	-	+	-	+	-	+
+	-	+	-	+	-	+	-
-	+	-	+	-	+	-	+
+	-	+	-	+	-	+	-
-	+	-	+	-	+	-	+
+	-	+	-	+	-	+	-
-	+	-	+	-	+	-	+

Eine sehr wichtige Regel ist der Multiplikationssatz:

**Satz 2.2.** *Seien  $A$  und  $B$  quadratische  $n$ -reihige Matrizen. Dann gilt*

$$\det(A \cdot B) = \det(A) \cdot \det(B).$$

Genauso wie früher gilt:

**Satz 2.3.** *Eine quadratische Matrix  $A$  ist genau dann invertierbar, wenn  $\det(A) \neq 0$  ist.*

(Denn ist  $A$  invertierbar mit Inverser  $A^{-1}$  so ist

$$1 = \det(E_n) = \det(A \cdot A^{-1}) = \det(A) \cdot \det(A^{-1}),$$

weshalb  $\det(A)$  nicht Null sein kann. Ist umgekehrt  $\det(A) \neq 0$ , so hat die Zeilenstufenform von  $A$  lauter von Null verschiedene Einträge auf der Diagonalen, also sind die Spalten von  $A$  linear unabhängig.)

### 3. Eigenwerte und Diagonalisierbarkeit

**3.1. Einführung.** Eigenwerte und Eigenvektoren sind ein ausgesprochen wichtiges Instrument zur Untersuchung und Beschreibung vieler aus Natur und Technik stammender Probleme. Einige davon sind:

- In jedem Körper lässt sich ein System von drei Achsen finden (senkrecht aufeinander stehende noch dazu), um die der Körper ohne zu taumeln rotieren kann, also ohne Unwucht. Das Auffinden dieser Achsen ist ein Eigenwertproblem.
- Analog hat ein Körper, an dem äußere Kräfte zerren, drei Hauptspannungsrichtungen.
- In der Akustik sind Obertöne (oder auch Eigenschwingungen) Eigen„werte“ (richtiger: Eigenfunktionen) einer schwingenden Saite.
- Die Hauptkomponentenanalyse der Statistik ist ebenfalls ein Eigenwertproblem.
- Gekoppelte lineare Differentialgleichungen lassen sich mit Eigenwerttheorie lösen.

Ganz allgemein gilt, dass sich mittels Eigenwerttheorie lineare Phänomene vereinfacht darstellen und bearbeiten lassen.

Wir beginnen mit einem Beispiel zu Differentialgleichungen:

**Beispiel 3.1.** Gegeben sei folgendes System gekoppelter Differentialgleichungen:

$$\begin{aligned} \dot{u}_1 &= u_1 + 2u_2 \\ \dot{u}_2 &= 3u_1 + 2u_2 \end{aligned} \quad \text{bzw.} \quad \dot{\vec{u}} = \underbrace{\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}}_A \vec{u}$$

Die Idee zur Lösung dieses Systems besteht darin, einen Vektor  $\vec{x} \neq \vec{0}$  mit  $A\vec{x} = \lambda\vec{x}$  zu suchen, denn dann gilt für  $\vec{u}(t) := e^{\lambda t}\vec{x}$ :

$$\dot{\vec{u}}(t) = \lambda e^{\lambda t}\vec{x} = Ae^{\lambda t}\vec{x} = A\vec{u}(t),$$

das heißt  $\vec{u}(t)$  ist eine Lösung der Differentialgleichung. Wir müssen also eine Lösung  $\vec{x} \neq \vec{0}$  des Gleichungssystems

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}$$

mit einem noch unbestimmten Parameter  $\lambda$  finden. Dieses Gleichungssystem ist äquivalent zu

$$\begin{pmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

und dieses hat genau dann eine vom Nullvektor verschiedene Lösung, wenn die Matrix

$$A - \lambda E_2 = \begin{pmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{pmatrix}$$

nicht invertierbar ist, also wenn gilt:

$$0 = \det \begin{pmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{pmatrix} = (1-\lambda)(2-\lambda) - 6 = \lambda^2 - 3\lambda - 4 = (\lambda+1)(\lambda-4)$$

Die beiden Lösungen  $\lambda_1 = -1$  und  $\lambda_2 = 4$  dieser Gleichung nennt man die Eigenwerte von  $A$ . Zu diesen Eigenwerten suchen wir nun Vektoren  $\vec{x}$  mit  $A\vec{x} = \lambda\vec{x}$ , die man dann Eigenvektoren (zum Eigenwert  $\lambda$ ) nennt.

- $\lambda_1 = -1$ : Zu lösen ist  $A\vec{x} = -\vec{x}$ , also

$$\begin{pmatrix} 2 & 2 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad x_2 = -x_1$$

Alle Lösungen dieses Gleichungssystems sind demnach Vielfache des Vektors  $\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ .

- $\lambda_2 = 4$ : Hier müssen wir  $A\vec{x} = 4\vec{x}$  lösen, das heißt

$$\begin{pmatrix} -3 & 2 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad 2x_2 = 3x_1$$

Somit ist  $v_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$  ein Eigenvektor zu  $\lambda_2 = 4$ , und alle weiteren sind Vielfache von  $v_2$ .

Nun zurück zur Differentialgleichung: Wir haben nunmehr gesehen, dass sowohl  $e^{-t} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$  als auch  $e^{4t} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$  das System lösen. Aber dann gilt das auch für jede Linearkombination, und die allgemeine Lösung des Differentialgleichungssystems ist

$$\vec{u}(t) = \alpha_1 e^{-t} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \alpha_2 e^{4t} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

mit Koeffizienten  $\alpha_1, \alpha_2$  (die man bestimmt, sobald Anfangsbedingungen  $\vec{u}(0)$  festgelegt sind).

**3.2. Eigenwerte, Eigenvektoren und Eigenräume.** Im einleitenden Beispiel hatten wir schon von den Eigenwerten einer Matrix gesprochen; es folgt die allgemeine Definition.

**Definition 3.2.** Sei  $A$  eine  $n \times n$ -Matrix. Eine Zahl  $\lambda \in \mathbb{C}$  heißt Eigenwert von  $A$ , falls es einen Vektor  $\vec{x} \neq \vec{0}$  gibt, so dass

$$A\vec{x} = \lambda\vec{x}$$

gilt. Jeder solche Vektor  $\vec{x}$  heißt dann ein Eigenvektor zum Eigenwert  $\lambda$ .

Sind  $\vec{v}, \vec{w}$  beides Eigenvektoren zum Eigenwert  $\lambda$ , so gilt

$$A(\vec{v} + \vec{w}) = A\vec{v} + A\vec{w} = \lambda\vec{v} + \lambda\vec{w} = \lambda(\vec{v} + \vec{w})$$

sowie für einen Skalar  $0 \neq \alpha \in \mathbb{C}$

$$A(\alpha\vec{v}) = \alpha A\vec{v} = \alpha\lambda\vec{v} = \lambda(\alpha\vec{v}),$$

folglich sind  $\vec{v} + \vec{w}$  und  $\alpha\vec{v}$  wieder Eigenvektoren zum Eigenwert  $\lambda$ .

**Definition 3.3.** Die Menge  $\mathcal{E}(A; \lambda)$  aller Eigenvektoren zum Eigenwert  $\lambda$  zusammen mit dem Nullvektor nennt man den Eigenraum zum Eigenwert  $\lambda$ .

Die der Definition voranstehende Überlegung bedeutet gerade, dass  $\mathcal{E}(A; \lambda)$  ein Unterraum von  $\mathbb{C}^n$  ist.

In den nächsten Unterabschnitten wenden wir uns der Frage zu, wie man Eigenwerte, Eigenvektoren und Eigenräume bestimmt.

**3.3. Bestimmung der Eigenwerte.** Damit  $\lambda$  ein Eigenwert der Matrix  $A$  ist, muss es einen Vektor  $\vec{x} \neq \vec{0}$  geben mit

$$A\vec{x} = \lambda\vec{x}.$$

Diese Gleichung ist aber äquivalent zu

$$(A - \lambda E_n)\vec{x} = \vec{0},$$

die genau dann eine Lösung  $\vec{x} \neq \vec{0}$  hat, wenn  $A - \lambda E_n$  nicht invertierbar ist, also Determinante Null hat.

**Definition 3.4.** Sei  $A$  eine  $n \times n$ -Matrix. Dann heißt

$$\chi_A(t) := \det(A - tE_n)$$

das charakteristische Polynom von  $A$ .

Ausführlicher geschrieben:

$$\chi_A(t) = \det \begin{pmatrix} a_{11} - t & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} - t & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} - t & a_{34} & \dots & a_{3n} \\ \vdots & \vdots & & \ddots & & \vdots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,n-1} - t & a_{n-1,n} \\ a_{n1} & a_{n2} & \dots & a_{n,n-1} & a_{nn} - t \end{pmatrix}$$

Die Eigenschaften der Determinante implizieren insbesondere, dass  $\chi_A(t)$  ein Polynom vom Grad  $n$  ist (in der Variablen  $t$ ), denn einer der Summanden bei der Berechnung der Determinante ist

das Produkt

$$(a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t) = (-1)^n t^n + \cdots$$

**Satz 3.5.** Sei  $A$  eine  $n \times n$ -Matrix.

- (a)  $\lambda \in \mathbb{C}$  ist genau dann ein Eigenwert von  $A$ , wenn  $\lambda$  eine Nullstelle des charakteristischen Polynoms  $\chi_A(t)$  ist.  
 (b) Ein Vektor  $\vec{x} \neq \vec{0}$  ist genau dann ein Eigenvektor von  $A$  zum Eigenwert  $\lambda$ , wenn  $\vec{x}$  eine Lösung des linearen Gleichungssystems

$$(A - \lambda E_n)\vec{x} = \vec{0}$$

ist. Der Lösungsraum dieses Gleichungssystems ist der Eigenraum  $\mathcal{E}(A; \lambda)$  von  $\lambda$ .

Aus diesem Satz ergibt sich folgendes Rezept:

1. Berechne das charakteristische Polynom  $\chi_A(t)$  der Matrix.
2. Bestimme die Nullstellen  $\lambda_1, \dots, \lambda_n$  von  $\chi_A(t)$ : ein Polynom  $n$ -ten Grades hat stets  $n$  komplexe Nullstellen, wobei auch mehrfache Nullstellen auftreten können.
3. Löse die linearen Gleichungssysteme

$$(A - \lambda_i E_n)\vec{x} = \vec{0}$$

für  $i = 1, \dots, n$ , also

$$\begin{pmatrix} a_{11} - \lambda_i & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda_i & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} - \lambda_i & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,n-1} - \lambda_i & a_{n-1,n} \\ a_{n1} & a_{n2} & \dots & a_{n,n-1} & a_{nn} - \lambda_i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Dieses Verfahren illustrieren wir anhand einiger Beispiele.

**Beispiele 3.6.** (a) Sei  $A = \begin{pmatrix} 1 & 4 \\ 1 & -2 \end{pmatrix}$ .

1. Schritt:

$$\begin{aligned} \chi_A(t) &= \det(A - tE_2) = \det \begin{pmatrix} 1-t & 4 \\ 1 & -2-t \end{pmatrix} \\ &= (1-t)(-2-t) - 4 = t^2 + t - 6 = (t-2)(t+3) \end{aligned}$$

2. Schritt: Die Nullstellen von  $\chi_A(t) = (t-2)(t+3)$  sind  $\lambda_1 = 2$  und  $\lambda_2 = -3$ .

3. Schritt: Bestimmung der Eigenvektoren. Wir beginnen mit  $\lambda_1 = 2$ :

$$\begin{aligned} (A - \lambda_1 E_2)\vec{x} = \vec{0} &\Leftrightarrow \begin{pmatrix} 1-2 & 4 \\ 1 & -2-2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} -1 & 4 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow x_1 = 4x_2 \end{aligned}$$

Eine Lösung ist etwa  $\vec{v}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ . Nun zu  $\lambda_2 = -3$ :

$$\begin{aligned} (A - \lambda_2 E_2)\vec{x} = \vec{0} &\Leftrightarrow \begin{pmatrix} 1+3 & 4 \\ 1 & -2+3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow x_1 = -x_2 \end{aligned}$$

Eine Lösung hier ist zum Beispiel  $\vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ . Wir machen noch die Probe:

$$A\vec{v}_1 = \begin{pmatrix} 1 & 4 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix} = 2\vec{v}_1 \quad \checkmark$$

$$A\vec{v}_2 = \begin{pmatrix} 1 & 4 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -3 \\ 3 \end{pmatrix} = -3\vec{v}_1 \quad \checkmark$$

(b) Sei  $B = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & -1 \\ 2 & -2 & 1 \end{pmatrix}$ .

1. Schritt:

$$\begin{aligned} \chi_B(t) &= \det \begin{pmatrix} 3-t & 0 & 0 \\ 1 & 2-t & -1 \\ 2 & -2 & 1-t \end{pmatrix} = (3-t)[(2-t)(1-t) - 2] \\ &= (3-t)(t^2 - 3t) = -t(t-3)^2 \end{aligned}$$

2. Schritt: Die Nullstellen sind  $\lambda_1 = 0$  und  $\lambda_2 = 3$  (doppelt).

3. Schritt: Bestimmung der Eigenvektoren/Eigenräume. Wir beginnen mit der einfachen Nullstelle  $\lambda_1 = 0$ .

$$\begin{aligned} (B - \lambda_1 E_3)\vec{x} = \vec{0} &\Leftrightarrow \begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & -1 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow x_1 = 0 \text{ und } 2x_2 = x_3. \end{aligned}$$

Eine Lösung ist  $\vec{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$ , und der Eigenraum von  $\lambda_1 = 0$  ist  $\mathcal{E}(B; 0) = \text{Spann}\{\vec{v}_1\}$ , das heißt der Unterraum aller Vielfachen von  $\vec{v}_1$  (manche Quellen verwenden auch Lin statt Spann).

Nun zu  $\lambda_2 = 3$ :

$$\begin{aligned} (B - \lambda_2 E_3)\vec{x} = \vec{0} &\Leftrightarrow \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & -1 \\ 2 & -2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow x_1 = x_2 + x_3 \end{aligned}$$

Hier gibt es zwei linear unabhängige Eigenvektoren, etwa  $\vec{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$  und  $\vec{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ . Der Eigenraum ist die Menge aller Linearkombinationen von  $\vec{v}_2$  und  $\vec{v}_3$ , also  $\mathcal{E}(B; 3) = \text{Spann}\{\vec{v}_2, \vec{v}_3\}$  und hat die Dimension 2.

(c) Sei  $C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & 2 & 2 \end{pmatrix}$ .

1. Schritt:

$$\begin{aligned} \chi_C(t) &= \det \begin{pmatrix} 1-t & 1 & 0 \\ 1 & -t & -1 \\ -1 & 2 & 2-t \end{pmatrix} = -t(1-t)(2-t) + 1 + 2(1-t) - (2-t) \\ &= -t^3 + 3t^2 - 3t + 1 = (1-t)^3 \end{aligned}$$

2. Schritt: Es gibt nur eine Nullstelle, nämlich die dreifache Nullstelle  $\lambda = 1$ .

**3. Schritt:** Bestimmung des Eigenraumes zu  $\lambda = 1$ :

$$(A - \lambda E_3)\vec{x} = \vec{0} \Leftrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & -1 \\ -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ \Leftrightarrow x_2 = 0 \text{ und } x_1 = x_3.$$

Jeder Eigenvektor ist ein Vielfaches von  $\vec{v} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ . Insbesondere hat  $\mathcal{E}(C; 1)$  die Dimension 1.

In Beispiel 3.6(b) hatten wir einen doppelten Eigenwert mit zwei linear unabhängigen Eigenvektoren, aber in Teil (c) war der Eigenwert  $\lambda = 1$  dreifach, die Dimension des Eigenraums jedoch nur 1.

Sei nun  $A$  eine  $n \times n$ -Matrix. Die Dimension des Eigenraums zum Eigenwert  $\lambda$  ist stets kleiner oder gleich der Vielfachheit von  $\lambda$  als Nullstelle des charakteristischen Polynoms. Da die Anzahl der linear unabhängigen Lösungen des Gleichungssystems  $(A - \lambda E_n)\vec{x} = \vec{0}$  gerade  $n - \text{Rang}(A - \lambda E_n)$  ist, gilt

$$\dim \mathcal{E}(A; \lambda) = n - \text{Rang}(A - \lambda E_n).$$

Ein Polynom vom Grad  $n$  hat bekanntlich höchstens  $n$  paarweise verschiedene Nullstellen; folglich kann eine  $n \times n$ -Matrix auch nur  $n$  verschiedene Eigenwerte haben.

Seien nun  $\lambda, \mu$  zwei verschiedene Eigenwerte von  $A$  und  $\vec{v}$  ein Eigenvektor zum Eigenwert  $\lambda$  sowie  $\vec{w}$  ein Eigenvektor zum Eigenwert  $\mu$ . Dann sind  $\vec{v}, \vec{w}$  linear unabhängig, denn gäbe es ein  $\alpha$  mit  $\vec{w} = \alpha \vec{v}$ , so wäre notwendigerweise  $\alpha \neq 0$ , und es gälte

$$\vec{0} = A\vec{w} - \mu\vec{w} = A\alpha\vec{v} - \mu\alpha\vec{v} = \alpha\lambda\vec{v} - \mu\alpha\vec{v} = \alpha(\lambda - \mu)\vec{v}$$

was wegen  $\lambda \neq \mu$  und  $\vec{v} \neq \vec{0}$  nicht sein kann. Allgemeiner gilt:

**Satz 3.7.** Seien  $\vec{v}_1, \dots, \vec{v}_m$  Eigenvektoren zu den paarweise verschiedenen Eigenwerten  $\lambda_1, \dots, \lambda_m$  von  $A$ . Dann sind die Vektoren  $\vec{v}_1, \dots, \vec{v}_m$  linear unabhängig.

Schließlich noch einige Rechenhinweise.

**3.8.** Sei  $A = (a_{ik})$  eine  $2 \times 2$  Matrix. Dann hat das charakteristische Polynom die Gestalt

$$\chi_A(t) = t^2 - (a_{11} + a_{22})t + (a_{11}a_{22} - a_{12}a_{21}).$$

Den konstanten Koeffizienten erkennen wir als die Determinante von  $A$ . Der Koeffizient von  $t$  ist die Summe der Diagonalelemente der Matrix; diese nennt man die Spur von  $A$ . Es gilt also

$$\chi_A(t) = t^2 - \text{Spur}(A)t + \det(A).$$

Für eine  $n \times n$ -Matrix  $A = (a_{ik})$  mit  $n > 2$  gilt immer noch

$$\chi_A(t) = (-1)^n t^n + (-1)^{n-1} \text{Spur}(A) + c_{n-2} t^{n-2} + \dots + c_1 t + \det(A),$$

wobei  $\text{Spur}(A) = a_{11} + a_{22} + \dots + a_{nn}$  die Summe der Diagonaleinträge ist. (Die anderen Koeffizienten  $c_i$  kann man durch geeignete Unterdeterminanten, sogenannte Minoren, ausdrücken.)

**3.9.** Sei  $A = (a_{ik})$  eine reelle Matrix, das heißt mit Einträgen  $a_{ik} \in \mathbb{R}$ . Dann ist  $\chi_A(t)$  ein reelles Polynom  $n$ -ten Grades. Ist nun  $\lambda$  eine komplexe Nullstelle von  $\chi_A(t)$ , so ist auch  $\bar{\lambda}$  eine Nullstelle, denn wegen der reellen Koeffizienten gilt  $\chi_A(\bar{\lambda}) = \overline{\chi_A(\lambda)} = 0$ . Ist nun  $\vec{x}$  ein Eigenvektor zum komplexen Eigenwert  $\lambda$ , so ist der zu  $\vec{x}$  komplex konjugierte Vektor  $\bar{\vec{x}}$  ein Eigenvektor zum Eigenwert  $\bar{\lambda}$ , denn wegen der reellen Koeffizienten von  $A$  gilt

$$A\bar{\vec{x}} = \bar{A\vec{x}} = \bar{\lambda\vec{x}} = \bar{\lambda}\bar{\vec{x}} = \bar{\lambda}\bar{\vec{x}}.$$

**3.4. Diagonalisierung.** Zu Beginn hatten wir schon darauf hingewiesen, dass es nützlich sein kann, lineare Abbildungen in möglichst einfacher Form darzustellen. Will man zum Beispiel einen starren Körper studieren, lohnt sich unter Umständen die Wahl eines Koordinatensystem, dessen Achsen die Hauptträgheitsachsen sind. Eine andere Anwendung besteht in der vereinfachten Berechnung von Potenzen von Matrizen.

**Definition 3.10.** Zwei  $n \times n$ -Matrizen  $A$  und  $B$  heißen *ähnlich*, wenn es eine invertierbare  $n \times n$ -Matrix  $T$  gibt, so dass

$$B = T^{-1}AT$$

gilt. Die Matrix  $T$  nennt man *Transformationsmatrix*, die Transformation  $A \mapsto T^{-1}AT$  nennt man eine *Ähnlichkeitstransformation*.

Eigenwerte ändern sich nicht unter Ähnlichkeitstransformationen: Sei  $A$  eine  $n \times n$ -Matrix,  $T$  eine invertierbare  $n \times n$ -Matrix und  $B = T^{-1}AT$ . Sei weiterhin  $\lambda$  ein Eigenwert von  $A$  mit zugehörigem Eigenvektor  $\vec{v}$ . Dann gilt

$$B(T^{-1}\vec{v}) = T^{-1}AT(T^{-1}\vec{v}) = T^{-1}A(TT^{-1})\vec{v} = T^{-1}A\vec{v} = T^{-1}\lambda\vec{v} = \lambda(T^{-1}\vec{v})$$

Mit anderen Worten,  $\vec{w} := T^{-1}\vec{v}$  ist ein Eigenvektor von  $B$  zum Eigenwert  $\lambda$ . Damit ist gezeigt:

**Satz 3.11.** Seien  $A$  und  $B$  zwei ähnliche  $n \times n$ -Matrizen mit Transformationsmatrix  $T$ . Sei ferner  $\lambda$  ein Eigenwert von  $A$  und  $\vec{v}$  ein Eigenvektor zum Eigenwert  $\lambda$ . Dann ist  $\lambda$  auch ein Eigenwert von  $B$  und  $\vec{w} := T^{-1}\vec{v}$  ein Eigenvektor von  $B$  zu  $\lambda$ .

Das Ziel ist es, eine Transformation auf eine möglichst einfache Matrix zu finden. Die einfachsten Matrizen sind sogenannte Diagonalmatrizen:

**Definition 3.12.** Eine  $n \times n$ -Matrix  $D = (d_{ik})$  heißt *Diagonalmatrix*, wenn alle Einträge außerhalb der Diagonalen Null sind, das heißt  $d_{ik} = 0$  für  $i \neq k$ .

Eine solche Diagonalmatrix schreibt man dann gelegentlich auch in der Form

$$D = \text{diag}(d_{11}, d_{22}, \dots, d_{nn}).$$

Die Eigenwerte einer Diagonalmatrix  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  sind gerade die Diagonaleinträge, denn es gilt offenbar

$$D\vec{e}_k = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_k & & \\ & & & \ddots & \\ 0 & & & & \lambda_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \lambda_k \\ \vdots \\ 0 \end{pmatrix} = \lambda_k \vec{e}_k.$$

**Definition 3.13.** Eine  $n \times n$ -Matrix  $A$  heißt *diagonalisierbar*, wenn es eine Diagonalmatrix  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  und eine invertierbare Matrix  $T$  gibt, so dass

$$T^{-1}AT = D.$$

Da ähnliche Matrizen die gleichen Eigenwerte haben, stehen auf der Diagonalen von  $D = T^{-1}AT$  gerade die Eigenwerte von  $A$ .

Wir wenden uns nun der Frage zu, welche Matrizen diagonalisierbar sind. Sicherlich nicht alle:



**Beispiel.** Sei  $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ . Angenommen es gäbe eine invertierbare Matrix  $T$  mit  $T^{-1}AT = \text{diag}(\lambda_1, \lambda_2)$ , so müssten  $\lambda_1, \lambda_2$  die Eigenwerte von  $A$  sein. Wegen

$$\det(A - tE_2) = \det \begin{pmatrix} 1-t & 1 \\ 0 & 1-t \end{pmatrix} = (1-t)^2$$

besitzt  $A$  nur den einen Eigenwert 1, also hätten wir  $D = E_2$ . Aber aus  $T^{-1}AT = E_2$  folgt durch Linksmultiplikation mit  $T$  zuerst  $AT = T$  und dann durch Rechtsmultiplikation mit  $T^{-1}$  schließlich  $A = E_2$ . Aber  $A \neq E_2$ . Folglich muss die Annahme,  $A$  sei diagonalisierbar, falsch gewesen sein.

Was ist in diesem Beispiel passiert? Wir haben für die Matrix  $A$  den doppelten Eigenwert  $\lambda = 1$ , aber der Eigenraum  $\mathcal{E}(A; 1) = \{\alpha \vec{e}_1 \mid \alpha \in \mathbb{C}\}$  hat nur die Dimension 1. Das kann für diagonalisierbare Matrizen aber nicht sein: Nehmen wir an, die  $n \times n$ -Matrix sei diagonalisierbar mit Transformationsmatrix  $T$ ,

$$T^{-1}AT = D = \text{diag}(\lambda_1, \dots, \lambda_n).$$

Durch Linksmultiplikation mit  $T$  erhalten wir aus dieser Gleichung

$$AT = TD.$$

Seien nun  $\vec{v}_1, \dots, \vec{v}_n$  die Spalten von  $T$ , also  $T = (\vec{v}_1, \dots, \vec{v}_n)$ . Multiplikation von rechts mit einer Diagonalmatrix  $D$  hat den Effekt, dass die Spalten von  $T$  mit den Diagonaleinträgen von  $D$  multipliziert werden, also ist

$$TD = (\lambda_1 \vec{v}_1, \lambda_2 \vec{v}_2, \dots, \lambda_n \vec{v}_n).$$

Andererseits ist, wieder nach den Regeln der Matrixmultiplikation,

$$AT = A(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n) = (A\vec{v}_1, A\vec{v}_2, \dots, A\vec{v}_n).$$

Folglich muss

$$A\vec{v}_1 = \lambda_1 \vec{v}_1, \dots, A\vec{v}_k = \lambda_k \vec{v}_k, \dots, A\vec{v}_n = \lambda_n \vec{v}_n$$

gelten, das heißt, die  $k$ -te Spalte  $\vec{v}_k$  von  $T$  ist ein Eigenvektor von  $A$  zum Eigenwert  $\lambda_k$ . Insbesondere hat  $A$  also  $n$  linear unabhängige Eigenvektoren, denn die Spalten einer invertierbaren Matrix sind linear unabhängig. In unserem obigen Beispiel war das nicht der Fall, also konnte die Beispielmatrix auch nicht diagonalisierbar sein. Damit haben wir eine notwendige Bedingung für Diagonalisierbarkeit gefunden:  $A$  muss  $n$  linear unabhängige Eigenvektoren besitzen.

Aber ist diese auch ausreichend?

Nehmen wir also umgekehrt an, zu  $A$  gebe es  $n$  linear unabhängige Eigenvektoren  $\vec{v}_1, \dots, \vec{v}_n$  zu den  $n$  Eigenwerten  $\lambda_1, \dots, \lambda_n$ ; diese müssen nicht alle verschieden sein. Schreiben wir diese als Spalten in eine Matrix  $T$ , so ist  $T = (\vec{v}_1, \dots, \vec{v}_n)$  invertierbar, und es gilt

$$AT = A(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n) = (A\vec{v}_1, A\vec{v}_2, \dots, A\vec{v}_n) = (\lambda_1 \vec{v}_1, \lambda_2 \vec{v}_2, \dots, \lambda_n \vec{v}_n).$$

Aber

$$\underbrace{(\lambda_1 \vec{v}_1, \dots, \lambda_n \vec{v}_n)}_{AT} = \underbrace{(\vec{v}_1, \dots, \vec{v}_n)}_T \cdot \underbrace{\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}}_D,$$

das heißt  $T^{-1}AT = D = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Die Bedingung ist also auch hinreichend.

Zusammengefasst:

**Satz 3.14.** Eine  $n \times n$ -Matrix  $A$  ist genau dann diagonalisierbar, wenn sie  $n$  linear unabhängige Eigenvektoren besitzt. Sind  $\vec{v}_1, \dots, \vec{v}_n$  linear unabhängige Eigenvektoren, so transformiert die Matrix  $T = (\vec{v}_1, \dots, \vec{v}_n)$ , die die  $\vec{v}_k$  als Spalten hat,  $A$  auf Diagonalgestalt.

Der Satz liefert nicht nur ein Kriterium für Diagonalisierbarkeit, sondern auch ein konkretes Verfahren zur Diagonalisierung einer  $n \times n$ -Matrix  $A$ , wenn dies möglich ist.

### 3.15. Diagonalisierungsverfahren

1. Man berechne die Eigenwerte von  $A$
2. Zu jedem der verschiedenen Eigenwerte bestimme man den Eigenraum. Stimmen für jeden Eigenwert  $\lambda$  dessen Vielfachheit als Nullstelle und die Dimension des Eigenraums überein, erhält man insgesamt  $n$  linear unabhängige Eigenvektoren  $\vec{v}_1, \dots, \vec{v}_n$  und  $A$  ist diagonalisierbar, ansonsten nicht.
3. Man schreibe die Eigenvektoren als Spalten in eine Matrix  $T$ ; das Produkt  $T^{-1}AT$  ist dann eine Diagonalmatrix mit den Eigenwerten als Diagonalelementen.

**Beispiel 3.16.** Wir bestimmen eine Matrix  $T$ , die

$$B = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & -1 \\ 2 & -2 & 1 \end{pmatrix},$$

die Matrix aus Beispiel 3.6, diagonalisiert. Das charakteristische Polynom ist  $\chi_B(t) = -t(t-3)^2$  mit einer einfachen Nullstelle  $\lambda_1 = 0$  und einer doppelten Nullstelle  $\lambda_2 = 3$ . Als Eigenraumbasen haben wir

$$\begin{aligned} \lambda_1 = 0: \quad \vec{v}_1 &= \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \\ \lambda_2 = 3: \quad \vec{v}_2 &= \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

berechnet. Da wir drei linear unabhängige Eigenvektoren gefunden haben, ist  $B$  diagonalisierbar mit

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

Man sollte zur Probe nachrechnen, dass

$$T^{-1}BT = \frac{1}{3} \begin{pmatrix} -1 & 1 & 1 \\ 1 & 2 & -1 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & -1 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

gilt.

Hingegen ist die Matrix  $C$  aus Beispiel 3.6 nicht diagonalisierbar, denn der Eigenraum des einzigen Eigenwerts ist nur eindimensional.

Wenn die Eigenwerte der Matrix paarweise verschieden sind, gibt es laut Satz 3.7  $n$  linear unabhängige Eigenvektoren. Also gilt:

**Satz 3.17.** *Besitzt eine  $n \times n$ -Matrix  $n$  paarweise verschiedene Eigenwerte, so ist sie diagonalisierbar.*

**Warnung:** Dieser Satz liefert eine hinreichende Bedingung, keine notwendige. Selbst wenn die Matrix mehrfache Eigenwerte hat, kann sie diagonalisierbar sein.

**Beispiel 3.18.** Wir untersuchen die Matrix

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}$$

auf Diagonalisierbarkeit. Das charakteristische Polynom ist

$$\chi_A(t) = \det \begin{pmatrix} -t & 0 & -2 \\ 1 & 2-t & 1 \\ 1 & 0 & 3-t \end{pmatrix} = -t(2-t)(3-t) = (2-t)(t^2 - 3t + 2) = (2-t)^2(1-t)$$

mit einer einfachen Nullstelle  $\lambda_1 = 1$  und einer doppelten Nullstelle  $\lambda_2 = 2$ . Wir müssen die Eigenräume bestimmen:

$$A - \lambda_1 E_3 = \begin{pmatrix} -1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ & 1 & -1 \\ & & 0 \end{pmatrix} \quad \text{führt zu} \quad \vec{v}_1 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix},$$

und

$$A - \lambda_2 E_3 = \begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{pmatrix} \quad \text{liefert} \quad \vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

die linear unabhängig sind. Wir haben also drei linear unabhängige Eigenvektoren und die Matrix  $A$  ist diagonalisierbar. Die Matrix

$$T = (\vec{v}_1, \vec{v}_2, \vec{v}_3) = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad \text{mit Inverser} \quad \begin{pmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}$$

transformiert  $A$  auf Diagonalgestalt:

$$T^{-1}AT = \begin{pmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 2 \end{pmatrix}.$$

#### 4. Hermitesche und unitäre Matrizen

Wir erinnern zunächst an das komplexe Standard-Skalarprodukt: für zwei Vektoren  $\vec{x}, \vec{y} \in \mathbb{C}^n$  ist dies definiert durch

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^* \vec{y} = \sum_{i=1}^n \bar{x}_i y_i;$$

sind die Vektoren reell, ist dies gerade das Standard-Skalarprodukt des  $\mathbb{R}^n$ . Das komplexe Skalarprodukt definiert die komplexe Norm

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \left( \sum_{i=1}^n \bar{x}_i x_i \right)^{1/2},$$

die reelle Werte annimmt. Man nennt zwei Vektoren  $\vec{v}, \vec{w}$  *orthogonal*, wenn ihr Skalarprodukt Null ist, also wenn  $\langle \vec{v}, \vec{w} \rangle = 0$  gilt. (In manchen Quellen wird dafür im Fall komplexer Vektoren auch der Begriff unitär verwendet.)

**Definition 4.1.** Eine komplexe Matrix  $A = (a_{ik})$  heißt hermitesch, wenn gilt:

$$A = A^*,$$

also  $a_{ik} = \bar{a}_{ki}$  für alle  $i$  und  $k$ .

Insbesondere sind die Diagonaleinträge einer hermiteschen Matrix reell, denn für sie gilt  $a_{ii} = \bar{a}_{ii}$ .

**Beispiele 4.2.** Die Matrix

$$\begin{pmatrix} 4 & 2j \\ -2j & 7 \end{pmatrix}$$

ist hermitesch, die Matrix

$$\begin{pmatrix} 1+j & 2+j \\ 2-j & 2 \end{pmatrix}$$

jedoch nicht (der erste Diagonaleintrag ist nicht reell).

