

Probabilistische Datenwissenschaft für die Psychologie

Dirk Ostwald

Inhaltsverzeichnis

Willkommen	3
Zitation	3
Lizenz	3
I. Mathematische Grundlagen	4
1. Sprache und Logik	5
1.1. Mathematik ist eine Sprache	5
1.2. Grundbausteine	6
1.3. Aussagenlogik	7
1.4. Äquivalenzumformungen	12
1.5. Beweistechniken	13
1.6. Selbstkontrollfragen	14
2. Mengen	15
2.1. Grundlegende Definitionen	15
2.2. Verknüpfungen	18
2.3. Spezielle Mengen	20
2.4. Selbstkontrollfragen	24
3. Summen, Produkte, Potenzen	25
3.1. Summen	25
3.2. Produkte	28
3.3. Potenzen	29
3.4. Selbstkontrollfragen	31
4. Funktionen	32
4.1. Definition und Eigenschaften	32
4.2. Funktionentypen	36
4.3. Elementare Funktionen	39
4.4. Selbstkontrollfragen	42
5. Folgen, Grenzwerte, Stetigkeit	43
5.1. Folgen	43
5.2. Grenzwerte	45
5.3. Stetigkeit	47
6. Differentialrechnung	50
6.1. Definitionen und Rechenregeln	50
6.2. Analytische Optimierung	55
6.3. Differentialrechnung multivariater reellwertiger Funktionen	57
6.4. Selbstkontrollfragen	65

7. Integralrechnung	66
7.1. Unbestimmte Integrale	66
7.2. Bestimmte Integrale	68
7.3. Uneigentliche Integrale	75
7.4. Mehrdimensionale Integrale	76
7.5. Selbstkontrollfragen	78
8. Vektoren	79
8.1. Reeller Vektorraum	79
8.2. Euklidischer Vektorraum	84
8.3. Lineare Unabhängigkeit	90
8.4. Vektorraumbasen	92
8.5. Selbstkontrollfragen	95
9. Matrizen	97
9.1. Definition	97
9.2. Grundlegende Matrixoperationen	99
9.3. Matrixmultiplikation	104
9.4. Matrixinversion	108
9.5. Determinanten	111
9.6. Spezielle Matrizen	115
9.7. Literaturhinweise	119
9.8. Selbstkontrollfragen	119
9.9. Eigenanalyse	119
9.10. Eigenvektoren und Eigenwerte	120
9.11. Orthonormalzerlegung	125
9.12. Singulärwertzerlegung	128
9.13. Literaturhinweise	129
9.14. Selbstkontrollfragen	129
10. Deskriptivstatistiken	130
10.1. Häufigkeitsverteilungen	132
10.2. Histogramme	134
10.3. Empirische Verteilungsfunktionen	137
10.4. Quantile und Boxplots	140
10.5. Maße der Zentralen Tendenz	143
10.6. Maße der Variabilität	152
Referenzen	159

Willkommen

Herzlich willkommen zur Arbeitsversion von *Probabilistische Datenwissenschaft für die Psychologie (PDWP)*, einem Lehrbuch zur datenwissenschaftlichen Methodenlehre am Institut für Psychologie der Otto-von-Guericke-Universität Magdeburg.

Zitation

Ostwald, D. (2024) Probabilistische Datenwissenschaft für die Psychologie. [10.5281/zenodo.10730199](https://doi.org/10.5281/zenodo.10730199)

Lizenz

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Part I.

Mathematische Grundlagen

1. Sprache und Logik

1.1. Mathematik ist eine Sprache

Mathematik ist die Sprache der naturwissenschaftlichen Modellbildung. So entspricht zum Beispiel der Ausdruck

$$F = ma \tag{1.1}$$

im Sinne des zweiten Newtonschen Axioms einer Theorie zur Bewegung von Objekten unter der Einwirkung von Kräften (Newton (1687)). Gleichermäßen entspricht der Ausdruck

$$\max_{q(z)} \int q(z) \ln \left(\frac{p(y, z)}{q(z)} \right) dz \tag{1.2}$$

im Sinne der Variational Inference der zeitgenössischen Theorie zur Funktionsweise des Gehirns (Friston (2005)). Mathematische Symbolik dient dabei insbesondere der genauen Kommunikation wissenschaftlicher Erkenntnisse und zielt darauf ab, komplexe Sachverhalte exakt und effizient zu beschreiben. Wie beim reflektierten Umgang mit jeder Form von Sprache steht also die Frage “Was soll das heißen?” als Leitfrage im Umgang mit mathematischen Inhalten und Symbolismen immer im Vordergrund.

Als Sprachgebäude weist die Mathematik einige Besonderheiten auf. Zum einen sind ihre Inhalte oft abstrakt. Dies rührt daher, dass sich die Mathematik um eine möglichst breite Allgemeinverständlichkeit und Anwendbarkeit bemüht. Mathematische Zugänge zu den Phänomenen der Welt sind dabei an einer möglichst einfachen Transferierbarkeit von Erkenntnissen in andere Kontexte interessiert. Um dies zu ermöglichen, versucht die Mathematik möglichst genau und verständlich, also im Sinne präziser Begriffe zu arbeiten. Sie geht dabei insbesondere streng hierarchisch vor, so dass an späterer Stelle eingeführte Begriffe oft ein gutes Verständnis der ihnen zugrundeliegenden und an früherer Stelle eingeführten Begriffe voraussetzen.

Die Genauigkeit der mathematischen Sprache impliziert eine hohe Informationsdichte. Sie ist daher eher nüchtern und lässt überflüssiges weg, so dass in mathematischen Texten im besten Fall *alles* für die Kommunikation einer Idee relevant ist. Als Rezipient:in mathematischer Texte nimmt man die Informationsdichte mathematischer Texte anhand des hohen Verbrauchs kognitiver Energie beim Lesen eines Textes wahr. Dieser hohe Energieverbrauch gebietet insbesondere Ruhe und Langsamkeit bei einem auf ein gutes Verständnis abzielenden Lesen. Als Leitsatz im Umgang mit mathematischen Texten mag dabei folgendes Zitat dienen: “Einen mathematischen Text kann man nicht lesen wie einen Roman, man muss ihn sich erarbeiten” (Unger (2000)). Nach dem Lesen eines kurzen mathematischen Textes sollte man sich immer kritisch fragen, ob man das Gelesene wirklich verstanden hat oder ob man zur Klärung des Sachverhaltes weitere Quellen heranziehen sollte. Auch ist es hilfreich, sich im Sinne des berühmten Zitats “What I cannot create, I do not understand” von Richard Feynman eigene Aufzeichnungen anzufertigen und mathematische Sprachgebäude selbst nachzukonstruieren.

Möchte man sich also die Welt der naturwissenschaftliche Modellbildung erschließen, so ist es hilfreich, beim Umgang mit ihrer mathematischen Ausdrucksweise und Symbolik die gleichen Strategien wie beim Erlernen einer Fremdsprache anzuwenden. Hierzu gehört neben dem Eintauchen in den entsprechenden Sprachraum, also der ständigen Exposition mit mathematischen Ausdrucksweisen, sicherlich auch zunächst einmal das Auswendiglernen von Begriffen und das Übersetzen von Texten in die Alltagssprache. Ein tiefes und sicheres Verständnis mathematischer Modellbildung ergibt sich dann insbesondere durch die Anwendung mathematischer Herangehensweisen in schriftlicher und mündlicher Form.

1.2. Grundbausteine

Im Folgenden stellen wir mit den Begriffen der *Definition*, des *Theorems* und des *Beweises* drei Grundbausteine mathematischer Kommunikation vor, die uns durchgängig begleiten.

Definition

Eine *Definition* ist eine Grundannahme eines mathematischen Systems, die innerhalb dieses Systems weder begründet noch deduktiv abgeleitet wird. Definitionen können nur nach ihrer Nützlichkeit innerhalb eines mathematischen Systems bewertet werden. Eine Definition lernt man am besten erst einmal auswendig und hinterfragt sie erst dann, wenn man ihren Nutzen in der Anwendung verstanden hat oder von diesem nicht überzeugt ist. Etwas Entspannung und Ruhe beim Umgang mit auf den ersten Blick komplexen Definitionen ist generell hilfreich. Um zu kennzeichnen, dass wir ein Symbol als etwas definieren, nutzen wir die Schreibweise “:=”. Zum Beispiel definiert der Ausdruck “ $a := 2$ ” das Symbol a als die Zahl Zwei. Definitionen enden in diesem Text immer mit dem Symbol •.

Theorem

Ein *Theorem* ist eine mathematische Aussage, die mittels eines Beweises als wahr (richtig) erkannt werden kann. Das heißt, ein Theorem wird immer aus Definitionen und/oder anderen Theoremen hergeleitet. Theoreme sind in diesem Sinne die empirischen Ergebnisse der Mathematik. Im Deutschen werden Theoreme oft auch als *Sätze* bezeichnet. In der angewandten, datenanalytischen Mathematik sind Theoreme häufig für Berechnungen hilfreich. Es lohnt sich also, sie auswendig zu lernen, da sie meist die Grundlage für Datenauswertung und Dateninterpretation bilden. Oft tauchen in Theoremen Gleichungen auf. Diese ergeben sich dabei aus den Voraussetzungen des Theorems. Um Gleichungen zu kennzeichnen nutzen wir das Gleichheitszeichen “=”. So besagt also beispielsweise der Ausdruck “ $a = 2$ ” in einem gegebenen Kontext, dass aufgrund bestimmter Voraussetzungen das Symbol oder die Variable a den Wert zwei hat. Theoreme enden in diesem Text immer mit dem Symbol ◦.

Beweis

Ein *Beweis* ist eine logische Argumentationskette, die auf bekannte Definitionen und Theoreme zurückgreift, um die Wahrheit (Richtigkeit) eines Theorems zu belegen. Kurze Beweise tragen dabei oft zum Verständnis eines Theorems bei, lange Beweise eher nicht. Beweise sind also insbesondere die Antwort auf die Frage, warum eine mathematische Aussage gilt (“Warum ist das so?”). Beweise lernt man nicht auswendig. Wenn Beweise

kurz sind, ist es sinnvoll, sie durchzuarbeiten, da sie meist als bekannt vorausgesetzte Inhalte wiederholen. Wenn sie lang sind, ist es sinnvoller sie zunächst zu übergehen, um sich nicht in Details zu verlieren und vom eigentlichen Weg durch das mathematische Gebäude abzukommen. Beweise enden in diesem Text immer mit dem Symbol \square .

Neben Definitionen, Theoremen und Beweisen gibt es mit *Axiomen*, *Lemmata*, *Korollaren* und *Vermutungen* noch weitere typische Bausteine mathematischer Texte. Wir werden diese Begriffe nicht verwenden und geben deshalb für sie nur einen kurzen Überblick.