Zerlegt man eine hermitesche Matrix  $A$  in Real- und Imaginärteil, also

$$A = B + jC \quad \text{mit reellen Matrizen } B \text{ und } C,$$

so folgt

$$B + jC = A = A^* = B^\top - jC^\top.$$

Folglich gilt

- (i)  $B = B^\top$ , das heißt  $B$  ist symmetrisch,
- (ii)  $C = -C^\top$ , das heißt  $C$  ist antisymmetrisch.

Insbesondere ist eine reelle hermitesche Matrix symmetrisch.

*Bemerkung.* Für eine hermitesche Matrix  $A$  und komplexe Vektoren  $\vec{x}, \vec{y}$  gilt offenbar

$$\langle A\vec{x}, \vec{y} \rangle = (A\vec{x})^* \vec{y} = \vec{x}^* A^* \vec{y} = \vec{x}^* A \vec{y} = \langle \vec{x}, A\vec{y} \rangle. \quad (4.1)$$

**Definition 4.3.** Eine Matrix  $A$  heißt unitär, wenn gilt:

$$A^* A = E$$

Ist  $A$  reell, so bedeutet dies  $A^\top A = E$ , und man nennt  $A$  orthogonal.

Insbesondere ist also  $A$  invertierbar mit Inverser  $A^{-1} = A^*$ .

**Beispiel 4.4.** Sei

$$A = \frac{1}{5} \begin{pmatrix} 3 & 4j \\ -4 & 3j \end{pmatrix}.$$

Dann gilt

$$A^* = \frac{1}{5} \begin{pmatrix} 3 & -4 \\ -4j & -3j \end{pmatrix} \quad \text{und} \quad A^* A = \frac{1}{25} \begin{pmatrix} 9+16 & 12j-12j \\ -12j+12j & -(4j)^2 - (3j)^2 \end{pmatrix} = E_2.$$

Für eine unitäre Matrix  $A$  ist auch ihre komplex Transponierte unitär, denn es gilt  $A^{**} A^* = A A^* = A A^{-1} = E$

Aus dem Multiplikationssatz 2.2 und  $\det(A) = \det(A^\top)$  folgt für eine unitäre Matrix  $A$

$$\begin{aligned} |\det(A)|^2 &= \overline{\det(A)} \cdot \det(A) = \det(\bar{A}) \cdot \det(A) \\ &= \det(A^*) \cdot \det(A) = \det(A^* A) = \det(E) = 1, \end{aligned}$$

also

$$|\det(A)| = 1.$$

Weiterhin gilt für beliebige Vektoren  $\vec{x}, \vec{y}$

$$\langle A\vec{x}, A\vec{y} \rangle = (A\vec{x})^*(A\vec{y}) = \vec{x}^* A^* A \vec{y} = \vec{x}^* \vec{y} = \langle \vec{x}, \vec{y} \rangle$$

beziehungsweise

$$\|A\vec{x}\|^2 = (A\vec{x})^*(A\vec{x}) = \vec{x}^* A^* A \vec{x} = \vec{x}^* \vec{x} = \|\vec{x}\|^2.$$

Für orthogonale (also reell unitäre) Matrizen heißt dies gerade, dass  $A$  eine längen- und winkeltreue Abbildung von  $\mathbb{R}^n$  nach  $\mathbb{R}^n$  definiert.

*Bemerkung.* Die Gleichung  $A^*A = E$  bedeutet Folgendes: Seien  $\vec{v}_1, \dots, \vec{v}_n$  die Spalten von  $A$ , dann gilt

$$\langle \vec{v}_i, \vec{v}_k \rangle = \begin{cases} 1 & \text{für } i = k, \\ 0 & \text{für } i \neq k. \end{cases}$$

Mit anderen Worten, die Spalten von  $A$  bilden eine Basis von  $\mathbb{C}^n$ , die aus paarweise orthogonalen normierten Vektoren besteht; dies nennt man eine *Orthonormalbasis* von  $\mathbb{C}^n$ .

Sei nun  $A$  eine hermitesche Matrix und  $\lambda$  ein Eigenwert mit zugehörigem Eigenvektor  $\vec{v}$ . Dann gilt

$$\lambda \langle \vec{v}, \vec{v} \rangle = \langle \vec{v}, \lambda \vec{v} \rangle = \langle \vec{v}, A\vec{v} \rangle = \langle A\vec{v}, \vec{v} \rangle = \bar{\lambda} \langle \vec{v}, \vec{v} \rangle,$$

also  $\bar{\lambda} = \lambda$  und damit  $\lambda \in \mathbb{R}$ . Sei ferner  $\mu \neq \lambda$  ein weiterer Eigenwert mit zugehörigem Eigenvektor  $\vec{w}$ , so gilt wegen  $\lambda \in \mathbb{R}$  und (4.1)

$$(\lambda - \mu) \langle \vec{v}, \vec{w} \rangle = \langle \lambda \vec{v}, \vec{w} \rangle - \langle \vec{v}, \mu \vec{w} \rangle = \langle A\vec{v}, \vec{w} \rangle - \langle \vec{v}, A\vec{w} \rangle$$

und es folgt  $\langle \vec{v}, \vec{w} \rangle = 0$ . Andererseits gilt für jeden Vektor  $\vec{x}$  mit  $\langle \vec{v}, \vec{x} \rangle = 0$

$$\langle \vec{v}, A\vec{x} \rangle = \langle A\vec{v}, \vec{x} \rangle = \langle \lambda \vec{v}, \vec{x} \rangle = \lambda \langle \vec{v}, \vec{x} \rangle = 0,$$

das heißt  $A$  bildet jeden zu  $\vec{v}$  orthogonalen Vektor wieder auf einen solchen ab. Dies kann man benutzen, um induktiv eine Orthonormalbasis aus Eigenvektoren von  $A$  zu konstruieren.

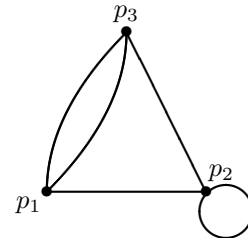
Zusammengefasst:

**Satz 4.5.** Sei  $A$  eine hermitesche Matrix.

- (a) Die Eigenwerte von  $A$  sind reell.
- (b) Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal.
- (c)  $A$  ist diagonalisierbar.

Genauer gesagt gilt sogar folgendes: es gibt eine unitäre Transformationsmatrix  $T$ , die  $A$  diagonalisiert.

**Beispiel 4.6.** Gegeben sei ein Wegdiagramm mit drei Knoten  $p_1, p_2, p_3$  und zwei Wegen, die  $p_1$  mit  $p_2$ , und je einem Weg, der  $p_1$  mit  $p_3$  beziehungsweise  $p_2$  mit  $p_3$  verbindet, sowie einem Rundweg von  $p_3$  nach  $p_3$ , der keinen weiteren Knoten enthält, so wie in der nebenstehenden Skizze. Dieses Diagramm kann man in einer quadratischen Matrix kodieren, die ebensoviel Zeilen wie das Diagramm Knoten hat: an die Position  $(i, k)$  der Matrix schreibt man die Anzahl der Kanten, die den Knoten  $p_i$  mit dem Knoten  $p_k$  verbinden; für unser Diagramm also



$$A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Eine solche Matrix nennt man die *Inzidenzmatrix* des Diagramms; wenn es (wie hier) keine Einbahnstraßenregelung gibt, ist sie symmetrisch und damit diagonalisierbar. Wenn man nun die

Anzahl der Wege der Länge 2 von  $p_1$  nach  $p_2$  sucht, so wird diese vom Eintrag an der Matrix  $A^2$  an der Stelle  $(1, 2)$  (oder der Stelle  $(2, 1)$ ) angegeben, und allgemeiner liefern die Einträge der  $k$ -ten Potenz die Anzahl der Wege der Länge  $k$ . Um Potenzen einfacher zu berechnen, diagonalisieren wir die Matrix. Es ist

$$\chi_A(t) = -t^3 + t^2 - 6t = -t(t+2)(t-3)$$

mit Eigenwerten  $\lambda_1 = 0$ ,  $\lambda_2 = -2$  und  $\lambda_3 = 3$ . Als normierte Eigenvektoren erhalten wir

$$\vec{v}_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

als orthogonale Transformationsmatrix also

$$T = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & -\sqrt{3} & \sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \\ -2 & 0 & \sqrt{2} \end{pmatrix}$$

Es ist dann  $T^\top AT = D := \text{diag}(0, -2, 3)$ , also  $A = TDT^\top$ , und damit

$$\begin{aligned} A^n &= TD^nT^\top = \frac{1}{6} \begin{pmatrix} 1 & -\sqrt{3} & \sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \\ -2 & 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & & \\ & (-2)^n & \\ & & 3^n \end{pmatrix} \begin{pmatrix} 1 & 1 & -2 \\ -\sqrt{3} & \sqrt{3} & 0 \\ \sqrt{2} & \sqrt{2} & \sqrt{2} \end{pmatrix} \\ &= \frac{1}{6} \begin{pmatrix} 2 \cdot 3^n + 3 \cdot (-2)^n & 2 \cdot 3^n - 3 \cdot (-2)^n & 2 \cdot 3^n \\ 2 \cdot 3^n - 3 \cdot (-2)^n & 2 \cdot 3^n + 3 \cdot (-2)^n & 2 \cdot 3^n \\ 2 \cdot 3^n & 2 \cdot 3^n & 2 \cdot 3^n \end{pmatrix} \end{aligned}$$

Damit erhalten wir geschlossene Formeln für die Zahl der Wege. Es ist zum Beispiel

$$A^{10} = \begin{pmatrix} 20195 & 19171 & 19683 \\ 19171 & 20195 & 19683 \\ 19683 & 19683 & 19683 \end{pmatrix},$$

und es gibt 19171 Wege von  $p_1$  nach  $p_2$  der Länge 10.

**Definition 4.7.** Eine hermitesche Matrix  $A$  heißt positiv (semi)definit, falls für alle Vektoren  $\vec{x} \neq \vec{0}$  gilt:

$$\vec{x}^* A \vec{x} > 0 \quad (\vec{x}^* A \vec{x} \geq 0)$$

$A$  heißt negativ (semi)definit, wenn  $-A$  positiv (semi)definit ist.

Hermitesche Matrizen, die weder positiv noch negativ (semi)definit sind (und davon gibt es viele!), nennt man auch *indefinit*. Die Bedeutung der positiven Definitheit einer Matrix liegt darin, dass man  $\sqrt{\langle \vec{x}, A \vec{x} \rangle}$  wieder als „Länge“ betrachten kann. Die Tatsache, dass man hermitesche Matrizen orthogonal diagonalisieren kann bedeutet wiederum, dass man diesen Längenbegriff in den euklidischen Standardabstand transformieren kann, jedenfalls bis auf Streckungen der Form  $x'_i = \lambda_i x_i$ . Diesen Vorgang nennt man dann *Hauptachsentransformation*.

**Beispiel 4.8.** Wir betrachten die Matrix  $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$  mit  $a, b, c \in \mathbb{R}$ ,  $a \neq 0$ . Diese Matrix ist reell symmetrisch, also hermitesch. Es gilt

$$\vec{x}^\top A \vec{x} = (x_1, x_2) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = ax_1^2 + 2bx_1x_2 + cx_2^2.$$

Da  $a \neq 0$  ist, wird daraus durch quadratische Ergänzung

$$a \left( x_1^2 + \frac{2b}{a} x_1 x_2 + \frac{b^2}{a^2} x_2^2 \right) + cx_2^2 - \frac{b^2}{a} x_2^2 = a \left( x_1 + \frac{b}{a} x_2 \right)^2 + \frac{ac - b^2}{a} x_2^2 = ay_1^2 + \frac{\det(A)}{a} y_2^2$$

wenn wir  $y_1 = x_1 + \frac{b}{a}$  und  $y_2 = x_2$  setzen. Die Matrix  $A$  ist also genau dann positiv definit, wenn  $a > 0$  und  $\det(A) > 0$  ist.

**Anwendung** (Hesse-Matrix). Eine Funktion  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , die genügend oft differenzierbar ist, hat im Punkt  $\vec{p} \in \mathbb{R}^n$  ein lokales Minimum, wenn gilt:

- alle partiellen Ableitungen  $\frac{\partial f}{\partial x_i}(\vec{p})$  im Punkt  $\vec{p}$  sind Null, also  $\text{grad } f(\vec{p}) = 0$ , und
- die sogenannte Hesse-Matrix der partiellen Ableitungen zweiter Ordnung

$$\text{Hesse}(f)(\vec{p}) := \left( \frac{\partial^2 f}{\partial x_i \partial x_k}(\vec{p}) \right)_{ik} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{p}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\vec{p}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\vec{p}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\vec{p}) & \frac{\partial^2 f}{\partial x_2^2}(\vec{p}) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\vec{p}) \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\vec{p}) & \dots & \dots & \frac{\partial^2 f}{\partial x_n^2}(\vec{p}) \end{pmatrix}$$

ist positiv definit.

Sei nun  $A$  eine hermitesche Matrix. Da hermitesche Matrizen orthogonal diagonalisierbar sind, gibt es eine Matrix  $T$  mit

$$T^* A T = \text{diag}(\lambda_1, \dots, \lambda_n) =: D, \quad \lambda_i \in \mathbb{R}.$$

Ist  $A$  zusätzlich noch positiv definit, so müssen alle  $\lambda_i$  positiv sein, denn wäre etwa  $\lambda_k \leq 0$ , so gälte für  $\vec{x} := T^* \vec{e}_k$ :

$$\vec{x}^* A \vec{x} = (T \vec{e}_k)^* A (T \vec{e}_k) = \vec{e}_k^* T^* A T \vec{e}_k = \vec{e}_k^* D \vec{e}_k = \lambda_k \leq 0.$$

Sind umgekehrt alle  $\lambda_i > 0$ , so ist  $A$  auch positiv definit: jeder Vektor  $\vec{x}$  ist eine Linearkombination

$$\vec{x} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \dots + \alpha_n \vec{v}_n$$

der Spalten von  $T$ , da diese eine Orthonormalbasis bilden, und wegen  $A \vec{v}_i = \lambda_i \vec{v}_i$  gilt

$$\vec{x}^* A \vec{x} = \alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_2 + \dots + \alpha_n^2 \lambda_n > 0.$$

(Genauso argumentiert man für positiv semidefinit.)

Daraus kann man ein weiteres Kriterium für positive Definitheit reeller, symmetrischer Matrizen ableiten. Sei dazu  $A$  eine reelle, symmetrische, positiv definite  $n \times n$ -Matrix. Dann ist  $A$  diagonalisierbar, und da alle Eigenwerte  $\lambda_i > 0$  sind, ist auch  $\det(A) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n > 0$ . Für ein  $k$  mit  $1 \leq k \leq n$  sei  $A_k$  die Untermatrix, die aus  $A$  durch Streichen der letzten  $n - k$  Zeilen und Spalten entsteht:

$$A_k = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix}.$$

Ist  $A$  positiv definit, so sicherlich auch jede der Matrizen  $A_k$  (man betrachte Vektoren, bei denen nur die ersten  $k$  Einträge  $\neq 0$  sind). Folglich müssen nach dem gleichen Argument wie eben auch alle Abschnittsdeterminanten

$$\Delta_k = \det(A_k)$$

positiv sein. Die Umkehrung gilt ebenfalls: sind alle  $\Delta_k > 0$ , so ist  $A$  positiv definit.

Zusammengefasst:

**Satz 4.9.** (a) Sei  $A$  eine hermitesche Matrix.  $A$  ist genau dann positiv (semi)definit, wenn alle Eigenwerte positiv (nicht negativ) sind.  
 (b) Sei  $A$  eine reelle symmetrische Matrix.  $A$  ist genau dann positiv definit, wenn alle Abschnittsdeterminanten  $\Delta_k = \det(A_k)$  positiv sind.

**Vorsicht:** Will man mit dem Kriterium in (b) feststellen, ob eine reelle, symmetrische Matrix *negativ* definit ist, muss man es auf  $-A$  anwenden! Das bedeutet dann eben **nicht**, dass alle Abschnittsdeterminanten negativ sind, sondern dass sie abwechselnd negativ und positiv sind:  $(-1)^k \Delta_k > 0$ .

Die Lage der Eigenwerte einer komplexen  $n \times n$ -Matrix kann man in etwa abschätzen: Angenommen  $\vec{x}$  ist ein Eigenvektor zum Eigenwert  $\lambda$  von  $A$ . Für die  $i$ -te Koordinate  $x_i$  heißt das  $\sum_k a_{ik} x_k = \lambda x_i$ . Sei nun  $x_i$  die Koordinate, für die  $|x_i|$  maximal unter allen Koordinaten von  $\vec{x}$  ist. Dann gilt  $|x_i| > 0$  und

$$(\lambda - a_{ii})x_i = \sum_{\substack{k=1 \\ k \neq i}}^n a_{ik} x_k,$$

also wenn man Beträge nimmt

$$|\lambda - a_{ii}| = \left| \sum_{\substack{k=1 \\ k \neq i}}^n \frac{a_{ik} x_k}{x_i} \right| \leq \sum_{\substack{k=1 \\ k \neq i}}^n |a_{ik}|.$$

Mit anderen Worten: zu jedem Eigenwert  $\lambda$  gibt es ein Diagonalelement  $a_{ii}$  der Matrix, so dass  $\lambda$  in der Kreisscheibe um  $a_{ii}$  mit Radius  $\sum_{k \neq i} |a_{ik}|$  liegt. Aus dieser Eigenschaft folgt nun das sogenannte Zeilensummenkriterium:

**Satz 4.10** (Zeilensummenkriterium). *Sei  $A$  eine hermitesche  $n \times n$ -Matrix. Gilt für jedes  $i$ ,  $1 \leq i \leq n$ ,*

$$a_{ii} > \sum_{\substack{k=1 \\ k \neq i}}^n |a_{ik}| \quad (\text{bzw. } a_{ii} \geq \sum_{\substack{k=1 \\ k \neq i}}^n |a_{ik}|)$$

*so ist  $A$  positiv definit (positiv semidefinit). Im ersten Fall spricht man auch vom starken, im zweiten vom schwachen Zeilensummenkriterium.*

**Beispiele 4.11.** (a) Die reelle und symmetrische (also hermitesche) Matrix

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

erfüllt das starke Zeilensummenkriterium und ist daher positiv definit. Zur (eigentlich überflüssigen) Kontrolle berechnen wir die Eigenwerte von  $A$ :

$$\chi_A(t) = \det \begin{pmatrix} 3-t & 1 & -1 \\ 1 & 2-t & 0 \\ -1 & 0 & 2-t \end{pmatrix} = (3-t)(2-t)^2 - 2(2-t) = (2-t)(t-1)(t-4)$$

mit Nullstellen 1, 2, 4, die alle positiv sind.

(b) Die ebenfalls reelle und symmetrische Matrix

$$B = \begin{pmatrix} 11 & 3 & -1 \\ 3 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

erfüllt weder das starke noch das schwache Zeilensummenkriterium, denn der Diagonaleintrag 2 der zweiten Zeile ist kleiner als die Summe der Nichtdiagonaleinträge  $3 = 3 + 0$ . Daraus können wir gar nichts schließen; der Satz sagt schlicht nichts über die Matrix  $B$  aus. Daher müssen wir



entweder die Eigenwerte oder die Abschnittdeterminanten berechnen; wir tun beides.

$$\chi_B(t) = \det \begin{pmatrix} 11 & 3 & -1 \\ 3 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix} = (11-t)(2-t)^2 - 10(2-t) = (2-t)(t-1)(t-12)$$

hat die Nullstellen 1, 2, 12. Da diese positiv sind, ist die Matrix positiv definit. Das gleiche Ergebnis liefern die Abschnittdeterminanten: es gilt

$$\Delta_1 = 11 > 0, \quad \Delta_2 = \det \begin{pmatrix} 11 & 3 \\ 3 & 2 \end{pmatrix} = 13 > 0, \quad \Delta_3 = \det(B) = 24 > 0.$$

(c) Wir berechnen die lokalen Extrema der Funktion  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  gegeben durch

$$f(x, y) = xy(1 - x - y).$$

Notwendige Bedingung ist das Verschwinden der partiellen Ableitungen:

$$0 = \frac{\partial f}{\partial x} = y(1 - 2x - y), \quad (4.2)$$

$$0 = \frac{\partial f}{\partial y} = x(1 - x - 2y). \quad (4.3)$$

Damit  $\frac{\partial f}{\partial x} = 0$  ist, muss also  $y = 0$  oder  $2x + y = 1$  gelten. Für  $y = 0$  folgt aber aus der zweiten Bedingung (4.3)  $x(1 - x) = 0$ , also  $x = 0$  oder  $x = 1$ . Wir erhalten die beiden Punkte  $\vec{p}_1 = (0, 0)$  und  $\vec{p}_2 = (1, 0)$ . Ist nun  $y \neq 0$ , so muss gelten:

$$2x + y = 1 \quad \text{und} \quad x(1 - x - 2y) = 0.$$

Hier gibt es wieder zwei Möglichkeiten:  $x = 0$  und  $2x + y = 1$ , was zum Punkt  $\vec{p}_3 = (0, 1)$  führt, oder  $x + 2y = 1$  und  $2x + y = 1$ . Dieses lineare Gleichungssystem reduzieren wir in (erweiterter) Matrixnotation per Gaußverfahren

$$\left( \begin{array}{cc|c} 2 & 1 & 1 \\ 1 & 2 & 1 \end{array} \right) \rightarrow \left( \begin{array}{cc|c} 1 & 2 & 1 \\ 0 & -3 & -1 \end{array} \right) \quad \text{mit der Lösung} \quad \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix}.$$

Als Kandidaten für Extrema haben wir also die vier Punkte

$$\vec{p}_1 = (0, 0), \quad \vec{p}_2 = (1, 0), \quad \vec{p}_3 = (0, 1), \quad \vec{p}_4 = (1/3, 1/3);$$

an diesen Stellen müssen wir die Hesse-Matrix bestimmen. Es gilt

$$\frac{\partial^2 f}{\partial x^2} = -2y, \quad \frac{\partial^2 f}{\partial x \partial y} = 1 - 2x - 2y, \quad \frac{\partial^2 f}{\partial y^2} = -2x.$$

$\text{Hesse}(f)(\vec{p}_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  ist indefinit, denn die Eigenwerte sind 1 und  $-1$ ,

$\text{Hesse}(f)(\vec{p}_2) = \begin{pmatrix} 0 & -1 \\ -1 & -2 \end{pmatrix}$  hat Eigenwerte  $-1 \pm \sqrt{2}$  und ist daher ebenfalls indefinit,

$\text{Hesse}(f)(\vec{p}_3) = \begin{pmatrix} -2 & -1 \\ -1 & 0 \end{pmatrix}$  hat die gleichen Eigenwerte wie für  $\vec{p}_2$ , ist also auch indefinit,

$\text{Hesse}(f)(\vec{p}_4) = \begin{pmatrix} -2/3 & -1/3 \\ -1/3 & -2/3 \end{pmatrix}$  ist negativ definit, denn die Eigenwerte  $-1, -1/3$  sind negativ.

Folglich hat die Funktion  $f$  in  $\vec{p}_4 = (1/3, 1/3)$  ein lokales Maximum.

Nicht nur hermitesche, sondern auch unitäre Matrizen lassen sich orthogonal diagonalisieren:

**Satz 4.12.** Sei  $A$  eine unitäre  $n \times n$ -Matrix. Dann gilt:

- (a) Jeder Eigenwert  $\lambda$  von  $A$  hat Betrag  $|\lambda| = 1$ .
- (b) Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal.
- (c) Es gibt eine unitäre Matrix  $T$ , so dass  $T^*AT$  Diagonalgestalt hat.

Eigenschaft (a) ist einfach nachzurechnen: da unitäre Matrizen das Skalarprodukt erhalten, gilt für einen Eigenvektor  $\vec{v}$  zum Eigenwert  $\lambda$

$$\langle \vec{v}, \vec{v} \rangle = \langle A\vec{v}, A\vec{v} \rangle = \langle \lambda\vec{v}, \lambda\vec{v} \rangle = (\lambda\vec{v})^*(\lambda\vec{v}) = \bar{\lambda}\lambda\vec{v}^*\vec{v} = \bar{\lambda}\lambda\langle \vec{v}, \vec{v} \rangle,$$

also  $|\lambda|^2 = \bar{\lambda}\lambda = 1$  und damit  $|\lambda| = 1$ .

(b) bedeutet: Ist  $\vec{v}$  ein Eigenvektor zum Eigenwert  $\lambda$  und  $\vec{w}$  ein Eigenvektor zum Eigenwert  $\mu \neq \lambda$ , so gilt  $\langle \vec{v}, \vec{w} \rangle = 0$ . Wäre nämlich  $\langle \vec{v}, \vec{w} \rangle \neq 0$ , so folgte aus

$$\langle \vec{v}, \vec{w} \rangle = \langle A\vec{v}, A\vec{w} \rangle = \langle \lambda\vec{v}, \mu\vec{w} \rangle = \bar{\lambda}\mu\langle \vec{v}, \vec{w} \rangle$$

zunächst  $\bar{\lambda}\mu = 1$ , durch Multiplikation mit  $\lambda$  also  $\lambda\bar{\lambda}\mu = \lambda$  und mit (a) schließlich  $\mu = \lambda$ , was ausgeschlossen war.

Teil (c) des Satzes besagt, dass es eine Orthonormalbasis bestehend aus Eigenvektoren gibt; diese bilden die Spalten der Transformationsmatrix  $T$ .

**Beispiel 4.13.** Die Matrix  $A = \frac{1}{2} \begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix}$  ist orthogonal (also unitär):

$$A^T A = \frac{1}{4} \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 3+1 & -\sqrt{3}+\sqrt{3} \\ -\sqrt{3}+\sqrt{3} & 1+3 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = E_2.$$

Wir berechnen erst die Eigenwerte: es ist

$$\chi_A(t) = \left( \frac{\sqrt{3}}{2} - t \right)^2 + \frac{1}{4} = t^2 - \sqrt{3}t + 1 \text{ mit Nullstellen } \lambda_1 = \frac{1}{2}(\sqrt{3} + j), \lambda_2 = \frac{1}{2}(\sqrt{3} - j).$$

Zu den Eigenvektoren:

$$A - \lambda_1 E_2 = \frac{1}{2} \begin{pmatrix} -j & -1 \\ 1 & -j \end{pmatrix} \text{ liefert den normierten Eigenvektor } \vec{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} j \\ 1 \end{pmatrix},$$

$$A - \lambda_2 E_2 = \frac{1}{2} \begin{pmatrix} j & -1 \\ 1 & j \end{pmatrix} \text{ liefert den normierten Eigenvektor } \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix}.$$

Aus den Eigenvektoren bilden wir die Transformationsmatrix  $T$ , indem wir sie als Spalten schreiben:

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} j & 1 \\ 1 & j \end{pmatrix}$$

Dann ist  $T$  unitär und es gilt

$$T^* A T = \frac{1}{2} \begin{pmatrix} \sqrt{3} + j & 0 \\ 0 & \sqrt{3} - j \end{pmatrix}$$

wie verlangt. (Die gegebene Matrix  $A$  beschreibt eine Drehung im  $\mathbb{R}^2$  um den Winkel  $30^\circ$ . Man beachte, dass die Eigenwerte der reellen Matrix  $A$  komplex sind, ebenso wie die Transformationsmatrix im Allgemeinen komplex sein wird.)

*Bemerkung.* Hermitesche und unitäre Matrizen sind Spezialfälle sogenannter „normaler“ Matrizen: eine Matrix  $A$  heißt *normal*, wenn sie mit ihrer konjugiert Transponierten vertauscht, also wenn  $AA^* = A^*A$  gilt. Zu solchen Matrizen gibt es stets eine Orthonormalbasis, die aus Eigenvektoren besteht, das heißt  $A$  ist orthogonal diagonalisierbar, und umgekehrt ist jede komplexe, orthogonal diagonalisierbare Matrix normal.

# II Kurven

## 1. Kurven im Raum

Anschaulich stellen wir uns Kurven im Raum als Bahnen von Massepunkten vor, die sich durch den Raum bewegen, Beschreiben lassen sich solche Bahnen als die Ortsvektoren des Massepunktes zum Zeitpunkt  $t$ .

**Definition 1.1.** Es sei  $\vec{r}: [a, b] \rightarrow \mathbb{R}^n$  eine stetig differenzierbare Funktion, also

$$\vec{r}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}, \quad t \in [a, b],$$

wobei die Koordinatenfunktionen  $x_i(t)$  stetig differenzierbare Funktionen seien. Die Menge der Bildpunkte

$$C = \vec{r}([a, b]) = \{(x_1(t), \dots, x_n(t)) \mid t \in [a, b]\}$$

nennt man eine Kurve mit Anfangspunkt  $\vec{r}(a)$  und Endpunkt  $\vec{r}(b)$ . Die Funktion  $\vec{r}$  nennt man eine Parametrisierung der Kurve  $C$ .

Im Fall von  $n = 2$  und  $n = 3$  benutzen wir auch die üblichen Koordinaten  $(x, y)$  beziehungsweise  $(x, y, z)$ .

**Beispiele 1.2.** (a) Die Abbildung

$$\vec{r}(t) = \begin{pmatrix} t \\ 1+t \\ 1-t \end{pmatrix}, \quad 0 \leq t \leq 2, \quad (1.1)$$

parametrisiert die Strecke von  $(0, 1, 1)$  nach  $(1, 3, -1)$ .

(b) Die Parametrisierung

$$\vec{r}(t) = \begin{pmatrix} R \cos(t) \\ R \sin(t) \end{pmatrix}, \quad 0 \leq t < \pi, \quad (1.2)$$

beschreibt den Halbkreis in der oberen Halbebene um  $(0, 0)$  vom Radius  $R$  (ohne den Endpunkt  $(-R, 0)$ ).

(c) Die durch

$$\vec{r}(t) = \begin{pmatrix} R \cos(\omega t) \\ R \sin(\omega t) \\ ct \end{pmatrix}, \quad 0 \leq t, \quad (1.3)$$

parametrisierte Kurve ist eine Schraubenlinie, wobei die Konstanten  $R, \omega$  und  $c$  bestimmte Eigenschaften der Kurve festlegen:  $R$  ist der Radius des Kreises, der durch Projektion auf die  $xy$ -Ebene

entsteht,  $\omega$  regelt die Geschwindigkeit, mit der die Kurve durchlaufen wird, und  $c$  wie schnell die „Höhe“ zunimmt.

Eine Kurve  $C$  besitzt im allgemeinen viele verschiedene Parametrisierungen. So wird etwa die Kurve aus Beispiel (a) ebenso durch

$$\begin{pmatrix} 2t \\ 1 + 2t \\ 1 - t^2 \end{pmatrix}, \quad 0 \leq t \leq 1, \quad (1.1')$$

beschrieben, die Kurve aus (b) auch durch

$$\begin{pmatrix} -t \\ \sqrt{1-t^2} \end{pmatrix}, \quad -1 \leq t \leq 1, \quad (1.2')$$

und die Schraubenlinie durch

$$\begin{pmatrix} R \cos\left(\frac{\omega}{c}t\right) \\ R \sin\left(\frac{\omega}{c}t\right) \\ t \end{pmatrix}, \quad 0 \leq t. \quad (1.3')$$

In der Interpretation von Kurven als Bahnen von Massenpunkten entsprechen verschiedene Parametrisierungen verschiedene Geschwindigkeiten, mit denen die Kurve durchlaufen wird. Die Geschwindigkeit ist wie üblich die Zeitableitung des Ortes, also die Ableitung nach dem Parameter, nur sind dies hier Vektoren. Dabei sind Grenzwerte und Ableitungen vektorwertiger Funktionen komponentenweise erklärt: Sei

$$\vec{r}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

die Parameterdarstellung einer Kurve in  $\mathbb{R}^n$ , so ist

$$\lim_{t \rightarrow t_0} \vec{r}(t) = \begin{pmatrix} \lim_{t \rightarrow t_0} x_1(t) \\ \lim_{t \rightarrow t_0} x_2(t) \\ \vdots \\ \lim_{t \rightarrow t_0} x_n(t) \end{pmatrix}$$

und

$$\dot{\vec{r}}(t) = \frac{d}{dt} \vec{r}(t) = \lim_{h \rightarrow 0} \frac{1}{h} [\vec{r}(t+h) - \vec{r}(t)] = \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix};$$

diesen Vektor  $\dot{\vec{r}}(t)$  interpretieren wir als Geschwindigkeitsvektor und genauso  $\ddot{\vec{r}}(t)$  als Beschleunigungsvektor.

**Beispiel 1.3.** In den obigen Beispielen 1.2 sind die Geschwindigkeitsvektoren gegeben durch

$$(a) \quad \dot{\vec{r}}(t) = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad (b) \quad \dot{\vec{r}}(t) = \begin{pmatrix} -R \sin(t) \\ R \cos(t) \end{pmatrix}, \quad (c) \quad \dot{\vec{r}}(t) = \begin{pmatrix} -R\omega \sin(\omega t) \\ R\omega \cos(\omega t) \\ c \end{pmatrix}.$$

In den Parametrisierungen (1.1')-(1.3') hingegen sind hat man

$$(a) \quad \dot{\vec{r}}(t) = \begin{pmatrix} 2 \\ 2 \\ -2 \end{pmatrix}, \quad (b) \quad \dot{\vec{r}}(t) = \begin{pmatrix} -1 \\ -t \\ \sqrt{1-t^2} \end{pmatrix}, \quad (c) \quad \dot{\vec{r}}(t) = \begin{pmatrix} -\frac{R\omega}{c} \sin\left(\frac{\omega}{c}t\right) \\ \frac{R\omega}{c} \cos\left(\frac{\omega}{c}t\right) \\ 1 \end{pmatrix}.$$

In (a) wird bei der zweiten Parametrisierung die Kurve doppelt so schnell durchlaufen wie in der ersten und bei (c) geht es um den Faktor  $1/c$  schneller. (In (b) existiert die Ableitung der zweiten Parametrisierung nur für  $-1 < t < 1$ .)

Besonders geeignet sind Parametrisierungen, bei denen man weder stehenbleibt noch umkehrt:

**Definition 1.4.** Eine Parametrisierung  $\vec{r}(t)$ ,  $a \leq t \leq b$ , heißt regulär, wenn für alle  $t \in [a, b]$  gilt:

$$\dot{\vec{r}}(t) \neq \vec{0}.$$

Die Parametrisierungen in den obigen Beispielen sind alle regulär.

**Satz 1.5.** Sei  $\vec{r}(t)$ ,  $a \leq t \leq b$ , eine reguläre Parametrisierung. Dann ist der Vektor  $\dot{\vec{r}}(t)$  parallel zur Kurventangente im Punkt  $\vec{r}(t)$ .

Das liegt daran, wie man die Ableitung bildet, nämlich durch Grenzwertbildung:

$$\lim_{h \rightarrow 0} \frac{1}{h} [\vec{r}(t+h) - \vec{r}(t)]$$

Hierbei ist  $\vec{r}(t+h) - \vec{r}(t)$  die Sekante zwischen den Punkten  $\vec{r}(t+h)$  und  $\vec{r}(t)$ ; im Grenzübergang wird daraus eine Tangente.

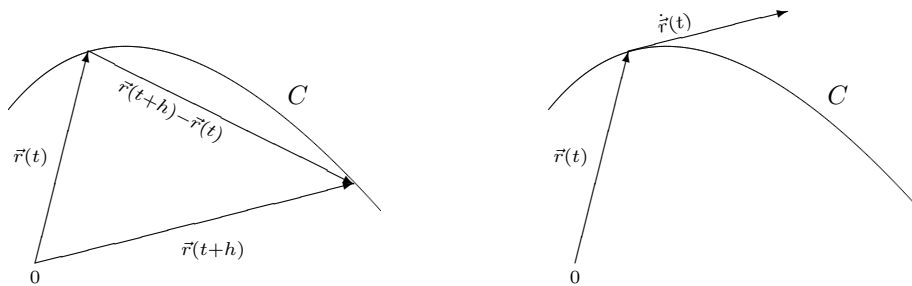


ABBILDUNG 1. Sekante und Tangente

Darauf basierend definiert man:

**Definition 1.6.** Sei  $\vec{r}(t)$ ,  $a \leq t \leq b$ , eine reguläre Parametrisierung einer Kurve  $C$ .

(a) Dann heißt

$$\vec{T}(t) := \frac{1}{\|\dot{\vec{r}}(t)\|} \dot{\vec{r}}(t)$$

der Tangenteneinheitsvektor im Punkt  $\vec{r}(t)$ .

(b) Ist zusätzlich  $\vec{T}(t) \neq \vec{0}$ , so heißt

$$\vec{N}(t) := \frac{1}{\|\dot{\vec{T}}(t)\|} \dot{\vec{T}}(t)$$

der Hauptnormaleneinheitsvektor im Punkt  $\vec{r}(t)$ .

Der Name „Hauptnormaleneinheitsvektor“ gründet sich darauf, dass  $\vec{N}(t)$  senkrecht auf  $\vec{T}(t)$  steht: Da  $\vec{T}(t)$  die Länge 1 hat, erhält man durch Differenzieren

$$0 = \frac{d}{dt}(1) = \frac{d}{dt} \langle \vec{T}(t), \vec{T}(t) \rangle = \langle \dot{\vec{T}}(t), \vec{T}(t) \rangle + \langle \vec{T}(t), \dot{\vec{T}}(t) \rangle = 2 \langle \dot{\vec{T}}(t), \vec{T}(t) \rangle \quad (1.4)$$

Der Vektor  $\vec{N}(t)$  zeigt in die (Haupt)krümmungsrichtung der Kurve (s.u.).

**Beispiel 1.7.** Wir betrachten die Schraubenlinie

$$\vec{r}(t) = \begin{pmatrix} t \\ 2 \cos(t) \\ 2 \sin(t) \end{pmatrix}, \quad 0 \leq t.$$

Dann ist

$$\begin{aligned} \dot{\vec{r}}(t) &= \begin{pmatrix} 1 \\ -2 \sin(t) \\ 2 \cos(t) \end{pmatrix}, \quad \|\dot{\vec{r}}(t)\| = \sqrt{1 + 4 \sin^2(t) + 4 \cos^2(t)} = \sqrt{1 + 4} = \sqrt{5}, \\ \vec{T}(t) &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \sin(t) \\ 2 \cos(t) \end{pmatrix}, \\ \vec{N}(t) &= \frac{1}{2} \begin{pmatrix} 0 \\ -2 \cos(t) \\ -2 \sin(t) \end{pmatrix}. \end{aligned}$$

## 2. Bogenlänge

Sei  $C$  eine Kurve in der Ebene mit regulärer Parametrisierung  $\vec{r}: [a, b] \rightarrow \mathbb{R}^2$ . Um die Länge der Kurve zu berechnen, approximieren wir sie durch Sekanten. Dazu teilen wir das Intervall in  $n$  gleichlange Teilintervalle  $[t_{i-1}, t_i]$ ,  $1 \leq i \leq n$ , mit  $t_0 = a$ ,  $t_n = b$  und  $t_i - t_{i-1} = \frac{b-a}{n} =: \Delta t$ :

$$\begin{array}{ccccccccccc} | & & | & & | & & | & & | & & | \\ a = t_0 & t_1 & t_2 & & & & & & & & t_{n-1} & t_n = b \end{array}$$

Die Strecken zwischen den Punkten  $\vec{r}(t_{i-1})$  und  $\vec{r}(t_i)$ , so wie in der Skizze, approximieren die Kurve, und zwar um so besser, je mehr Unterteilungspunkte man nimmt, also je größer  $n$  ist.

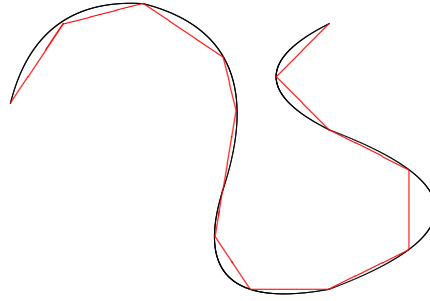


ABBILDUNG 2. Approximation einer Kurve durch Geradenstücke

Dann beträgt die Länge der Strecke von  $\vec{r}(t_{i-1})$  nach  $\vec{r}(t_i)$  gerade

$$\Delta r_i = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2}, \text{ wobei } \begin{cases} \Delta x_i = x(t_i) - x(t_{i-1}), \\ \Delta y_i = y(t_i) - y(t_{i-1}). \end{cases}$$

Nun lassen sich Sekanten zwischen zwei Punkten auch als Tangenten an einer Zwischenstelle beschreiben (Mittelwertsatz), das heißt es gibt Parameterwerte  $\xi_i, \eta_i \in [t_{i-1}, t_i]$  mit

$$\Delta x_i = \dot{x}(\xi_i)\Delta t, \quad \Delta y_i = \dot{y}(\eta_i)\Delta t$$

und daher ist

$$\Delta r_i = \sqrt{\dot{x}(\xi_i)^2 + \dot{y}(\eta_i)^2} \Delta t$$

Summiert man alle diese Längen auf, erhält man als Approximation an die Länge der Kurve

$$L \approx \sum_{i=1}^n \sqrt{\dot{x}(\xi_i)^2 + \dot{y}(\eta_i)^2} \Delta t.$$

Lässt man nun  $n$  wachsen, also  $\Delta t$  schrumpfen, so wird daraus im Grenzübergang  $\Delta t \rightarrow 0$  ein Integral:

$$L = \int_a^b \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2} dt.$$

Wir fassen diese Überlegungen in einem Satz zusammen:

**Satz 2.1.** Sei  $\vec{r}: [a, b] \rightarrow \mathbb{R}^2$ ,  $\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$ , eine reguläre Parameterdarstellung der ebenen Kurve  $C$ . Die Länge von  $C$  beträgt

$$L(C) = \int_a^b \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2} dt = \int_a^b \|\dot{\vec{r}}(t)\| dt.$$

**Spezialfall 2.2.** Sei  $f: [a, b] \rightarrow \mathbb{R}$  eine stetig differenzierbare Funktion. Der Graph  $\Gamma_f$  von  $f$  ist eine ebene Kurve mit Parameterdarstellung

$$\vec{r}(t) = \begin{pmatrix} t \\ f(t) \end{pmatrix}, \quad a \leq t \leq b.$$

Es folgt

$$\dot{\vec{r}}(t) = \begin{pmatrix} 1 \\ f'(t) \end{pmatrix}, \text{ also } \|\dot{\vec{r}}(t)\| = \sqrt{1 + f'(t)^2}$$

und die Länge des Graphen ist

$$L = \int_a^b \sqrt{1 + f'(t)^2} dt.$$

**Beispiele 2.3.** (a) Die durch  $\vec{r}(t) = \begin{pmatrix} t - \sin(t) \\ 1 - \cos(t) \end{pmatrix}$  parametrisierte Kurve nennt man eine *Zykloide*. (Diese Kurve löst das sogenannte *Brachistochronenproblem*, also die Frage, wie die Bahn aussieht, auf der eine Kugel am schnellsten zu Tale rollt.) Es ist

$$\dot{\vec{r}}(t) = (1 - \cos(t), \sin(t))$$

$$\|\dot{\vec{r}}(t)\|^2 = 1 - 2\cos(t) + \cos^2(t) + \sin^2(t) = 2 - 2\cos(t) = 4\sin^2(t/2),$$

wobei wir  $\sin^2(t) + \cos^2(t) = 1$  sowie die aus den Additionstheoremen abgeleitete Formel

$$2\sin^2(x) = 1 - \cos(2x)$$

benutzt haben, also

$$\|\dot{\vec{r}}(t)\| = |2\sin(t/2)|$$

und daher

$$L = 2 \int_0^{2\pi} |\sin(t/2)| dt = 2 \int_0^{2\pi} \sin(t/2) dt = 4[-\cos(t/2)]_0^{2\pi} = 8.$$

(b) Sei  $a$  eine natürliche Zahl  $\geq 2$  und  $C$  die durch

$$\vec{r}(t) = \begin{pmatrix} a \sin(t) - \sin(at) \\ a \cos(t) - \cos(at) \end{pmatrix}, \quad 0 \leq t \leq 2\pi,$$

parametrisierte Kurve. Dann gilt

$$\dot{\vec{r}}(t) = \begin{pmatrix} a \cos(t) - a \cos(at) \\ -a \sin(t) + a \sin(at) \end{pmatrix}$$

und

$$\begin{aligned} \|\dot{\vec{r}}(t)\|^2 &= a^2(\cos^2(t) - 2\cos(t)\cos(at) + \cos^2(at) + \sin^2(t) - 2\sin(t)\sin(at) + \sin^2(at)) \\ &= 2a^2 \left[ 1 - \cos(t) \left( \cos((a-1)t)\cos(t) - \sin((a-1)t)\sin(t) \right) \right. \\ &\quad \left. + \sin(t) \left( \sin((a-1)t)\cos(t) + \cos((a-1)t)\sin(t) \right) \right] \\ &= 2a^2 \left[ 1 - \cos^2(t)\cos((a-1)t) - \sin^2(t)\cos((a-1)t) \right] \\ &= 2a^2 \left[ 1 - \cos\left(\frac{a-1}{2}t\right) \right] \\ &= 4a^2 \sin^2\left(\frac{a-1}{2}t\right) \end{aligned}$$

Es folgt

$$L = \int_0^{2\pi} \|\dot{\vec{r}}(t)\| dt = 2a \int_0^{2\pi} \left| \sin\left(\frac{a-1}{2}t\right) \right| dt = \frac{4a}{a-1} \int_0^{(a-1)\pi} |\sin(u)| du = 4a \int_0^\pi \sin(u) du = 8a.$$

(c) Sei  $C$  der Graph der Funktion  $y = \sqrt{x^3} = x^{3/2}$ ,  $0 \leq x \leq 1$ . Gemäß der Formel für Funktionsgraphen gilt

$$L = \int_0^1 \sqrt{1 + (y')^2} dx = \int_0^1 \sqrt{1 + \frac{9}{4}x} dx = \frac{1}{2} \int_0^1 \sqrt{4 + 9x} dx;$$

mit der Substitution  $u = 4 + 9x$ , also  $du = 9dx$ , wird daraus

$$= \frac{1}{18} \int_4^{13} \sqrt{u} du = \frac{1}{18} \left[ \frac{2}{3} u^{3/2} \right]_4^{13} = \frac{1}{27} (13\sqrt{13} - 8).$$

(d) Wir betrachten die Spirale  $\vec{r}(t) = \begin{pmatrix} t \sin(t) \\ t \cos(t) \end{pmatrix}$ ,  $0 \leq t \leq 2k\pi$ , also  $k$  volle Umdrehungen. Es ist

$$\dot{\vec{r}}(t) = \begin{pmatrix} \sin(t) + t \cos(t) \\ \cos(t) - t \sin(t) \end{pmatrix}$$



und

$$\|\dot{\vec{r}}(t)\|^2 = \sin^2(t) + 2t \sin(t) \cos(t) + t^2 \cos^2(t) - 2t \sin(t) \cos(t) + t^2 \sin^2(t) = 1 + t^2;$$

zu berechnen ist also das Integral  $\int_0^{2k\pi} \sqrt{1+t^2} dt$ . Eine Stammfunktion von  $\sqrt{1+t^2}$  (die man mit Hilfe der Substitution  $t = \sinh(u)$  findet) ist

$$\frac{1}{2}t\sqrt{1+t^2} + \frac{1}{2}\ln(t + \sqrt{1+t^2}),$$

so dass man für die Länge

$$\begin{aligned} L &= \frac{1}{2} \left[ \sqrt{1+t^2} + \ln(t + \sqrt{1+t^2}) \right]_0^{2k\pi} \\ &= k\pi \sqrt{1+4k^2\pi^2} + \frac{1}{2} \ln(2k\pi + \sqrt{1+4k^2\pi^2}) \end{aligned}$$

erhält.

Man kann die Bogenlänge auch als Parameter benutzen: Ist nämlich  $\vec{r}: [a, b] \rightarrow \mathbb{R}^2$  eine reguläre Parameterdarstellung, so definiert

$$L(t) := \int_0^t \|\dot{\vec{r}}(u)\| du$$

eine monoton wachsende Funktion mit  $L(a) = 0$ , die das Intervall  $[a, b]$  umkehrbar eindeutig auf das Intervall  $[0, L(b)]$  abbildet und daher eine Umkehrfunktion  $L^{-1}: [0, L(b)] \rightarrow [a, b]$  besitzt. Definiere nun

$$\vec{c}(s) := \vec{r}(L^{-1}(s)), \quad 0 \leq s \leq L(b).$$

Diese Parameterdarstellung hat dann die Eigenschaft, dass die Kurve zum Zeitpunkt  $s$  einen Weg der Länge  $s$  zurückgelegt hat. Mit der Kettenregel rechnet man aus

$$\dot{\vec{c}}(s) = \frac{d}{ds} \vec{c}(s) = \frac{d}{dt} \vec{r}(L^{-1}(s)) \cdot \left( \frac{d}{ds} L^{-1}(s) \right) = \dot{\vec{r}}(t) \cdot \left( \frac{d}{dt} L(t) \right)^{-1} = \frac{\dot{\vec{r}}(t)}{\|\dot{\vec{r}}(t)\|}$$

mit  $s = L(t)$ . Folglich gilt  $\|\dot{\vec{c}}(s)\| = 1$  und daher  $\vec{T}(s) = \dot{\vec{c}}(s)$ .

**Beispiel 2.4.** Eine Parametrisierung des Kreises um Null mit Radius  $R$  nach der Bogenlänge ist gegeben durch

$$\vec{c}(s) = R \begin{pmatrix} \cos(s/R) \\ \sin(s/R) \end{pmatrix}, \quad 0 \leq s \leq 2\pi R.$$

Die gleichen Überlegungen zur Bogenlänge kann man auch für Raumkurven anstellen:

**Satz 2.5.** Sei  $\vec{r}: [a, b] \rightarrow \mathbb{R}^3$ ,  $\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$ , eine reguläre Parameterdarstellung der Kurve  $C$ .  
Die Länge von  $C$  beträgt

$$L(C) = \int_a^b \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2} dt = \int_a^b \|\dot{\vec{r}}(t)\|^2 dt.$$

**Beispiel 2.6.** Sei  $\vec{r}(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \\ t \end{pmatrix}$ ,  $0 \leq t \leq 2k\pi$ , eine Schraubenlinie mit  $2k$  Umläufen. Es ist

$$\|\dot{\vec{r}}(t)\|^2 = (-\sin(t))^2 + \cos^2(t) + 1 = 2,$$

also

$$L = \int_0^{2k\pi} \sqrt{2} dt = 2k\pi\sqrt{2}.$$

Auch Raumkurven kann man natürlich nach der Bogenlänge parametrisieren:

**Beispiel 2.7.** Sei  $\vec{r}(t) = \begin{pmatrix} 3t \\ 4 \cos(t) \\ 4 \sin(t) \end{pmatrix}$ ,  $0 \leq t$ , eine Schraubenlinie mit Anfangspunkt  $(0, 0, 0)$ . Dann

gilt

$$\|\dot{\vec{r}}(t)\|^2 = 3^2 + (-4 \sin(t))^2 + (4 \cos(t))^2 = 9 + 16(\sin^2(t) + \cos^2(t)) = 25,$$

also

$$s = s(t) = \int_0^t \|\dot{\vec{r}}(u)\| du = \int_0^t 5 du = 5t$$

und damit  $t = t(s) = \frac{s}{5}$ . Es folgt

$$\vec{c}(s) = \vec{r}(t(s)) = \begin{pmatrix} 3s/5 \\ 4 \cos(s/5) \\ 4 \sin(s/5) \end{pmatrix}.$$

### 3. Krümmung

Wie stark eine Kurve gekrümmt ist, drückt sich dadurch aus, wie schnell sich ihre Richtung verändert, wenn man die Kurve durchläuft. Bei ebenen Kurven kann man dies durch die relative Änderung des Neigungswinkels der Tangente bezüglich der Bogenlänge messen:

**Definition 3.1.** Sei  $C$  eine ebene, reguläre Kurve,  $\varphi$  der Neigungswinkel der Tangente und  $s$  die Bogenlänge. Dann nennt man

$$\kappa = \frac{d\varphi}{ds} = \lim_{\Delta s \rightarrow 0} \frac{\Delta\varphi}{\Delta s}$$

die Krümmung der Kurve (sofern der Grenzwert existiert).

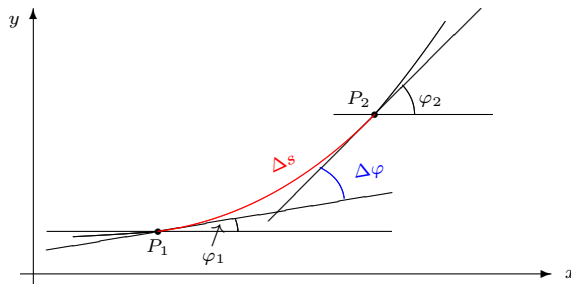


ABBILDUNG 3. Definition der Krümmung für ebene Kurven

Die Krümmung einer ebenen Kurve ist also eine reelle Zahl; ihr Vorzeichen drückt dabei aus, in welche Richtung sich die Kurve krümmt: ist  $\kappa < 0$ , so nimmt der Neigungswinkel ab, und man spricht von Rechtskrümmung (denn die Kurve „biegt rechts ab“), genauso nennt man die Kurve linksgekrümmt in einem Punkt, wenn dort  $\kappa > 0$  ist.

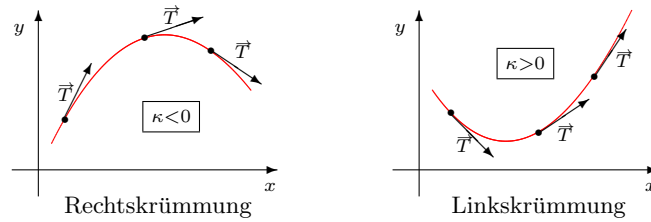


ABBILDUNG 4. Rechts- und Linkskrümmung ebener Kurven

Für Raumkurven ist die Unterscheidung nach Rechts- und Linkskrümmung nicht mehr möglich, genausowenig kann man von einem Neigungswinkel der Tangente sprechen, da es keine ausgezeichnete Referenzgerade gibt. Als Maß für die Krümmung benutzt man statt dessen die relative Änderung des Tangenteneinheitsvektors  $\vec{T}$  bezüglich der Bogenlänge  $s$ : In einer engen Kurve ist die Richtungsänderung pro zurückgelegte Wegstrecke groß, ist die Kurve jedoch beinahe gerade, so ändert sich die Richtung kaum. Diese geometrische Vorstellung führt zu folgender Definition:

**Definition 3.2.** Sei  $n > 2$  und  $C \subset \mathbb{R}^n$  eine reguläre Kurve,  $\vec{T}$  ihr Tangenteneinheitsvektor und  $s$  die Bogenlänge. Dann heißt

$$\kappa = \left\| \frac{d\vec{T}}{ds} \right\|$$

die Krümmung der Kurve.

Hier ist  $\kappa$  also stets eine nichtnegative reelle Zahl.

*Bemerkung.* (a) Im Fall einer ebenen Kurve entspricht, bis auf ein Vorzeichen, die Änderung des Neigungswinkels  $\varphi$  der Tangente der Richtungsänderung des Tangenteneinheitsvektors: für kleine Wegstrecken  $\Delta s$  gilt

$$\Delta\varphi \approx \pm \|\Delta\vec{T}\| \quad \text{mit Vorzeichen} \quad \begin{cases} \text{„+“ für Linkskrümmung,} \\ \text{„-“ für Rechtskrümmung.} \end{cases}$$

Daher gilt für eine ebene Kurve  $C$

$$\left\| \frac{d\vec{T}}{ds} \right\| = \begin{cases} \kappa, & \text{falls } C \text{ im Punkt } (x, y) \text{ linksgekrümmt ist,} \\ -\kappa, & \text{falls } C \text{ im Punkt } (x, y) \text{ rechtsgekrümmt ist,} \end{cases}$$

und Definition 3.2 ist die richtige Verallgemeinerung von Definition 3.1 auf höhere Dimensionen.

(b) Ist  $\vec{c}: [a, b] \rightarrow \mathbb{R}^n$  eine Parametrisierung nach der Bogenlänge, so erhält man  $\kappa$  auch als zweite Ableitung:

$$\left\| \frac{d\vec{T}}{ds} \right\| = \|\ddot{\vec{c}}(s)\|.$$

Diese Definitionen sind zwar geometrisch anschaulich, aber rechnerisch unangenehm, weil die Kurve meist nicht nach der Bogenlänge parametrisiert ist. Aber die Kettenregel verschafft Abhilfe: Sei  $\vec{r}(t)$ ,  $a \leq t \leq b$ , eine (beliebige) Parameterdarstellung der Kurve  $C$ . Es ist

$$\frac{d\vec{T}}{dt} = \frac{d\vec{T}}{ds} \frac{ds}{dt}, \quad \text{also} \quad \frac{d\vec{T}}{ds} = \frac{dt}{ds} \dot{\vec{T}} = \frac{\dot{\vec{T}}}{\frac{ds}{dt}} = \frac{\dot{\vec{T}}}{\|\dot{\vec{r}}\|}.$$

Folglich gilt:

$$\mathbf{3.3} \quad \left\| \frac{d\vec{T}}{ds} \right\| = \frac{\|\dot{\vec{T}}(t)\|}{\|\dot{\vec{r}}(t)\|}.$$

Damit kann man schon ein wenig besser rechnen.

**Beispiel 3.4.** Sei  $\vec{r}(t) = \begin{pmatrix} R \cos(t) \\ R \sin(t) \end{pmatrix}$ ,  $0 \leq t \leq 2\pi$ , die Parameterdarstellung eines Kreises vom Radius  $R$ . Dann gilt

$$\dot{\vec{r}}(t) = \begin{pmatrix} -R \sin(t) \\ R \cos(t) \end{pmatrix}, \quad \|\dot{\vec{r}}(t)\| = R, \quad \vec{T}(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}, \quad \dot{\vec{T}}(t) = \begin{pmatrix} -\cos(t) \\ -\sin(t) \end{pmatrix},$$

also

$$|\kappa| = |\kappa(t)| = \frac{\|\dot{\vec{T}}(t)\|}{\|\dot{\vec{r}}(t)\|} = \frac{1}{R}.$$

Benutzen wir die Parametrisierung nach der Bogenlänge  $\vec{c}(s) = R \begin{pmatrix} \cos(s/R) \\ \sin(s/R) \end{pmatrix}$ ,  $0 \leq s \leq 2\pi R$ ,

(siehe Beispiel 2.4), so erhalten wir  $\ddot{\vec{c}}(s) = \frac{1}{R} \begin{pmatrix} -\sin(s/R) \\ -\cos(s/R) \end{pmatrix}$  und damit ebenfalls  $\kappa = 1/R$ . (Da der Kreis linksgekrümmt ist, gilt  $\kappa = |\kappa|$ .)

**Beispiel 3.5.** Die Parabel  $y = x^2$  kann man durch  $\vec{r}(t) = \begin{pmatrix} t \\ t^2 \end{pmatrix}$  parametrisieren. Es gilt

$$\dot{\vec{r}}(t) = \begin{pmatrix} 1 \\ 2t \end{pmatrix} \text{ mit } \|\dot{\vec{r}}(t)\| = \sqrt{1+4t^2}, \text{ also } \vec{T}(t) = \frac{1}{\sqrt{1+4t^2}} \begin{pmatrix} 1 \\ 2t \end{pmatrix}$$

und

$$\dot{\vec{T}}(t) = \frac{-4t}{(1+4t^2)^{3/2}} \begin{pmatrix} 1 \\ 2t \end{pmatrix} + \frac{1}{\sqrt{1+4t^2}} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \frac{2}{(1+4t^2)^{3/2}} \begin{pmatrix} -2t \\ 1 \end{pmatrix}$$

und damit

$$|\kappa(t)| = \frac{\|\dot{\vec{T}}(t)\|}{\|\dot{\vec{r}}(t)\|} = \frac{2}{(1+4t^2)^{3/2}} \cdot \frac{1}{\sqrt{1+4t^2}} \cdot \sqrt{4t^2+1} = \frac{2}{(1+4t^2)^{3/2}}.$$

Diese Rechnung erscheint unnötig umständlich, in ihr steckt immer noch viel Redundanz. Für ebene Kurven und Raumkurven gibt es auch einfachere Berechnungsformeln:

**Satz 3.6.** Sei  $\vec{r}: [a, b] \rightarrow \mathbb{R}^n$  eine reguläre Parameterdarstellung.

(a) Sei  $n = 2$  und  $\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$ . Dann gilt

$$\kappa = \kappa(t) = \frac{\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}} = \frac{\det(\dot{\vec{r}}(t), \ddot{\vec{r}}(t))}{\|\dot{\vec{r}}(t)\|^3}.$$

(b) Sei  $n = 3$ . Dann gilt

$$\kappa = \kappa(t) = \frac{\|\dot{\vec{r}}(t) \times \ddot{\vec{r}}(t)\|}{\|\dot{\vec{r}}(t)\|^3}.$$

In der Tat gilt wegen  $\dot{\vec{r}}(t) = \|\dot{\vec{r}}(t)\| \vec{T}(t)$

$$\ddot{\vec{r}}(t) = \left( \frac{d}{dt} \|\dot{\vec{r}}(t)\| \right) \vec{T}(t) + \|\dot{\vec{r}}(t)\| \dot{\vec{T}}(t),$$

also

$$\dot{\vec{r}}(t) \times \ddot{\vec{r}}(t) = \|\dot{\vec{r}}(t)\| \vec{T}(t) \times \left( \left( \frac{d}{dt} \|\dot{\vec{r}}(t)\| \right) \vec{T}(t) + \|\dot{\vec{r}}(t)\| \dot{\vec{T}}(t) \right).$$

Nun ist  $\vec{T}(t) \times \vec{T}(t) = \vec{0}$ , und weil  $\dot{\vec{T}}(t)$  auf  $\vec{T}(t)$  senkrecht steht (siehe(1.4)), gilt

$$\|\vec{T}(t) \times \dot{\vec{T}}(t)\| = \underbrace{\|\vec{T}(t)\|}_{=1} \cdot \|\dot{\vec{T}}(t)\| = \|\dot{\vec{T}}(t)\|.$$

Es folgt

$$\|\dot{\vec{r}}(t) \times \ddot{\vec{r}}(t)\| = \|\dot{\vec{r}}(t)\|^2 \|\vec{T}(t) \times \dot{\vec{T}}(t)\| = \|\dot{\vec{r}}(t)\|^2 \|\dot{\vec{T}}(t)\|,$$

also

$$\kappa(t) = \frac{\|\dot{\vec{T}}(t)\|}{\|\dot{\vec{r}}(t)\|} = \frac{\|\dot{\vec{r}}(t) \times \ddot{\vec{r}}(t)\|}{\|\dot{\vec{r}}(t)\|^3}$$

und die Aussage (b). Teil (a) des Satzes folgt aus Teil (b), wenn man die ebene Kurve mittels

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ 0 \end{pmatrix}$$

als Raumkurve auffasst (dann ist nämlich  $\dot{\vec{r}}(t) \times \ddot{\vec{r}}(t) = (\dot{x}(t)\dot{y}(t) - \ddot{x}(t)\dot{y}(t)) \vec{e}_3$ ) und das Vorzeichen berücksichtigt.

**Beispiele 3.7.** (a) Sei  $C$  die durch  $\vec{r}(t) = \begin{pmatrix} t \cos(t) \\ t \sin(t) \end{pmatrix}$  gegebene Spirale. Dann ist  $\|\dot{\vec{r}}(t)\| = \sqrt{1+t^2}$  (siehe Beispiel 2.3 (d)) und

$$\begin{aligned} \det(\dot{\vec{r}}(t), \ddot{\vec{r}}(t)) &= \det \begin{pmatrix} \cos(t) - t \sin(t) & -2 \sin(t) - t \cos(t) \\ \sin(t) + t \cos(t) & 2 \cos(t) - t \sin(t) \end{pmatrix} \\ &= 2 \cos^2(t) - 3t \sin(t) \cos(t) + t^2 \sin^2(t) + 2 \sin^2(t) + 3t \sin(t) \cos(t) + t^2 \cos^2(t) \\ &= t^2 + 2, \end{aligned}$$

also

$$\kappa(t) = \frac{t^2 + 2}{(1+t^2)^{3/2}}.$$

(b) Wir betrachten die Zykloide

$$\vec{r}(t) = \begin{pmatrix} t - \sin(t) \\ 1 - \cos(t) \end{pmatrix} \quad \text{mit} \quad \begin{array}{ll} \dot{x}(t) = 1 - \cos(t), & \ddot{x}(t) = \sin(t), \\ \dot{y}(t) = \sin(t), & \ddot{y}(t) = \cos(t). \end{array}$$

Es folgt

$$\begin{aligned} \kappa(t) &= \frac{(1 - \cos(t)) \cos(t) - \sin(t) \sin(t)}{(1 - 2 \cos(t) + \cos^2(t) + \sin^2(t))^{3/2}} = \frac{\cos(t) - \cos^2(t) - \sin^2(t)}{(2 - 2 \cos(t))^{3/2}} \\ &= \frac{\cos(t) - 1}{(2(1 - \cos(t)))^{3/2}} = \frac{-2 \sin^2(t/2)}{(4 \sin^2(t/2))^{3/2}} = \frac{-2 \sin^2(t/2)}{(2|\sin(t/2)|)^3} = -\frac{1}{4|\sin(t/2)|}. \end{aligned}$$

(c) Für die Schraubenlinie

$$\vec{r}(t) = \begin{pmatrix} 4 \cos(t) \\ 4 \sin(t) \\ 3t \end{pmatrix}, \quad 0 \leq t,$$

gilt

$$\dot{\vec{r}}(t) \times \ddot{\vec{r}}(t) = \begin{pmatrix} -4 \sin(t) \\ 4 \cos(t) \\ 3 \end{pmatrix} \times \begin{pmatrix} -4 \cos(t) \\ -4 \sin(t) \\ 0 \end{pmatrix} = \begin{pmatrix} 12 \sin(t) \\ -12 \cos(t) \\ 16 \sin^2(t) + 16 \cos^2(t) \end{pmatrix} = \begin{pmatrix} 12 \sin(t) \\ -12 \cos(t) \\ 16 \end{pmatrix}$$

und daher

$$\kappa(t) = \frac{\sqrt{144 \sin^2(t) + 144 \cos^2(t) + 256}}{(16 \sin^2(t) + 16 \cos^2(t) + 9)^{3/2}} = \frac{\sqrt{400}}{125} = \frac{4}{25}.$$

**Spezialfall 3.8** (Graph einer Funktion). Sei  $C \subset \mathbb{R}^2$  der Graph einer (zweimal stetig differenzierbaren) Funktion  $f: [a, b] \rightarrow \mathbb{R}$ . Dann ist  $\begin{pmatrix} x \\ f(x) \end{pmatrix}$ ,  $a \leq x \leq b$ , eine Parametrisierung von  $C$ , und man erhält

$$\kappa = \frac{1 \cdot f''(x) - 0 \cdot f'(x)}{(1 + f'(x)^2)^{3/2}} = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}} = \frac{y''}{(1 + y'^2)^{3/2}}.$$

Insbesondere gilt

$$\begin{aligned} \kappa(x) > 0 &\Leftrightarrow f''(x) > 0 \quad (\text{Linkskrümmung}) \\ \kappa(x) < 0 &\Leftrightarrow f''(x) < 0 \quad (\text{Rechtskrümmung}) \end{aligned}$$

**Beispiele 3.9.** (a) Sei  $y = x^2$ . Mit der Formel aus 3.8 erhält man sofort  $\kappa = \frac{2}{(1 + 4x^2)^{3/2}}$ .

(b) Sei  $y = \cos(x)$ . Dann ist  $y' = -\sin(x)$  und  $y'' = -\cos(x)$ , also  $\kappa = \frac{-\cos(x)}{(1 + \sin^2(x))^{3/2}}$ .

**Spezialfall 3.10.** (Polarkoordinaten) Die Kurve  $C$  sei in Polarkoordinaten gegeben, also

$$\vec{r}(\varphi) = \begin{pmatrix} R(\varphi) \cos(\varphi) \\ R(\varphi) \sin(\varphi) \end{pmatrix}.$$

Dann erhält man mit der Formel aus dem Satz

$$\kappa(\varphi) = \frac{R(\varphi)^2 + 2R'(\varphi)^2 - R(\varphi)R''(\varphi)}{(R'(\varphi)^2 + R(\varphi)^2)^{3/2}}.$$

**Beispiel 3.11.** Wir wiederholen das Beispiel 3.7(a). In Polarkoordinaten hat die Spirale die Darstellung  $R(\varphi) = \varphi$ ; es folgt  $\kappa(\varphi) = \frac{\varphi^2 + 2}{(1 + \varphi^2)^{3/2}}$ .

#### 4. Evolute und Evolvente

**Definition 4.1.** Sei  $C$  eine hinreichend glatte ebene Kurve. Der Kreis, der die Kurve im Punkt  $P$  berührt und dort dieselbe Tangente und Krümmung aufweist, heißt Krümmungs- oder Schmiegkreis im Punkt  $P$ .