- *Axiome* sind unbeweisbare Theoreme, in dem Sinne, als dass sie als Grundannahmen zum Aufbau mathematischer Systeme dienen. Der Übergang zwischen Definitionen und Axiomen ist dabei oft fließend. Da wir mathematisch nicht besonders tief arbeiten, bevorzugen wir in den allermeisten Fällen den Begriff der Definition.
- Ein *Lemma* ist ein “Hilfstheorem”, also eine mathematische Aussage, die zwar bewiesen wird, aber nicht so bedeutend ist wie ein Theorem. Da wir einerseits auf bedeutende Inhalte fokussieren und andererseits mathematische Aussagen nicht diskriminieren wollen, verzichten wir auf diesen Begriff und nutzen stattdessen durchgängig den Begriff des Theorems.
- Ein *Korollar* ist eine mathematische Aussage, die sich durch einen einfachen Beweis aus einem Theorem ergibt. Da die “Einfachheit” mathematischer Beweise eine relative Eigenschaft ist, verzichten wir auf diesen Begriff und nutzen stattdessen auch hier durchgängig den Begriff des Theorems.
- *Vermutungen* sind mathematische Aussagen, von denen unbekannt ist, ob sie beweisbar oder widerlegbar sind. Da wir im Bereich der angewandten Mathematik arbeiten, treffen wir nicht auf Vermutungen.

1.3. Aussagenlogik

Nachdem wir nun einige Grundbausteine mathematischer Modellbildung kennengelernt haben, wollen wir uns mit der *Aussagenlogik* einem einfachen System nähern, das es erlaubt, Beziehungen zwischen mathematischen Aussagen herzustellen und zu formalisieren. Die Aussagenlogik spielt zum Beispiel in der Definition von Mengenoperationen, bei Optimierungsbedingungen von Funktionen und in vielen Beweisen eine tragende Rolle. In der datenanalytischen Anwendung ist die Aussagenlogik die Grundlage der Booleschen Logik der Programmierung. In der mathematischen Psychologie schließlich ist die Aussagenlogik beispielsweise die Grundlage der Repräsentationstheorie des Messens.

Wir beginnen mit der Definition des Begriffs der mathematischen *Aussage*.

Definition 1.1 (Aussage). Eine *Aussage* ist ein Satz, dem eindeutig die Eigenschaft *wahr* oder *falsch* zugeordnet werden kann.

•

Das Adjektiv *wahr* kann auch als *richtig* verstanden werden. Wir kürzen wahr mit “w” und falsch mit “f” ab. Im Körper der reellen Zahlen ist zum Beispiel die Aussage $1+1=2$ wahr und die Aussage $1+1=3$ falsch. Man beachte, dass die Binärität des Wahrheitsgehalts von Aussagen eine Grundannahme der Aussagenlogik und damit formal wissenschaftlich und

nicht empirisch zu verstehen ist. Wahrheitsgehalte beziehen sich nicht auf Definitionen, Definitionen sind immer wahr.

Eine erste Möglichkeit, mit Aussagen zu arbeiten, ist, sie zu negieren. Dies führt auf folgende Definition.

Definition 1.2 (Negation). A sei eine Aussage. Dann ist die *Negation von A* die Aussage, die falsch ist, wenn A wahr ist und die wahr ist, wenn A falsch ist. Die Negation von A wird mit $\neg A$, gesprochen als “nicht A ”, bezeichnet.

•

Beispielsweise ist die Negation der Aussage “Die Sonne scheint” die Aussage “Die Sonne scheint nicht”. Die Negation der Aussage $1 + 1 = 2$ ist die Aussage $1 + 1 \neq 2$ und die Negation der Aussage $x > 1$ ist die Aussage $x \leq 1$. Tabellarisch stellt man die Definition der Negation einer Aussage A wie folgt dar:

Tabelle 1.1. Wahrheitstafel der Negation

A	$\neg A$
w	f
f	w

Tabellen dieser Form nennt man *Wahrheitstafeln*. Sie sind ein beliebtes Hilfsmittel in der Aussagenlogik, das wir im Folgenden oft einsetzen werden.

Möchte man zwei Aussagen logisch verbinden, so bieten sich zunächst die Begriffe der *Konjunktion* und *Disjunktion* an.

Definition 1.3 (Konjunktion). A und B seien Aussagen. Dann ist die *Konjunktion von A und B* die Aussage, die dann und nur dann wahr ist, wenn A und B beide wahr sind. Die Konjunktion von A und B wird mit $A \wedge B$, gesprochen als “ A und B ”, bezeichnet.

•

Die Definition der Konjunktion impliziert folgende Wahrheitstafel:

Tabelle 1.2. Wahrheitstafel der Konjunktion

A	B	$A \wedge B$
w	w	w
w	f	f
f	w	f
f	f	f

Als Beispiel sei A die Aussage $2 \geq 1$ und B die Aussage $2 > 1$. Da sowohl A und B wahr sind, ist auch die Aussage $2 \geq 1 \wedge 2 > 1$ wahr. Als weiteres Beispiel sei A die Aussage $1 \geq 1$ und B die Aussage $1 > 1$. Hier ist nur A wahr und B falsch. Also ist die Aussage $1 \geq 1 \wedge 1 > 1$ falsch.

Definition 1.4 (Disjunktion). A und B seien Aussagen. Dann ist die *Disjunktion von A und B* die Aussage, die dann und nur dann wahr ist, wenn mindestens eine der beiden Aussagen A und B wahr ist. Die Disjunktion von A und B wird mit $A \vee B$, gesprochen als “ A oder B ”, bezeichnet.

•

Die Definition der Disjunktion impliziert folgende Wahrheitstafel:

Tabelle 1.3. Wahrheitstafel der Disjunktion

A	B	$A \vee B$
w	w	w
w	f	w
f	w	w
f	f	f

$A \vee B$ ist also insbesondere auch dann wahr, wenn A und B beide wahr sind. Damit ist das hier betrachtete “oder” genauer ein “und/oder”. Man nennt die Disjunktion daher auch ein “nicht-exklusives oder”. Als Beispiel sei A die Aussage $2 \geq 1$ und B die Aussage $2 > 1$. A ist wahr und B ist wahr. Also ist die Aussage $2 \geq 1 \vee 2 > 1$ wahr. Sei nun wiederum A die Aussage $1 \geq 1$ wahr und B die Aussage $1 > 1$. Dann ist A wahr und B falsch. Also ist die Aussage $1 \geq 1 \vee 1 > 1$ wahr.

Eine Möglichkeit, Aussagen in einen mechanischen logischen Zusammenhang zu stellen, ist die *Implikation*. Diese ist wie folgt definiert.

Definition 1.5 (Implikation). A und B seien Aussagen. Dann ist die *Implikation*, bezeichnet mit $A \Rightarrow B$, die Aussage, die dann und nur dann falsch ist, wenn A wahr und B falsch ist. A heißt dabei die *Voraussetzung (Prämisse)* und B der *Schluss (Konklusion)* der Implikation. $A \Rightarrow B$ spricht man als “aus A folgt B ”, “ A impliziert B ”, oder “wenn A , dann B ”.

•

Die Definition der Implikation kann durch folgende Wahrheitstafel dargestellt werden:

Tabelle 1.4. Wahrheitstafel der Implikation

A	B	$A \Rightarrow B$
w	w	w
w	f	f
f	w	w
f	f	w

Ein intuitives Verständnis der Definition der Implikation im Sinne obiger Wahrheitstafel ergibt sich am ehesten, indem man sie als Versuch liest, die intuitive Vorstellung einer Folgerung im Kontext der Aussagenlogik abzubilden und zu formalisieren. Um dies nachzuvollziehen, liest man die Wahrheitszustände der obigen Tabelle am besten in der

Reihenfolge Wahrheitszustand von A , Wahrheitszustand von $A \Rightarrow B$ und betrachtet schließlich den Wahrheitszustand von B . Liest man die Wahrheitstafel auf diese Weise, so sieht man, dass wenn A wahr ist und $A \Rightarrow B$ wahr ist, B wahr ist. Konstruiert man basierend auf einer wahren Aussage also (zum Beispiel durch das Umformen von Gleichungen) eine wahre Implikation, so folgt, dass auch B wahr ist. Ist dies nicht möglich, wenn also gilt, dass, wenn A wahr ist, $A \Rightarrow B$ immer falsch ist, dann ist auch B falsch. So mag man Aussagen widerlegen. Schließlich sieht man, dass wenn A falsch ist und $A \Rightarrow B$ wahr ist, B wahr oder falsch sein kann. Nur aus einer wahren Voraussetzung folgt also bei wahrer Implikation immer eine wahre Konklusion. Insbesondere genügt die Definition der Implikation damit der Forderung “Aus Falschem folgt beliebiges (ex falso sequitur quodlibet)”. Man kann aus falschen Aussagen also mithilfe der Implikation nichts Sinnvolles folgern.

Im Kontext der Implikation ergeben sich die Begriffe der *hinreichenden* und der *notwendigen* Bedingungs Aussagen: Wenn $A \Rightarrow B$ wahr ist, sagt man, dass “ A hinreichend für B ist” und dass “ B notwendig für A ist”. Diese Sprechweise erklärt sich im Kontext der Implikation folgendermaßen: Wenn $A \Rightarrow B$ wahr ist, gilt dass, wenn A wahr ist, auch B wahr ist. Die Wahrheit von A reicht also für die Wahrheit von B aus. A ist also hinreichend (ausreichend) für B . Weiterhin gilt, dass wenn $A \Rightarrow B$ wahr ist, dass wenn B falsch ist, dann auch A falsch ist. Die Wahrheit von B ist also für die Wahrheit von A notwendig.

Eine sehr häufig auftretender Zusammenhang zwischen zwei Aussagen ist ihre *Äquivalenz*.

Definition 1.6 (Äquivalenz). A und B seien Aussagen. Die *Äquivalenz von A und B* ist die Aussage, die dann und nur dann wahr ist, wenn A und B beide wahr sind oder wenn A und B beide falsch sind. Die Äquivalenz von A und B wird mit $A \Leftrightarrow B$ bezeichnet und gesprochen als “ A genau dann wenn B ” oder “ A ist äquivalent zu B ”.

•

Die Definition der Äquivalenz impliziert folgende Wahrheitstafel:

Tabelle 1.5. Wahrheitstafel der Äquivalenz

A	B	$A \Leftrightarrow B$
w	w	w
w	f	f
f	w	f
f	f	w

Die Definition des Begriffes der *logischen Äquivalenz* erlaubt es unter anderem, die Äquivalenz zweier Aussagen mithilfe von Implikationen nachzuweisen.

Definition 1.7 (Logische Äquivalenz). Zwei Aussagen heißen *logisch äquivalent*, wenn ihre Wahrheitstabellen gleich sind.

•

Als Beispiele für logische Äquivalenzen, die häufig in Beweisargumentationen genutzt werden, zeigen wir die Aussagen folgenden Theorems.

Theorem 1.1 (Logische Äquivalenzen). *A und B seien zwei Aussagen. Dann sind folgende Aussagen logisch äquivalent:*

- (1) $A \Leftrightarrow B$ und $(A \Rightarrow B) \wedge (B \Rightarrow A)$
- (2) $A \Rightarrow B$ und $(\neg B) \Rightarrow (\neg A)$

◦

Beweis. Nach Definition des Begriffs der logischen Äquivalenz müssen wir zeigen, dass die Wahrheitstabellen der betrachteten Aussagen gleich sind. Wir zeigen erst (1), dann (2).

(1) Wir erinnern an die Wahrheitstafel von $A \Leftrightarrow B$:

A	B	$A \Leftrightarrow B$
w	w	w
w	f	f
f	w	f
f	f	w

Wir betrachten weiterhin die Wahrheitstafel von $(A \Rightarrow B) \wedge (B \Rightarrow A)$:

A	B	$A \Rightarrow B$	$B \Rightarrow A$	$(A \Rightarrow B) \wedge (B \Rightarrow A)$
w	w	w	w	w
w	f	f	w	f
f	w	w	f	f
f	f	w	w	w

Der Vergleich der Wahrheitstafel von $A \Leftrightarrow B$ mit den ersten beiden und der letzten Spalte der Wahrheitstafel von $(A \Rightarrow B) \wedge (B \Rightarrow A)$ zeigt ihre Gleichheit.

(2) Wir erinnern an die Wahrheitstafel von $A \Rightarrow B$:

A	B	$A \Rightarrow B$
w	w	w
w	f	f
f	w	w
f	f	w

Wir betrachten weiterhin die Wahrheitstafel von $(\neg B) \Rightarrow (\neg A)$:

A	B	$\neg B$	$\neg A$	$(\neg B) \Rightarrow (\neg A)$
w	w	f	f	w
w	f	w	f	f
f	w	f	w	w
f	f	w	w	w

Der Vergleich der Wahrheitstafel von $A \Rightarrow B$ mit den ersten beiden und der letzten Spalte der Wahrheitstafel von $(\neg B) \Rightarrow (\neg A)$ zeigt ihre Gleichheit. \square

Die erste Aussage von Theorem 1.1 besagt, dass die Aussage “ A und B sind äquivalent” logisch äquivalent zur Aussage “Aus A folgt B ” und zur Aussage “Aus B folgt A ” ist. Dies ist die Grundlage für viele sogenannte *direkte Beweise* mithilfe von Äquivalenzumformungen. Die zweite Aussage von Theorem 1.1 besagt, dass die Aussage “Aus A folgt B ” logisch äquivalent zur Aussage “Aus nicht B folgt nicht A ” ist. Dies ist die Grundlage für die Technik des *indirekten Beweises*. Wir betrachten diese Beweistechniken im folgenden Abschnitt genauer. Zunächst fassen wir die Bedeutungen der in diesem Abschnitt eingeführte Symbole noch einmal in untenstehender Tabelle zusammen.

Tabelle 1.10. Symbolübersicht. A und B sind Aussagen, d.h. A und B sind entweder wahr oder falsch.

Symbol	Bedeutung	Bemerkung
$\neg A$	Nicht A	Wahr, wenn A falsch ist und umgekehrt
$A \wedge B$	A und B	Nur wahr, wenn A und B beide wahr sind
$A \vee B$	A und/oder B	Wahr, wenn mindestens eine der Aussagen wahr ist
$A \Rightarrow B$	Aus A folgt B	B ist notwendig für A , A ist hinreichend für A
$A \Leftrightarrow B$	A ist äquivalent zu B	Es gelten $A \Rightarrow B$ und $B \Rightarrow A$

1.4. Äquivalenzumformungen

Mathematische Probleme führen oft auf den Fall, dass Information über eine unbekannte Variable implizit mithilfe einer Gleichung oder einer Ungleichung dargestellt wird. Um die Information über die entsprechende Variable explizit darzustellen, also im Falle von Gleichungen ihren Wert zu bestimmen oder im Falle von Ungleichungen den Zahlenbereich zu ermitteln, in dem der Wert der Variable liegt, nutzt man *Äquivalenzumformungen*. Wir betrachten hier lediglich die aus der Schulmathematik bekannten Äquivalenzumformungen bezüglich Gleichungen und Ungleichungen reeller Variablen. Äquivalenzumformungen von Gleichungen und Ungleichungen haben dabei die Eigenschaft, dass sich der Wahrheitsgehalt der durch eine Gleichung oder Ungleichung formulierten Aussage bei Anwendung der entsprechenden Umformung nicht ändert. Dabei werden stets beide Seiten der Gleichung oder Ungleichung umgeformt. Damit es sich bei der Anwendung einer mathematischen Operation auf eine Gleichungs- oder Ungleichungsaussage A in eine Gleichungs- bzw. Ungleichungsaussage B um eine Äquivalenzumformung der Form $A \Leftrightarrow B$ handelt, müssen bekanntlich sowohl $A \Rightarrow B$ als auch $B \Rightarrow A$ gelten. Dies impliziert, dass es sich bei Äquivalenzumformungen um umkehrbare (invertierbare) Operationen handelt.

Bei Gleichungen sind zulässige Äquivalenzumformungen insbesondere

- die Addition einer Zahl auf beiden Seiten der Gleichung
- die Subtraktion einer Zahl auf beiden Seiten der Gleichung
- die Multiplikation mit einer Zahl auf beiden Seiten der Gleichung
- die Division durch eine von Null verschiedene Zahl auf beiden Seiten der Gleichung
- die Anwendung einer invertierbaren Funktion auf beiden Seiten der Gleichung

Beispiel

Betrachten wir beispielsweise unter Vorgriff auf Kapitel 4.3 die Aussage

$$2 \exp(x) - 2 = 0. \quad (1.3)$$

Dann gelten

$$\begin{aligned}
 & 2 \exp(x) - 2 = 0 \\
 \Leftrightarrow & 2 \exp(x) - 2 + 2 = +2 \\
 \Leftrightarrow & 2 \exp(x) = 2 \\
 \Leftrightarrow & \frac{1}{2} \cdot 2 \exp(x) = \frac{1}{2} \cdot 2 \\
 \Leftrightarrow & \exp(x) = 1 \\
 \Leftrightarrow & \ln(\exp(x)) = \ln(1) \\
 \Leftrightarrow & x = 0.
 \end{aligned} \tag{1.4}$$

Zusammengefasst gilt also

$$2 \exp(x) - 2 \Leftrightarrow x = 0. \tag{1.5}$$

Bei Ungleichungen sind zulässige Äquivalenzumformungen insbesondere

- die Addition einer Zahl auf beiden Seiten der Ungleichung
- die Subtraktion einer Zahl auf beiden Seiten der Ungleichung
- die Multiplikation mit einer Zahl auf beiden Seiten der Gleichung, wobei die Multiplikation mit einer negativen Zahl das Ungleichungszeichen umkehrt,
- die Division durch eine von Null verschiedenen positiven Zahl auf beiden Seiten der Gleichung, wobei die Division mit einer negativen Zahl das Ungleichungszeichen umkehrt,
- die Anwendung einer invertierbaren monotonen Funktion auf beiden Seiten der Ungleichung

Beispiel

Betrachten wir beispielsweise die Aussage

$$-5x - 2 \geq 8. \tag{1.6}$$

Dann gelten

$$\begin{aligned}
 & -5x - 2 \geq 8 \\
 \Leftrightarrow & -5x - 2 + 2 \geq 8 + 2 \\
 \Leftrightarrow & -5x \geq 10 \\
 \Leftrightarrow & -\frac{1}{5} \cdot 5x \geq \frac{1}{5} \cdot 10 \\
 \Leftrightarrow & -x \geq 2 \\
 \Leftrightarrow & x \leq 2.
 \end{aligned} \tag{1.7}$$

Zusammengefasst gilt also

$$-5x - 2 \geq 8 \Leftrightarrow x \leq 2. \tag{1.8}$$

1.5. Beweistechniken

In diesem Abschnitt wollen wir mit den Begriffen der *direkten* und *indirekten Beweise* sowie des *Beweises durch Widerspruch* drei fundamentale Beweistechniken skizzieren. Dabei

wird vor allem die erste im Folgenden immer wieder zur Begründung von Theoremen herangezogen werden.

Es gilt dabei

- *Direkte Beweise* nutzen Äquivalenzumformungen, um $A \Rightarrow B$ zu zeigen.
- *Indirekte Beweise* nutzen die logische Äquivalenz von $A \Rightarrow B$ und $(\neg B) \Rightarrow (\neg A)$.
- *Beweise durch Widerspruch* zeigen, dass $(\neg B) \wedge A$ falsch ist.

Damit ausgestattet wollen wir nun folgendes Theorem mithilfe eines direkten Beweises, eines indirekten Beweises und eines Beweises durch Widerspruch beweisen (vgl. Arens et al. (2018)).

Theorem 1.2 (Quadrate positiver Zahlen). *Es seien a und b zwei positive Zahlen. Dann gilt $a^2 < b^2 \Rightarrow a < b$.*

◦

Beweis. Wir geben zunächst einen *direkten Beweis*. Dazu sei $a^2 < b^2$ die Aussage A und $a < b$ die Aussage B . Dann gilt

$$a^2 < b^2 \Leftrightarrow 0 < b^2 - a^2 \Leftrightarrow 0 < (b + a)(b - a) \Leftrightarrow 0 < (b - a) \Leftrightarrow a < b. \quad (1.9)$$

Wir geben nun einen *indirekten Beweis*. Es sei $a^2 \geq b^2$ die Aussage $\neg A$. Weiterhin sei $a \geq b$ die Aussage $\neg B$. Dann gilt

$$a \geq b \Leftrightarrow a^2 \geq ab \wedge ab \geq b^2 \Leftrightarrow a^2 \geq b^2. \quad (1.10)$$

Schließlich geben wir einen *Beweis durch Widerspruch*. Wir zeigen, dazu, dass die Annahme $(\neg B) \wedge A$ auf eine falsche Aussage führt. Es gilt

$$a \geq b \wedge a^2 < b^2 \Leftrightarrow a^2 \geq ab \wedge a^2 < b^2 \Leftrightarrow ab \leq a^2 < b^2. \quad (1.11)$$

Weiterhin gilt

$$a \geq b \wedge a^2 < b^2 \Leftrightarrow ab \geq b^2 \wedge a^2 < b^2 \Leftrightarrow a^2 < b^2 \leq ab. \quad (1.12)$$

Insgesamt gilt dann also die falsche Aussage

$$ab \leq a^2 < b^2 \leq ab \Leftrightarrow ab < ab. \quad (1.13)$$

□

1.6. Selbstkontrollfragen

1. Erläutern Sie die Besonderheiten der mathematischen Sprache.
2. Was sind wesentliche Tätigkeiten zum Erlernen einer Sprache?
3. Erläutern Sie den Begriff der Definition.
4. Erläutern Sie den Begriff des Theorems.
5. Erläutern Sie den Begriff des Beweises.
6. Geben Sie die Definition einer mathematischen Aussage wieder.
7. Geben Sie die Definition der Negation einer mathematischen Aussage wieder.
8. Geben Sie die Definition der Konjunktion zweier mathematischer Aussagen wieder.
9. Geben Sie die Definition der Disjunktion zweier mathematischer Aussagen wieder.
10. Geben Sie die Definition der Implikation wieder.
11. Geben Sie die Definition der Äquivalenz wieder.
12. Geben Sie die Definition der logischen Äquivalenz wieder.
13. Erläutern Sie die Begriffe des direkten Beweises, des indirekten Beweises und des Beweises durch Widerspruch.

2. Mengen

2.1. Grundlegende Definitionen

Mengen fassen mathematische Objekte wie beispielsweise Zahlen zusammen und bilden die Grundlage der modernen Mathematik. Wir beginnen mit folgender Definition.