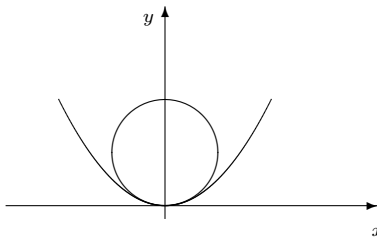


ABBILDUNG 5. Krümmungskreis

Der Radius des Krümmungskreises ist  $\frac{1}{|\kappa|}$ , da die Krümmungen übereinstimmen sollen. Um den Mittelpunkt  $(x_M, y_M)$  des Krümmungskreises zum Punkt  $P(x, y)$  zu bestimmen, geht man von  $P$

um  $1/\kappa$  in Normalenrichtung, also

$$\begin{pmatrix} x_M \\ y_M \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{\kappa} \cdot \frac{\vec{n}}{\|\vec{n}\|}$$

mit

$$\kappa = \frac{\dot{x}\ddot{y} - \ddot{x}\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}, \quad \vec{n} = \begin{pmatrix} -\dot{y} \\ \dot{x} \end{pmatrix}, \quad \|\vec{n}\| = \sqrt{\dot{x}^2 + \dot{y}^2}.$$

Dann gilt  $\frac{1}{\kappa \cdot \|\vec{n}\|} = \frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \ddot{x}\dot{y}}$ , also

$$x_M = x - \frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \ddot{x}\dot{y}} \dot{y}, \quad y_M = y + \frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \ddot{x}\dot{y}} \dot{x}. \quad (4.1)$$

Im Spezialfall eines Graphen einer Funktion  $y = f(x)$  erhält man

$$x_M = x - \frac{1 + y'^2}{y''} y', \quad y_M = y + \frac{1 + y'^2}{y''}. \quad (4.2)$$

**Beispiele 4.2.** (a) Sei  $C$  der Graph der Funktion  $y = f(x) = x^2$ . Im Scheitelpunkt  $(0, 0)$  der Parabel gilt  $x = y = y' = 0$  und  $y'' = 2$ , also

$$x_M = 0, \quad y_M = \frac{1}{2}$$

wie erwartet.

(b) Wir betrachten die Zykloide

$$\vec{r}(t) = R \begin{pmatrix} t - \sin(t) \\ 1 - \cos(t) \end{pmatrix}.$$

Dann gilt

$$\begin{aligned} \frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \ddot{x}\dot{y}} &= \frac{R^2(1 - \cos(t))^2 + R^2 \sin^2(t)}{R^2(1 - \cos(t))\cos(t) - R^2 \sin^2(t)} \\ &= \frac{1 - 2\cos(t) + \cos^2(t) + \sin^2(t)}{\cos(t) - \cos^2(t) - \sin^2(t)} = \frac{2 - 2\cos(t)}{\cos(t) - 1} = -2 \end{aligned}$$

und

$$\begin{aligned} x_M &= R(t - \sin(t)) - (-2) \cdot R \sin(t) = R(t + \sin(t)) \\ y_M &= R(1 - \cos(t)) + (-2) \cdot R(1 - \cos(t)) = -R(1 - \cos(t)) \end{aligned}$$

sind die Koordinaten des Krümmungskreismittelpunkts zum Punkt  $(x, y)$ .

**Definition 4.3.** Die von den Krümmungskreismittelpunkten einer Kurve  $C$  gebildete Kurve heißt die *Evolute* von  $C$ ; die Kurve  $C$  selbst heißt dann die *Evolvente* (oder auch *Involute*) der betreffenden Evolute.

**Beispiel 4.4.** Wie oben ausgerechnet ist die durch

$$R \begin{pmatrix} t + \sin(t) \\ -1 - \cos(t) \end{pmatrix}$$

parametrisierte Kurve die Evolute der Zykloide

$$R \begin{pmatrix} t - \sin(t) \\ 1 - \cos(t) \end{pmatrix}.$$

Verschiebt man die Evolute nun um  $\pi R$  nach links und um  $2R$  nach oben, ersetzt also  $x$  durch

$$\tilde{x} = x + \pi R = R(t + \pi - \sin(t))$$

und  $y$  durch

$$\tilde{y} = y + 2R = R(1 + \cos(t))$$

und benutzt den neuen Parameter  $\tilde{t} = t + \pi$ , so erhält man wegen  $\sin(t + \pi) = -\sin(t)$  und  $\cos(t + \pi) = -\cos(t)$

$$\tilde{x} = R(\tilde{t} - \sin(\tilde{t})), \quad \tilde{y} = R(1 - \cos(\tilde{t})).$$

Mit anderen Worten, die Evolute einer Zykloide ist eine (verschobene) Zykloide, und die Evolvente einer Zykloide ist ebenfalls eine (verschobene) Zykloide.

Evolventen bezeichnet man auch als Abwickelkurven oder Fadenkurven, denn sie entstehen aus ihrer Ausgangskurve durch Abwickeln eines Fadens: Schlingt man um die Ausgangskurve einen (straff gespannten) Faden, so beschreibt das Fadenende beim Abwickeln, wobei der Faden stets straff gehalten werden muss, die Evolvente.

**Beispiel 4.5.** Wir behaupten, dass die Kurve  $C$  mit Parameterdarstellung

$$R \begin{pmatrix} \cos(\varphi) + \varphi \sin(\varphi) \\ \sin(\varphi) - \varphi \cos(\varphi) \end{pmatrix}$$

die Evolvente eines Kreises um den Nullpunkt mit Radius  $R$  ist. Dazu müssen wir nachrechnen, dass die Evolute dieser Kurve der Kreis mit Radius  $R$  ist. Es gilt

$$\begin{aligned} \dot{x} &= R(-\sin(\varphi) + \sin(\varphi) + \varphi \cos(\varphi)) = R\varphi \cos(\varphi), & \ddot{x} &= R(\cos(\varphi) - \varphi \sin(\varphi)), \\ \dot{y} &= R(\cos(\varphi) - \cos(\varphi) + \varphi \sin(\varphi)) = R\varphi \sin(\varphi), & \ddot{y} &= R(\sin(\varphi) + \varphi \cos(\varphi)), \end{aligned}$$

also

$$\frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \ddot{x}\dot{y}} = \frac{R^2\varphi^2(\cos^2(\varphi) + \sin^2(\varphi))}{R^2\varphi \cos(\varphi)(\sin(\varphi) + \varphi \cos(\varphi)) - R^2\varphi \sin(\varphi)(\cos(\varphi) - \varphi \sin(\varphi))} = 1.$$

Es folgt

$$\begin{aligned} x_M &= R(\cos(\varphi) + \varphi \sin(\varphi)) - 1 \cdot R\varphi \sin(\varphi) = R\cos(\varphi), \\ y_M &= R(\sin(\varphi) - \varphi \cos(\varphi)) + 1 \cdot R\varphi \cos(\varphi) = R\sin(\varphi), \end{aligned}$$

und dies ist die Parametrisierung eines Kreises mit Radius  $R$  um den Ursprung.

Weitere Eigenschaften von Evolventen und Evoluten sind:

- Die Tangente in einem Punkt der Evolute ist die Normale im entsprechenden Punkt der Evolvente.
- Die Bogenlänge zwischen zwei Punkten der Evolute ist die Differenz der Krümmungsradien der entsprechenden Punkte der Evolvente.

Evolventen spielen im Maschinenbau eine Rolle bei der Konstruktion von Getrieben: Bei der Evolventenverzahnung von Zahnrädern werden die Flanken der Zähne von Evolventen eines Kreises gebildet; dies erlaubt eine gleichmäßige Übertragung von Drehmomenten durch eine konstante Übersetzung.

## 5. Vektorfelder

Unter einem *Vektorfeld* auf einem Gebiet  $B \subset \mathbb{R}^n$  versteht man eine vektorwertige Funktion

$$\vec{F}: B \rightarrow \mathbb{R}^n, \quad \vec{p} \mapsto \vec{F}(\vec{p}) = \begin{pmatrix} F_1(\vec{p}) \\ \vdots \\ F_n(\vec{p}) \end{pmatrix},$$

das heißt, jedem Punkt  $\vec{p}$  wird ein Vektor  $\vec{F}(\vec{p})$  mit Komponenten  $F_i(\vec{p})$  zugeordnet. Es genügt hier, die Fälle  $n = 2$  oder  $n = 3$  zu betrachten; wir benutzen dann auch wieder die Koordinaten  $x, y$  beziehungsweise  $x, y, z$ .



Jede der Komponenten  $F_i$  ist ihrerseits eine Funktion  $F_i: B \rightarrow \mathbb{R}$ . Solche Funktionen, die einem Punkt des Raumes einen Skalar zuordnen, nennt man auch *Skalarfelder*.

Als motivierende Beispiele für Vektorfelder können Kraftfelder, elektrische Felder, magnetische Felder oder Strömungsfelder dienen. Ein zweidimensionales Strömungsfeld liefert etwa ein fließendes Gewässer. Typische Beispiele für Skalarfelder sind Dichtefunktionen oder Temperaturverteilungen. Ist  $\phi: \mathbb{R}^3 \rightarrow \mathbb{R}$  ein (partiell) differenzierbares Skalarfeld, so ist sein Gradient

$$\text{grad } \phi = \left( \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial z} \right)^\top$$

ein Vektorfeld; statt  $\text{grad } \phi$  schreibt man auch  $\vec{\nabla} \phi$ . Ein Vektorfeld dieser Form, also ein Vektorfeld  $\vec{F}$ , zu dem es ein Skalarfeld  $\phi$  gibt mit  $\text{grad } \phi = \vec{F}$ , nennt man ein *Gradientenfeld* und  $\phi$  eine *Stammfunktion* von  $\vec{F}$ .

**Definition 5.1.** Sei  $\vec{F}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  ein Vektorfeld.

(a) Die Divergenz des Vektorfeldes ist das Skalarfeld

$$\text{div } \vec{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}.$$

(b) Die Rotation von  $\vec{F}$  ist das Vektorfeld

$$\text{rot } \vec{F} = \begin{pmatrix} \frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \\ \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} \\ \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \end{pmatrix}.$$

**Erläuterungen:** (a) Die Divergenz misst, ob aus einem Volumenelement etwas heraus- oder etwas in es hineinfließt. Ist in einem Bereich  $\text{div } \vec{F} > 0$ , so hat das Feld dort Quellen, ist  $\text{div } \vec{F} < 0$ , so hat es Senken. Im Fall  $\text{div } \vec{F} = 0$  nennt man das Feld quellen- und senkenfrei.

(b) Die Rotation des Feldes ist ein Maß für seine Verwirbelung, denn sie beschreibt die Veränderung des Feldes quer zur Strömungsrichtung. Eine suggestive Schreibweise für die Rotation ist

$$\vec{\nabla} \times \vec{F} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \times \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix},$$

wobei man den ersten Vektor als „Differentialoperator“ auffasst, der auf den zweiten wirkt.

**Rechenregeln 5.2.** Seien  $\vec{F}, \vec{G}$  Vektorfelder,  $\phi$  ein Skalarfeld,  $\vec{v}$  ein konstanter Vektor und  $c$  eine Konstante.

1.  $\text{div } \vec{v} = 0$
2.  $\text{div}(\phi \vec{F}) = \langle \text{grad } \phi, \vec{F} \rangle + \phi \text{div } \vec{F}$
3.  $\text{div}(c \vec{F}) = c \text{div } \vec{F}$
4.  $\text{div}(\vec{F} + \vec{G}) = \text{div } \vec{F} + \text{div } \vec{G}$
5.  $\text{rot } \vec{v} = \vec{0}$
6.  $\text{rot}(\phi \vec{F}) = \text{grad } \phi \times \vec{F} + \phi \text{rot } \vec{F}$
7.  $\text{rot}(c \vec{F}) = c \text{rot } \vec{F}$
8.  $\text{rot}(\vec{F} + \vec{G}) = \text{rot } \vec{F} + \text{rot } \vec{G}$
9.  $\text{rot grad } \phi = \vec{0}$
10.  $\text{div rot } \vec{F} = 0$

Mit Hilfe der Rotation kann man ein Kriterium formulieren, wann ein Vektorfeld eine Stammfunktion hat:

**Satz 5.3.** Sei  $\vec{F}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  ein wirbelfreies Feld. Dann gibt es ein Skalarfeld  $\phi$  mit  $\text{grad } \phi = \vec{F}$ .

Der Satz sagt nur, dass es ein wirbelfreies Feld auf  $\mathbb{R}^3$  eine Stammfunktion hat, aber wie findet man sie? Natürlich durch Integration...

**Beispiel 5.4.** Sei  $\vec{F}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  gegeben durch

$$\vec{F}(x, y, z) = \begin{pmatrix} 2xyz + 4y \\ x^2 + 4x - z^3 \\ x^2y - 3yz^2 \end{pmatrix}.$$

Dann gilt

$$\left. \begin{array}{l} \frac{\partial F_3}{\partial y} = x^2 - 3z^2 = \frac{\partial F_2}{\partial z} \\ \frac{\partial F_1}{\partial z} = 2xy = \frac{\partial F_3}{\partial x} \\ \frac{\partial F_2}{\partial x} = 2xz + 4 = \frac{\partial F_1}{\partial y} \end{array} \right\} \Rightarrow \text{rot } \vec{F} = \vec{0}.$$

Gesucht ist eine Funktion  $\phi(x, y, z)$  mit  $\text{grad } \phi = \vec{F}$ , also

$$\begin{aligned} \frac{\partial \phi}{\partial x} = 2xyz + 4y &\Rightarrow \phi = x^2yz + 4xy + \psi_1(y, z) \\ \frac{\partial \phi}{\partial y} = x^2z + 4x - z^3 &\Rightarrow \phi = x^2yz + 4xy - yz^3 + \psi_2(x, z) \\ \frac{\partial \phi}{\partial z} = x^2y - 3xyz^2 &\Rightarrow \phi = x^2yz - yz^3 + \psi_3(x, y) \end{aligned}$$

Durch Vergleich dieser drei Kandidaten findet man die Stammfunktion

$$\phi(x, y, z) = x^2yz + 4xy - yz^3.$$

Es gibt im Übrigen viele Stammfunktionen, denn addiert man zu  $\phi$  eine Konstante, so ändert das nicht am Gradienten.

**Warnung:** Ist  $\vec{F}$  nur auf einem Teilbereich  $B \subset \mathbb{R}^3$  definiert, ist der Satz nicht ohne weiteres anwendbar!

## 6. Kurvenintegrale

Als motivierendes Beispiel dient die Arbeit, die man verrichten muss, um ein Teilchen in einem Kraftfeld zu bewegen: Wirkt die Kraft genau entgegen der Bewegungsrichtung, so gilt bekanntlich Arbeit=Kraft·Weg; haben der Kraftvektor und die Bewegungsrichtung einen Winkel, so nimmt man die Projektion der Kraft auf diese Richtung, also das Skalarprodukt mit dem Einheitsvektor in Bewegungsrichtung. Bewegt man nun das Teilchen entlang einer Kurve  $C$ , so muss man die Komponente von des Feldes in Richtung der Kurventangente berücksichtigen: Sei dazu  $\vec{F}$  ein Kraftfeld und  $\vec{r}: [a, b] \rightarrow \mathbb{R}^n$  eine Parametrisierung von  $C$ . Wie früher zerlegen wir das Intervall in gleichlange Teilintervalle  $[t_{i-1}, t_i]$  mit  $a = t_0 < t_1 < \dots < t_n = b$  und  $t_i - t_{i-1} = \frac{b-a}{n} =: \Delta t$ . Liegen  $\vec{r}(t_{i-1})$  und  $\vec{r}(t_i)$  nahe genug beieinander, so ist die Arbeit bei der Bewegung von  $\vec{r}(t_{i-1})$  nach  $\vec{r}(t_i)$  näherungsweise

$$\langle \vec{F}(\vec{r}(t_i)), \Delta \vec{r}_i \rangle \quad \text{wobei} \quad \Delta \vec{r}_i = \vec{r}(t_i) - \vec{r}(t_{i-1}).$$

Beim Grenzübergang  $\Delta t \rightarrow 0$  entsteht daraus in mittlerweile bekannter Weise ein Integral, das sogenannte Wegintegral.

**Definition 6.1.** Sei  $\vec{F}(x, y, z)$  ein Vektorfeld in  $\mathbb{R}^3$  und  $C$  eine durch  $\vec{r}: [a, b] \rightarrow \mathbb{R}^3$  regulär parametrisierte Kurve. Dann heißt

$$\int_C \langle \vec{F}, d\vec{r} \rangle = \int_a^b \langle \vec{F}(\vec{r}(t)), \dot{\vec{r}}(t) \rangle dt$$

das Weg- oder Kurvenintegral des Vektorfeldes  $\vec{F}$  entlang  $C$ .

Bevor wir die Definition anhand von Beispielen verdeutlichen, ist noch einiges anzumerken:

1. Man kann zeigen, dass diese Definition nicht von der gewählten Parametrisierung abhängt (dies folgt aus der Transformationsformel).
2. Jedoch hängt im Allgemeinen der Wert des Kurvenintegrals nicht nur von Anfangs- und Endpunkt, sondern auch vom Verlauf der Kurve ab!
3. Stimmen Anfangs- und Endpunkt der Kurve  $C$  überein, nennt man  $C$  eine *geschlossene* Kurve und schreibt auch

$$\oint_C \langle \vec{F}, d\vec{r} \rangle$$

für das Kurvenintegral.

4. Die Definition überträgt sich wörtlich auf ebene Felder  $\vec{F}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ .
5. Sei  $C$  eine Kurve, die aus zwei Stücken zusammengesetzt ist, also  $C = C_1 \cup C_2$ . so dass der Endpunkt von  $C_1$  der Anfangspunkt von  $C_2$  ist. Dann gilt

$$\int_C \langle \vec{F}, d\vec{r} \rangle = \int_{C_1} \langle \vec{F}, d\vec{r} \rangle + \int_{C_2} \langle \vec{F}, d\vec{r} \rangle .$$

6. Sei  $C^*$  die Kurve, die aus  $C$  durch Umkehr der Durchlaufrichtung entsteht. Dann gilt

$$\int_{C^*} \langle \vec{F}, d\vec{r} \rangle = - \int_C \langle \vec{F}, d\vec{r} \rangle .$$

Um ein Kurvenintegral  $\int_C \langle \vec{F}, d\vec{r} \rangle$  zu berechnen, muss man also mehrere Schritte ausführen:

- Man finde eine Parametrisierung  $\vec{r}(t)$ ,  $a \leq t \leq b$ , der Kurve  $C$ .
- Man berechne die Vektoren  $\vec{F}(\vec{r}(t))$  und  $\dot{\vec{r}}(t)$ ; dies sind beides Vektoren in einer einzigen Variablen  $t$ , dem Kurvenparameter.
- Man bilde das Skalarprodukt  $\langle \vec{F}(\vec{r}(t)), \dot{\vec{r}}(t) \rangle$ . Dies ist eine Funktion  $[a, b] \rightarrow \mathbb{R}$  in der Variablen  $t$ .
- Schließlich integriere man diese Funktion von  $a$  nach  $b$ .

**Beispiel 6.2.** Sei  $\vec{F}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  das Vektorfeld definiert durch

$$\vec{F}(x, y, z) = \begin{pmatrix} x + y^2 \\ xyz \\ y - z \end{pmatrix} .$$

Wir berechnen nun das Wegintegral über drei verschiedene Wege, die alle vom Punkt  $(0, 0, 0)$  zum Punkt  $(1, 1, 1)$  führen:

- (a) Sei  $C$  die Strecke von  $(0, 0, 0)$  nach  $(1, 1, 1)$ . Eine Parametrisierung von  $C$  ist gegeben durch

$$\vec{r}(t) = \begin{pmatrix} t \\ t \\ t \end{pmatrix}, \quad 0 \leq t \leq 1. \quad \text{Es gilt}$$

$$\vec{F}(\vec{r}(t)) = \begin{pmatrix} t + t^2 \\ t^3 \\ t - t \end{pmatrix} = \begin{pmatrix} t + t^2 \\ t^3 \\ 0 \end{pmatrix}, \quad \dot{\vec{r}}(t) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \text{also} \quad \langle \vec{F}(\vec{r}(t)), \dot{\vec{r}}(t) \rangle = t + t^2 + t^3$$

und damit

$$\int_C \langle \vec{F}, d\vec{r} \rangle = \int_0^1 (t + t^2 + t^3) dt = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}.$$

(b) Sei  $C$  die Zusammensetzung aus drei Teilstrecken:

- (i) der Strecke  $C_1$  von  $(0, 0, 0)$  nach  $(1, 0, 0)$ ,
- (ii) der Strecke  $C_2$  von  $(1, 0, 0)$  nach  $(1, 1, 0)$ ,
- (iii) der Strecke  $C_3$  von  $(1, 1, 0)$  nach  $(1, 1, 1)$ .

Parametrisierungen für diese Kurven sind

$$\vec{r}_1(t) = \begin{pmatrix} t \\ 0 \\ 0 \end{pmatrix} \text{ für } C_1, \quad \vec{r}_2(t) = \begin{pmatrix} 1 \\ t \\ 0 \end{pmatrix} \text{ für } C_2, \quad \vec{r}_3(t) = \begin{pmatrix} 1 \\ 1 \\ t \end{pmatrix} \text{ für } C_3,$$

jeweils mit  $0 \leq t \leq 1$ . Dann gilt  $\dot{\vec{r}}_1(t) = \vec{e}_1$ ,  $\dot{\vec{r}}_2(t) = \vec{e}_2$ ,  $\dot{\vec{r}}_3(t) = \vec{e}_3$ , sowie

$$\begin{aligned} \langle \vec{F}(\vec{r}_1(t)), \dot{\vec{r}}_1(t) \rangle &= \left\langle \begin{pmatrix} t \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle = t, \\ \langle \vec{F}(\vec{r}_2(t)), \dot{\vec{r}}_2(t) \rangle &= \left\langle \begin{pmatrix} 1+t^2 \\ 0 \\ t \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle = 0, \\ \langle \vec{F}(\vec{r}_3(t)), \dot{\vec{r}}_3(t) \rangle &= \left\langle \begin{pmatrix} 2 \\ t \\ 1-t \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle = 1-t, \end{aligned}$$

also

$$\begin{aligned} \int_C \langle \vec{F}, d\vec{r} \rangle &= \int_{C_1} \langle \vec{F}, d\vec{r} \rangle + \int_{C_2} \langle \vec{F}, d\vec{r} \rangle + \int_{C_3} \langle \vec{F}, d\vec{r} \rangle \\ &= \int_0^1 t dt + \int_0^1 0 dt + \int_0^1 (1-t) dt = \frac{1}{2} + 0 + 1 - \frac{1}{2} = 1. \end{aligned}$$

(c) Sei schließlich  $C$  der durch  $\vec{r}(t) = \begin{pmatrix} t \\ t \\ t^2 \end{pmatrix}$ ,  $0 \leq t \leq 1$ , parametrisierte parabolische Weg (er verläuft auf dem Schnitt des Paraboloids  $z = x^2 + y^2$  mit der Ebene  $x = y$ ). Hier gilt nun

$$\vec{F}(\vec{r}(t)) = \begin{pmatrix} t+t^2 \\ t^4 \\ t-t^2 \end{pmatrix}, \quad \dot{\vec{r}}(t) = \begin{pmatrix} 1 \\ 1 \\ 2t \end{pmatrix},$$

also

$$\langle \vec{F}(\vec{r}(t)), \dot{\vec{r}}(t) \rangle = \left\langle \begin{pmatrix} t+t^2 \\ t^4 \\ t-t^2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2t \end{pmatrix} \right\rangle = t^4 - 2t^3 + 3t^2 + t$$

und damit

$$\int_C \langle \vec{F}, d\vec{r} \rangle = \int_0^1 (t^4 - 2t^3 + 3t^2 + t) dt = \frac{1}{5} - \frac{1}{2} + 1 + \frac{1}{2} = \frac{6}{5}.$$

Diese Beispiel illustriert den Punkt 2. der Anmerkungen im Anschluss an die Definition, dass es durchaus auf den Verlauf des Weges ankommt, welchen Wert das Integral annimmt. Wann kommt es aber nicht darauf an?

**Definition 6.3.** Ein Vektorfeld  $\vec{F}: B \rightarrow \mathbb{R}^3$  heißt konservativ, wenn für je zwei Kurven  $C_1, C_2$  in  $B$ , die beide den gleichen Anfangspunkt und den gleichen Endpunkt haben, gilt:

$$\int_{C_1} \langle \vec{F}, d\vec{r} \rangle = \int_{C_2} \langle \vec{F}, d\vec{r} \rangle$$

Man spricht dann davon, dass das Kurvenintegral wegunabhängig sei, da sein Wert nur von Anfangs- und Endpunkt abhängt.

Die oben gestellte Frage lautet also: welche Felder sind konservativ? Ist zum Beispiel  $\vec{F} = \text{grad } \phi$  ein Gradientenfeld und  $\vec{r}(t)$ ,  $a \leq t \leq b$ , die Parameterdarstellung einer Kurve, so gilt nach der Kettenregel

$$\int_a^b \langle \text{grad } \phi(\vec{r}(t)), \dot{\vec{r}}(t) \rangle dt = \int_a^b \left( \frac{d}{dt} \phi(\vec{r}(t)) \right) dt = \phi(\vec{r}(b)) - \phi(\vec{r}(a))$$

und das Wegintegral hängt nur vom Anfangspunkt  $\vec{r}(a)$  und vom Endpunkt  $\vec{r}(b)$  ab. Weiß man umgekehrt, dass ein Feld  $\vec{F}$  konservativ ist, so kann man eine Stammfunktion durch Integration erhalten: man wählt einen festen Punkt  $\vec{p}_0 = (x_0, y_0, z_0)$  in  $B$  sowie zu jedem Punkt  $\vec{p} = (x, y, z)$  einen Weg  $C_{\vec{p}}$  von  $\vec{p}_0$  nach  $\vec{p}$  in  $B$  (wir setzen stillschweigend voraus, dass das in  $B$  geht) und definiert  $\phi(x, y, z) := \int_{C_{\vec{p}}} \langle \vec{F}, d\vec{r} \rangle$ ; die Wegunabhängigkeit besagt gerade, dass es auf den Verlauf des Weges nicht ankommt.

Wir fassen zusammen:

**Satz 6.4.** Für ein Vektorfeld  $\vec{F}: B \rightarrow \mathbb{R}^n$  sind folgende Aussagen äquivalent:

- (i)  $\vec{F}$  ist ein Gradientenfeld.
- (ii)  $\vec{F}$  ist konservativ.

Beide Aussagen sind wiederum äquivalent zu

- (iii) Für jede geschlossene Kurve  $C$  in  $B$  ist  $\oint_C \langle \vec{F}, d\vec{r} \rangle = 0$ .

**Beispiel 6.5.** Sei  $\vec{F}: \mathbb{R}^3 \setminus \{z\text{-Achse}\}$  das Vektorfeld

$$\frac{1}{x^2 + y^2} \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix};$$

dieses Feld ist auf der  $z$ -Achse nicht definiert, denn dort ist  $x^2 + y^2 = 0$ . Sei ferner  $C$  die Kreislinie in der  $xy$ -Ebene vom Radius 1 um den Nullpunkt, parametrisiert durch  $\vec{r}t = \begin{pmatrix} \cos(t) \\ \sin(t) \\ 0 \end{pmatrix}$ ,  $0 \leq t \leq 2\pi$ .

Dann gilt

$$\oint_C \langle \vec{F}, d\vec{r} \rangle = \int_0^{2\pi} \left\langle \begin{pmatrix} -\sin(t) \\ \cos(t) \\ 0 \end{pmatrix}, \begin{pmatrix} -\sin(t) \\ \cos(t) \\ 0 \end{pmatrix} \right\rangle dt = \int_0^{2\pi} (\sin^2(t) + \cos^2(t)) dt = 2\pi.$$

Das Feld ist also nicht konservativ und daher kein Gradientenfeld, obwohl es wirbelfrei ist; die Warnung war also berechtigt!

Das Kriterium des Satzes funktioniert zwar ganz gut um zu zeigen, dass ein Feld nicht konservativ ist, aber für die umgekehrte Richtung ist es nicht praktikabel (wie soll man jemals alle Wegintegrale testen?) Um zu einem anwendbaren Kriterium zu gelangen, muss man etwas mehr über die Geometrie des Definitionsbereichs des Feldes wissen: Man nennt ein Gebiet  $B$  *einfach zusammenhängend*, wenn sich jeder geschlossene Weg in  $B$  zu einem Punkt zusammenziehen lässt.

**Beispiele.** (a)  $\mathbb{R}^2$  und  $\mathbb{R}^3$  sind einfach zusammenhängend.

(b)  $B = \mathbb{R}^2 \setminus \{(0, 0)\}$  ist nicht einfach zusammenhängend, denn eine Kreislinie, die den Nullpunkt umläuft, kann man nicht in  $B$  zusammenziehen. Allgemeiner ist ein Teilgebiet  $B$  der Ebene, das Löcher hat, nicht einfach zusammenhängend.

(c) Genausowenig ist  $\mathbb{R}^3 \setminus \{z\text{-Achse}\}$  einfach zusammenhängend.

**Satz 6.6.** Ein Linienintegral  $\int_C \langle \vec{F}, d\vec{r} \rangle$  ist genau dann wegunabhängig, wenn in einem einfach zusammenhängenden Gebiet  $B$ , das die Kurve  $C$  enthält,

$$\operatorname{rot} \vec{F} = 0$$

gilt.

Insbesondere hat ein wirbelfreies Feld auf einem einfach zusammenhängenden Gebiet eine Stammfunktion, und durch ein Wegintegral können wir eine bestimmen.

**Beispiel 6.7.** Sei  $\vec{F}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  das durch

$$\begin{pmatrix} 3x^2y \\ x^3 + z^2 \\ 2yz \end{pmatrix}$$

definierte Feld. Da  $\mathbb{R}^3$  einfach zusammenhängend ist, reicht es nachzuprüfen, ob  $\vec{F}$  wirbelfrei ist: wegen

$$\frac{\partial F_3}{\partial y} = 2z = \frac{\partial F_2}{\partial z}, \quad \frac{\partial F_1}{\partial z} = 0 = \frac{\partial F_3}{\partial x}, \quad \frac{\partial F_2}{\partial x} = 3x^2 = \frac{\partial F_1}{\partial y}$$

ist das der Fall. Wir überprüfen aber trotzdem noch Bedingung (iii) von oben für den geschlossenen Weg

$$\vec{r}(t) = \begin{pmatrix} t(t^2 - 1) \\ -t(t - 1) \\ t(t - 1)^2 \end{pmatrix}, \quad 0 \leq t \leq 1.$$

Es ist

$$\vec{F}(\vec{r}(t)) = \begin{pmatrix} -3t^2(t - 1)^2 \cdot t(t - 1) \\ t^3(t - 1)^3 + t^2(t - 1)^4 \\ -2t^2(t - 1)^3 \end{pmatrix}, \quad \dot{\vec{r}}(t) = \begin{pmatrix} 3t^2 - 1 \\ -2t + 1 \\ 3t^2 - 4t + 1 \end{pmatrix}$$

und damit

$$\begin{aligned} \oint_C \langle \vec{F}, d\vec{r} \rangle &= \int_0^1 \left\langle \begin{pmatrix} -3t^2(t - 1)^2 \cdot t(t - 1) \\ t^3(t - 1)^3 + t^2(t - 1)^4 \\ -2t^2(t - 1)^3 \end{pmatrix}, \begin{pmatrix} 3t^2 - 1 \\ -2t + 1 \\ 3t^2 - 4t + 1 \end{pmatrix} \right\rangle dt \\ &= \int_0^1 (-11t^{10} + 10t^9 + 27t^8 - 32t^7 + 14t^6 - 42t^5 + 55t^4 - 24t^3 + 3t^2) dt = 0. \end{aligned}$$

Eine Stammfunktion finden wir, indem wir entlang der Strecke  $C$  von  $(0, 0, 0)$  nach  $(x, y, z)$  integrieren: dieser Weg wird durch

$$\vec{r}(t) = \begin{pmatrix} tx \\ ty \\ tz \end{pmatrix}, \quad 0 \leq t \leq 1$$

parametrisiert mit

$$\langle \vec{F}(\vec{r}(t)), \dot{\vec{r}}(t) \rangle = \left\langle \begin{pmatrix} 3t^2x^2 \cdot ty \\ t^3x^3 + t^2z^2 \\ 2t^2yz \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle = 4t^3x^3y + 3t^2yz^2$$

und wir setzen

$$\phi(x, y, z) = \int_C \langle \vec{F}, d\vec{r} \rangle = \int_0^1 (4t^3 x^3 y + 3t^2 y z^2) dt = x^3 y + y z^2.$$

# III Grundlagen der Statistik

Wahrscheinlichkeitstheorie und Statistik sind mathematische Fachrichtungen, die sich mit dem Zufall und seinen Auswirkungen befassen.

Die Wahrscheinlichkeitstheorie sucht nach Gesetzen, denen Ereignisse unterliegen, die nicht durch Naturgesetze bestimmt sind, deren Ausgang also dem Zufall unterliegt. Mit Methoden der Statistik versucht man, den in der wirklichen Welt vorkommenden Ereignissen ihre Zufälligkeit anzusehen und zu beschreiben.

Die Wahrscheinlichkeitstheorie hat eine lange Geschichte. Bei den Griechen mit ihrer Betonung axiomatischer Mathematik und exakter Beweise war sie weniger beliebt: Platon hielt Argumente, die sich der Wahrscheinlichkeit bedienen, für Hochstapelei. Die Römer, eher praktisch veranlagt, sahen durchaus den Nutzen („Wahrscheinlichkeit ist das Handbuch des Lebens“, Cicero zugeschrieben), haben aber, vielleicht auch wegen ihres für Rechenoperationen ungeeigneten Zahlensystems, nie eine solche Theorie entwickelt. Den eigentlichen Beginn der Wahrscheinlichkeitstheorie markiert das „Buch der Glücksspiele“ (Liber de Ludo Aleae) von Gerolamo Cardano aus dem Jahr 1524, wie überhaupt das Interesse am Glücksspiel eine wesentliche Motivation für die Befassung mit dem Zufall war (und ist).

Zur Illustration beginnen wir mit einigen Beispielen.

**Beispiel.** Zur Kontrolle, ob Sportler ein verbotenes, leistungsförderendes Mittel anwenden, gibt es verschiedene Tests. Nehmen wir an, ein bestimmter Test kann in der Hälfte der Fälle, in denen das Mittel benutzt wurde, seine Anwendung nachweisen. Wie für jeden Test soll es aber auch hier vorkommen, dass der Test positiv ist, obwohl das Mittel nicht angewandt wurde („falsch positiv“); dies sei in 1% aller Anwendungen des Tests der Fall. Heißt das nun, dass ein positiv getesteter Sportler zu 99% schuldig ist?