Definition 2.1 (Mengen). Nach Cantor (1895) ist eine *Menge* definiert als “eine Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denken (welche die Elemente der Menge genannt werden) zu einem Ganzen”. Wir schreiben

$$m \in M \text{ bzw. } m \notin M \quad (2.1)$$

um auszudrücken, dass m ein Element bzw. kein Element von M ist.

•

Zur Definition von Mengen gibt es mindestens folgende Möglichkeiten:

- Auflisten der Elemente in geschweiften Klammern, z.B. $M := \{1, 2, 3\}$.
- Angabe der Eigenschaften der Elemente, z.B. $M := \{x \in \mathbb{N} | x < 4\}$.
- Gleichsetzen mit einer anderen eindeutig definierten Menge, z.B. $M := \mathbb{N}_3$.

Die Schreibweise $\{x \in \mathbb{N} | x < 4\}$ wird gelesen als “ $x \in \mathbb{N}$, für die gilt, dass $x < 4$ ist”, wobei die Bedeutung von \mathbb{N} im Folgenden noch zu erläutern sein wird. Es ist wichtig zu erkennen, dass Mengen *ungeordnete* mathematische Objekte sind, dass heißt die Reihenfolge der Auflistung der Elemente einer Menge spielt keine Rolle. Zum Beispiel bezeichnen $\{1, 2, 3\}$, $\{1, 3, 2\}$ und $\{2, 3, 1\}$ dieselbe Menge, nämlich die Menge der ersten drei natürlichen Zahlen.

Grundlegende Beziehungen zwischen mehreren Mengen werden in der nächsten Definition festgelegt.

Definition 2.2 (Teilmengen und Mengengleichheit). M und N seien zwei Mengen.

- Eine Menge M heißt *Teilmenge* einer Menge N , wenn für jedes Element $m \in M$ gilt, dass auch $m \in N$. Ist M eine Teilmenge von N , so schreibt man

$$M \subseteq N \quad (2.2)$$

und nennt M *Untermenge* von N und N *Obermenge* von M .

- Eine Menge M heißt *echte Teilmenge* einer Menge N , wenn für jedes Element $m \in M$ gilt, dass auch $m \in N$, es aber zumindest ein Element $n \in N$ gibt, für das gilt $n \notin M$. Ist M eine echte Teilmenge von N , so schreibt man

$$M \subset N. \quad (2.3)$$

- Zwei Mengen M und N heißen *gleich*, wenn für jedes Element $m \in M$ gilt, dass auch $m \in N$, und wenn für jedes Element $n \in N$ gilt, dass auch $n \in M$. Sind die Mengen M und N gleich, so schreibt man

$$M = N. \quad (2.4)$$

•

Beispiel

Betrachten wir zum Beispiel die Mengen $M := \{1\}$, $N := \{1, 2\}$, und $O := \{1, 2\}$. Dann gilt mit obigen Definitionen, dass $M \subset N$, weil $1 \in M$ und $1 \in N$, aber $2 \in N$ und $2 \notin M$. Weiterhin gilt, dass $N \subseteq O$, weil $1 \in N$ und $1 \in O$ sowie $2 \in N$ und $2 \in O$ und es kein Element von O gibt, welches nicht in N ist. Ebenso gilt $O \subseteq N$, weil $1 \in O$ und $1 \in N$ sowie $2 \in O$ und $2 \in N$ und es kein Element von N gibt, welches nicht in O ist. Schließlich gilt sogar $N = O$, weil für jedes Element $n \in N$ gilt, dass auch $n \in O$, und gleichzeitig für jedes Element $o \in O$ gilt, dass auch $o \in N$. Wir stellen diese Zusammenhänge schematisch mit Hilfe von *Venn-Euler-Diagrammen* in Abbildung 2.1 dar.

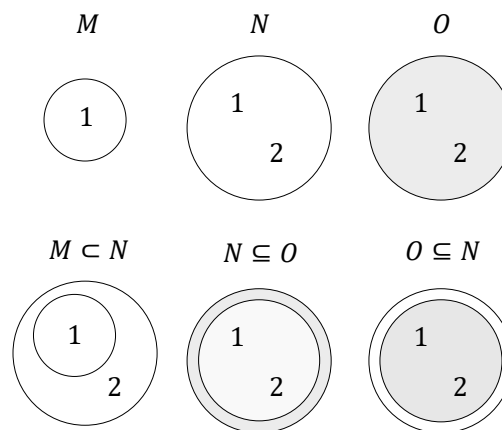


Abbildung 2.1. Venn-Euler Diagramme der Teilmengeneigenschaften von Mengen $M := \{1\}$, $N := \{1, 2\}$ und $O := \{1, 2\}$.

Eine wichtige Eigenschaft einer Menge ist die Anzahl der in ihr enthaltenen Elemente. Diese wird als *Kardinalität* der Menge bezeichnet.

Definition 2.3 (Kardinalität). Die Anzahl der Elemente einer Menge M heißt *Kardinalität* und wird mit $|M|$ bezeichnet.

•

Eine besondere Menge ist die Menge ohne Elemente.

Definition 2.4. Eine Menge mit Kardinalität Null heißt *leere Menge* und wird mit \emptyset bezeichnet.

•

Als Beispiele seien $M := \{1, 2, 3\}$, $N = \{a, b, c, d\}$ und $O := \{\}$. Dann gelten $|M| = 3$, $N = 4$ und $|O| = 0$.

Zu jeder Menge kann man die Menge aller Teilmengen dieser Menge betrachten. Dies führt auf den wichtigen Begriff der *Potenzmenge*.

Definition 2.5 (Potenzmenge). Die Menge aller Teilmengen einer Menge M heißt *Potenzmenge von M* und wird mit $\mathcal{P}(M)$ bezeichnet.

•

Man beachte, dass die leere Untermenge von M und M selbst auch immer Elemente von $\mathcal{P}(M)$ sind. Wir betrachten zunächst vier Beispiele zum Begriff der Potenzmenge.

- $M_0 := \emptyset$ sei die leere Menge. Dann gilt

$$\mathcal{P}(M_0) = \emptyset. \quad (2.5)$$

- M_1 sei die einelementige Menge $M_1 := \{a\}$. Dann gilt

$$\mathcal{P}(M_1) = \{\emptyset, \{a\}\}. \quad (2.6)$$

- Es sei $M_2 := \{a, b\}$. Dann hat M_2 sowohl ein- als auch zweielementige Teilmengen und es gilt

$$\mathcal{P}(M_2) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}. \quad (2.7)$$

- Schließlich sei $M_3 := \{a, b, c\}$. Dann hat M_3 unter anderem ein-, zwei-, und dreielementige Teilmengen und es gilt

$$\mathcal{P}(M_3) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}. \quad (2.8)$$

Theorem 2.1 (Kardinalität der Potenzmenge). Gegeben sei eine Menge M mit Kardinalität $|M| = n$ und $\mathcal{P}(M)$ sei ihre Potenzmenge. Dann gilt $|\mathcal{P}(M)| = 2^n$

◦

Beweis. Um die Aussage des Theorems zu beweisen assoziieren wir jedes Element P der Potenzmenge von M eindeutig mit einer binären Folge der Länge n , wobei der Eintrag an der i ten Stelle repräsentiert, ob das i te Element von M ein Element von P ist oder nicht. Seien beispielsweise $M := \{m_1, m_2, m_3\}$ und $P := \{m_2, m_3\}$. Dann entspricht P die binäre Folge 011. Der leeren Menge $P := \emptyset$ entspricht die binäre Folge 000 und der Ausgangsmenge $P = M$ entspricht die binäre Folge 111. Es ergibt sich also die Frage, wieviele eindeutige binäre Folgen der Länge n es gibt. Da es für jedes Element der Folge zwei mögliche Zustände gibt, ergeben sich n Faktoren $2 \cdot 2 \cdots 2$, also 2^n . \square

In den obigen Beispielen haben wir die Fälle

- $|M_0| = 0 \Rightarrow |\mathcal{P}(M_0)| = 2^0 = 1,$
- $|M_1| = 1 \Rightarrow |\mathcal{P}(M_1)| = 2^1 = 2,$
- $|M_2| = 2 \Rightarrow |\mathcal{P}(M_2)| = 2^2 = 4,$
- $|M_3| = 3 \Rightarrow |\mathcal{P}(M_3)| = 2^3 = 8,$

wovon man sich durch Nachzählen der Elemente der entsprechenden Potenzmengen oben überzeugt.

2.2. Verknüpfungen

Zwei Mengen können auf unterschiedliche Weise miteinander verknüpft werden. Das Ergebnis einer solchen Verknüpfung ist eine weitere Menge. Wir bezeichnen die Verknüpfung zweier Mengen als *Mengenoperation* und geben folgende Definitionen.

Definition 2.6 (Mengenoperationen). M und N seien zwei Mengen.

- Die *Vereinigung von M und N* ist definiert als die Menge

$$M \cup N := \{x | x \in M \vee x \in N\}, \quad (2.9)$$

wobei \vee gemäß Definition 1.4 als *nicht-exklusives oder*, also als und/oder, zu verstehen ist.

- Der *Durchschnitt von M und N* ist definiert als die Menge

$$M \cap N := \{x | x \in M \wedge x \in N\}. \quad (2.10)$$

Wenn für M und N gilt, dass $M \cap N = \emptyset$, dann heißen M und N *disjunkt*.

- Die *Differenz von M und N* ist definiert als die Menge

$$M \setminus N := \{x | x \in M \wedge x \notin N\}. \quad (2.11)$$

Die Differenz M und N heißt, insbesondere bei $M \subseteq N$, auch das *Komplement von N bezüglich M* und wird mit N^c bezeichnet.

- Die *symmetrische Differenz von M und N* ist definiert als die Menge

$$M \Delta N := \{x | (x \in M \vee x \in N) \wedge x \notin M \cap N\}, \quad (2.12)$$

Die symmetrische Differenz kann also als *exklusives oder* verstanden werden.

•

Beispiel

Als Beispiel betrachten wir die Mengen $M := \{1, 2, 3\}$ und $N := \{2, 3, 4, 5\}$. Dann gelten

- $M \cup N = \{1, 2, 3, 4, 5\}$, weil $1 \in M$, $2 \in M$, $3 \in M$, $4 \in N$ und $5 \in N$.
- $M \cap N = \{2, 3\}$, weil nur für 2 und 3 gilt, dass $2 \in M$, $3 \in M$ und auch $2 \in N$, $3 \in N$. Für 1 gilt lediglich, dass $1 \in M$ und für 4 und 5 gelten lediglich, dass $4 \in N$ und $5 \in N$.
- $M \setminus N = \{1\}$, weil $1 \in M$, aber $1 \notin N$ und $2 \in M$, aber auch $2 \in N$.
- $N \setminus M = \{4, 5\}$, weil $2 \in N$ und $3 \in N$, aber auch $2 \in M$ und $3 \in M$. Dies zeigt insbesondere, dass die Differenz von M und N *nicht* symmetrisch ist, also dass *nicht* zwangsläufig gilt, dass $M \setminus N$ gleich $N \setminus M$ ist.
- $M \Delta N = \{1, 4, 5\}$, weil $1 \in M$, aber $1 \notin \{2, 3\}$, $2 \in M$, aber $2 \in \{2, 3\}$, $3 \in M$, aber $3 \in \{2, 3\}$, $4 \in N$, aber $4 \notin \{2, 3\}$ und $5 \in N$, aber $5 \notin \{2, 3\}$.

Abbildung 2.2 visualisiert die in diesem Beispiel betrachteten Mengenoperationen.

Schließlich wollen wir noch den Begriff der Partition einer Menge einführen.

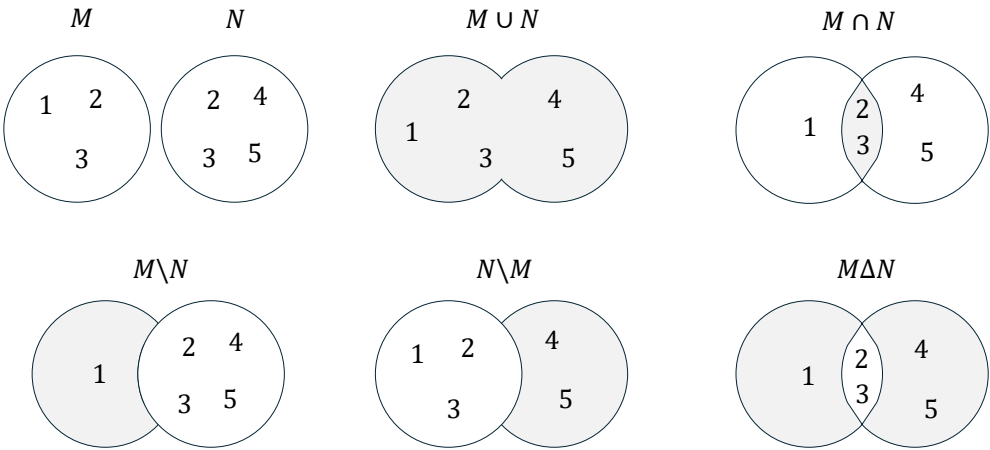


Abbildung 2.2. Venn-Euler Diagramme der im Beispiel für $M := \{1, 2, 3\}$ und $N := \{2, 3, 4, 5\}$ betrachteten Mengenoperationen. Die grau hinterlegten Flächen entsprechen jeweils den sich ergebenden Mengen.

Definition 2.7 (Partition). M sei eine Menge und $P := \{N_i\}$ sei eine Menge von Mengen N_i mit $i = 1, \dots, n$, so dass gilt

$$M = \cup_{i=1}^n N_i \wedge N_i \cap N_j = \emptyset \text{ f\"ur } i = 1, \dots, n, j = 1, \dots, n, i \neq j. \tag{2.13}$$

Dann heit P eine *Partition* von M .

•

Intuitiv entspricht die Partition einer Menge also dem Aufteilen der Menge in disjunkte Teilmengen. Partitionen sind generell nicht eindeutig, d.h. es gibt meist verschiedene Mglichkeiten eine gegebene Menge zu partitionieren.

Als Beispiel betrachten wir die Menge $M := \{1, 2, 3, 4, 5, 6\}$. Dann sind $P_1 := \{\{1\}, \{2, 3, 4, 5, 6\}\}$, $P_2 := \{\{1, 2, 3\}, \{4, 5, 6\}\}$ und $P_3 := \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ drei mgliche Partitionen von M . [Abbildung 2.3](#) visualisiert die in diesem Beispiel betrachteten Partitionen.

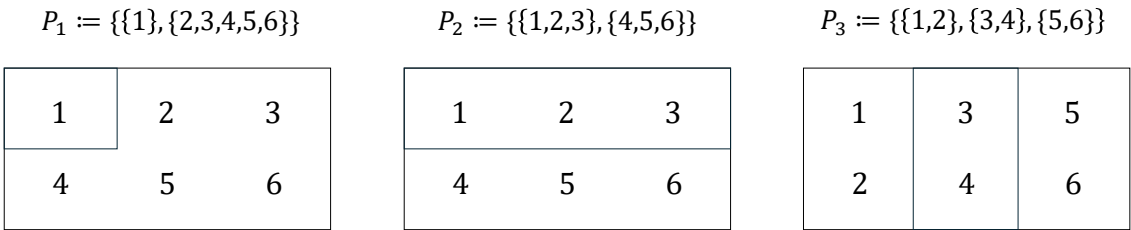


Abbildung 2.3. Diagramme der im Beispiel fr $M := \{1, 2, 3, 4, 5, 6\}$ betrachteten Partitionen.

2.3. Spezielle Mengen

Zahlenmengen

In der Naturwissenschaft versucht man, in der Vorstellung intuitiv als diskret oder kontinuierlich identifizierte Phänomene der Welt mit Zahlen zu beschreiben. Je nach Art des Phänomens bieten sich dazu verschiedene Zahlenmengen an. Die Mathematik stellt unter anderem die in folgender Definition gegebenen Zahlen bereit.

Definition 2.8 (Zahlenmengen). Es bezeichnen

- $\mathbb{N} := \{1, 2, 3, \dots\}$ die *natürlichen Zahlen*,
- $\mathbb{N}_n := \{1, 2, 3, \dots, n\}$ die *natürlichen Zahlen der Ordnung n* ,
- $\mathbb{N}^0 := \mathbb{N} \cup \{0\}$ die *natürlichen Zahlen* und Null,
- $\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ die *ganzen Zahlen*,
- $\mathbb{Q} := \{\frac{p}{q} | p, q \in \mathbb{Z}, q \neq 0\}$ die *rationalen Zahlen*,
- \mathbb{R} die *reellen Zahlen*, und
- $\mathbb{C} := \{a + ib | a, b \in \mathbb{R}, i := \sqrt{-1}\}$ die *komplexen Zahlen*.

•

Die natürlichen und ganzen Zahlen eignen sich insbesondere zum Quantifizieren diskreter Phänomene. Die rationalen und insbesondere die reellen Zahlen eignen sich zum Quantifizieren kontinuierlicher Phänomene. \mathbb{R} umfasst dabei die rationalen Zahlen und die sogenannten *irrationalen Zahlen* $\mathbb{R} \setminus \mathbb{Q}$. Rationale Zahlen sind Zahlen, die sich, durch Brüche ganzer Zahlen ausdrücken lassen. Dies sind alle ganzen Zahlen sowie die negativen und positiven Dezimalzahlen wie beispielsweise $-\frac{9}{10} = -0.9$, $\frac{1}{3} = 1.\overline{33}$ und $\frac{196}{100} = 1.96$. Irrationale Zahlen sind Zahlen, die sich nicht als rationale Zahlen ausdrücken lassen. Beispiele für irrationale Zahlen sind die *Eulersche Zahl* $e \approx 2.71$, die *Kreiszahl* $\pi \approx 3.14$ und die Quadratwurzel von 2, $\sqrt{2} \approx 1.41$.

Die reellen Zahlen enthalten als Teilmengen die natürlichen, ganzen und die rationalen Zahlen. Es gibt also sehr viele reelle Zahlen. Tatsächlich hat Cantor (1892) bewiesen, dass es mehr reelle Zahlen als natürliche Zahlen gibt, obwohl es sowohl unendlich viele reelle Zahlen als auch unendlich viele natürliche Zahlen gibt. Diese Eigenschaft der reellen Zahlen bezeichnet man als die *Überabzählbarkeit* der reellen Zahlen. Insbesondere gilt also

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}. \quad (2.14)$$

Zwischen zwei reellen Zahlen gibt es unendlich viele weitere reelle Zahlen. Positiv-Unendlich ∞ und Negativ-Unendlich $-\infty$ sind keine Zahlen, mit denen in der Standardmathematik gerechnet werden könnte. Sie gehören auch nicht zu den in obiger Definition gegebenen Zahlenmengen, es gelten also $\infty \notin \mathbb{R}$ und $-\infty \notin \mathbb{R}$. Komplexe Zahlen eignen sich zur Beschreibung zweidimensionaler kontinuierlicher Phänomene. Dabei werden die Werte der ersten Dimension im reellen Teil a und die Werte der zweiten Dimension im komplexen Teil b einer komplexen Zahl repräsentiert. Komplexe Zahlen kommen insbesondere bei der Modellierung physikalischer Phänomene und im Bereich der Fourieranalyse zum Einsatz.

Wichtige Teilmengen der reellen Zahlen sind die sogenannten *Intervalle*. Wir geben folgende Definitionen.

Intervalle

Definition 2.9. Zusammenhängende Teilmengen der reellen Zahlen heißen *Intervalle*. Für $a, b \in \mathbb{R}$ unterscheidet man

- das *abgeschlossene Intervall*

$$[a, b] := \{x \in \mathbb{R} | a \leq x \leq b\}, \quad (2.15)$$

- das *offene Intervall*

$$]a, b[:= \{x \in \mathbb{R} | a < x < b\}, \quad (2.16)$$

- und die *halboffenen Intervalle*

$$]a, b] := \{x \in \mathbb{R} | a < x \leq b\} \text{ und } [a, b[:= \{x \in \mathbb{R} | a \leq x < b\}. \quad (2.17)$$

•

Exemplarisch stellen wir in Abbildung 2.4 die Intervalle $[1, 2]$, $]1, 2]$, $[1, 2[$ und $]1, 2[$ graphisch dar. Dabei stellt man sich die reellen Zahlen als kontinuierlicher Zahlenstrahl vor und muss jeweils beachten, ob die linken bzw. rechten Endpunkte Teil des Intervalls sind oder nicht. Wie oben erwähnt sind Positiv-Unendlich (∞) und Negativ-Unendlich ($-\infty$) keine Elemente von \mathbb{R} . Es gilt also immer $] - \infty, b]$ oder $] - \infty, b[$ bzw. $]a, \infty[$ oder $[a, \infty[$, sowie $\mathbb{R} =] - \infty, \infty[$.

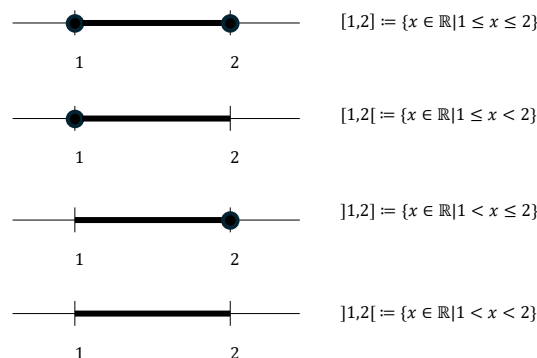


Abbildung 2.4. Darstellung von Intervallen am Zahlenstrahl. Der schwarze Punkt signalisiert dabei, dass die entsprechende Zahl Teil des Intervalls ist.

Kartesische Produkte

Oft möchte man mehrere unabhängige Eigenschaften eines Phänomens gleichzeitig quantitativ beschreiben. Zu diesem Zweck können die oben definierten eindimensionalen Zahlenmenge durch Bildung *Kartesischer Produkte* auf mehrdimensionale Zahlenmengen erweitert werden. Die Elemente Kartesischer Produkte nennt man *geordnete Tupel* oder *Vektoren*.