Die Antwort lautet: so leicht kann man es sich nicht machen. Angenommen man wüsste (aus Erfahrung, aus anderen Quellen, usw.), dass 10 % aller Sportler dieses Mittel benutzen. Werden nun 1000 Sportler getestet, so wären 100 schuldig, und der Test hätte 50 dieser Betrüger entlarvt. Aber von den 900 Unschuldigen hätte der Test auch 9 für schuldig erklärt (1%). Die positive Probe sagt also nur, dass ein positiv Getesteter mit Wahrscheinlichkeit  $50/59 = 84,7\%$  tatsächlich gedopt hat.

**Beispiel.** In einer Urne befinden sich 3000 weiße und 2000 schwarze Kugeln. Wir ziehen 100 mal eine Kugel (mit Zurücklegen). Welches Ergebnis ist zu erwarten? Und welches, wenn wir 1000 oder eine Million mal ziehen?

Jakob Bernoulli fand heraus, dass es stets möglich ist, genügend viele Kugeln zu ziehen, um „beinahe sicher“ zu sein, ein Ergebnis „beliebig nah“ an 60% zu erzielen. Zu klären ist dabei natürlich, was „beinahe sicher“ und „beliebig nah“ bedeuten soll. Will man zum Beispiel, dass die Wahrscheinlichkeit, mit dem Experiment zwischen 58 und 62 Prozent zu landen, bei 99,9% liegt, muss man 25 550 mal ziehen. Dies ist offenbar ein übertriebener Standard. Reicht einem jedoch 95%-ige Gewissheit, ein Ergebnis zwischen 55 und 65 Prozent zu erzielen, kommt man mit 370 Ziehungen aus. (Praktische Bedeutung haben solche Fragen in der Wahlforschung.)



In den beiden bisherigen Beispielen war die Wahrscheinlichkeit des Eintretens des interessanten Falles bekannt. Aber wenn das nicht der Fall ist? Dann will man sie dennoch ermitteln, und versucht das gemeinhin durch Stichproben oder Testreihen. Aber was sagen diese wirklich aus?

**Beispiel.** In einem Werk werden Röhren produziert. Der Kunde verlangt Röhren mit einem Durchmesser von 300mm, mit einer Toleranz von  $\pm 1$ mm. Man entnimmt der laufenden Produktion 6 Röhren und misst nach, mit folgendem Ergebnis: 299,1 – 300,4 – 298,0 – 301,5 – 297,5 – 300,5 mit dem Durchschnittswert 299,5. Kann man mit ausreichender Sicherheit sagen, dass die Vorgabe eingehalten wird, oder wird man einen unzufriedenen Kunden bekommen?

Das Vorgehen in diesen Fällen ist wie geschildert: man stellt eine Hypothese auf (etwa „die Produktion ist in Ordnung“) und versucht mittels einer Stichprobe herauszufinden, ob die Hypothese vernünftig ist, das heißt, dass man sie mit nur geringer Wahrscheinlichkeit (typisch wäre  $< 5\%$ ) irrtümlich aufrechterhält.

Weitere typische Probleme, die mit Methoden der Wahrscheinlichkeitstheorie und Statistik behandelt werden können, sind:

- Wie lange hält ein Bauteil durch, bevor es versagt?
- Wie wirksam ist ein Medikament?
- Wie hoch sind die Kosten einer Mobilitätsgarantie bei Neuwagen? Wie berechnet man Versicherungsprämien?
- Wie bewertet man Gesundheitsrisiken? (Aktuelles Beispiel: wie aussagekräftig sind erhöhte Fallzahlen von Krebserkrankungen in der Nähe von Nuklearanlagen?)
- Wie entscheidet man, ob das Ergebnis eines Handelns Produkt des Zufalls oder Ausdruck von Können ist? (Erfolg einer Anlagestrategie, Torquote eines Fußballers)

Bevor wir uns solchen Fragen zuwenden können, müssen allerdings erst einmal einige Grundlagen geklärt werden.

## 1. Zufallsexperimente

Unter einem Zufallsexperiment versteht man ein Experiment, dessen Ausgang nicht durch Naturgesetze vorherbestimmt ist, sondern vom Zufall anhängt.

Um zu beschreiben, welche möglichen Ausgänge ein solches Experiment haben kann, bedient man sich der Sprache der Mengen:

**Definition 1.1.** Die Ergebnismenge (auch: der Ergebnisraum) eines Zufallsexperiments ist die Menge  $\Omega$  aller möglichen Ergebnisse, die das Experiment haben kann.

**Beispiel 1.2.** (a) Beim Wurf eines Standardwürfels ist die Ergebnismenge  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

(b) Beim gleichzeitigen Wurf zweier Würfel ist  $\Omega = \{(a, b) \mid a, b \in \{1, 2, 3, 4, 5, 6\}\}$ .

(c) Die Lebensdauer eines Akkus kann man ebenfalls als Zufallsexperiment betrachten; hier ist die Ergebnismenge die Menge  $\mathbb{R}_+$  der positiven reellen Zahlen.

Oft interessiert man sich nicht nur für die einzelnen Ergebnisse, sondern für Zusammenfassungen von solchen, also für Teilmengen von  $\Omega$ :

**Definition 1.3.** Sei  $\Omega$  die Ergebnismenge eines Zufallsexperiments. Ein Ereignis ist eine Teilmenge  $A$  von  $\Omega$ . Die einelementigen Teilmengen nennt man Elementarereignisse, die Menge aller Ereignisse den Ereignisraum.

**Beispiel 1.4.** Beim französischen Roulette gibt es die Zahlenfelder 1 bis 36 sowie die 0, die eine gewisse Sonderrolle spielt, mit

- Ergebnismenge:  $\Omega = \{0, 1, 2, \dots, 35, 36\}$ ,
- Elementarereignisse:  $\{0\}, \{1\}, \dots, \{36\}$ .
- Ereignisse: zum Beispiel

$$G = \{2, 4, 6, 8, \dots, 34, 36\} \text{ („gerade“)}$$

$$R = \{1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36\} \text{ („rot“)}$$

$$M = \{1, 2, \dots, 17, 18\} \text{ („manque“)}$$

**Beispiel 1.5.** Bei der Lebensdauer eines Akkus sind

- $\Omega = \mathbb{R}_+$ ,
- Elementarereignisse: positive reelle Zahlen,
- tpische Ereignisse zum Beispiel Intervalle wie  $[2, 10]$ .

Die Kombinationen verschiedener Ereignisse werden durch Operationen auf Teilmengen beschrieben: Ist  $\Omega$  die Ergebnismenge eines Zufallexperiments und sind  $A, B$  Ereignisse, so bedeuten

- $A \cup B$ :  $A$  oder  $B$  tritt ein (nicht ausschließendes oder),
- $A \cap B$ :  $A$  und  $B$  treten beide gleichzeitig ein. Ist  $A \cap B = \emptyset$ , so nennt man  $A$  und  $B$  *unvereinbar*. Der Grund dafür liegt in:
- $\emptyset$  ist das unmögliche Ereignis,
- $\Omega$  das sichere Ereignis,
- $\bar{A} = \Omega \setminus A$  das Komplementärereignis, also  $A$  ist nicht eingetreten,
- $A \setminus B = A \cap \bar{B}$ :  $A$  tritt ein,  $B$  aber nicht,

**Beispiele 1.6.** Wir setzen das Roulette-Beispiel 1.4 fort: Sei  $U = \{1, 3, 5, \dots, 33, 35\}$  das Ereignis „ungerade“, dann gilt

$$\bar{G} = U \cup \{0\}$$

$$\bar{U} = G \cup \{0\}$$

$$G \cap U = \emptyset$$

$$G \cap M \cap R = \{12, 14, 16, 18\}$$

Beim Würfeln mit zwei Würfeln ist

$$A := \{(a, b) \mid a, b \geq 4\}$$

das Ereignis, dass die Summe der gewürfelten Augen mindestens 8 beträgt.

Für das Rechnen mit Mengen, also mit Ereignissen, gelten bestimmte Regeln:

**1.7 (Rechenregeln für Mengen).** Seien  $A, B$  und  $C$  Mengen. Dann gilt

- |     |   |                     |
|-----|---|---------------------|
| (a) | $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$                | (Distributivgesetz) |
| (b) | $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$ | (De Morgan)         |
|     | $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$ |                     |

Diese Regeln sind wahrscheinlich wohlbekannt, ansonsten bleibt ihr Nachweis dem Leser zur Übung überlassen.

## 2. Wahrscheinlichkeitsmaße

In diesen Abschnitt erklären wir endlich, was wir unter Wahrscheinlichkeit verstehen wollen. Erstmals widerspruchsfrei wurde dies in den 30er Jahren von Kolmogorov geklärt, der Wahrscheinlichkeit axiomatisch beschrieb, nämlich als eine Funktion auf dem Ereignisraum, die gewisse Eigenschaften erfüllt.

**Definition 2.1.** Ein Wahrscheinlichkeitsmaß (oder kurz eine Wahrscheinlichkeit) auf einer Ergebnismenge  $\Omega$  ist eine Funktion  $P$ , die jedem Ereignis  $A$  eine reelle Zahl  $P(A)$  zuordnet, so dass gilt:

1. Für jedes Ereignis  $A$  ist  $0 \leq P(A) \leq 1$ .
2.  $P(\Omega) = 1$  und  $P(\emptyset) = 0$ .
3. Sind  $A_1, A_2, A_3, \dots$  paarweise unvereinbare Ereignisse (also  $A_i \cap A_j = \emptyset$  für  $i \neq j$ ), so gilt

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

Aus diesen Axiomen lassen sich einige nützliche Eigenschaften leicht ableiten. Angenommen wir haben zwei Ereignisse  $A, B$ . Was ist die Wahrscheinlichkeit, dass  $A$  oder  $B$  eintritt? Dies kann man mittels des Mengenkalküls ermitteln: Sei

$$C = (A \cup B) \setminus (A \cap B) = \underbrace{(A \setminus (A \cap B))}_{=: C_1} \cup \underbrace{(B \setminus (A \cap B))}_{=: C_2},$$

dann ist  $A \cup B = C \cup (A \cap B)$ , wobei  $C$  und  $A \cap B$  unvereinbar sind. Ferner gilt  $A = C_1 \cup (A \cap B)$  und  $B = C_2 \cup (A \cap B)$ , wobei  $A \cap B, C_1$  und  $C_2$  paarweise unvereinbar sind. Aus dem dritten Axiom folgt daher

$$\begin{aligned} P(A \cup B) &= P(C) + P(A \cap B) \\ &= P(C_1) + P(C_2) + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Insbesondere folgt daraus wegen  $\Omega = A \cup \bar{A}$  und dem zweiten Axiom

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

also

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

Diese auch intuitiv plausiblen Ergebnisse fassen wir in einem Satz zusammen:

**Satz 2.2** (Additionssatz). Sei  $P$  ein Wahrscheinlichkeitsmaß auf einer Ergebnismenge  $\Omega$  und seien  $A, B \subset \Omega$  Ereignisse. Dann gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Insbesondere gilt

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

Mit anderen Worten: die Wahrscheinlichkeit, dass  $A$  oder  $B$  eintritt, ist die Summe der Wahrscheinlichkeiten von  $A$  und  $B$ , abzüglich der Wahrscheinlichkeit, dass beide gleichzeitig auftreten. Genauso plausibel ist die nächste Folgerung

**2.3.** Für  $A \subset B$  gilt  $P(A) \leq P(B)$ .

Demn wegen  $A \subset B$  ist  $B = A \cup (B \setminus A)$  mit  $A \cap (B \setminus A) = \emptyset$ , also

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

**Beispiel 2.4.** Sei  $\Omega$  eine endliche Menge, etwa  $\Omega = \{\omega_1, \dots, \omega_n\}$ . Ein Wahrscheinlichkeitsmaß  $P$  auf  $\Omega$  besteht dann aus  $n$  reellen Zahlen  $p_i = P(\omega_i) \geq 0$  mit  $p_1 + \dots + p_n = 1$ . Für ein Ereignis  $A \subset \Omega$  gilt dann

$$P(A) = \sum_{\omega_i \in A} p_i.$$

**Definition 2.5.** Ein Zufallsexperiment heißt Laplace-Experiment, wenn gilt:

- (i) Die Ergebnismenge  $\Omega$  ist endlich.
- (ii) Alle Elementarereignisse sind gleichwahrscheinlich.

Ist also  $\Omega = \{\omega_1, \dots, \omega_n\}$ , so gilt für jedes  $\omega_j$

$$P(\omega_j) = \frac{1}{\#\Omega} = \frac{1}{n}$$

und für jedes Ereignis  $A \subset \Omega$

$$P(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{\#A}{\#\Omega}.$$

**Beispiele 2.6.** (a) Für den Würfel ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$  und daher die Wahrscheinlichkeit jedes Elementarereignisses  $1/6$ ; beim Roulette gilt  $\Omega = \{0, 1, \dots, 36\}$  mit  $\#\Omega = 37$ , also  $p_i = 1/37$  für jedes  $i$ .

(b) Beim gleichzeitigen Wurf zweier Würfel ist  $\Omega = \{(a, b) \mid a, b \in \{1, 2, 3, 4, 5, 6\}\}$  mit 36 Elementen. Sei nun  $A$  das Ereignis „die Summe der Augenzahlen ist 6“, in Mengenschreibweise

$$A = \{(a, b) \in \Omega \mid a + b = 6\}.$$

Dann ist

$$A = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$$

mit 5 Elementen, folglich gilt  $P(A) = \frac{5}{36}$ .

### 3. Ein wenig Kombinatorik

Um die Wahrscheinlichkeiten bei Zufallsexperimenten zu berechnen, die aus der wiederholten Durchführung eines Experiments bestehen, muss man wissen, wie man alle möglichen Ausgänge solcher sogenannter mehrstufiger Zufallsexperimente, die zu einem Ereignis gehören, zählt. Dies ist die Domäne der Kombinatorik.

Seien etwas allgemeiner  $k$  Zufallsexperimente gegeben, mit Ergebnismengen  $\Omega_1, \dots, \Omega_k$ . Führt man diese  $k$  Zufallsexperimente gleichzeitig oder nacheinander aus, ist dies wieder ein Zufallsexperiment, deren Ergebnismenge  $\Omega'$  die Menge aller geordneten  $k$ -Tupel mit Einträgen in den Mengen  $\Omega_i$  ist:

$$\Omega' = \{(\omega_1, \omega_2, \dots, \omega_k) \mid \omega_i \in \Omega_i\}.$$

Diese Menge nennt man auch das *Produkt* der Mengen  $\Omega_i$ , geschrieben als

$$\Omega_1 \times \Omega_2 \times \cdots \times \Omega_k = \prod_{i=1}^k \Omega_i.$$

Wird das gleiche Zufallsexperiment  $k$  mal wiederholt, sind alle  $\Omega_i$  gleich, also  $\Omega_1 = \cdots = \Omega_k =: \Omega$ , und man schreibt  $\Omega' = \Omega^n$ .

Sind alle Ergebnismengen  $\Omega_i$  endlich mit  $\#\Omega_i = n_i$ , so gilt

$$\#(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_k) = n_1 \cdot n_2 \cdot \cdots \cdot n_k,$$

denn für den ersten Eintrag  $\omega_1$  kann man aus  $n_1$  Elementen wählen, und zu jeder dieser Wahlen gibt es  $n_2$  Möglichkeiten, ein  $\omega_2$  zu wählen, und so fort. Mit anderen Worten:

**Satz 3.1.** *Seien  $\Omega_i$ ,  $1 \leq i \leq k$ , endliche Mengen mit  $\#\Omega_i = n_i$ . Dann gilt*

$$\# \prod_{i=1}^k \Omega_i = \prod_{i=1}^k \#\Omega_i.$$

**Beispiel 3.2.** Aus einer Urne mit  $n$  verschiedenen Kugeln werden nacheinander  $k$  von ihnen gezogen, wobei man nach jeder Ziehung die Kugel wieder zurücklegt. Der obige Satz beantwortet die Frage, wieviele mögliche Resultate es gibt: Numeriert man die Kugeln von 1 bis  $n$ , so ist die Ergebnismenge dieses Experiments „ $k$  mal ziehen mit Zurücklegen“ nämlich die Menge aller geordneten  $k$ -Tupel  $(\omega_1, \dots, \omega_k)$  mit  $\omega_i \in \{1, 2, \dots, n\}$ , und es gibt  $n^k$  Möglichkeiten. (Bei diesem Beispiel kommt es auf die Reihenfolge an! Soll es nicht darauf ankommen, ist das ein anderes Experiment, siehe unten.)

Will man Reihenfolgen ignorieren, muss man wissen, in wievielen verschiedenen Reihenfolgen man die Elemente einer endlichen Menge ordnen kann; man nennt so etwas eine *Anordnung* oder *Permutation* der endlichen Menge.

**Satz 3.3.** *Die Menge  $X$  habe  $n$  Elemente  $x_1, \dots, x_n$ . Dann lassen sich diese Elemente auf*

$$n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n = \prod_{k=1}^n k$$

*Weisen anordnen. (Das Symbol  $n!$  spricht man „ $n$  Fakultät“.)*

Am einfachsten kann man sich das durch folgende Überlegung plausibel machen: Für das erste Element  $x_1$  stehen  $n$  Plätze zur Verfügung, für das zweite dann noch  $n - 1$  (denn ein Platz ist bereits durch  $x_1$  belegt), für das dritte  $n - 2$ , und so weiter, bis schließlich für das letzte Element noch genau ein Platz frei ist. Das ergibt  $n(n-1)(n-2) \cdot 2 \cdot 1 = n!$  Möglichkeiten. (Streng genommen sollte man einen Induktionsbeweis führen.)

**Beispiele 3.4.** (a) Es gibt  $3! = 6$  Möglichkeiten, die Buchstaben  $A, B, C$  anzuordnen, nämlich

$$ABC, \quad ACB, \quad BAC, \quad ,BCA, \quad CAB, \quad CBA.$$

(b) Eine 7-elementige Menge hat  $7! = 5040$  Permutationen.

Wir ziehen nun wieder  $k$  Kugeln aus einer Urne, in der sich  $n$  verschiedene Kugeln befinden, aber diesmal, ohne die Kugeln zurückzulegen; es muss dann natürlich  $k \leq n$  sein. Bei der ersten Ziehung können wir aus  $n$  Kugeln wählen, bei der zweiten gibt es noch  $n - 1$  Möglichkeiten, bei der dritten  $n - 2$ , und so weiter, bis zur  $k$ -ten Ziehung, bei der noch  $n - k + 1$  Kugeln in der Urne sind. Die

Anzahl der möglichen Resultate, unter Berücksichtigung der Reihenfolge, ist demnach

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Wir fassen zusammen:

**Satz 3.5.** *Werden aus  $n$  verschiedenen Objekten  $k$  unter Berücksichtigung der Reihenfolge ausgewählt, so ist die Zahl der möglichen Resultate*

- (a)  $\frac{n!}{(n-k)!}$ , wenn nicht zurückgelegt wird ( $k \leq n$ ),
- (b)  $n^k$ , wenn zurückgelegt wird ( $k \in \mathbb{N}$  beliebig).

**Beispiel 3.6** (Das Geburtstagsproblem). Wie groß muss eine Gruppe von Personen sein, damit die Wahrscheinlichkeit, dass zwei von ihnen am gleichen Tag Geburtstag haben, mindestens 50% ist? Da es sinnvoll erscheint anzunehmen, dass alle Tage des Jahres als Geburtstage gleichwahrscheinlich sind, handelt es sich bei diesem Problem um ein Laplace-Experiment. Um die Antwort zu finden, muss man also für gegebenes  $k$  die Zahl der Elemente des Ereignisses

$A$ : (Mindestens) zwei aus  $k$  Personen haben am gleichen Tag Geburtstag

bestimmen, also auf wieviele Weisen es geschehen kann, dass von  $k$  Personen zwei am gleichen Tag Geburtstag haben. Zu Vereinfachung gehen wir dabei von 365 Tagen im Jahr aus (ignorieren also Schaltjahre). Leichter zu zählen als das Ereignis  $A$  ist das Komplementärereignis

$\bar{A}$ : alle  $k$  Personen haben verschiedene Geburtstage,

denn dies entspricht einem  $k$ -maligen Ziehen (von Geburtstagen) ohne Zurücklegen, mit dem Ergebnis

$$\#\bar{A} = \frac{365!}{(365-k)!}.$$

Da es insgesamt  $365^k$  mögliche Geburtstagsverteilungen gibt, folgt

$$P(\bar{A}) = \frac{\#\text{Fälle, in denen } A \text{ eintritt}}{\#\text{insgesamt mögliche Fälle}} = \frac{365!}{365^k (365-k)!}$$

und

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

Man erhält so zum Beispiel für

$$\begin{aligned} k = 22: & \approx 0.476 = 47.6\%, \\ k = 23: & \approx 0.507 = 50.7\%, \\ k = 29: & \approx 0.681 = 68.1\%, \\ k = 57: & \approx 0.990 = 99.0\%. \end{aligned}$$

Es genügt also eine Gruppe von 23 Personen.

Bisher hatten wir beim Ziehen auf die Reihenfolge Wert gelegt. Ist sie unwichtig (wie etwa beim Zahlenlotto), so ist das ein anderes Zufallsexperiment. Um beim Ziehen ohne Zurücklegen die Zahl der möglichen Ergebnisse zu erhalten, muss man noch durch die Zahl der Anordnungen des Ergebnisses teilen. Für das Ziehen mit Zurücklegen ist die Angelegenheit ein klein wenig komplizierter, lässt sich aber auf den ersten Fall zurückführen: Das Ziehen von  $k$  Kugeln aus einer Urne mit  $n$  verschiedenen Kugeln mit Zurücklegen ist, wenn man die Reihenfolge ignoriert, analog zum Verteilen von Markierungen zwischen Kugeln verschiedener Farbe: Stellen wir uns eine Reihe von Plätzen vor, auf die wir Kugeln und Markierungen verteilen, so haben wir  $n-1$  Markierungen und  $k$  Kugeln, also  $n-1+k$  Plätze, und das einzige, worauf es ankommt, ist, wo die Markierungen beziehungsweise die Kugeln sitzen. Damit ist dieses Experiment gleichbedeutend mit der Auswahl von  $k$  aus  $n+k-1$  Objekten **ohne** Zurücklegen.

Mit diesen Überlegungen sollte der folgende Satz plausibel sein:

**Satz 3.7.** Werden aus  $n$  verschiedenen Objekten  $k$  ohne Berücksichtigung der Reihenfolge ausgewählt, so ist die Zahl der möglichen Resultate

- (a)  $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ , wenn nicht zurückgelegt wird ( $k \leq n$ ),  
 (b)  $\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$ , wenn zurückgelegt wird ( $k$  beliebig).

Die in diesem Satz auftauchenden Symbole  $\binom{n}{k}$  nennt man *Binomialkoeffizienten* und spricht sie „ $n$  über  $k$ “. Eine andere Formulierung von Teil (a) ist, dass eine  $n$ -elementige Menge  $X$  genau  $\binom{n}{k}$   $k$ -elementige Teilmengen besitzt. Da eine  $k$ -elementige Teilmenge  $A$  durch ihr  $(n-k)$ -elementiges Komplement  $X \setminus A$  eindeutig bestimmt ist, gilt

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

was auch sofort aus der definierenden Formel hervorgeht. Nützlich ist oft die folgende Formel:

**3.8.** Für natürliche Zahlen  $1 \leq k \leq n$  gilt

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

Dies rechnet man einfach nach: es ist

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

Mit Hilfe dieser Formel kann man leicht eine Liste von Binomialkoeffizienten in einem *Pascalschen Dreieck* anfertigen: in jeder Zeile stehen die Binomialkoeffizienten  $\binom{n}{0}, \dots, \binom{n}{n}$ , und die nächste Zeile erhält man, indem man je zwei aufeinanderfolgende Einträge addiert und zwischen diesen beiden Zahlen (eine Zeile tiefer) plaziert.

$n$	$\binom{n}{k}$									
0	1									
1		1		1						
2			1	2	1					
3			1	3	3	1				
4			1	4	6	4	1			
5			1	5	10	10	5	1		
6		1	6	15	20	15	6	1		

Die Spiegelsymmetrie dieses Schemas entspricht der schon erwähnten Identität

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

**Beispiel 3.9** (Zahlenlotto). Die Zahlenlotterie „6 aus 49“ ist ein Zufallsexperiment, bei dem 6 Zahlen aus der Menge  $\{1, 2, \dots, 49\}$  ohne Zurücklegen gezogen werden. Es gibt also

$$\binom{49}{6} = 13\,983\,816$$

verschiedene Möglichkeiten.

Wie sieht jedoch die Situation bei 3 Richtigen aus, also wieviele Möglichkeiten gibt es, dass von 6 getippten Zahlen 3 unter den 6 gezogenen Zahlen vorkommen? Nennen wir dieses Ereignis  $A_3$ ; es entspricht der Auswahl einer 3-elementigen Teilmenge einer 6-elementigen Menge und andererseits (unabhängig davon) der Auswahl einer 3-elementigen Teilmenge (den Nieten) aus einer 43-elementigen Menge (den nicht gezogenen Zahlen). Man erhält somit

$$\#A_3 = \binom{6}{3} \cdot \binom{43}{3} = 246\,820$$

Möglichkeiten.

Bei 4 Richtigen erhält man analog

$$\#A_4 = \binom{6}{4} \cdot 432 = 13\,545,$$

bei 5 Richtigen  $\#A_5 = 6 \cdot 43 = 258$  (eine von den 6 Richtigen (nicht), eine von den 43 Nieten), und für 6 Richtige gibt es schließlich genau eine Möglichkeit. Für das Ereignis  $A$ : „mindestens drei Richtige“ gibt es also

$$\#A = \#(A_3 \cup A_4 \cup A_5 \cup A_6) = \#A_1 + \#A_2 + \#A_3 + \#A_4 = 260\,624$$

verschiedene Möglichkeiten; die Wahrscheinlichkeit, mindestens drei Richtige zu haben ist demnach

$$P(A) = \frac{\#A}{13\,983\,816} \approx 0.0186 = 1.86\%,$$

ziemlich gering also. Noch schlimmer wird es bei 4 Richtigen, dort ergibt sich eine Wahrscheinlichkeit von 0.09 Prozent.

Ein weiteres, häufig auftretendes Problem ist das folgende: Angenommen wir haben  $k$  Kugeln, die wir auf  $n$  Urnen verteilen wollen, so dass sich in der  $i$ -ten Urne  $k_i$  Kugeln befinden (dann ist  $k_1 + k_2 + \dots + k_n = k$ ). Um zu ermitteln, wieviele mögliche Aufteilungen es gibt, ordnen wir die  $k$  Kugeln zunächst beliebig an – dafür gibt es  $k!$  Möglichkeiten – und legen die ersten  $k_1$  Kugeln in die erste Urne, die nächsten  $k_2$  in die zweite, und so weiter. Jede der  $k_i!$  Permutationen des Inhalts der  $i$ -ten Urne liefert das gleiche Ergebnis, so dass wir die gewünschte Zahl erhalten, wenn wir durch  $k_1! \cdot k_2! \cdot \dots \cdot k_n!$  teilen. Dies zeigt:

**Satz 3.10.** Sei  $k = k_1 + k_2 + \dots + k_n$ . Die Anzahl der Möglichkeiten,  $k$  Objekte auf  $n$  Abteilungen mit  $k_1, \dots, k_n$  Elementen aufzuteilen, ist

$$\frac{k!}{k_1! k_2! \dots k_n!}.$$

**Beispiel 3.11.** Ein Restaurant habe vier Tische, einen mit zwei Plätzen, zwei mit vier Plätzen und einen mit sechs Plätzen. Dann kann man 16 Personen auf

$$\frac{16!}{2!4!4!6!} = 25\,225\,200$$

Weisen auf die vier Tische verteilen.



#### 4. Bedingte Wahrscheinlichkeit

Oft tritt die Frage auf, wie wahrscheinlich ein Ereignis  $A$  unter der Bedingung ist, dass ein anderes Ereignis  $B$  bereits eingetreten ist. In diesem Fall hängt die Eintrittswahrscheinlichkeit von  $A$  von der Wahrscheinlichkeit von  $B$  ab.

**Beispiel 4.1.** Eine Familie hat zwei Kinder. Wir betrachten die Frage nach dem Geschlecht der Kinder als Zufallsexperiment; der Ergebnisraum ist dann

$$\Omega = \{(J, J), (J, M), (M, J), (M, M)\},$$

wobei  $J$  für Junge und  $M$  für Mädchen stehe. Dann ist die Wahrscheinlichkeit, dass beide Kinder Mädchen sind, die Wahrscheinlichkeit des Ereignisses  $\{(M, M)\}$ , also  $1/4$ .

Wie ist aber die Wahrscheinlichkeit, dass beides Mädchen sind, wenn wir schon wissen, dass eines der Kinder ein Mädchen ist? Wenn man nicht aufpasst, könnte man glauben, diese sei  $1/2$ , denn es geht ja nur noch um das zweite Kind. Aber das ist falsch, denn es ist ja nicht gesagt, ob das ältere oder das jüngere Kind ein Mädchen ist. Am einfachsten untersucht man dies, indem man die richtige Ergebnismenge betrachtet: Ist ein Kind ein Mädchen, so ist die Ergebnismenge für diese Frage

$$\Omega' = \Omega \setminus \{(J, J)\} = \{(J, M), (M, J), (M, M)\}$$

mit drei Elementen, so dass die gesuchte Wahrscheinlichkeit  $1/3$  ist.

Die Wahrscheinlichkeit  $1/2$  erhält man, wenn man schon weiss, dass das erstgeborene Kind ein Mädchen ist, denn dann reduziert sich die Ergebnismenge auf  $\{(M, J), (M, M)\}$ .

Eine letzte Variante: Wie hoch ist die Wahrscheinlichkeit, dass beide Kinder Mädchen sind, wenn eines der Kinder ein Mädchen namens Ida ist? Hier scheint einem die Intuition zu sagen, der Name spielt doch keine Rolle, die Wahrscheinlichkeit ist nach wie vor  $1/3$ . Weit gefehlt! Betrachten wir zunächst das Experiment ohne die Zusatzbedingung „eines der Kinder ist ein Mädchen namens Ida“: Schreibt man  $I$  für „Mädchen namens Ida“ und  $M\bar{I}$  für „Mädchen, dessen Name nicht Ida ist“, so erhält man als Ergebnismenge

$$\Omega = \{(J, J), (J, MI), (J, M\bar{I}), (MI, J), (M\bar{I}, J), (MI, MI), (MI, M\bar{I}), (M\bar{I}, MI), (M\bar{I}, M\bar{I})\}.$$

Durch die Bedingung, eines der Kinder sei ein Mädchen namens Ida, reduziert sich die Ergebnismenge zu

$$\Omega' = \{(J, MI), (MI, J), (MI, M\bar{I}), (M\bar{I}, MI)\},$$

jedenfalls wenn man ausschließt, dass die Eltern den Namen Ida zweimal vergeben. Lässt man eine mögliche Präferenz der Eltern bei der Namensgebung für die Erstgeborene außer acht, kann man annehmen, dass diese vier Fälle gleich wahrscheinlich sind. In zwei von ihnen sind beide Kinder Mädchen, die Wahrscheinlichkeit ist also  $1/2$ . Wenn das nicht überzeugt, der stelle sich vor, in einem riesigen Raum seien 30 Millionen Familien versammelt, die zwei Kinder haben, von denen eines ein Mädchen ist. Von vorn wissen wir schon, dass etwa 10 Millionen Familien zwei Mädchen und etwa 20 Millionen nur ein Mädchen haben werden. Jetzt versammeln wir nur die Familien, in denen eines der Mädchen Ida heißt. Nun ist Ida ein seltener Name, er kommt unter 200 000 Mädchen etwa einmal vor (eine grobe, aber realistische Schätzung). Von den Ein-Mädchen-Familien bleiben daher etwa 100 übrig, aber eben auch 100 von den Zwei-Mädchen-Familien, 50, weil die Erstgeborene Ida heißt, und 50, weil die Zweitgeborene so heißt. (Eine solche Simulation ist oft hilfreich zum Verstehen, auch wenn sie natürlich keine Beweiskraft hat.)

Die Frage nach der Berechnung der Wahrscheinlichkeit von Ereignissen, die von anderen Ereignissen abhängen, wurde systematisch zum ersten Mal von dem englischen Geistlichen Thomas Bayes (1701-1761) untersucht. Hier noch ein weiteres Beispiel:

**Beispiel 4.2.** Eine Firma hat einen Metalldetektor für die Gepäckkontrolle entwickelt, der 90% der Koffer, die metallische Gegenstände enthalten, entdeckt, aber auch mit Wahrscheinlichkeit 5% Alarm schlägt, wenn sich in dem Koffer kein Metall befindet. Die Wahrscheinlichkeit, dass ein beliebiger Koffer etwas metallisches enthält, sei 20%.

**1. Frage:** Mit welcher Wahrscheinlichkeit löst der Detektor bei einem beliebigen Koffer aus?

**2. Frage:** Mit welcher Wahrscheinlichkeit ist tatsächlich Metall im Koffer, wenn der Detektor das behauptet?

Bei dem Zufallsexperiment „Untersuchung eines Koffers“ hat man es mit zwei Ereignissen zu tun, die voneinander abzuhängen scheinen:

*A:* Der Detektor schlägt an.

*B:* Es ist Metall im Koffer.

Bekannt ist  $P(B) = 0.2$ , aber auch die folgenden Wahrscheinlichkeiten:

- die Wahrscheinlichkeit, dass der Detektor anschlägt (Ereignis  $A$ ), wenn Metall im Koffer ist (Ereignis  $B$ ) – wir wollen sie im Vorgriff auf spätere Notation mit  $P(A|B)$  bezeichnen – beträgt  $P(A|B) = 0.9$ ,
- die Wahrscheinlichkeit, dass der Detektor anschlägt (Ereignis  $A$ ), wenn kein Metall im Koffer ist (Ereignis  $\bar{B}$ ), ist  $P(A|\bar{B}) = 0.05$

Da entweder Metall im Koffer ist oder nicht, erhalten wir die gesuchte Wahrscheinlichkeit, indem wir die Wahrscheinlichkeiten  $P(A|B)$  und  $P(A|\bar{B})$  mit den Wahrscheinlichkeiten ihrer Bedingungen  $B, \bar{B}$  gewichten und die Ergebnisse addieren:

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) = 0.9 \cdot 0.2 + 0.05 \cdot 0.8 = 0.22.$$

Die zweite Frage ist die nach der Wahrscheinlichkeit, dass  $B$  eintritt, falls  $A$  eingetreten ist, also  $P(B|A)$ ; dies berechnen wir gleich.