Definition 2.10 (Kartesische Produkte). M und N seien zwei Mengen. Dann ist das *Kartesische Produkt der Mengen M und N* die Menge aller geordneten Tupel (m, n) mit $m \in M$ und $n \in N$, formal

$$M \times N := \{(m, n) | m \in M, n \in N\}. \quad (2.18)$$

Das Kartesische Produkt einer Menge M mit sich selbst wird bezeichnet mit

$$M^2 := M \times M. \quad (2.19)$$

Seien weiterhin M_1, M_2, \dots, M_n Mengen. Dann ist das *Kartesische Produkt der Mengen* M_1, \dots, M_n die Menge aller geordneten n -Tupel (m_1, \dots, m_n) mit $m_i \in M_i$ für $i = 1, \dots, n$, formal

$$\prod_{i=1}^n M_i := M_1 \times \dots \times M_n := \{(m_1, \dots, m_n) | m_i \in M_i \text{ für } i = 1, \dots, n\}. \quad (2.20)$$

Das n -fache Kartesische Produkt einer Menge M mit sich selbst wird bezeichnet mit

$$M^n := \prod_{i=1}^n M := \{(m_1, \dots, m_n) | m_i \in M\}. \quad (2.21)$$

•

Im Gegensatz zu Mengen sind die in Definition 2.10 eingeführten Tupel *geordnet*. Das heißt, für Mengen gilt beispielsweise $\{1, 2\} = \{2, 1\}$, aber für Tupel gilt $(1, 2) \neq (2, 1)$.

Beispiel

Es seien $M := \{1, 2\}$ und $N := \{1, 2, 3\}$. Dann ist das Kartesische Produkt $M \times N$ gegeben durch

$$M \times N := \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\} \quad (2.22)$$

und das Kartesische Produkt $N \times M$ ist gegeben durch

$$N \times M := \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2)\} \quad (2.23)$$

Das Kartesische Produkt ist also im Allgemeinen nicht kommutativ, es gilt also nicht notwendigerweise, dass $M \times N = N \times M$. Man mag sich die in diesem Beispiel konstruierten Mengen $M \times N \neq N \times M$ beispielsweise anhand untenstehender Tabellen Tabelle 2.1 und Tabelle 2.2 konstruieren.

Tabelle 2.1. Kartesisches Produkt $M \times N$

(m, n)	$n = 1$	$n = 2$	$n = 3$
$m = 1$	(1, 1)	(1, 2)	(1, 3)
$m = 2$	(2, 1)	(2, 2)	(2, 3)

Tabelle 2.2. Kartesisches Produkt $N \times M$

(n, m)	$m = 1$	$m = 2$
$n = 1$	(1, 1)	(1, 2)
$n = 2$	(2, 1)	(2, 2)
$n = 3$	(3, 1)	(3, 2)

\mathbb{R} hoch n

Wie oben beschrieben eignen sich insbesondere die reellen Zahlen zur Beschreibung kontinuierlicher Phänomene. Zur simultanen Beschreibung mehrere Aspekte eines kontinuierlichen Phänomens bietet sich entsprechend die *Menge der reellen Tupel n -ter Ordnung*, kurz \mathbb{R} hoch n an.

Definition 2.11 (Menge der reellen Tupel n -ter Ordnung). Das n -fache Kartesische Produkt der reellen Zahlen mit sich selbst wird bezeichnet mit

$$\mathbb{R}^n := \prod_{i=1}^n \mathbb{R} := \{x := (x_1, \dots, x_n) | x_i \in \mathbb{R}\} \quad (2.24)$$

und wird “ \mathbb{R} hoch n ” gesprochen. Wir schreiben die Elemente von \mathbb{R}^n als Spalten

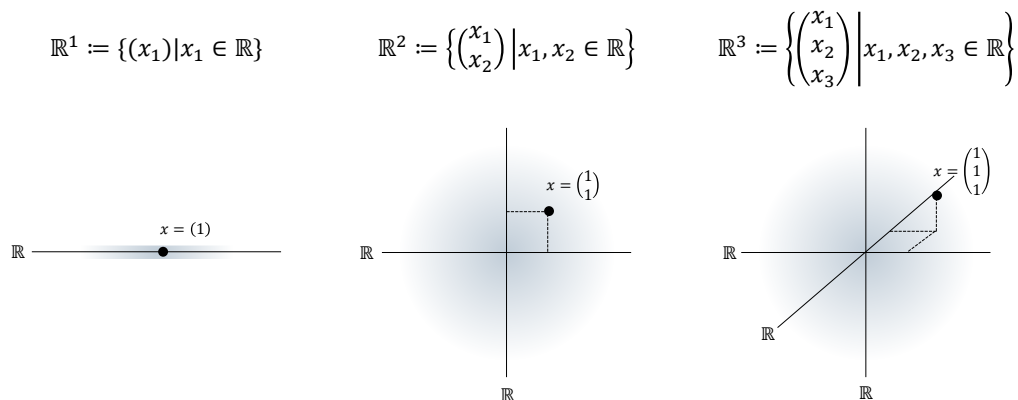
$$x := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (2.25)$$

und nennen sie n -dimensionale Vektoren. Zu Abgrenzung nennen wir die Elemente von $\mathbb{R}^1 = \mathbb{R}$ auch *Skalare*.

•

Beispiele

Vertraute Beispiele von \mathbb{R}^n sind \mathbb{R}^1 als Menge der reellen Zahlen, \mathbb{R}^2 als Menge der reellen Tupel im Modell der zweidimensionalen Ebene und \mathbb{R}^3 als Menge der reellen Tripel im Modell des dreidimensionalen Raumes wie in ?@fig-rhochn visualisiert.



Ein Beispiel für ein $x \in \mathbb{R}^4$ ist

$$x = \begin{pmatrix} 0.16 \\ 1.76 \\ 0.23 \\ 7.11 \end{pmatrix}. \quad (2.26)$$

2.4. Selbstkontrollfragen

1. Geben Sie die Definition einer Menge nach Cantor (1895) wieder.
2. Nennen Sie drei Möglichkeiten zur Definition einer Menge.
3. Erläutern Sie die Ausdrücke $m \in M$, $m \notin N$, $M \subseteq N$, $M \subset N$ für zwei Mengen M und N .
4. Geben Sie die Definition der Kardinalität einer Menge wieder.
5. Geben Sie die Definition der Potenzmenge einer Menge wieder.
6. Es sei $M := \{1, 2\}$. Bestimmen Sie $\mathcal{P}(M)$.
7. Es seien $M := \{1, 2\}$, $N := \{1, 4, 5\}$. Bestimmen Sie $M \cup N$, $M \cap N$, $M \setminus N$, $M \Delta N$.
8. Erläutern Sie die Symbole \mathbb{N} , \mathbb{N}_n , und \mathbb{N}^0 .
9. Erläutern Sie die Unterschiede zwischen \mathbb{N} und \mathbb{Z} und zwischen \mathbb{R} und \mathbb{Q} .
10. Geben Sie die Definition abgeschlossener, offener, und halboffener Intervalle wieder.
11. Es seien M und N Mengen. Erläutern Sie die Notation $M \times N$.
12. Geben Sie die Definition von \mathbb{R}^n wieder.

3. Summen, Produkte, Potenzen

Diese Einheit führt Schreibweisen für die Grundrechenarten ein.

3.1. Summen

Definition 3.1 (Summenzeichen). Es bezeichnet

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n. \quad (3.1)$$

Dabei stehen

- Σ für das griechische *Sigma*, mnemonisch für *Summe*,
- das Subskript $i = 1$ für den *Laufindex* und den *Startindex*,
- das Superskript n für den *Endindex* und
- x_1, x_2, \dots, x_n für die *Summanden*.

•

Für eine sinnvolle Benutzung des Summenzeichens ist es essentiell, mithilfe des Subskripts und des Superskripts den Anfang und das Ende der Summation festzulegen. Die genaue Bezeichnung des Laufindex ist dagegen für den Wert der Summe irrelevant, es gilt

$$\sum_{i=1}^n x_i = \sum_{j=1}^n x_j. \quad (3.2)$$

Manchmal wird der Laufindex auch als Element einer *Indexmenge* angegeben. Ist z.B. die Indexmenge $I := \{1, 5, 7\}$ definiert, so ist

$$\sum_{i \in I} x_i := x_1 + x_5 + x_7. \quad (3.3)$$

Im Folgenden wollen wir kurz einige Beispiele für die Benutzung des Summenzeichens betrachten.

- *Summation vordefinierter Summanden*. Es seien $x_1 := 2$, $x_2 := 10$, $x_3 := -4$. Dann gilt

$$\sum_{i=1}^3 x_i = x_1 + x_2 + x_3 = 2 + 10 - 4 = 8. \quad (3.4)$$

- *Summation gewichteter vordefinierter Summanden.* Es seien wiederum $x_1 := 2$, $x_2 := 10$, $x_3 := -4$. Weiterhin seien die *Wichtungskoeffizienten* $a_1 := \frac{1}{2}$, $a_2 := \frac{1}{5}$, $a_3 := 2$ definiert. Dann gilt

$$\sum_{i=1}^3 a_i x_i = a_1 x_1 + a_2 x_2 + a_3 x_3 = \frac{1}{2} \cdot 2 + \frac{1}{5} \cdot 10 + 2 \cdot (-4) = 1 + 2 - 8 = -5. \quad (3.5)$$

Ausdrücke der Form $\sum_{i=1}^n a_i x_i$ werden auch als *Linearkombinationen* der x_1, \dots, x_n mit den *Koeffizienten* oder *Wichtungsparametern* a_1, \dots, a_n bezeichnet.

- *Summation natürlicher Zahlen.* Es gilt

$$\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5 = 15. \quad (3.6)$$

- *Summation gerader natürlicher Zahlen.* Es gilt

$$\sum_{i=1}^5 2i = 2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 + 2 \cdot 4 + 2 \cdot 5 = 2 + 4 + 6 + 8 + 10 = 30. \quad (3.7)$$

- *Summation ungerader natürlicher Zahlen.* Es gilt

$$\sum_{i=1}^5 (2i-1) = 2 \cdot 1 - 1 + 2 \cdot 2 - 1 + 2 \cdot 3 - 1 + 2 \cdot 4 - 1 + 2 \cdot 5 - 1 = 1 + 3 + 5 + 7 + 9 = 25. \quad (3.8)$$

Der Umgang mit dem Summenzeichen kann oft durch die Anwendung folgender Rechenregeln vereinfacht werden.

Theorem 3.1 (Rechenregeln für Summen).