Gegeben sei ein Zufallsexperiment mit Ergebnismenge  $\Omega$ , einem Wahrscheinlichkeitsmaß  $P$  und zwei Ereignisse  $A, B$  mit  $P(B) > 0$ . Will man die Wahrscheinlichkeit von  $A$  bestimmen unter der Voraussetzung, dass  $B$  eingetreten ist, so hat man eine neue Ergebnismenge, nämlich  $B$ . Das Wahrscheinlichkeitsmaß  $P$  muss dann renormiert werden, so dass das sichere Ereignis  $B$  wieder den Wert 1 zugewiesen bekommt. Dies erreicht man mit Division durch  $P(B)$  und erhält so ein Wahrscheinlichkeitsmaß auf  $B$ . Diese Überlegung ist die Grundlage für die folgende Definition.

**Definition 4.3.** Sei  $\Omega$  die Ergebnismenge eines Zufallsexperiments und  $P$  ein Wahrscheinlichkeitsmaß auf  $\Omega$ . Seien  $A, B$  Ereignisse und  $P(B) > 0$ . Unter der bedingten Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$  versteht man den Wert

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

*Bemerkung.* Auch für bedingte Wahrscheinlichkeiten gilt natürlich die Regel von den Komplementäreignissen, das heißt

$$P(\bar{A}|B) = 1 - P(A|B)$$

wie im Additionssatz 2.2, denn man hat ja nichts weiter getan, als die Ergebnismenge auf  $B$  zu reduzieren.

**Beispiel** (Fortsetzung von 4.2). Wir hatten  $P(B) = 0.2$  und  $P(A|B) = 0.9$ , daraus schließen wir

$$P(A \cap B) = P(A|B)P(B) = 0.9 \cdot 0.2 = 0.18$$

und damit

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.18}{0.22} \approx 0.818 = 81.8\%.$$

Wir haben auch noch

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}|\bar{B})P(\bar{B}) = (1 - P(A|\bar{B}))P(\bar{B}) = 0.95 \cdot 0.8 = 0.76$$

und daher

$$P(\overline{B}|\overline{A}) = \frac{P(\overline{A} \cap \overline{B})}{P(\overline{A})} = \frac{P(\overline{A} \cap \overline{B})}{1 - P(A)} = \frac{0.76}{0.78} \approx 0.974,$$

das heißt wenn der Detektor anschlägt, geschieht dies in 81.8% der Fälle zu recht, und wenn er nicht anschlägt, ist mit Wahrscheinlichkeit 97.4% auch kein Metall im Koffer.

Aus all diesen Beispielen geht hervor, dass man um die Wahrscheinlichkeit des gemeinsamen Eintretens zweier Ereignisse zu berechnen, im allgemeinen die Einzelwahrscheinlichkeiten nicht einfach multiplizieren darf; dies geht nur, wenn die Ereignisse nicht voneinander abhängig sind.

**Definition 4.4.** (a) Zwei Ereignisse  $A, B$  eines Zufallsexperiments heißen *stochastisch unabhängig*, wenn  $P(A|B) = P(A)$  (oder äquivalent  $P(B|A) = P(B)$ ), also

$$P(A \cap B) = P(A)P(B)$$

*gilt.*

(b)  $n$  Ereignisse  $A_1, \dots, A_n$  heißen (*vollständig*) *stochastisch unabhängig*, wenn für jede Auswahl  $\{i_1, \dots, i_k\}$ ,  $2 \leq k \leq n$ , von  $k$  verschiedenen Zahlen  $i_j \in \{1, 2, \dots, n\}$

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

*gilt.*

**Beispiele 4.5.** (a) Beim Würfeln mit einem Würfel ist die Ergebnismenge bekanntlich  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Wir untersuchen die Ereignisse

$A$ : „gerade Augenzahl“, also  $A = \{2, 4, 6\}$  mit  $P(A) = 1/2$ ,

$B$ : „Augenzahl  $> 3$ “, also  $B = \{4, 5, 6\}$  mit  $P(B) = 1/2$ ,

$C$ : „weder 1 noch 6 Augen“, also  $C = \{2, 3, 4, 5\}$  mit  $P(C) = 2/3$

auf stochastische Unabhängigkeit. Es gilt

$$\begin{aligned} P(A \cap B) &= P(\{4, 6\}) = \frac{1}{3} \neq P(A)P(B) = \frac{1}{4} \\ P(A \cap C) &= P(\{2, 4\}) = \frac{1}{3} = P(A)P(C) \\ P(B \cap C) &= P(\{4, 5\}) = \frac{1}{3} = P(B)P(C) \\ P(A \cap B \cap C) &= P(\{4\}) = \frac{1}{6} = P(A)P(C)P(C) \end{aligned}$$

Folglich sind  $A$  und  $C$  sowie  $B$  und  $C$  stochastisch unabhängig, aber  $A, B, C$  nicht vollständig unabhängig.

(b) Beim Würfeln mit zwei Würfeln ist  $\Omega = \{(i, j) \mid i, j \in \{1, \dots, 6\}\}$ . Wir betrachten hier die Ereignisse

$A$ : „1. Würfel zeigt eine 6“,

$B$ : „2. Würfel zeigt eine 6“,

$C$ : „beide Würfel zeigen die gleiche Zahl“

mit  $P(A) = P(B) = P(C) = 1/6$ . Es ist  $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/36$ . was jeweils dem Produkt der Einzelwahrscheinlichkeiten entspricht, so dass je zwei der Ereignisse unabhängig sind. Es gilt jedoch

$$P(A \cap B \cap C) = P(\{(6, 6)\}) = \frac{1}{36} \neq \frac{1}{216} = P(A)P(B)P(C),$$

so dass  $A, B, C$  nicht vollständig unabhängig sind.

(c) Wir betrachten den  $n$ -fach wiederholten Wurf einer Münze. Sei  $p$  die Wahrscheinlichkeit von Kopf und damit  $1 - p$  die Wahrscheinlichkeit von Zahl. Seien ferner  $A_j$  und  $K_k$  die Ereignisse

$A_j$ : beim  $j$ -ten Wurf erscheint Kopf,

$K_k$ : unter den  $n$  Würfeln tritt  $k$  mal Kopf auf.

Die Ereignisse  $A_j$  sind offenbar unabhängig. Ist nun  $\{\omega\} \in K_k$  ein Elementarereignis, bei dem bei den Würfeln Nummer  $i_1, \dots, i_k$  ( $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ) Kopf und bei den übrigen Würfeln  $i_{k+1}, \dots, i_n$  Zahl gewürfelt wurde, so ist

$$\{\omega\} = A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap \bar{A}_{i_{k+1}} \cap \bar{A}_{i_{k+2}} \cap \dots \cap \bar{A}_{i_n}$$

mit

$$P(\{\omega\}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}) \cdot P(\bar{A}_{i_{k+1}})P(\bar{A}_{i_{k+2}}) \cdots P(\bar{A}_{i_n}) = p^k(1-p)^{n-k}.$$

Es ist aber  $\#K_k = \binom{n}{k}$ , so dass

$$P(K_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

ist.

*Bemerkung.* Wendet man die Formel  $P(A \cap B) = P(A|B)P(B)$  auf drei Ereignisse an, erhält man

$$P(A \cap B \cap C) = P(A|B \cap C)P(B \cap C) = P(A|B \cap C)P(B|C)P(C).$$

Fortgesetzte Anwendung auf  $n$  Ereignisse  $A_1, \dots, A_n$  liefert dann

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= P(A_n | A_{n-1} \cap \dots \cap A_1) P(A_{n-1} | A_{n-2} \cap \dots \cap A_1) \cdots P(A_3 | A_2 \cap A_1) P(A_2 | A_1) P(A_1) \\ &= \left( \prod_{i=2}^n P(A_i | A_i \cap A_{i-1} \cap \dots \cap A_1) \right) P(A_1). \end{aligned}$$

Diese Formel nennt man auch den *Multiplikationssatz*.

**Beispiel 4.6.** In einer Urne befinden sich 26 blaue und 4 rote Kugeln. Man entnimmt eine Stichprobe von 4 Kugeln. Wie groß ist die Wahrscheinlichkeit, dass alle 4 Kugeln blau sind? Dazu definieren wir die Ereignisse

$A$ : alle Kugeln sind blau,

$A_i$ : die  $i$ -te Kugel ist blau ( $i = 1, 2, 3, 4$ ),

dann ist  $A = A_1 \cap A_2 \cap A_3 \cap A_4$ . Das Experiment entspricht viermaligem Ziehen ohne Zurücklegen, so dass wir folgende Wahrscheinlichkeiten für die vier Ziehungen haben:

- (i) In der ersten Ziehung gibt es 26 blaue und 4 rote Kugeln, also ist  $P(A_1) = \frac{26}{30}$ .
- (ii) In der zweiten Ziehung gibt es, wenn die erste Kugel blau war, noch 25 blaue unter 29 Kugeln, folglich ist  $P(A_2|A_1) = \frac{25}{29}$ .
- (iii) In der dritten Ziehung gibt es unter der Annahme, dass  $A_1$  und  $A_2$  eingetreten sind, noch 24 blaue unter 28 Kugeln, also gilt  $P(A_3|A_1 \cap A_2) = \frac{24}{28}$ .
- (iv) In der vierten Ziehung schließlich gibt es, wenn die ersten drei Kugeln blau waren, noch 23 blaue unter 27 Kugeln, also ist  $P(A_4|A_1 \cap A_2 \cap A_3) = \frac{23}{27}$ .

Mit dem Multiplikationssatz erhält man damit

$$P(A) = P(A_4|A_1 \cap A_2 \cap A_3)P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1) = \frac{26}{30} \cdot \frac{25}{29} \cdot \frac{24}{28} \cdot \frac{23}{27} \approx 0.546 = 54.6\%.$$

Sei nun  $\Omega$  die Ergebnismenge eines Zufallsexperiments und seien  $B_1, \dots, B_n$  Ereignisse, so dass gilt:

- (i)  $B_i \cap B_j = \emptyset$  für  $i \neq j$  (paarweise Unvereinbarkeit),
  - $\Omega = \bigcup_{i=1}^n B_i$ .

Man sagt dann,  $\Omega$  sei die *disjunkte Vereinigung* der  $B_i$ . Ist nun  $A$  ein weiteres Ereignis, so ist

$$A = \bigcup_{i=1}^n (A \cap B_i)$$

die disjunkte Vereinigung der Schnitte  $A \cap B_i$ , und da sich die Wahrscheinlichkeiten unvereinbarer Ereignisse addieren, folgt

$$P(A) = \sum_{i=1}^n P(A \cap B_i). \quad (4.1)$$

**Satz 4.7** (Satz von der totalen Wahrscheinlichkeit, Formel von Bayes).

Die Ergebnismenge  $\Omega$  eines Zufallsexperiments sei die disjunkte Vereinigung von Ereignissen  $B_1, \dots, B_n$  mit  $P(B_i) > 0$ . Dann gilt für jedes Ereignis  $A$

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

sowie für jedes  $k$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

Nach der Definition bedingter Wahrscheinlichkeit und (4.1) gilt nämlich

$$\sum_{i=1}^n P(A|B_i)P(B_i) = \sum_{i=1}^n \frac{P(A \cap B_i)}{P(B_i)} P(B_i) = \sum_{i=1}^n P(A \cap B_i) = P(A);$$

dies zeigt die erste Formel. Die zweite Formel folgt aus der ersten, denn es ist

$$P(B_k|A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

**Beispiel 4.8.** In einer Automobilfabrik gibt es drei Fließbänder, an denen das gleiche Fahrzeugmodell montiert wird. Das erste Band liefert die Hälfte der Fahrzeugproduktion, das zweite 30% und das dritte die verbleibenden 20%, wobei beim ersten Band erfahrungsgemäß 3%, beim zweiten 4% und beim dritten 1% der Fahrzeuge Mängel aufweisen.

(a) Wie groß ist die Wahrscheinlichkeit, dass ein beliebiges Fahrzeug Mängel hat?

Wir betrachten die Untersuchung eines beliebig der Produktion entnommenen Fahrzeugs als Zufallsexperiment; seien dazu  $A$  und  $B_i$  ( $i = 1, 2, 3$ ) die Ereignisse

$A$ : das Fahrzeug hat Mängel,

$B_i$ : das Fahrzeug stammt vom Band Nummer  $i$ .

Bekannt sind aus der Beschreibung die Wahrscheinlichkeiten  $P(A|B_1) = \frac{3}{100}$ ,  $P(A|B_2) = \frac{4}{100}$ ,  $P(A|B_3) = \frac{1}{100}$ ,  $P(B_1) = \frac{1}{2}$ ,  $P(B_2) = \frac{3}{10}$  und  $P(B_3) = \frac{2}{10}$ . Es ist dann

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) = \frac{3}{200} + \frac{12}{1000} + \frac{2}{1000} = \frac{29}{1000} = 2.9\%.$$

(b) Ein defektes Fahrzeug sei ausgewählt. Mit welcher Wahrscheinlichkeit stammt es von Band Nummer  $i$ ?

Hier benutzen wir die zweite Formel: es ist

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{\frac{3}{200}}{\frac{29}{1000}} = \frac{15}{29} \approx 51.7\%,$$

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{\frac{12}{1000}}{\frac{29}{1000}} = \frac{12}{29} \approx 41.4\%,$$

$$P(B_3|A) = \frac{P(A|B_3)P(B_3)}{P(A)} = \frac{\frac{2}{1000}}{\frac{29}{1000}} = \frac{2}{29} \approx 6.9\%.$$

### 5. Zufallsvariable und Verteilungsfunktion

Bei Zufallsexperimenten kommt es oft nicht so sehr auf den Ausgang des Experiments selbst an, sondern vielmehr auf einen daraus abgeleiteten Wert, der meist in Form einer reellen Zahl daherkommt. Mit anderen Worten, man interessiert sich für Abbildungen  $X$  von der Ergebnismenge  $\Omega$  in eine Menge  $\Omega'$ , die man meist als Menge von Messgrößen auffassen kann.

**Beispiele 5.1.** (a) Die Ergebnismenge des Ziehens von Lotterielosen ist zunächst die Menge der Lose selbst. An dem (physischen) Los ist man allerdings kaum interessiert, sondern nur an dem damit verbundenen Gewinn (der jedoch meist Null ist). In diesem Fall würde  $X$  den Losen den dazugehörigen Gewinn zuordnen.

(b) Beim einfachen Münzwurf hat man als Ergebnismenge  $\{\text{Kopf, Zahl}\}$ , wobei man Kopf und Zahl auch die Zahlen 1 und 0 zuordnen kann. Beim  $n$ -fachen Münzwurf kann  $X$  die Anzahl der Würfle mit dem Ausgang „Kopf“ sein, was in der eben beschriebenen Kodierung durch 0 und 1 die Summe der einzelnen Ergebnisse wäre; genauso ist aber auch die Anzahl der Ergebnisse „Zahl“ möglich.

(c) Beim wiederholten Wurf zweier Würfel kann  $X$  die Anzahl der Würfel mit gleicher Augenzahl bei beiden Würfeln sein, oder die Anzahl der Würfel mit Summe der Augenzahlen  $\geq 5$ .

(d) Ebenso ein Zufallsexperiment ist der Wurf einer Nadel auf eine Ebene, die wir uns als  $xy$ -Ebene vorstellen wollen. Dann definiert der Winkel der Nadel mit der  $x$ -Achse eine solche Abbildung  $X$ , mit Werten im Intervall  $[0, 2\pi]$ .

**Definition 5.2.** Sei  $\Omega$  die Ergebnismenge eines Zufallsexperiments. Eine Zufallsvariable ist eine Funktion  $X$  auf der Menge  $\Omega$ , deren Werte reelle Zahlen sind.

Die Menge aller Werte, die von der Funktion auch tatsächlich angenommen werden, nennt man das *Bild* von  $X$ , das auch in vielen Büchern als *Wertebereich* bezeichnet und daher mit dem Buchstaben  $W$  versehen wird, also

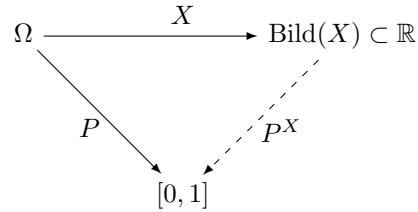
$$W := \text{Bild}(X) = \{X(\omega) \mid \omega \in \Omega\} \subset \mathbb{R}.$$

Sei nun  $P$  ein Wahrscheinlichkeitsmaß auf  $\Omega$  und  $X$  eine Zufallsvariable auf  $\Omega$  mit Bild  $W$ . Für eine Teilmenge  $A \subset W$  kann man nun die Wahrscheinlichkeit bestimmen, dass der Wert der Zufallsvariablen in  $A$  liegt<sup>1</sup>, als die Wahrscheinlichkeit des Ereignisses  $\{\omega \in \Omega \mid X(\omega) \in A\}$ . Mit anderen Worten, man erhält ein Wahrscheinlichkeitsmaß auf  $W = \text{Bild}(X)$ : Für eine Teilmenge  $A \subset W$  (also ein Ereignis) setzt man

$$P^X(A) = P(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

<sup>1</sup>Ohne weitere Einschränkung geht das nur für diskrete Zufallsvariablen, wie sei im nächsten Abschnitt behandelt werden. Ist  $X$  nicht diskret, so gilt dies aber immerhin noch für Intervalle  $A$ .

Ein kleines Diagramm zur Veranschaulichung:



Um die Notation nicht zu überfrachten, schreiben wir meist wieder  $P$  statt  $P^X$ , wenn es aus dem Kontext klar ist, welche Zufallsvariable betrachtet wird.

Oft ist man nicht daran interessiert, ob ein bestimmter Wert angenommen wird, sondern ob ein bestimmter Wert nicht überschritten wird. Diese Wahrscheinlichkeiten werden durch sogenannte Verteilungsfunktionen beschrieben.

**Definition 5.3.** Sei  $X$  eine Zufallsvariable und  $P$  ein Wahrscheinlichkeitsmaß auf der Ergebnismenge  $\Omega$ . Dann heißt die Funktion

$$F = F^X : \mathbb{R} \longrightarrow [0, 1]$$

definiert durch

$$F(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\})$$

die Verteilungsfunktion von  $X$ .

Man benutzt auch vielfach die suggestive Schreibweise

$$P(X \leq x) = F(x).$$

Verteilungsfunktionen haben folgende Eigenschaften, die sich unmittelbar aus den Wahrscheinlichkeitsaxiomen ergeben:

**5.4.** Sei  $F$  die Verteilungsfunktion einer Zufallsvariablen  $X$ . Dann gilt:

- (a)  $F$  ist monoton wachsend,
- (b)  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,
- (c)  $\lim_{x \rightarrow \infty} F(x) = 1$ ,
- (d)  $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$ ,
- (e)  $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$ ,

## 6. Diskrete Zufallsvariablen

Wir werden uns zunächst nur mit Zufallsvariablen beschäftigen, deren Bild sogar eine abzählbare Teilmenge  $\{x_0, x_1, x_2, \dots\}$  von  $\mathbb{R}$  ist, wie zum Beispiel die natürlichen Zahlen, oder jede endliche Teilmenge von  $\mathbb{R}$ . Solche Zufallsvariablen nennt man *diskret*. In den einleitenden Beispielen 5.1 sind die in (a) – (c) betrachteten Zufallsvariablen diskret, aber auch in Beispiel (d) kann man eine diskrete Zufallsvariable etwa dadurch erklären, dass man das Intervall  $[0, 2\pi]$  in endlich viele Teilintervalle  $I_1, \dots, I_n$  aufteilt und  $X(\omega)$  als die Nummer  $i$  des Intervalls definiert, in dem der beobachtete Winkel liegt. (Damit trägt man der Tatsache Rechnung, dass die Messung des Winkels sowieso nicht beliebig genau sein kann.)

Sei nun  $P$  ein Wahrscheinlichkeitsmaß und  $X$  eine diskrete Zufallsvariable auf einer Ergebnismenge  $\Omega$ . Wie vorher erhalten wir ein Wahrscheinlichkeitsmaß auf  $W$ ; für die Wahrscheinlichkeit eines

einzelnen Wertes  $x \in W$  schreibt man auch  $P(X(\omega) = x)$  oder noch kürzer als  $P(X = x)$ , wenn dies zu keinen Missverständnissen führt. Damit gilt

$$P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\}) = \sum_{\substack{\omega \in \Omega \text{ mit} \\ X(\omega) = x}} P(\{\omega\}).$$

**Beispiel 6.1.** Beim Würfeln mit zwei Würfeln ist

$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\} = \{(\omega_1, \omega_2) \mid \omega_i \in \{1, \dots, 6\}\}.$$

Sei  $X$  die Summe der gewürfelten Augenzahlen, also  $X(\omega_1, \omega_2) = \omega_1 + \omega_2$ . Dann ist  $W = \text{Bild}(X)$  die Menge  $\{2, 3, \dots, 12\}$ , und man hat folgende Wahrscheinlichkeiten:

$x$	$\{\omega \mid X(\omega) = x\}$	$P(x)$
2	(1, 1)	1/36
3	(2, 1), (1, 2)	2/36
4	(3, 1), (2, 2), (1, 3)	3/36
5	(4, 1), (3, 2), (2, 3), (1, 4)	4/36
6	(5, 1), (4, 2), (3, 3), (2, 4), (1, 5)	5/36
7	(6, 1), (5, 2), (4, 3), (3, 4), (2, 5), (1, 6)	6/36
8	(6, 2), (5, 3), (4, 4), (3, 5), (2, 6)	5/36
9	(6, 3), (5, 4), (4, 5), (3, 6)	4/36
10	(6, 4), (5, 5), (4, 6)	3/36
11	(6, 5), (5, 6)	2/36
12	(6, 6)	1/36

Man erhält also

$$P(X = k) = \frac{6 - |k - 7|}{36}.$$

Sei  $P$  ein Wahrscheinlichkeitsmaß und  $X$  eine diskrete Zufallsvariable auf  $\Omega$  mit  $W = \text{Bild}(X) = \{x_0, x_1, \dots\}$ , so dass  $x_i < x_{i+1}$ . Schreibt man  $p_i := P(X = i)$ , so muss offenbar  $\sum_{i \geq 0} p_i = 1$  sein. Für die zugehörige Verteilungsfunktion gilt

$$F(x) = \sum_{i: x_i \leq x} P(X = x_i) = \sum_{i: x_i \leq x} p_i.$$

Insbesondere ist  $F(x) = 0$ , wenn  $x < x_0$  ist. Für einen endlichen Wertebereich  $\{x_0, x_1, \dots, x_n\}$  hat man

$$F(x) = \begin{cases} 0 & \text{für } x < x_0, \\ \sum_{i=0}^k p_i & \text{für } x_k \leq x < x_{k+1}, k = 0, 1, 2, \dots, n-1, \\ 1 & \text{für } x_n \leq x. \end{cases}$$

Für diskrete Zufallsvariablen sind die Verteilungsfunktionen also Treppenfunktionen, die an den Stellen  $x_i$  Sprünge haben (können).

**Beispiel 6.2.** Wie groß ist die Wahrscheinlichkeit, beim Wurf zweier Würfel mindestens 9 Augen zu erzielen? Dazu betrachten wir die Zufallsvariable  $X$  von Beispiel 6.1 mit  $X(\omega_1, \omega_2) = \omega_1 + \omega_2$ . Aus der Tabelle entnehmen wir die Wahrscheinlichkeiten  $p_k = P(X = k)$  der einzelnen Werte und erhalten

$$P(X \geq 9) = p_9 + p_{10} + p_{11} + p_{12} = \frac{4 + 3 + 2 + 1}{36} = \frac{5}{18}.$$

Als nächstes wollen wir einige spezielle Verteilungen untersuchen.

#### A. Laplace- oder Gleichverteilung.

Eine Zufallsvariable  $X$  mit  $W = \text{Bild}(X) = \{x_1, \dots, x_n\}$  heißt gleichverteilt, wenn gilt:

$$p_i := P(X = x_i) = \frac{1}{n} \quad (i = 1, 2, \dots, n).$$



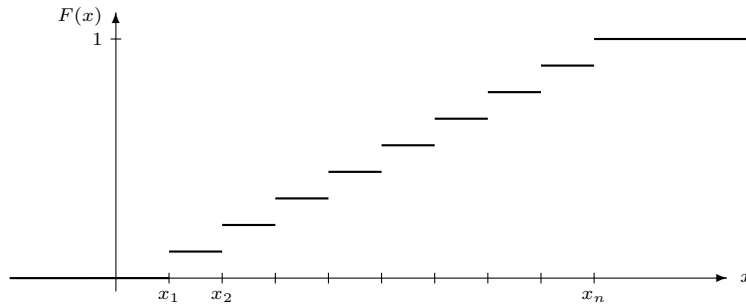


ABBILDUNG 6. Verteilungsfunktion der Laplace-Verteilung

Die zugehörige Verteilungsfunktion ist

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < x_1 \\ \frac{k}{n} & x_k \leq x < x_{k+1}, \quad k = 1, 2, \dots, n-1, \\ 1 & x_n \leq x. \end{cases}$$

**Beispiel 6.3.** Der Wurf eines Würfels mit  $W = \{1, 2, 3, 4, 5, 6\}$  und  $X = \text{Augenzahl}$  ist gleichverteilt.

### B. Binomialverteilung.

Eine Zufallsvariable  $X$  mit  $W = \{0, 1, 2, \dots, n\}$  heißt binomialverteilt, wenn für ein festes  $p$  mit  $0 < p < 1$  gilt:

$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} =: b(i, n, p).$$

Aus dem binomischen Lehrsatz ergibt sich, dass dies tatsächlich eine Wahrscheinlichkeit ist, denn es gilt

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

Diese Verteilung hat die beiden Parameter  $n$  und  $p$ ; man nennt eine so verteilte Zufallsvariable daher auch  $\mathcal{B}_{n,p}$ -verteilt.

Die zugehörige Verteilungsfunktion ist

$$F(k) = P(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}, \quad 0 \leq k \leq n.$$

Zufallsexperimente, die aus  $n$ -maliger Wiederholung eines Experiments mit zwei möglichen Ausgängen 1 (für Erfolg) mit Wahrscheinlichkeit  $p$  und 0 (für Misserfolg) mit Wahrscheinlichkeit  $1-p$  bestehen, sind binomialverteilt. Genauer gesagt ist die Zufallsvariable  $X$ , die die Anzahl der erfolgreichen Ausgänge zählt, binomialverteilt: dies hatten wir in Beispiel 4.5 bereits ausgerechnet.

**Beispiel 6.4** (Qualitätskontrolle durch eine Stichprobe I). Ein Lieferant gibt an, dass seine Lieferung eine Ausschussquote von 2% hat. Das will man überprüfen, indem man 100 mal ein Einzelstück der Lieferung zufällig herausgreift und prüft (Stichprobe mit Zurücklegen). Die Lieferung soll zurückgewiesen werden, wenn mehr als 4 mal ein defektes Teil gezogen wird. Die Entnahme einer solchen Stichprobe ist  $\mathcal{B}_{n,p}$ -verteilt mit  $p = 0.02$  und  $n = 100$ . Die Wahrscheinlichkeit,  $k$  defekte Einzelstücke zu finden, ist dann

$$p_k = \binom{100}{k} (0.02)^k (0.98)^{100-k}.$$

Gefragt ist  $P(X > 4)$  (Ausschlusskriterium). Man errechnet

$$p_0 = 0.1326, \quad p_1 = 0.2706, \quad p_2 = 0.2734, \quad p_3 = 0.1822, \quad p_4 = 0.0902$$

und daher

$$P(X > 4) = 1 - P(X \leq 4) = 1 - (p_0 + p_1 + p_2 + p_3 + p_4) = 0.051.$$

Die Lieferung wird also mit Wahrscheinlichkeit 5.1% abgelehnt.

### C. Geometrische Verteilung.

Eine Zufallsvariable  $X$  mit Bild  $W = \{1, 2, 3, \dots\}$  heißt geometrisch verteilt, wenn für jedes  $k \in W$

$$p_k = P(X = k) = p(1-p)^{k-1}$$

für ein festes  $p$ ,  $0 < p < 1$ , gilt.

Diese Verteilung liegt dann vor, wenn bei der wiederholten Durchführung eines Experiments mit zwei Ausgängen  $\omega, \bar{\omega}$  mit Wahrscheinlichkeiten  $p$  beziehungsweise  $1-p$  die Wahrscheinlichkeit gefragt ist, dass in der  $k$ -ten Wiederholung das Ergebnis  $\omega$  zum ersten Mal auftritt, denn  $(k-1)$ -mal  $\bar{\omega}$  hintereinander liefert die Wahrscheinlichkeit  $(1-p)^{k-1}$ , die dann noch mit der Eintrittswahrscheinlichkeit  $p$  von  $\omega$  multipliziert werden muss.

Wie prüfen wieder nach, dass diese Verteilung den Wahrscheinlichkeitsaxiomen genügt: Da  $0 < p < 1$  ist, folgt aus der Summenformel für die geometrische Reihe

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

Ferner gilt für die Verteilungsfunktion

$$F(k) = P(X \leq k) = \sum_{i=1}^k p(1-p)^{i-1} = p \sum_{j=0}^{k-1} (1-p)^j = p \cdot \frac{1-(1-p)^k}{1-(1-p)} = 1 - (1-p)^k.$$

Dies ist auch klar, denn das Komplementärereignis zu  $\{\omega \mid X(\omega) \leq k\}$  ist das Ereignis, dass jede der ersten  $k$  Wiederholungen das Ergebnis  $\bar{\omega}$  hat.

**Beispiel 6.5.** Was ist die Wahrscheinlichkeit, fünfmal hintereinander keine 1 zu würfeln? Die Wahrscheinlichkeit, eine 1 zu würfeln, ist  $p = 1/6$ , die Wahrscheinlichkeit des Komplementärereignisses („keine 1“) also  $5/6$ . Hier ist  $X$  die Zufallsvariable, die einer Folge von Würfeln den ersten Versuch, in dem eine 1 auftritt, zuweist; gesucht ist die Wahrscheinlichkeit  $P(X \geq 6)$ . Es gilt

$$P(X \geq 6) = 1 - P(X \leq 5) = 1 - \sum_{i=1}^5 \frac{1}{6} \left(\frac{5}{6}\right)^{i-1} = 1 - 0.598 = 0.402 = 40.2\%.$$

### D. Hypergeometrische Verteilung.

Gegeben sei eine Menge von  $N$  Objekten, die in zwei Sorten auftreten,  $d$  von der Sorte 1 und  $N-d$  von der Sorte 2. Man entnimmt eine Stichprobe von Umfang  $n$ , ohne Zurücklegen. Wie ermittelt man die Wahrscheinlichkeit, genau  $k$  Objekte ( $k \leq d$ ) der Sorte 1 gezogen zu haben?

Numerieren wir die Elemente Ausgangsmenge mit  $1, 2, \dots, N$  durch, so können wir die Ergebnismenge dieses Experiments als die Menge

$$\Omega = \{A \subset \{1, 2, \dots, N\} \mid \#A = n\}$$

der  $n$ -elementigen Teilmengen von  $\{1, 2, \dots, N\}$  schreiben; nach Satz 3.7 ist  $\#\Omega = \binom{N}{n}$ . Als Zufallsvariable  $X$  nehmen wir die Abbildung, die einer Stichprobe  $S \in \Omega$  die Zahl der in ihr vorkommenden Objekte der Sorte 1 zuordnet; der Wertebereich dieser Zufallsvariablen ist dann

$$W = \text{Bild}(X) = \{0, 1, 2, \dots, \min(n, d)\}.$$

Es bezeichne  $K$  das Ereignis

$$K = \{S \mid S \text{ enthält } k \text{ Objekte der Sorte 1}\}.$$

Gefragt ist dann  $P(X = k) = P(K)$ . Die Anzahl der Möglichkeiten, aus  $d$  Objekten  $k$  auszuwählen, beträgt  $\binom{d}{k}$ . Außerdem müssen noch  $n - k$  Objekte der zweiten Sorte gewählt werden, dafür gibt es  $\binom{N-d}{n-k}$  Möglichkeiten. Folglich ist  $\#K = \binom{d}{k} \binom{N-d}{n-k}$  und daher

$$p_k := P(X = k) = \frac{\binom{d}{k} \binom{N-d}{n-k}}{\binom{N}{n}} =: \mathcal{H}_{N,d,n}(k).$$

Diese Verteilung nennt man die hypergeometrische Verteilung mit Parametern  $N$ ,  $d$  und  $n \leq N$ , und die Zufallsvariable  $H$  hypergeometrisch oder  $\mathcal{H}_{N,d,n}$ -verteilt.

Um einzusehen, dass es sich dabei wirklich um ein Wahrscheinlichkeitsmaß auf  $\text{Bild}(X)$  handelt, muss man zeigen, dass die Summe aller  $p_k$  ( $0 \leq k \leq \min(n, d)$ ) Eins ergibt. Dazu zeigen wir zunächst

$$\sum_{k=0}^{\min(n,d)} \binom{d}{k} \binom{N-d}{n-k} = \binom{N}{n} \quad (6.1)$$

mit Hilfe des binomischen Lehrsatzes: diesem Satz zufolge gilt

$$(1+x)^d = \sum_{i=0}^d \binom{d}{i} x^i \quad \text{sowie} \quad (1+x)^{N-d} = \sum_{j=0}^d \binom{N-d}{j} x^j;$$

multipliziert man diese beiden Ausdrücke erhält man

$$\begin{aligned} \sum_{s=0}^N \binom{N}{s} x^s &= (1+x)^N = (1+x)^d \cdot (1+x)^{N-d} \\ &= \left( \sum_{i=0}^d \binom{d}{i} x^i \right) \cdot \left( \sum_{j=0}^d \binom{N-d}{j} x^j \right) = \sum_{s=0}^N \left( \sum_{k=0}^{\min(n,d)} \binom{d}{k} \binom{N-d}{n-k} \right) x^s \end{aligned}$$

und die Behauptung ergibt sich durch Koeffizientenvergleich. Folglich ist

$$\sum_{k=0}^{\min(n,d)} p_k = \sum_{k=0}^{\min(n,d)} \frac{\binom{d}{k} \binom{N-d}{n-k}}{\binom{N}{n}} = \frac{\sum_{k=0}^{\min(n,d)} \binom{d}{k} \binom{N-d}{n-k}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1.$$

Die zugehörige Verteilungsfunktion ist

$$F(k) = P(X \leq k) = \sum_{i=0}^k \frac{\binom{d}{i} \binom{N-d}{n-i}}{\binom{N}{n}}.$$

Ist nun  $N$  sehr groß im Vergleich zu  $n$  (dies schreibt man häufig als  $N \gg n$ ), kann man davon ausgehen, dass bei einer genügend großen Stichprobe der Anteil der Sorte 1 in etwa dem Anteil der Sorte 1 der gesamten Menge entspricht, also

$$\frac{k}{n} \approx \frac{d}{N}.$$

In diesem Fall hat man die Näherung

$$\frac{\binom{d}{k} \binom{N-d}{n-k}}{\binom{N}{n}} \approx \binom{n}{k} p^k (1-p)^{n-k} \quad \text{mit } p = \frac{d}{N};$$

das heißt die hypergeometrische Verteilung  $\mathcal{H}_{N,d,n}$  wird durch eine Binomialverteilung  $\mathcal{B}_{n,p}$  mit  $p = \frac{d}{N}$  approximiert.

**Beispiel 6.6** (Qualitätskontrolle II). Eine Firma hat mit dem Lieferanten gewisser Bauteile vereinbart, dass eine Ausschussquote von 5 Prozent akzeptabel ist. Um eine Lieferung von 100 Teilen zu überprüfen, entnimmt der Kunde eine Stichprobe von 10 Bauteilen mit der Absicht, die Lieferung zurückzuschicken, wenn mehr als ein Bauteil defekt ist. Mit welcher Wahrscheinlichkeit wird eine Lieferung mit 7 defekten Bauteilen abgelehnt?

Das Zufallsexperiment ist hier die Auswahl einer Stichprobe im Umfang von 10 Elementen. Die Zufallsvariable  $X$  ordnet einer Stichprobe die Anzahl der in ihr enthaltenen defekten Bauteile zu, sie ist  $\mathcal{H}_{100,7,10}$ -verteilt. Gefragt ist

$$P(X \geq 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - \frac{\binom{7}{0} \binom{93}{10} + \binom{7}{1} \binom{93}{9}}{\binom{100}{10}} = 1 - \frac{20753821}{24253880} \approx 0.144.$$

Die Wahrscheinlichkeit ist also nur 14.4%!

(Die Approximation durch die Binomialverteilung mit  $p = d/N = 0.07$  liefert

$$1 - \left( \binom{10}{0} (0.93)^{10} + \binom{10}{1} \cdot 0.07 \cdot (0.93)^9 \right) \approx 15.2\%;$$

gar nicht so schlecht.)

In folgender Tabelle listen wir die auf gleiche Weise errechneten Wahrscheinlichkeiten auf, dass eine Lieferung mit  $d$  defekten Teilen abgelehnt wird.

$d = 2$	0.9%	$d = 8$	18.2%
$d = 3$	2.6%	$d = 9$	22.1%
$d = 4$	4.9%	$d = 10$	26.2%
$d = 5$	7.7%	$d = 16$	50.0%
$d = 6$	10.9%	$d = 33$	90.4%
$d = 7$	14.4%	$d = 39$	95.5%

Mit 50prozentiger Wahrscheinlichkeit kann bei diesem Test der Lieferant also eine miserable Lieferung mit 16 defekten Bauteilen einschmuggeln.

### E. Poissonverteilung.

Eine Zufallsvariable  $X$  mit  $W = \text{Bild}(X) = \mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$  heißt poissonverteilt mit Parameter  $\lambda > 0$ , wenn

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

ist.

Benutzt man die Potenzreihenentwicklung  $e^x = \sum_{k \geq 0} \frac{x^k}{k!}$  der Exponentialfunktion, so ergibt sich

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$$

also definieren die  $p_k$  ein Wahrscheinlichkeitsmaß auf  $W$ .

Die zugehörige Verteilungsfunktion ist

$$F(k) = P(X \leq k) = e^{-\lambda} \sum_{i=0}^k \frac{\lambda^i}{i!}.$$

Die Poissonverteilung ist unter bestimmten Voraussetzungen eine gute Approximation an die Binomialverteilung: Wenn  $n$  sehr groß und  $p$  sehr klein ist, gilt mit  $\lambda = np$

$$b(k, n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

Das sieht man mit der Stirlingschen Formel  $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ ,  $n \gg 0$ :

$$\begin{aligned} p_k &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)! k!} p^k (1-p)^{n-k} \\ &\approx \frac{\sqrt{2\pi n}}{\sqrt{2\pi(n-k)}} e^{-k} \left(\frac{n}{n-k}\right)^n \frac{(np)^k}{k!} \left(1 - \frac{np}{n}\right)^{n-k} \\ &\approx \sqrt{\frac{n}{n-k}} e^{-k} \left(\frac{n-k+k}{n-k}\right)^{n-k} \frac{(np)^k}{k!} \left(1 - \frac{np}{n}\right)^{n-k} \\ &= \underbrace{\sqrt{\frac{n}{n-k}}}_{\approx 1} e^{-k} \underbrace{\left(1 + \frac{k}{n-k}\right)^{n-k}}_{\approx e^k} \underbrace{\frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}}_{\approx e^{-\lambda}} \approx \frac{\lambda^k}{k!} e^{-\lambda}. \end{aligned}$$

Als Faustregel gilt, dass die Approximation akzeptabel ist, wenn  $np < 10$  und  $n > 1500p$  ist. Für  $n = 1000$  und  $p = 0.01$  ist dies sicherlich erfüllt; in folgender Tabelle listen wir die Quotienten  $p_k/q_k$  der beiden Wahrscheinlichkeiten

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} = \binom{1000}{k} (0.01)^k (0.99)^{1000-k} \quad \text{und} \quad q_k = \frac{(np)^k}{k!} e^{-np} = \frac{10^k}{k!} e^{-10}$$

auf:

$k$	1	2	3	4	5	6	7	8	9	10
$p_k/q_k$	0.960	0.969	0.977	0.984	0.989	0.995	0.999	1.001	1.004	1.005

11	12	13	14	15	16	17	18	19	20
1.005	1.004	1.001	0.999	0.995	0.989	0.983	0.976	0.969	0.960

Die Poisson-Verteilung liefert also Voraussagen über die Anzahl ( $k$ ) des Eintretens seltener, zufälliger und voneinander unabhängiger Ereignisse.

Genauso hat man für  $n \gg 0$ ,  $\lambda = np$  und  $p = \frac{\lambda}{N}$  klein:

$$\frac{\binom{d}{k} \binom{N-d}{n-k}}{\binom{N}{n}} \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

das heißt die Poissonverteilung approximiert für (sehr) große  $N$  die hypergeometrische Verteilung.

**Beispiel 6.7.** Sei  $X$  die Anzahl der monatlichen Piratenangriffe vor dem Horn von Afrika. Dann kann man annehmen, dass  $X$  poissonverteilt ist mit  $\lambda = 10$  (Erklärung folgt später). Mit welcher Wahrscheinlichkeit gibt es in einem Monat wenigstens einen Angriff?

Gesucht ist

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{10^0}{0!} e^{-10} = 1 - e^{-10} \approx 99.99\%.$$

**Beispiel 6.8.** Die Anzahl der Anrufe, die pro Minute bei einer Hotline eingehen, sei poissonverteilt mit  $\lambda = 5$ . Das Callcenter kann in jeder Minute 7 Anrufe bearbeiten, weitere Anrufe müssen in die Warteschleife. Wir wollen die drei folgenden Fragen beantworten:

1. Was ist die wahrscheinlichste Zahl von Anrufen pro Minute?
2. Mit welcher Wahrscheinlichkeit muss in einer Minute mindestens ein Anrufer in die Warteschleife?
3. Wie groß müsste die Kapazität des Callcenters mindestens sein, damit mit Wahrscheinlichkeit 95% kein Anrufer warten muss?

Es ist  $p_k = P(X = k) = \frac{5^k}{k!} e^{-5}$ ; für die erste Frage betrachten wir den Quotienten

$$\frac{p_{k+1}}{p_k} = \frac{5}{k+1} \begin{cases} < 1 & \text{für } k \geq 5, \\ > 1 & \text{für } k < 4. \end{cases}$$

Daher gilt  $p_k \leq p_5$  für jedes  $k$ ; die wahrscheinlichste Zahl ist also 5. Für 2. ist  $P(X > 7)$  gefragt; wir rechnen aus:

$$P(X \geq 7) = 1 - P(X \leq 7) = 1 - e^{-5} \sum_{k=0}^7 \frac{5^k}{k!} = 1 - e^{-5} \frac{2701}{21} \approx 0.133 = 13.3\%.$$

In der dritten Frage ist eine möglichst kleine Zahl  $m$  gesucht, so dass  $P(X \leq m) \geq 0.95$  ist. So wie oben rechnet man aus

$$\begin{aligned} P(X \leq 8) &\approx 0.93, \\ P(X \leq 9) &\approx 0.968. \end{aligned}$$

Das Callcenter sollte also wenigstens 9 Anrufe pro Minute bearbeiten können.

## 7. Erwartungswerte und Varianz diskreter Zufallsvariablen

Diskrete Zufallsvariablen nehmen jeden ihrer Werte mit einer gewissen Wahrscheinlichkeit an. Daher können wir der Verteilung einen gewichteten Mittelwert zuordnen.

**Definition 7.1.** Sei  $X$  eine diskrete Zufallsvariable mit Wertebereich  $W = \{x_0, x_1, x_2, \dots\}$  und Wahrscheinlichkeitsmaß  $P$ . Sei  $p_i = P(X = x_i)$ . Wenn die Reihe  $\sum_{i=0}^{\infty} p_i |x_i|$  konvergiert, so heißt

$$E(X) := \sum_{i=0}^{\infty} p_i x_i$$

der Erwartungswert von  $X$ .

*Bemerkung.* Die Bedingung  $\sum_{i=0}^{\infty} p_i |x_i| < \infty$  garantiert nicht nur, dass der Erwartungswert existiert, sondern auch, dass es nicht darauf ankommt, in welcher Reihenfolge man summiert.

Der Erwartungswert ist also der Wert, den die Zufallsvariable im Mittel annimmt.

**Beispiel 7.2.** Bei einer Tombola werden 500 Lose verkauft. Der Hauptgewinn sei 1000 €, dazu gebe es noch zwei Gewinne à 100 € und 4 Gewinne à 10 €. Der Erwartungswert des Gewinns  $X$  ist

$$\begin{aligned} E(X) &= 0 \cdot P(X = 0) + 10 \cdot P(X = 10) + 100 \cdot P(X = 100) + 1000 \cdot P(X = 1000) \\ &= 10 \cdot \frac{4}{500} + 100 \cdot \frac{2}{500} + 1000 \cdot \frac{1}{500} = 2.48 \end{aligned}$$

Der Erwartungswert, obwohl eine wichtige Kenngröße einer Wahrscheinlichkeitsverteilung, sagt für sich genommen noch recht wenig über die Verteilung aus, genauso wenig wie der Mittelwert einer Zahlenreihe über die ganze Zahlenreihe. Man möchte auch wissen, wie die Werte um diesen Erwartungswert (beziehungsweise den Mittelwert) herum streuen. Ein Maß für diese Streuung ist die Varianz: Dazu betrachtet man das Quadrat der Abweichung (um Vorzeichen zu umgehen) der Werte vom Erwartungswert als neue Zufallsvariable und bestimmt deren Erwartungswert.

**Definition 7.3.** Die Varianz einer diskreten Zufallsvariablen  $X$  ist

$$\text{Var}(X) := E((X - E(X))^2).$$

Die Standardabweichung von  $X$  ist definiert als

$$\sigma(X) := \sqrt{\text{Var}(X)}.$$

**Erläuterung:** Ist also  $P$  ein Wahrscheinlichkeitsmaß,  $\{x_0, x_1, x_2, \dots\}$  das Bild von  $X$  und  $p_i = P(X = x_i)$ , so ist

$$\text{Var}(X) = \sum_{i=0}^{\infty} p_i (x_i - E(X))^2. \quad (7.1)$$

Allgemeiner kann man durch Rechenoperationen aus Zufallsvariablen neue Zufallsvariablen erzeugen. Sind zum Beispiel  $X$  und  $Y$  zwei Zufallsvariablen auf der gleichen Ergebnismenge  $\Omega$  und  $a, b \in \mathbb{R}$  Konstanten, so ist  $Z = aX + bY$  wieder eine Zufallsvariable auf  $\Omega$ : für ein  $\omega \in \Omega$  ist  $Z(\omega) = a \cdot X(\omega) + b \cdot Y(\omega)$ .

*Bemerkung.* (a) Eigentlich müsste man zur Definition noch hinzufügen: „sofern die definierende Reihe konvergiert“; gemeint ist die Reihe (7.1).

(b) Haben die Werte der Zufallsvariablen eine bestimmte Interpretation als Maßeinheiten, wie zum Beispiel Meter, oder (Temperatur-)Grade, so ist die Maßeinheit der zugehörigen Varianz das Quadrat dieser Einheit. Durch das Wurzelziehen wird das wieder „korrigiert“; die Standardabweichung hat also dieselbe Maßeinheit wie die Zufallsvariable selbst.

**Satz 7.4.** Ist die Varianz der diskreten Zufallsvariablen  $X$  definiert, so gilt

$$\text{Var}(X) = E(X^2) - E(X)^2.$$

*Beweis.*

$$\begin{aligned} \text{Var}(X) &= \sum_{i=0}^{\infty} p_i (x_i - E(X))^2 \\ &= \sum_{i=0}^{\infty} p_i (x_i^2 - 2x_i E(X) + E(X)^2) \\ &= \underbrace{\sum_{i=0}^{\infty} p_i x_i^2}_{=E(X^2)} - 2E(X) \underbrace{\sum_{i=0}^{\infty} p_i x_i}_{=E(X)} + E(X)^2 \underbrace{\sum_{i=0}^{\infty} p_i}_{=1} \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2. \end{aligned} \quad \square$$

**Beispiel** (Fortsetzung von Beispiel 7.2). Für die Tombola hat man als Varianz gemäß der Definition

$$\text{Var}(X) = \frac{493}{500}(0 - 2.48)^2 + \frac{4}{500}(10 - 2.48)^2 + \frac{2}{500}(100 - 2.48)^2 + \frac{1}{500}(1000 - 2.48)^2 = 2034.6496$$

Andererseits ist

$$E(X^2) = 0^2 \frac{493}{500} + 10^2 \frac{4}{500} + 100^2 \frac{2}{500} + 1000^2 \frac{1}{500} = 2040.8$$

so dass die Formel des Satzes  $E(X^2) - E(X)^2 = 2040.8 - (2.48)^2 = 2034.6496$  wie erwartet das gleiche Ergebnis liefert. (Die Standardabweichung ist  $\sigma(X) \approx 45$ .)

In der Folge bestimmen wir Erwartungswert und Varianz der im letzten Abschnitt beschriebenen diskreten Verteilungen.

**A. Diskrete Gleichverteilung.** Sei  $X$  gleichverteilt mit Werten in  $\{x_1, \dots, x_n\}$  und sei  $p_i = P(X = x_i) = 1/n$ ,  $1 \leq i \leq n$ . Dann gilt

$$E(X) = \frac{x_1}{n} + \dots + \frac{x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n x_i \right)^2.$$

Im Spezialfall  $x_i = i$ , also  $\text{Bild}(X) = \{1, 2, \dots, n\}$ , erhält man

$$E(X) = \frac{1}{2} \sum_{i=1}^n i = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2},$$

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n i \right)^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} - \frac{1}{n^2} \frac{n^2(n+1)^2}{4} = \frac{n^2-1}{12}$$

mit Hilfe der üblichen Summenformeln.

**B. Binomialverteilung.** Sei  $X$  binomialverteilt mit Parametern  $p, n$ , also Werten in  $\{0, 1, \dots, n\}$ . Dann ist  $p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$ , also

$$\begin{aligned} E(X) &= \sum_{i=0}^n i \cdot p_i = \sum_{i=1}^n i \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i \frac{n(n-1) \cdots (n-i+1)}{i \cdot (i-1)!} \cdot p \cdot p^{i-1} (1-p)^{n-i} \\ &= np \sum_{i=1}^n \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ &= np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=1} = np \end{aligned}$$

denn die letzte Summe berechnet die Wahrscheinlichkeit des sicheren Ereignisses, und

$$\begin{aligned} E(X^2) &= \sum_{i=0}^n i^2 \cdot p_i = \sum_{i=1}^n i^2 \binom{n}{i} p^i (1-p)^{n-i} \\ &= np \sum_{k=0}^{n-1} (k+1) \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &= np \underbrace{\sum_{k=1}^{n-1} k \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=(n-1)p \text{ (wie oben)}} + np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=1} \\ &= n(n-1)p^2 + np = n^2p^2 - np^2 + np \end{aligned}$$

und damit gilt

$$\text{Var}(X) = E(X^2) - E(X)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$



**C. Geometrische Verteilung.** Sei  $\text{Bild}(X) = \{1, 2, 3, \dots\}$  und  $p_k = P(X = k) = p(1-p)^{k-1}$ . Dann gilt

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum_{j=0}^{\infty} (j+1)(1-p)^j = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}, \\ \text{Var}(X) &= E(X^2) - E(X)^2 = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} - \frac{1}{p^2} = -\frac{1}{p^2} + p \sum_{j=0}^{\infty} (j+1)^2 (1-p)^j \\ &= -\frac{1}{p^2} + p \underbrace{\sum_{j=0}^{\infty} (j+1)(j+2)(1-p)^j}_{=\frac{2}{(1-(1-p))^2}} - p \underbrace{\sum_{j=0}^{\infty} (j+1)(1-p)^j}_{=\frac{1}{(1-(1-p))^2}} = \frac{1-p}{p^2}. \end{aligned}$$

Hierbei wurde die Formel

$$\sum_{k=0}^{\infty} (k+1)(k+1) \cdots (k+n)x^k = \frac{n!}{(1-x)^{n+1}} \quad \text{für } |x| < 1 \text{ und } n \in \mathbb{N}$$

benutzt, die man wiederum aus der Formel  $\frac{d^n}{dx^n} \left( \frac{x^n}{1-x} \right) = \frac{n!}{(1-x)^{n+1}}$  erhält.

**D. Hypergeometrische Verteilung.** Sei  $X$  hypergeometrisch verteilt mit Parametern  $N$  (Grundmenge),  $d$  (interessierende Objekte) und  $n$  (Umfang der Stichprobe); es ist  $P(X = k) = \mathcal{H}_{N,d,n}(k) = \frac{\binom{d}{k} \binom{N-d}{n-k}}{\binom{N}{n}}$ . Sei ferner  $m = \min(n, d)$ . Dann gilt

$$\begin{aligned} \binom{N}{n} E(X) &= \sum_{k=0}^m k \binom{d}{k} \binom{N-d}{n-k} \\ &= \sum_{k=1}^m d \binom{d-1}{k-1} \binom{(N-1)-(d-1)}{(n-1)-(k-1)} \\ &= d \sum_{\ell=0}^{m-1} \binom{d-1}{\ell} \binom{(N-1)-(d-1)}{(n-1)-\ell} = d \binom{N-1}{n-1} = d \frac{N}{n} \binom{N}{n}, \end{aligned}$$

denn teilt man die Summe der letzten Zeile durch  $\binom{N-1}{n-1}$ , so erhält man die Summe der Wahrscheinlichkeiten aller Elementarereignisse der  $\mathcal{H}_{N-1,d-1,n-1}$ -Verteilung. Es folgt

$$E(X) = d \frac{n}{N}.$$

Analog rechnet man aus:

$$\begin{aligned} \binom{N}{n} E(X^2) &= \sum_{k=0}^m k^2 \binom{d}{k} \binom{N-d}{n-k} = d \sum_{k=1}^m k \binom{d-1}{k-1} \binom{(N-1)-(d-1)}{(n-1)-(k-1)} \\ &= d \sum_{\ell=1}^{m-1} \ell \binom{d-1}{\ell} \binom{(N-1)-(d-1)}{(n-1)-\ell} + d \underbrace{\sum_{\ell=0}^{m-1} \binom{d-1}{\ell} \binom{(N-1)-(d-1)}{(n-1)-\ell}}_{=\frac{n}{N} \binom{N}{n} \text{ wie oben}} \\ &= d(d-1) \underbrace{\sum_{j=0}^{m-2} \binom{d-2}{j} \binom{(n-2)-(d-2)}{(n-2)-j}}_{=\binom{N-2}{n-2} = \frac{n(n-1)}{N(N-1)} \binom{N}{n}} + d \frac{n}{N} \binom{N}{n} \\ &= \binom{N}{n} \frac{dn}{N} \left[ \frac{(d-1)(n-1)}{N-1} + 1 \right] \end{aligned}$$

woraus dann wiederum

$$\text{Var}(X) = \frac{dn}{N^2(n-1)}(N-d)(N-n)$$

folgt.

**E. Poissonverteilung.** Sei  $X: \Omega \rightarrow \mathbb{N}_0$  poissonverteilt mit Parameter  $\lambda > 0$ . Dann gilt

$$E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

wobei wir die Reihendarstellung der Exponentialfunktion benutzt haben. Genauso rechnet man aus

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{k \lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} = \lambda e^{-\lambda} \left( \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} + \sum_{j=1}^{\infty} \frac{j \lambda^j}{j!} \right) \\ &= \lambda e^{-\lambda} \left( e^{\lambda} + \lambda \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} \right) = \lambda e^{-\lambda} (e^{\lambda} + \lambda e^{\lambda}) = \lambda + \lambda^2 \end{aligned}$$

woraus dann

$$\text{Var}(X) = E(X^2) - E(X)^2 = \lambda$$

folgt.

Zusammengefasst:

Verteilung von $X$	$E(X)$	$\text{Var}(X)$
Gleichverteilung auf $\{x_1, \dots, x_n\}$ auf $\{1, \dots, n\}$	$\frac{1}{n} \sum_{i=1}^n x_i$ $\frac{n+1}{2}$	$\frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n x_i \right)^2$ $\frac{n^2-1}{12}$
Binomialverteilung $\mathcal{B}_{n,p}$	$np$	$np(1-p)$
Geometrische Verteilung auf $\{1, 2, 3, \dots\}$ mit Parameter $p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometrische Verteilung $\mathcal{H}_{N,d,n}$	$\frac{dn}{N}$	$\frac{dn(N-d)(N-n)}{N^2(N-1)}$
Poissonverteilung mit Parameter $\lambda > 0$	$\lambda$	$\lambda$

*Bemerkung.* Für eine Zufallsvariable  $X$  und Konstanten  $a, b \in \mathbb{R}$  gilt

$$E(aX + b) = aE(X) + b, \quad \text{Var}(aX + b) = a^2 \text{Var}(X) + b \text{ falls } a \neq 0.$$

Außerdem hat man die sogenannte *Tschebycheff-Ungleichung*: Sie  $\varepsilon > 0$  und  $\text{Bild}(X) = \{x_k\}$ . Dann kann man die Varianz folgendermaßen abschätzen:

$$\begin{aligned} \text{Var}(X) &= E((x - E(X))^2) = \sum_{\substack{k \\ |x_k - E(X)| > \varepsilon}} p_k (x_k - E(X))^2 + \sum_{\substack{k \\ |x_k - E(X)| \leq \varepsilon}} p_k (x_k - E(X))^2 \\ &\geq \varepsilon^2 \sum_{\substack{k \\ |x_k - E(X)| > \varepsilon}} p_k = \varepsilon^2 P(|X - E(X)| > \varepsilon) = \varepsilon^2 (1 - P(|X - E(X)| \leq \varepsilon)) \end{aligned}$$

woraus folgt:

$$P(E(X) - \varepsilon \leq X \leq E(X) + \varepsilon) \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}. \quad (7.2)$$

## 8. Stetige Zufallsvariablen

Anwendungen aus der realen Welt zwingen einen dazu, auch andere als diskrete Zufallsvariablen zu betrachten: Wiederholte Messungen physikalischer Größen sind meist nicht exakt, sondern sind, wegen der eingeschränkten Genauigkeit von Messverfahren und -instrumenten, zufällig um einen Wert herum verteilt. Diese Verteilungen sind allerdings im allgemeinen nicht diskret, sondern können jeden beliebigen Wert eines Intervalls annehmen. Als Beispiel hatten wir schon früher das Werfen einer Nadel als Zufallsexperiment aufgefasst und den Winkel der Nadel mit einer vorgegebenen Gerade als Zufallsvariable, mit Werten im Intervall  $[0, 2\pi]$ .

Ist nun aber das Bild der Zufallsvariablen  $X$  nicht mehr diskret, also überabzählbar, so kann man nicht länger jedem einzelnen Wert eine positive Wahrscheinlichkeit zuordnen, denn dann wird die Summe aller Wahrscheinlichkeiten  $\sum_{x \in \text{Bild}(X)} P(X = x)$  unendlich groß. Man kann aber zeigen, dass man jedem echten Intervall, jeder abzählbaren Vereinigung von Intervallen und den Komplementen dieser Mengen Wahrscheinlichkeiten zuordnen kann, so dass mit diesen Mengen als Ereignissen die Axiome eines Wahrscheinlichkeitsmaßes erfüllt sind.<sup>2</sup>

Wir erklären hier stetige Verteilungen nicht axiomatisch, sondern über ein Konstruktionsprinzip:

**Definition 8.1.** Eine Zufallsvariable  $X$  mit Verteilungsfunktion  $F$  heißt *stetig verteilt*, wenn es eine stückweise stetige Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  gibt mit

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du.$$

Die Funktion  $f$  nennt man die *Verteilungsdichte* der Verteilungsfunktion.

Die Verteilungsdichte ersetzt die aus den Einzelwahrscheinlichkeiten bestehenden Koeffizientenfolgen  $(p_i)_{i \in \mathbb{N}}$  einer diskreten Verteilung; dabei wird aus der Summation ein Integral.

Mit dieser Definition ergeben sich leicht einige Eigenschaften:

<sup>2</sup>Diese Mengen bilden die sogenannte *Borelsche Sigma-Algebra*.

**Satz 8.2.** Sei  $X$  eine stetig verteilte Zufallsvariable mit Verteilungsdichte  $f$  und Verteilungsfunktion  $F$ .

- (a)  $F$  ist stetig, und ist  $f$  stetig im Punkt  $u_0$ , so gilt  $f(u_0) = F'(u_0)$ .
- (b)  $f(x) \geq 0$ .
- (c)  $\int_{-\infty}^{\infty} f(u) du = 1$ .
- (d)  $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(u) du$ .

(a) ist ein klassisches Resultat der Analysis. (b) folgt, da  $F$  als Verteilungsfunktion monoton wachsend ist, und (c) wegen

$$\int_{-\infty}^{\infty} f(u) du = \lim_{x \rightarrow \infty} \int_{-\infty}^x f(u) du = \lim_{x \rightarrow \infty} F(x) = 1$$

aus Eigenschaft 5.4(c) einer Verteilungsfunktion.

Zu beachten ist noch  $P(X = a) = \int_a^a f(u) du = 0$  für jedes  $a$ , so wie wir es in dem einleitenden Abschnitt schon vermutet hatten. (Das bedeutet aber nicht, dass  $a$  als Wert nicht angenommen werden kann.)

Die Verteilungsfunktion und Wahrscheinlichkeiten einer stetigen Verteilung kann man als Flächen unter dem Graphen der Verteilungsdichte veranschaulichen, so wie in Abbildung 7 angedeutet.

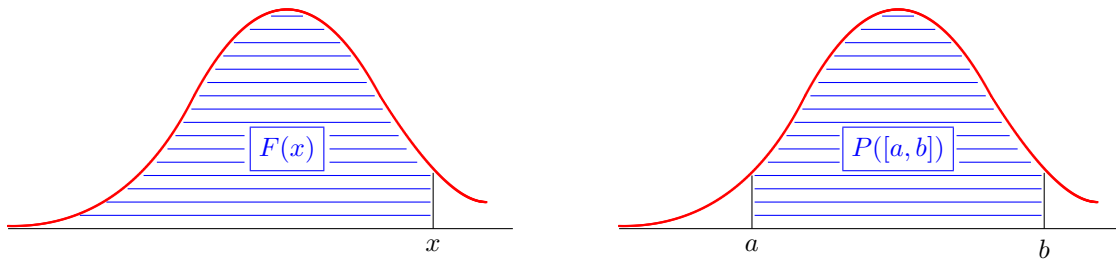


ABBILDUNG 7. Stetige Verteilung

Umgekehrt definiert jede stückweise stetige Funktion  $f$  mit  $f(x) \geq 0$  und  $\int_{-\infty}^{\infty} f(u) du = 1$  ein Wahrscheinlichkeitsmaß

$$P([a, b]) = F(b) - F(a) = \int_a^b f(u) du$$

auf der Menge der Intervalle (besser gesagt, auf der Borelschen Sigma-Algebra).

Erwartungswert und Varianz stetiger Zufallsvariablen kann man nun nicht mehr über die Summation über alle Werte definieren; dies ersetzt man durch Integration über die mit der Verteilungsdichte gewichteten Werte. Zusätzlich zur schon bekannten Notation  $E(X)$  und  $\text{Var}(X)$  haben sich auch die Bezeichnungen  $\mu(X)$  und  $\sigma^2(X)$  durchgesetzt.

**Definition 8.3.** Sei  $X$  eine stetige Zufallsvariable mit Verteilungsdichte  $f$ .

(a) Der Erwartungswert von  $X$  ist das Integral

$$\mu(X) = E(X) = \int_{-\infty}^{\infty} u f(u) du.$$

(b) Die Varianz ist

$$\sigma^2(X) = \text{Var}(X) = E((X - E(X))^2) = \int_{-\infty}^{\infty} f(u)(u - E(X))^2 du$$

(c) Die Standardabweichung oder Streuung ist  $\sigma(X) = \sqrt{\text{Var}(X)}$ .

(d) Für ein  $\varepsilon$  mit  $0 < \varepsilon < 1$  heißt die Zahl  $q_\varepsilon$ , für die

$$F(q_\varepsilon) = P(X \leq q_\varepsilon) = \varepsilon$$

gilt,  $\varepsilon$ -Quantil von  $X$ .

Die Standardabweichung ist also wie früher die Wurzel aus der Varianz.  $\varepsilon$ -Quantile sind neu; es folgt aus dem Zwischenwertsatz und der Monotonie von  $F$ , dass sie existieren und eindeutig bestimmt sind. Ebenfalls wie im diskreten Fall hat man:

**Satz 8.4.** Sei  $X$  eine stetige Zufallsvariable.

(a)  $\sigma^2(X) = E(X^2) - E(X)^2$ ,

(b) Für jedes  $\varepsilon > 0$  gilt

$$P(\mu(X) - \varepsilon \leq X \leq \mu(X) + \varepsilon) \geq 1 - \frac{\sigma^2(X)}{\varepsilon^2} \quad (\text{Tschebycheffsche Ungleichung}).$$

Als nächstes wollen wir einige wichtige Verteilungen vorstellen.

#### A. Die Rechteck- oder stetige Gleichverteilung.

Für ein Intervall  $[a, b]$  definiere eine Dichtefunktion durch

$$f(u) = \begin{cases} \frac{1}{b-a} & \text{für } u \in [a, b], \\ 0 & \text{sonst.} \end{cases}$$

Die zugehörige Verteilungsfunktion ist dann

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

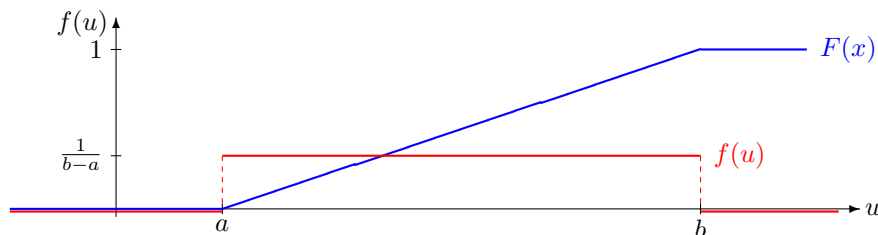


ABBILDUNG 8. Rechteck- oder stetige Gleichverteilung

Für Erwartungswert und Varianz erhält man

$$\mu(X) = E(X) = \frac{1}{b-a} \int_a^b u \, du = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}$$

sowie

$$\sigma^2(X) = \text{Var}(X) = \frac{1}{b-a} \int_a^b u^2 \, du - \left(\frac{a+b}{2}\right)^2 = \frac{1}{3} \frac{b^3 - a^3}{b-a} - \left(\frac{a+b}{2}\right)^2 = \frac{(a-b)^2}{12}.$$

### B. Exponentialverteilung.

Für ein  $\lambda > 0$  sei

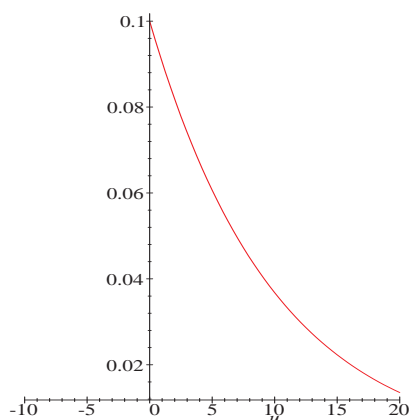
$$f_\lambda(u) = \begin{cases} 0 & \text{für } u \leq 0, \\ \lambda e^{-\lambda u} & \text{für } u > 0. \end{cases}$$

Mit partieller Integration rechnet man aus

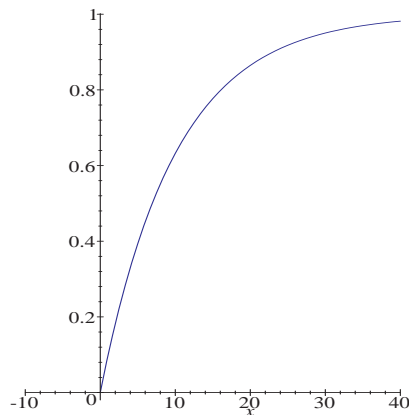
$$\int_0^x \lambda e^{-\lambda u} \, du = \left[-e^{-\lambda u}\right]_0^x = 1 - e^{-\lambda x}$$

daher ist die Verteilungsfunktion der Exponentialverteilung zum Parameter  $\lambda$  gegeben durch

$$F_\lambda(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ 1 - e^{-\lambda x} & \text{für } x > 0. \end{cases}$$



Verteilungsdichte  $f_{0.1}(u)$



Verteilungsfunktion  $F_{0.1}(x)$

ABBILDUNG 9. Exponentialverteilung mit  $\lambda = 0.1$

Wiederum mit partieller Integration erhält man

$$\begin{aligned} \mu(X) = E(X) &= \int_{-\infty}^{\infty} \lambda u e^{-\lambda u} \, du = \int_0^{\infty} \lambda u e^{-\lambda u} \, du = \left[-u e^{-\lambda u}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda u} \, du \\ &= -\lim_{u \rightarrow \infty} u e^{-\lambda u} - 0 + \left[\frac{-1}{\lambda} e^{-\lambda u}\right]_0^{\infty} = 0 - \frac{1}{\lambda} \lim_{u \rightarrow \infty} e^{-\lambda u} + \frac{1}{\lambda} = \frac{1}{\lambda} \end{aligned}$$

Ganz ähnlich berechnet man (etwa mit zweimaliger partieller Integration)

$$\mu(X^2) = E(X^2) = \int_0^{\infty} \lambda u^2 e^{-\lambda u} \, du = \frac{2}{\lambda^2},$$

woraus man schließlich

$$\sigma^2(X) = \text{Var}(X) = \mu(X^2) - \mu(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

erhält; die Standardabweichung ist folglich  $\sigma(X) = \frac{1}{\lambda}$ .