(1) *Summen gleicher Summanden*

$$\sum_{i=1}^n x = nx \quad (3.9)$$

(2) *Assoziativität bei Summen gleicher Länge*

$$\sum_{i=1}^n x_i + \sum_{i=1}^n y_i = \sum_{i=1}^n (x_i + y_i) \quad (3.10)$$

(3) *Distributivität bei Multiplikation mit einer Konstante*

$$\sum_{i=1}^n a x_i = a \sum_{i=1}^n x_i \quad (3.11)$$

(4) *Aufspalten von Summen mit $1 < m < n$*

$$\sum_{i=1}^n x_i = \sum_{i=1}^m x_i + \sum_{i=m+1}^n x_i \quad (3.12)$$

(5) *Umindizierung*

$$\sum_{i=0}^n x_i = \sum_{j=m}^{n+m} x_{j-m} \quad (3.13)$$

◦

Beweis. Man überzeugt sich von diesen Rechenregeln durch Ausschreiben der Summen und Anwenden der Rechenregeln von Addition und Multiplikation. Wir zeigen hier exemplarisch die Assoziativität bei Summen gleicher Länge und die Distributivität bei Multiplikation mit einer Konstante. Hinsichtlich ersterer haben wir

$$\begin{aligned}\sum_{i=1}^n x_i + \sum_{i=1}^n y_i &= x_1 + x_2 + \cdots + x_n + y_1 + y_2 + \cdots + y_n \\ &= x_1 + y_1 + x_2 + y_2 + \cdots + x_n + y_n \\ &= \sum_{i=1}^n (x_i + y_i).\end{aligned}\tag{3.14}$$

Hinsichtlich letzterer gilt

$$\begin{aligned}\sum_{i=1}^n ax_i &= ax_1 + ax_2 + \cdots + ax_n \\ &= a(x_1 + x_2 + \cdots + x_n) \\ &= a \sum_{i=1}^n x_i.\end{aligned}\tag{3.15}$$

□

Beispiele

Als erstes Beispiel für die Anwendung der in Theorem 3.1 festgehaltenen Rechenregeln betrachten wir die Auswertung eines *Mittelwertes* (manchmal auch *Durchschnitt* genannt). Dazu seien x_1, x_2, \dots, x_n reelle Zahlen. Der Mittelwert dieser Zahlen entspricht der Summe von x_1, x_2, \dots, x_n geteilt durch die Anzahl der Zahlen n . Dabei ist es nach Aussage (3) von Theorem 3.1 irrelevant, ob zunächst die Zahlen aufaddiert werden und dann die resultierende Summe durch n geteilt wird, oder die Zahlen jeweils einzeln durch n geteilt werden und die entsprechenden Ergebnisse dann aufaddiert werden. Genauer gilt durch Anwendung Theorem 3.1 (3) mit $a = 1/n$, dass

$$\frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n \frac{x_i}{n}.\tag{3.16}$$

So ist zum Beispiel der Mittelwert von $x_1 := 1, x_2 := 4, x_3 := 2, x_4 := 1$ gegeben durch

$$\frac{1}{4} \sum_{i=1}^4 x_i = \frac{1}{4}(1 + 4 + 2 + 1) = \frac{8}{4} = 2 = \frac{8}{4} = \frac{1}{4} + \frac{4}{4} + \frac{2}{4} + \frac{1}{4} = \sum_{i=1}^4 \frac{x_i}{4}.\tag{3.17}$$

Als zweites Beispiel betrachten wir die in Theorem 3.1 (5) festgehaltene Umindizierungsregel. Dazu seien $n := 3$ und $m := 2$, sowie $x_0 := 2, x_1 := 3, x_2 := 5$ und $x_3 := 10$. Dann gilt offenbar

$$\begin{aligned}\sum_{i=0}^3 x_i &= x_0 + x_1 + x_2 + x_3 \\ &= 2 + 3 + 5 + 10 \\ &= 20.\end{aligned}\tag{3.18}$$

Ebenso gilt aber auch

$$\begin{aligned}
 \sum_{j=m}^{n+m} x_{j-m} &= \sum_{j=2}^{3+2} x_{j-2} \\
 &= \sum_{j=2}^5 x_{j-2} \\
 &= x_{2-2} + x_{3-2} + x_{4-2} + x_{5-2} \\
 &= x_0 + x_1 + x_2 + x_3 \\
 &= 2 + 3 + 5 + 10 \\
 &= 20.
 \end{aligned} \tag{3.19}$$

Doppelsummen

Nicht selten trifft man in der Anwendung auf Ausdrücke, die mehrere Summationen hintereinander ausführen. Dabei gelten für jede der Summen die oben gelisteten Definitionen und Rechenregeln. Man macht sich die Bedeutung von Doppelsummen am besten dadurch klar, dass man sie von innen nach außen ausschreibt und dabei beachtet, dass für jede Iteration der inneren Summe der Laufindex der äußeren Summe konstant bleibt. Folgendes Beispiel mag dies verdeutlichen. Es gilt

$$\begin{aligned}
 \sum_{i=1}^2 \sum_{j=1}^3 (x_i + y_j) &= \sum_{i=1}^2 (x_i + y_1 + x_i + y_2 + x_i + y_3) \\
 &= (x_1 + y_1 + x_1 + y_2 + x_1 + y_3) + (x_2 + y_1 + x_2 + y_2 + x_2 + y_3).
 \end{aligned} \tag{3.20}$$

3.2. Produkte

Eine analoge Schreibweise zum Summenzeichen bietet das Produktzeichen für Produkte.

Definition 3.2 (Produktzeichen). Es bezeichnet

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n. \tag{3.21}$$

Dabei stehen

- \prod für das griechische *Pi*, mnemonisch für *Produkt*,
- das Subskript $i = 1$ für den *Laufindex* und den *Startindex*,
- das Superskript n für den *Endindex*,
- x_1, x_2, \dots, x_n für die *Produktterme*

•

Analog zum Summenzeichen gilt, dass das Produktzeichen nur mit Subskript und Superskripten zu Lauf- und Endindex Sinn ergibt. Die genaue Bezeichnung des Laufindex ist wiederum irrelevant, es gilt

$$\prod_{i=1}^n x_i = \prod_{j=1}^n x_j. \tag{3.22}$$

Auch hier wird in seltenen Fällen der Laufindex als Element einer Indexmenge angegeben. Ist z.B. die Indexmenge $J := \mathbb{N}_2^0$ definiert, so ist

$$\prod_{j \in J} x_j := x_0 \cdot x_1 \cdot x_2. \quad (3.23)$$

Ein Beispiel für die Benutzung des Produktzeichens ist etwa die Definition der *Fakultät* einer natürlichen Zahl n durch

$$n! := \prod_{i=1}^n i. \quad (3.24)$$

So ist etwa

$$3! := \prod_{i=1}^3 i = 1 \cdot 2 \cdot 3 = 6. \quad (3.25)$$

Auch für Produkte gibt es eine Reihe von Rechenregeln, die den Umgang mit ihnen oft vereinfachen. Wir listen einige in folgendem Theorem. Dabei machen wir vorgreifend Gebrauch der in Kapitel 3.3 definierten Schreibweise von *Potenzen*.

Theorem 3.2 (Rechenregeln für Produkte).

(1) *Produkte gleicher Faktoren*

$$\prod_{i=1}^n x = x^n \quad (3.26)$$

(2) *Potenzierung von Konstanten*

$$\prod_{i=1}^n ax_i = a^n \prod_{i=1}^n x_i \quad (3.27)$$

(3) *Aufspalten von Produkten mit $1 < m < n$*

$$\prod_{i=1}^n x_i = \prod_{i=1}^m x_i \prod_{j=m+1}^n x_j \quad (3.28)$$

(4) *Produkt von Produkten*

$$\prod_{i=1}^n x_i y_i = \prod_{i=1}^n x_i \prod_{i=1}^n y_i \quad (3.29)$$

◦

3.3. Potenzen

Produkte von Zahlen mit sich selbst können mithilfe der Potenzschreibweise abgekürzt werden.

Definition 3.3 (Potenz). Für $a \in \mathbb{R}$ und $n \in \mathbb{N}^0$ ist die n -te Potenz von a definiert durch

$$a^0 := 1 \text{ und } a^{n+1} := a^n \cdot a. \quad (3.30)$$

Weiterhin ist für $a \in \mathbb{R} \setminus 0$ und $n \in \mathbb{N}^0$ die negative n -te Potenz von a definiert durch

$$a^{-n} := (a^n)^{-1} := \frac{1}{a^n}. \quad (3.31)$$

a wird dabei *Basis* und n wird *Exponent* genannt.

•

Die Art der Definition von a^{n+1} mit Rückbezug auf die Potenz a^n in obiger Definition nennt man *rekursiv*. Die Definition $a^0 := 1$ nennt man dabei den *Rekursionsanfang*; er macht die rekursive Definition von a^{n+1} erst möglich. Die Definition $a^{n+1} := a^n \cdot a$ nennt man auch *Rekursionsschritt*. Folgende Rechenregeln vereinfachen das Rechnen mit Potenzen.

Theorem 3.3 (Rechenregeln für Potenzen). Für $a, b \in \mathbb{R}$ und $n, m \in \mathbb{Z}$ mit $a \neq 0$ bei negativen Exponenten gelten folgende Rechenregeln:

$$a^n a^m = a^{n+m} \quad (3.32)$$

$$(a^n)^m = a^{nm} \quad (3.33)$$

$$(ab)^n = a^n b^n \quad (3.34)$$

◦

Anstelle eines Beweises betrachten wir folgende Beispiele

Beispiel (1)

$$2^2 \cdot 2^3 = (2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) = 2^5 = 2^{2+3}, \quad (3.35)$$

Beispiel (2)

$$(3^2)^3 = (3 \cdot 3)^3 = (3 \cdot 3) \cdot (3 \cdot 3) \cdot (3 \cdot 3) = 3^6 = 3^{2 \cdot 3}, \quad (3.36)$$

Beispiel (3)

$$(2 \cdot 4)^2 = (2 \cdot 4) \cdot (2 \cdot 4) = (2 \cdot 2) \cdot (4 \cdot 4) = 2^2 \cdot 4^2. \quad (3.37)$$

In enger Beziehung zur Potenz steht die Definition der n ten Wurzel:

Definition 3.4 (n -te Wurzel). Für $a \in \mathbb{R}$ und $n \in \mathbb{N}$ ist die n -te Wurzel von a definiert als die reelle Zahl r mit

$$r^n = a. \quad (3.38)$$

•

Beim Rechnen mit Wurzeln ist die Potenzschreibweise von Wurzeln oft hilfreich, da sie die direkte Anwendung der Rechenregeln für Potenzen ermöglicht.

Theorem 3.4 (Potenzschreibweise der n -ten Wurzel). *Es sei $a \in \mathbb{R}$, $n \in \mathbb{N}$ und r die n -te Wurzel von a . Dann gilt*

$$r = a^{\frac{1}{n}} \quad (3.39)$$

◦

Beweis. Es gilt

$$\left(a^{\frac{1}{n}}\right)^n = a^{\frac{1}{n}} \cdot a^{\frac{1}{n}} \cdot \dots \cdot a^{\frac{1}{n}} = a^{\sum_{i=1}^n \frac{1}{n}} = a^1 = a. \quad (3.40)$$

Also gilt mit Definition 3.4, dass $r = a^{\frac{1}{n}}$. □

Das Rechnen mit Quadratwurzeln wird durch die Potenzschreibweise

$$\sqrt{x} = x^{\frac{1}{2}} \quad (3.41)$$

sehr erleichtert. So gilt zum Beispiel

$$\frac{2\pi}{\sqrt{2\pi}} = \frac{2\pi}{(2\pi)^{\frac{1}{2}}} = (2\pi)^1 \cdot (2\pi)^{-\frac{1}{2}} = (2\pi)^{1-\frac{1}{2}} = (2\pi)^{\frac{1}{2}} = \sqrt{2\pi}. \quad (3.42)$$

Folgender Zusammenhang zwischen Quadrat, Wurzel und Betrag wird des öfteren genutzt.