Die Exponentialverteilung tritt oft auf, wenn die Zufallsvariable Wartezeiten oder Lebensdauer beschreibt.

**Beispiel 8.5.** Die Entladedauer  $T$  eines Akkus sei exponentialverteilt mit  $\lambda = \frac{1}{100}$  pro Stunde. (Mit anderen Worten, der Erwartungswert für die Entladedauer sei 100 Stunden.) Wie groß ist die Wahrscheinlichkeit, dass der Akku mehr als 5 Tage, also 120 Stunden, durchhält?

Zu berechnen ist

$$P(X > 120) = 1 - P(X \leq 120) = 1 - F_{0,01}(120) = 1 - \left(1 - e^{-\frac{120}{100}}\right) = e^{-1.2} \approx 0.301.$$

Die Wahrscheinlichkeit beträgt also etwa 30%.

### C. Normalverteilung.

Die Normalverteilung ist eine der bekanntesten und wichtigsten Verteilungen überhaupt, da sie für viele andere Verteilungen unter geeigneten Voraussetzungen eine gute Näherung darstellt. Sie hat zwei parametrische Kenngrößen  $\mu$  und  $\sigma^2$ , die aus offensichtlichen Gründen (siehe unten) diese Bezeichnungen haben.

**Definition 8.6.** Eine Zufallsvariable  $X$  heißt normalverteilt mit Parametern  $\mu$  und  $\sigma^2$ , wenn sie eine Verteilungsdichte  $f_{\mu,\sigma^2}$  mit

$$f_{\mu,\sigma^2}(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-\mu)^2}{2\sigma^2}}$$

besitzt. In diesem Fall nennt man  $X$  auch kürzer  $N(\mu, \sigma^2)$ -verteilt.

Die Funktion  $f_{\mu,\sigma^2}$  ist sicherlich nichtnegativ, und es gilt auch

$$\int_{-\infty}^{\infty} f_{\mu,\sigma^2}(u) du = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du = 1, \quad (8.1)$$

auch wenn das mit unseren elementaren Mitteln nicht ohne weiteres einzusehen ist. Daher definiert diese Dichtefunktion tatsächlich eine stetige Verteilung. Sie ist gerade, ihr Graph also spiegelsymmetrisch zur vertikalen Achse. Für  $u \rightarrow \pm\infty$  geht sie gegen Null (wie jede Dichtefunktion), sie hat ein Maximum bei  $u = \mu$  und Wendepunkte an den Stellen  $\mu \pm \sigma$ .

Für eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable  $X$  gilt

$$\begin{aligned} \mu(X) = E(X) &= \int_{-\infty}^{\infty} u f_{\mu,\sigma^2}(u) du = \int_{-\infty}^{\infty} (\mu - (u - \mu)) f_{\mu,\sigma^2}(u) du \\ &= \mu \int_{-\infty}^{\infty} f_{\mu,\sigma^2}(u) du + \int_{-\infty}^{\infty} (u - \mu) f_{\mu,\sigma^2}(u) du. \end{aligned}$$

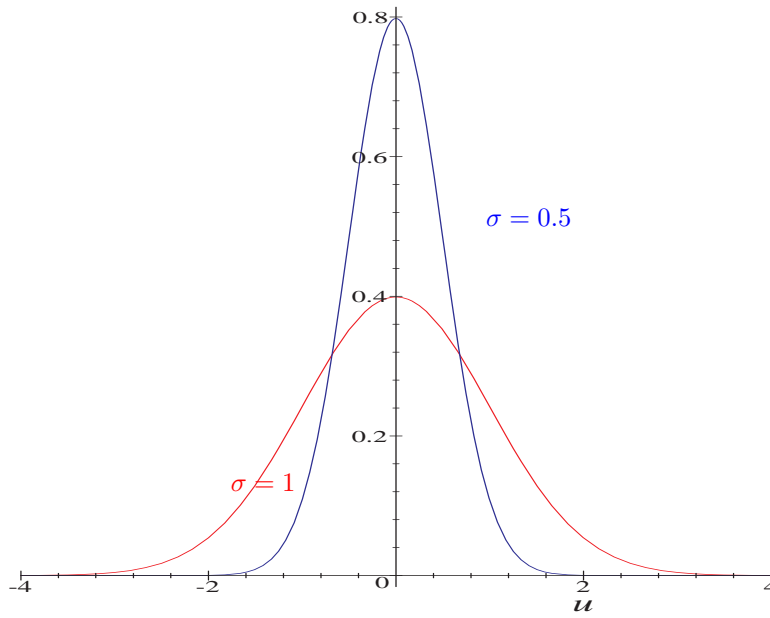
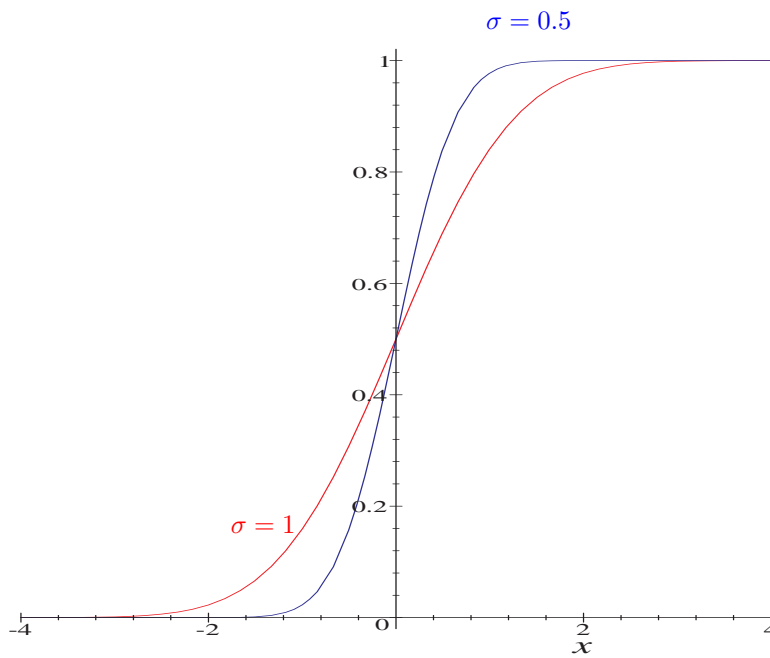
Das erste Integral hat den Wert 1 nach (8.1) und das zweite ist Null, da der Integrand eine ungerade Funktion ist. Es folgt also

$$\mu(X) = \mu.$$

Mit etwas mehr Aufwand erhält man auch

$$\sigma^2(X) = \sigma^2;$$

diese Ergebnisse rechtfertigen die Bezeichnungen der Parameter.

ABBILDUNG 10. Normalverteilungsdichte mit  $\mu = 0$ ,  $\sigma = 1$  und  $\sigma = 0.5$ ABBILDUNG 11. Verteilungsfunktion der Normalverteilung mit  $\mu = 0$ ,  $\sigma = 1$  und  $\sigma = 0.5$



Da das Integral

$$F_{\mu, \sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du$$

nicht elementar berechenbar ist, gibt es also auch keine geschlossene Formel für die Verteilungsfunktion. Wie wir gleich sehen werden, reicht es aber, die Verteilungsfunktionen für ein Paar von Parametern zu kennen, denn für andere Parameterwerte kann man die Verteilungsfunktion leicht aus der einen (bekannteren) gewinnen.

**Definition 8.7.** Die Normalverteilung zu dem Parametern  $\mu = 0$  und  $\sigma = 1$  heißt die Standardnormalverteilung. Ihre Verteilungsfunktion ist

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

Die Werte von  $\Phi(x)$  liegen in Tabellenform vor. Um zu rechtfertigen, dass ihre Kenntnis genügt, zeigen wir:

**Satz 8.8.** Sei  $X$  eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable. Dann genügt  $Y = \frac{X - \mu}{\sigma}$  der Standardnormalverteilung.

*Beweis.* Mir der Substitution  $t = \frac{u - \mu}{\sigma}$ , also  $du = \sigma dt$ , gilt

$$F_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-\mu)^2}{2\sigma^2}} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

□

**Beispiel 8.9.** Ein Unternehmen kauft Röhren mit 20cm Durchmesser; Abweichungen von bis zu 1mm werden toleriert. Man kann annehmen, dass die Röhren, die ein Produzent liefert, normalverteilt mit  $\mu = 20$ cm sind.

(a) Es sei  $\sigma = 2$ mm. Welcher Anteil der Röhren wird abgelehnt? Und wie sieht es bei 1.2mm aus?

Gefragt ist zuerst die Wahrscheinlichkeit

$$\begin{aligned} P(19.9 \leq X \leq 20.1) &= P(-0.1 \leq X - 20 \leq 0.1) = P\left(\frac{-0.1}{0.2} \leq \frac{X - 20}{0.2} \leq \frac{0.1}{0.2}\right) \\ &= \Phi\left(\frac{0.1}{0.2}\right) - \underbrace{\Phi\left(\frac{0.1}{0.2}\right)}_{=1-\Phi\left(\frac{0.1}{0.2}\right)} = 2\Phi(1/2) - 1 \approx 0.383 \end{aligned}$$

Es werden also 61.7% der Röhren abgelehnt. Bei einer Standardabweichung von 1.2mm erhält man mit dem gleichen Rechenverfahren

$$P(19.9 \leq X \leq 20.1) = P(-0.1 \leq X - 20 \leq 0.1) = 2\Phi\left(\frac{0.1}{0.12}\right) - 1 = 2\Phi(5/6) - 1 \approx 0.595$$

und es werden 41.5% abgelehnt.

(b) Wie groß ist  $\sigma$ , wenn 5% abgelehnt werden?

Da 95% akzeptiert werden, ist eine Zahl  $\sigma$  gesucht, so dass

$$0.95 = P\left(\frac{-0.1}{\sigma} \leq \frac{X - 20}{0.2} \leq \frac{0.1}{\sigma}\right) = 2\Phi\left(\frac{0.1}{\sigma}\right) - 1$$

gilt, also  $\Phi\left(\frac{0.1}{\sigma}\right) = 0.975$ . Die Tabelle liefert  $\Phi(1.96) = 0.975002$ ; es folgt  $\sigma = 0.051$ .

Wegen der Symmetrie der Dichtefunktion um die Stelle  $u = \mu$  gilt für beliebiges  $c > 0$

$$P(\mu - c\sigma < X < \mu + c\sigma) = P\left(-c < \frac{X - \mu}{c} < c\right) = 2\Phi(c) - 1.$$

Für einige Werte wichtige von  $c$  fassen wir diese Wahrscheinlichkeiten in einer Tabelle zusammen:

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 2\Phi(1) - 1 \approx 0.682 \\ P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) &= 2\Phi(1.96) - 1 \approx 0.95 \quad (\text{Statistische Signifikanz}) \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= 2\Phi(2) - 1 \approx 0.954 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &= 2\Phi(3) - 1 \approx 0.997 \end{aligned}$$

Die Normalverteilung ist auch deshalb so wichtig, weil man mit ihr die Binomialverteilung approximieren kann, wenn die Umstände geeignet sind.

**Satz 8.10** (Grenzwertsatz von De Moivre und Laplace). *Sei  $X$  binomialverteilt mit Parametern  $n$  und  $p$ . Falls  $n$  genügend groß ist, ist  $X$  näherungsweise normalverteilt mit Erwartungswert  $\mu = np$  und Varianz  $\sigma^2 = np(1-p)$ , das heißt für die Verteilungsfunktion  $F$  von  $X$  gilt*

$$F(k) \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right), \quad k = 0, 1, 2, \dots, n.$$

*Bemerkung.* 1. Die Parameter  $\mu$  und  $\sigma^2$  entsprechen also genau dem Erwartungswert und der Varianz der Binomialverteilung.

2. Man kann erwarten, dass die Approximation einigermaßen gut ist, wenn  $\sigma^2 > 9$  ist.

Für die Wahrscheinlichkeiten gilt also

$$P(k \leq X \leq \ell) \approx \Phi\left(\frac{\ell - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right).$$

Allerdings macht man, da die Verteilungsfunktion der Binomialverteilung stückweise konstant ist, einen kleinen Fehler: Links von  $k$  ist noch das ganze Intervall bis  $k-1$  dabei, während rechts von  $\ell$  nichts hinzukommt — die Formel berechnet die Wahrscheinlichkeit für das Intervall  $[k-1, \ell]$ . Daher ist es besser,

$$\Phi\left(\frac{\ell + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}}\right),$$

also das Intervall  $[k-1/2, \ell+1/2]$  zu benutzen. Diese Verbesserung der Näherung bezeichnet man als *Stetigkeitskorrektur*.

**Beispiel 8.11.** Was ist die Wahrscheinlichkeit, bei 1000 Versuchen mindestens 170 Einsen zu würfeln? Dieses Experiment ist binomialverteilt mit  $n = 1000$  und  $p = 1/6$ , die gesuchte Wahrscheinlichkeit also

$$P(X \geq 170) = 1 - P(X \leq 169) = 1 - \sum_{k=0}^{169} \binom{1000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{1000-k},$$

was ziemlich unangenehm auszurechnen ist. Die Näherung durch die Normalverteilung mit Parametern  $\mu = 1000/6$  und  $\sigma^2 = 1000 \cdot (1/6) \cdot (5/6)$  liefert

$$P(X \geq 170) = 1 - P(X \leq 169) \approx 1 - \Phi\left(\frac{169 - \frac{1000}{6}}{\sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = 1 - 0.5784 = 0.4216.$$

(Der mit der Binomialverteilung errechnete Wert ist 0.4016.)

Macht man daraus 10000 Versuche und fragt nach mindesten 1700 Einsen, erhält man

$$P(X \geq 1700) \approx 1 - \Phi\left(\frac{1699 - \frac{10000}{6}}{\sqrt{10000 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = 1 - 0.807 = 0.193.$$

Für den „tatsächlichen“ Wert

$$1 - \sum_{k=0}^{1699} \binom{10000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10000-k} = 1 - 0.8110 = 0.189$$

hat MAPLE auf einer Sun Workstation über eine Stunde lang gerechnet.

## 9. Zusammengesetzte Zufallsvariablen

Sei  $X: \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable und  $h: \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion. Dann ist auch die Komposition oder Hintereinanderausführung  $Y = h \circ X$  definiert durch  $Y(\omega) = h(X(\omega))$  wieder eine Zufallsvariable.

Ist nun  $X$  stetig mit Verteilungsdichte  $f$  und Verteilung  $F$  und  $h$  streng monoton wachsend, so kann man die Verteilungsfunktion  $G$  von  $Y$  leicht ausrechnen. Wegen der Monotonie von  $h$  gilt nämlich  $h(X(\omega)) \leq x$  genau dann, wenn  $X(\omega) \leq h^{-1}(x)$  ist, wobei  $h^{-1}$  die Umkehrfunktion zu  $h$  bezeichne (die wegen der strengen Monotonie auch existiert); es folgt

$$G(x) = P(Y \leq x) = P(h \circ X \leq x) = P(X \leq h^{-1}(x)) = F(h^{-1}(x)),$$

also  $G = F \circ h^{-1}$ . (Ist  $h$  hingegen streng monoton fallend, so gilt  $h(X(\omega)) \leq x$  genau dann, wenn  $X(\omega) \geq h^{-1}(x)$  ist, und es folgt

$$G(x) = P(Y \leq x) = P(h \circ X \leq x) = P(X \geq h^{-1}(x)) = 1 - F(h^{-1}(x)),$$

also  $G = 1 - f \circ h^{-1}$ .)

Nimmt man zusätzlich noch an, dass  $h$  differenzierbar ist, so kann man die Verteilungsdichte  $g$  von  $Y$  durch Differenzieren gewinnen: ist wieder  $h$  streng monoton (wachsend oder fallend), so gilt nach Kettenregel und Umkehrsatz

$$g(u) = G'(u) = \frac{d}{du} F(h^{-1}(u)) = F'(h^{-1}(u)) \frac{1}{|h'(h^{-1}(u))|} = \frac{f(h^{-1}(u))}{|h'(h^{-1}(u))|}.$$

Insbesondere gilt für  $h' > 0$

$$\int_{-\infty}^{\infty} u g(u) du = \int_{-\infty}^{\infty} u \frac{f(h^{-1}(u))}{h'(h^{-1}(u))} du = \int_{-\infty}^{\infty} h(t) f(t) \frac{h'(t)}{h'(t)} dt = \int_{-\infty}^{\infty} h(t) f(t) dt$$

wobei wir die Substitution  $u = h(t)$ , also  $du = h'(t)dt$  benutzt haben. Dies zeigt im Spezialfall  $h' > 0$  den ersten Teil des folgenden Satzes.

**Satz 9.1.** Sei  $X: \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable,  $h: \mathbb{R} \rightarrow \mathbb{R}$  und  $Y = h \circ X$ .

(a) Sei  $X$  stetig mit Verteilungsdichte  $f$ . Dann gilt

$$E(Y) = \int_{-\infty}^{\infty} h(t)f(t) dt.$$

(b) Sei  $X$  diskret mit  $\text{Bild}(X) = \{x_0, x_1, \dots\}$  und  $p_i = P(X = x_i)$ . Dann gilt

$$E(Y) = \sum_{i=0}^{\infty} h(x_i)p_i.$$

Der zweite Teil des Satzes geht unmittelbar aus der Definition des Erwartungswerts hervor; wir hatten ihn im Spezialfall  $h(x) = x^2$  bei der Herleitung der Formel für die Varianz auch schon benutzt.

**Beispiele 9.2.** 1. Sei  $X$  eine stetige Zufallsvariable und  $Y = aX + b$  für Konstanten  $a, b \in \mathbb{R}$ ,  $a > 0$ . Dann ist also  $h(t) = at + b$  mit  $h'(t) = a$  und  $h^{-1}(u) = \frac{u-b}{a}$ ; es folgt

$$g(u) = a \cdot f\left(\frac{u-b}{a}\right),$$

$$E(Y) = \int_{-\infty}^{\infty} (at+b)f(t) dt = a \cdot \underbrace{\int_{-\infty}^{\infty} t f(t) dt}_{=E(X)} + b \cdot \underbrace{\int_{-\infty}^{\infty} f(t) dt}_{=1} = aE(X) + b.$$

2. Sei  $X$  rechteckverteilt (gleichverteilt) auf dem Intervall  $[0, 2]$ . Dann ist die Verteilungsdichte  $f$  gegeben durch

$$f(u) = \begin{cases} \frac{1}{2} & \text{für } 0 \leq u \leq 2, \\ 0 & \text{sonst.} \end{cases}$$

Sei  $Y$  die Fläche eines Quadrats mit Kantenlänge  $X$ , also  $Y = X^2$ . Dann ist  $Y = h \circ X$  mit  $h(t) = t^2$  und somit

$$E(Y) = \int_{-\infty}^{\infty} h(t) \frac{1}{2} dt = \int_0^2 \frac{t^2}{2} dt = \left[ \frac{t^3}{6} \right]_0^2 = \frac{4}{3}.$$

Wir wenden uns nun Verknüpfungen von zwei oder mehr Zufallsvariablen durch arithmetischen Operationen zu. Dazu ist es hilfreich, sich eine Familie  $(X_1, \dots, X_n)$  von Zufallsvariablen als eine neue, vektorwertige Zufallsvariable vorzustellen, auch wenn das unserer früheren Definition, dass Zufallsvariable Werte in den reellen Zahlen haben sollen, nicht entspricht:  $n$ -wertige Zufallsvariablen (wenn wir sie so nennen wollen) haben Werte in  $\mathbb{R}^n$ . Dies ist aber ein Scheinproblem; wir könnten schlicht darauf verzichten, solche  $n$ -Tupel von Zufallsvariablen wieder als Zufallsvariablen zu bezeichnen.

Wir beginnen mit zwei Zufallsvariablen, die auf der gleichen Ergebnismenge definiert sein sollen.

**Definition 9.3.** Seien  $X, Y: \Omega \rightarrow \mathbb{R}$  zwei Zufallsvariablen. Unter ihrer gemeinsamen Verteilungsfunktion versteht man die auf  $\Omega \times \Omega$  definierte Funktion

$$F(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}).$$

Man schreibt meist kürzer

$$P(X \leq x, Y \leq y) \quad \text{statt} \quad P(\{X \leq x\} \cap \{Y \leq y\}).$$

Die Verallgemeinerung auf mehr als zwei Zufallsvariablen ist ganz analog: die gemeinsame Verteilungsfunktion von  $n$  Zufallsvariablen  $X_1, \dots, X_n$ , die alle auf der Ergebnismenge  $\Omega$  definiert sind, ist die (auf  $\Omega^n$  erklärte) Funktion

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Die Verteilungen der einzelnen Komponenten  $X_i$  nennt man die *Randverteilungen* und bezeichnet sie mit  $F_{X_i}$  ( $1 \leq i \leq n$ ).

**Satz 9.4** (Eigenschaften). *Seien  $X, Y: \Omega \rightarrow \mathbb{R}$  Zufallsvariablen und  $F$  ihre gemeinsame Verteilungsfunktion.*

- (a)  $\lim_{\substack{x \rightarrow -\infty \\ y \rightarrow -\infty}} F(x, y) = 0$  und  $\lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} F(x, y) = 1$ .
- (b)  $0 \leq F(x, y) \leq 1$ .
- (c)  $\lim_{y \rightarrow \infty} F(x, y) = F_X(x)$ ,  $\lim_{x \rightarrow \infty} F(x, y) = F_Y(y)$ .
- (d)  $F$  ist in beiden Argumenten monoton: für  $x_1 \leq x_2$  und  $y_1 \leq y_2$  gelten die Ungleichungen
 
$$F(x_1, y_1) \leq F(x_2, y_1) \leq F(x_2, y_2), \quad F(x_1, y_1) \leq F(x_1, y_2) \leq F(x_2, y_2).$$
- (e) Es gilt  $P(X > x, Y \leq y) = F_Y(y) - F(x, y)$  und  $P(X \leq x, Y > y) = F_X(x) - F(x, y)$ .
- (f) Für  $a < b$  und  $c < d$  gilt

$$P(a < X \leq b, c < Y \leq d) = (F(b, d) - F(b, c)) - (F(a, d) - F(a, c)).$$

Diese Eigenschaften sind nicht schwer nachzurechnen und rechtfertigen die Vorstellung von  $F(x, y)$  als der Verteilungsfunktion einer neuen, mehrwertigen Zufallsvariablen.

Nimmt man an, dass  $F(x, y)$  zweimal stetig differenzierbar ist (also insbesondere die Randverteilungen stetig), so gilt

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y \frac{\partial^2 F}{\partial u \partial v}(u, v) \, du \, dv.$$

Das legt folgende Definition nahe:

**Definition 9.5.** *Seien  $X, Y: \Omega \rightarrow \mathbb{R}$  stetige Zufallsvariablen. Ist die gemeinsame Verteilungsfunktion  $F(x, y)$  zweimal stetig differenzierbar, so nennt man*

$$f(u, v) = \frac{\partial^2 F}{\partial u \partial v}(u, v)$$

*die gemeinsame Verteilungsdichte von  $X$  und  $Y$ .*

Die individuellen Verteilungsdichten von  $X$  und  $Y$  nennt man die *Randverteilungsdichten* und bezeichnet sie mit  $f_X$  beziehungsweise  $f_Y$ . Die Eigenschaften fassen wir wieder zusammen:

**Satz 9.6** (Eigenschaften). Seien  $X, Y$  stetige Zufallsvariablen mit gemeinsamer Verteilung  $F(x, y)$  und gemeinsamer Verteilungsdichte  $f(u, v)$ . Dann gilt

$$(a) F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv.$$

$$(b) f(u, v) \geq 0.$$

$$(c) f_X(x) = \int_{-\infty}^{\infty} f(x, v) dv, f_Y(y) = \int_{-\infty}^{\infty} f(u, y) du.$$

$$(d) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) du dv = 1.$$

$$(e) \text{Für einen Bereich } B \subset \mathbb{R}^2 \text{ ist } P((X, Y) \in B) = \iint_B f(u, v) du dv.$$

Wir kommen nun zu einem zentralen Begriff, dem der Unabhängigkeit von Zufallsvariablen. Hat man zwei Zufallsvariablen auf der gleichen Ergebnismenge, so ist es häufig der Fall, dass das Ergebnis der einen mit dem der anderen zusammenhängt. Nimmt man bei Produktion und Vertrieb eines Gutes als eine Variable etwa den Ausschussanteil und als zweite die Reklamationskosten, so ist können sie sich durchaus beeinflussen.

**Definition 9.7.** Zwei Zufallsvariablen  $X, Y: \Omega \rightarrow \mathbb{R}$  mit gemeinsamer Verteilungsfunktion  $F(x, y)$  heißen stochastisch unabhängig, wenn

$$F(x, y) = F_X(x) \cdot F_Y(y)$$

gilt.

*Bemerkung.* 1.  $X$  und  $Y$  sind also genau dann stochastisch unabhängig, wenn

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

gilt, mit anderen Worten, wenn die Ereignisse  $\{X \leq x\}$  und  $\{Y \leq y\}$  unabhängig sind (daher der Name).

2. Sind  $X, Y$  stetig und haben eine gemeinsame Verteilungsdichte  $f(u, v)$ , so sind sie genau dann stochastisch unabhängig, wenn die gemeinsame Verteilungsdichte das Produkt der Randverteilungsdichten ist, also

$$f(u, v) = f_X(u) \cdot f_Y(v)$$

gilt.

3. Die Definition lässt sich leicht auf mehr als zwei Variablen verallgemeinern:  $X_1, \dots, X_n$  sind unabhängig, wenn

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$$

gilt. Insbesondere folgt dann, dass die Variablen  $X_i$  auch paarweise unabhängig sind.

**Beispiel 9.8.** Sei

$$f(u, v) = \begin{cases} 0 & \text{für } u \leq 0 \text{ oder } v \leq 0, \\ 2e^{-u-2v} & \text{für } u > 0 \text{ und } v > 0 \end{cases}$$

die gemeinsame Verteilungsdichte der Zufallsvariablen  $X, Y$ . Wir überprüfen zunächst, dass  $f$  tatsächlich als eine Dichtefunktion in Frage kommt: es gilt

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) du dv = \int_0^{\infty} \int_0^{\infty} 2e^{-u-2v} du dv = \int_0^{\infty} 2e^{-2v} [-e^{-u}]_0^{\infty} = \int_0^{\infty} 2e^{-2v} dv = [-e^{-2v}]_0^{\infty} = 1$$

wie verlangt.

Es ist  $f$  das Produkt seiner zwei Randverteilungen  $f_X(u) = e^{-u}$  und  $f_Y(v) = 2e^{-2v}$ ; dies sind beides Exponentialverteilungen mit den Parametern  $\lambda = 1$  beziehungsweise  $\lambda = 2$ , so dass man sich  $X$  und  $Y$  als Wartezeiten mit Erwartungswerten 1 und  $1/2$  vorstellen kann.

Was ist nun die Wahrscheinlichkeit  $P(X \geq Y)$ , also dass man auf das Ereignis mit Erwartungswert 1 mindestens so lange warten muss wie auf das Ereignis mit Erwartungswert  $1/2$ ?

Diese Wahrscheinlichkeit ist laut Satz 9.6 (e) gegeben als

$$P(X \geq Y) = \iint_B f(u, v) du dv \quad \text{mit} \quad B = \{(u, v) \mid u \geq v > 0\};$$

man rechnet daher aus

$$P(X \geq Y) = \int_0^\infty \int_v^\infty 2e^{-u-2v} du dv = \int_0^\infty 2e^{-2v} [e^{-u}]_v^\infty dv = \int_0^\infty 2e^{-3v} dv = \left[-\frac{2}{3}e^{-3v}\right]_0^\infty = \frac{2}{3}.$$

Als nächstes wollen wir einfache Funktionen wie Summe und Produkt zweier Zufallsvariablen betrachten. Seien also  $X, Y: \Omega \rightarrow \mathbb{R}$  zwei Zufallsvariablen, dann sind auch  $X + Y$  und  $X \cdot Y$  wieder Zufallsvariablen, und es stellt sich die Frage, ob und wie, gegebenenfalls mit Zusatzvoraussetzungen, man Kenngrößen wie Erwartungswert und Varianz dieser neuen Variablen aus denen der Verteilungen von  $X$  und  $Y$  erhalten kann.

Wie beginnen mit dem diskreten Fall. Sei also  $\text{Bild}(X) = \{x_0, x_1, \dots\}$ , und  $\text{Bild}(Y) = \{y_0, y_1, \dots\}$  mit  $p_i = P(X = x_i)$  und  $q_j = P(Y = y_j)$ . Für die gemeinsame Verteilung  $F$  sei noch

$$r_{ij} = P(X = x_i, Y = y_j).$$

$X$  und  $Y$  sind genau dann unabhängig, wenn  $r_{ij} = p_i \cdot q_j$  gilt.

**Satz 9.9.** Seien  $X_1, \dots, X_n$  diskrete Zufallsvariablen auf der gleichen Ergebnismenge  $\Omega$  und  $a_1, \dots, a_n \in \mathbb{R}$ . Dann gilt:

(a) Der Erwartungswert ist linear. das heißt,

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i).$$

(b) Sind  $X_1, \dots, X_n$  unabhängig, so gilt

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

(c) Sind  $X_1, \dots, X_n$  unabhängig, so gilt

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

**Beispiel 9.10.** Zwei unterscheidbare Würfel werden geworfen. Es seien  $X_1$  und  $X_2$  die dabei erzielten Augenzahlen der beiden Würfel.  $X_1$  und  $X_2$  sind sicherlich unabhängig. Seien ferner  $X = \max(X_1, X_2)$  mit  $\text{Bild}(X) = \{1, 2, 3, 4, 5, 6\}$  und  $Y$  die Anzahl der geraden Augenzahlen mit  $\text{Bild}(Y) = \{0, 1, 2\}$ . Dann gilt  $P(X = 4) = 7/36$ , denn das Ereignis  $X = 4$  besteht gerade aus den 7 Paaren  $(4, 1), (1, 4), (4, 2), (2, 4), (4, 3), (3, 4), (4, 4)$ . Weiterhin ist  $P(Y = 1) = 1/2$ , denn in der Hälfte aller möglichen Ausgänge ist genau eine der beiden Augenzahlen gerade. Andererseits ist

$$P(X = 4, Y = 1) = P(\{(4, 1), (1, 4), (4, 3), (3, 4)\}) = \frac{4}{36} \neq \frac{7}{36} \cdot \frac{1}{2} = P(X = 4) \cdot P(Y = 1).$$

Die Zufallsvariablen  $X$  und  $Y$  sind also nicht unabhängig.

Als Anwendung von Satz 9.9 erhält man noch

$$\begin{aligned} E(X_1 + X_2) &= E(X_1) + E(X_2) = 3.5 + 3.5 = 7, \\ \text{Var}(X_1 + X_2) &= \text{Var}(X_1) + \text{Var}(X_2) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}, \\ E(X_1 \cdot X_2) &= E(X_1) \cdot E(X_2) = (3.5) \cdot (3.5) = 12.25. \end{aligned}$$

Der obige Satz zeigt, wie man den Erwartungswert der Summe diskreter Zufallsvariablen bestimmen kann. Bevor wir die gleiche Frage für stetige Variablen behandeln können, muss erst einmal geklärt sein, welche Verteilungsdichte der Summe entspricht; dabei beschränken wir uns auf unabhängige Variablen. Wegen der Ähnlichkeit des Problems geben wir den diskreten Fall gleich mit an.

**Satz 9.11.** *Seien  $X$  und  $Y$  stochastisch unabhängige Zufallsvariablen.*

(a) *Seien  $X, Y$  diskret mit  $\text{Bild}(X) = \{x_0, x_1, \dots\}$  und  $\text{Bild}(Y) = \{y_0, y_1, \dots\}$ . Dann gilt*

$$P(X + Y = z) = \sum_{i=0}^{\infty} P(X = x_i)P(Y = z - x_i).$$

(b) *Seien  $X, Y$  stetig mit Verteilungsdichten  $f$  für  $X$  und  $g$  für  $Y$ . Die Verteilungsdichte für  $X + Y$  ist die Funktion*

$$h(u) = \int_{-\infty}^{\infty} f(t)g(u - t) dt = \int_{-\infty}^{\infty} f(u - t)g(t) dt.$$

*Diese Funktion schreibt man auch als  $h(u) = (f \star g)(u)$  und nennt sie die Faltung von  $f$  und  $g$ .*

Der diskrete Teil des Satzes ist leicht einzusehen: man summiert einfach die Wahrscheinlichkeiten aller Fälle auf, in denen die Summe der Zufallsvariablen  $X + Y$  den Wert  $z$  ergibt: wegen der Voraussetzung der Unabhängigkeit ist die Wahrscheinlichkeit jedes Summanden  $P(X = x_i, Y = z - x_i)$  gerade das Produkt  $P(X = x_i)P(Y = z - x_i)$ . Der stetige Fall übersteigt die hier zur Verfügung stehenden technischen Möglichkeiten, ebenso wie der Beweis des folgenden, wichtigen Spezialfalls:

**Satz 9.12.** *Sei  $X_i$  normalverteilt mit Erwartungswert  $\mu_i$  und Varianz  $\sigma_i^2$  für  $i = 1, 2$ . Dann ist die Summe  $X_1 + X_2$  ebenfalls normalverteilt, und zwar mit Erwartungswert  $\mu_1 + \mu_2$  und Varianz  $\sigma_1^2 + \sigma_2^2$ .*

Im Fall von normalverteilten Zufallsvariablen ist die Situation also wie im diskreten Fall.