Theorem 3.5 (Wurzel und Betrag). *Es sei $x \in \mathbb{R}$. Dann gilt*

$$\sqrt{x^2} = |x|. \quad (3.43)$$

◦

Beweis. Wir betrachten die Fälle $x \geq 0$ und $x < 0$. Sei zunächst $x \geq 0$. Dann gilt

$$x \geq 0 \Rightarrow x^2 \geq 0 \Rightarrow \sqrt{x^2} = x = |x|. \quad (3.44)$$

Sei nun $x < 0$. Dann gilt

$$x < 0 \Rightarrow x^2 > 0 \Rightarrow \sqrt{x^2} = -x = |x|. \quad (3.45)$$

□

3.4. Selbstkontrollfragen

1. Geben Sie die Definition des Summenzeichens wieder.
2. Berechnen Sie die Summen $\sum_{i=1}^3 2$, $\sum_{i=1}^3 i^2$, und $\sum_{i=1}^3 \frac{2}{3}i$.
3. Schreiben Sie die Summe $1 + 3 + 5 + 7 + 9 + 11$ mithilfe des Summenzeichens.
4. Schreiben Sie die Summe $0 + 2 + 4 + 6 + 8 + 10$ mithilfe des Summenzeichens.
5. Geben Sie die Definition des Produktzeichens wieder.
6. Geben Sie die Definition der n -ten Potenz von $a \in \mathbb{R}$ wieder.
7. Berechnen Sie $2^2 \cdot 2^3$ und 2^5 und geben Sie die zugehörige Potenzregel wieder.
8. Berechnen Sie 6^2 und $2^2 \cdot 3^2$ und geben Sie die zugehörige Potenzregel wieder.
9. Begründen Sie, warum die n -te Wurzel von a als $a^{\frac{1}{n}}$ geschrieben werden kann.
10. Berechnen Sie $(\sqrt{2})^{\frac{2}{3}}$, $9^{\frac{1}{2}}$, und $4^{-\frac{1}{2}}$.

4. Funktionen

Funktionen und Mengen bilden die Grundpfeiler mathematischer Modellierung. In dieser Einheit definieren wir den Begriff der Funktion, führen erste Eigenschaften von Funktionen ein und geben eine Übersicht über einige elementare Funktionen.

4.1. Definition und Eigenschaften

Definition 4.1 (Funktion). Eine *Funktion* oder *Abbildung* f ist eine Zuordnungsvorschrift, die jedem Element einer Menge D genau ein Element einer Menge Z zuordnet. D wird dabei *Definitionsmenge von f* und Z wird *Zielfmenge von f* genannt. Wir schreiben

$$f : D \rightarrow Z, x \mapsto f(x), \quad (4.1)$$

wobei $f : D \rightarrow Z$ gelesen wird als “die Funktion f bildet alle Elemente der Menge D eindeutig auf Elemente in Z ab” und $x \mapsto f(x)$ gelesen wird als “ x , welches ein Element von D ist, wird durch die Funktion f auf $f(x)$ abgebildet, wobei $f(x)$ ein Element von Z ist”. Der Pfeil \rightarrow steht für die Abbildung zwischen den Mengen D und Z , der Pfeil \mapsto steht für die Abbildung zwischen einem Element von D und einem Element von Z .

•

Es ist zentral, zwischen der *Funktion* f als Zuordnungsvorschrift und einem *Wert der Funktion* $f(x)$ als Element von Z zu unterscheiden. x ist das *Argument* der Funktion (der *Input* der Funktion), $f(x)$ der Wert, den die Funktion f für das Argument x annimmt (der *Output* der Funktion). Üblicherweise folgt in der Definition einer Funktion $f(x)$ die Definition der *funktionalen Form von f* , also einer Regel, wie aus x der Wert $f(x)$ zu bilden ist. Zum Beispiel wird in folgender Definition einer Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto f(x) := x^2 \quad (4.2)$$

die Definition der Potenz genutzt. Man beachte, dass Funktionen immer *eindeutig* sind, als dass sie jedem $x \in D$ bei jeder Anwendung der Funktion immer ein und dasselbe $f(x) \in Z$ zuordnen.

Abbildung 4.1 visualisiert einige Aspekte der Funktionsdefinition. Abbildung 4.1 A stellt dabei zunächst die zentralen Begriffe der Funktionsdefinition bildlich dar. Abbildung 4.1 B visualisiert ein erstes Beispiel für eine Funktion, bei der die den Argumenten zugeordneten Funktionswerte nicht durch eine Rechenregel, sondern durch direkte Definition bestimmt sind. Abbildung 4.1 C und Abbildung 4.1 D bedienen sich der Bildsprache um zwei Aspekte der Definition von Funktionen mithilfe von Gegenbeispielen hervorzuheben: Bei der Zeichnung in Abbildung 4.1 C handelt es sich nicht um die Darstellung einer Funktion, da nach Definition 4.1 eine Funktion *jedem Element* einer Menge D genau ein Element einer Zielfmenge Z zuordnet. Die hier angedeutete Zuordnungsvorschrift ordnet dem Element $2 \in D$ aber kein Element in Z zu und ist deshalb keine Funktion. Gleichsam gilt nach

Definition 4.1, dass eine Funktion jedem Element einer Menge D *genau ein Element einer Zielmenge* Z zuordnet. Die in Abbildung 4.1 D angedeutete Zuordnungsvorschrift ordnet $1 \in D$ aber sowohl das Element $a \in Z$ als auch das Element $b \in Z$ zu und ist deshalb mit Definition 4.1 nicht vereinbar.

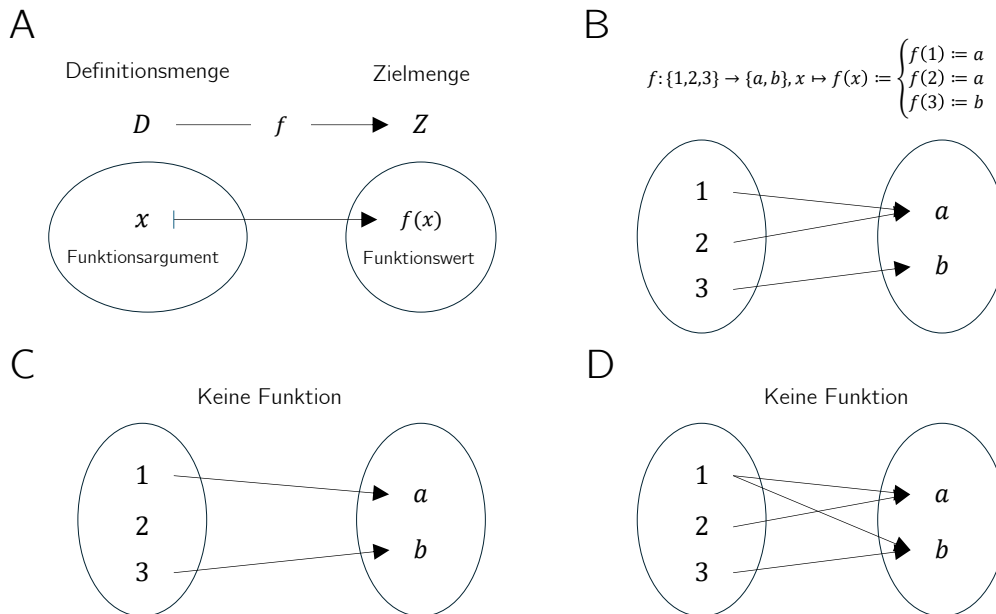


Abbildung 4.1. Aspekte der Funktionsdefinition

Funktionen setzen also Elemente von Mengen miteinander in Beziehung. Die Mengen dieser Elemente erhalten spezielle Bezeichnungen.

Definition 4.2 (Bildmenge, Wertebereich, Urbildmenge, Urbild). Es sei $f: D \rightarrow Z, x \mapsto f(x)$ eine Funktion und es seien $D' \subseteq D$ und $Z' \subseteq Z$. Die Menge

$$f(D') := \{z \in Z \mid \text{Es gibt ein } x \in D' \text{ mit } z = f(x)\} \quad (4.3)$$

heißt die *Bildmenge* von D' und $f(D) \subseteq Z$ heißt der *Wertebereich* von f . Weiterhin heißt die Menge

$$f^{-1}(Z') := \{x \in D \mid f(x) \in Z'\} \quad (4.4)$$

die *Urbildmenge* von Z' . Ein $x \in D$ mit $z = f(x) \in Z$ heißt ein *Urbild* von z .

•

Man beachte, dass der Wertebereich $f(D)$ von f und die Zielmenge Z von f nicht notwendigerweise identisch sein müssen.

Beispiel

Um die in Definition 4.2 eingeführten Begriffe zu verdeutlichen betrachten wir die in

Abbildung 4.2 A dargestellte Funktion

$$f : \{1, 2, 3, 4, 5\} \rightarrow \{a, b, c, d\}, x \mapsto f(x) := \begin{cases} f(1) := b \\ f(2) := d \\ f(3) := c \\ f(4) := c \\ f(5) := d \end{cases} . \quad (4.5)$$

Nach Definition 4.2 ist eine Bildmenge immer bezüglich einer Teilmenge D' der Definitionsmenge D definiert. Sei also wie in Abbildung 4.2 B dargestellt $D' := \{2, 3\} \subset D$. Dann ist die Bildmenge von D' die Menge der $z \in Z$, für die ein $x \in D'$ existiert, so dass $z = f(x)$. Diejenigen $z \in Z$ für die ein $x \in D'$ mit $z = f(x)$ existiert sind aber hier gerade $c, d \in Z$, da $f(2) = d$ und $f(3) = c$. Darüberhinaus gibt es keine $z \in Z$ mit $f(x) = z$ und $x \in \{2, 3\}$. Der Wertebereich $f(D)$ ist nach Definition 4.2 die Teilmenge der $z \in Z$, für die gilt, dass ein $x \in D$ existiert, für das $z = f(x)$ ist. Dies ist für alle Elemente von Z der Fall, außer für $a \in Z$, da für dieses kein $x \in D$ existiert mit $a = f(x)$. Der Wertebereich von f ist für die betrachtete Funktion also durch $f(D) := \{b, c, d\}$ gegeben.

Umgekehrt ist nach Definition 4.2 ist eine Urbildmenge immer bezüglich einer Teilmenge Z' der Zielmenge Z definiert. Sei also wie in Abbildung 4.2 C dargestellt $Z' := \{c, d\} \subset Z$. Dann ist die Urbildmenge von Z' die Menge der $x \in D$, für die gilt, dass $f(x) \in Z'$. Diejenigen Elemente von D , deren Funktionswerte unter f durch Elemente von Z' gegeben sind, sind gerade $\{2, 3, 4, 5\}$. $1 \in D$ dagegen ist kein Element von Z' , da $f(1) \in b \in Z$ abbildet und $b \notin Z'$. Nichtsdestotrotz ist natürlich $1 \in D$ ein Urbild von b .

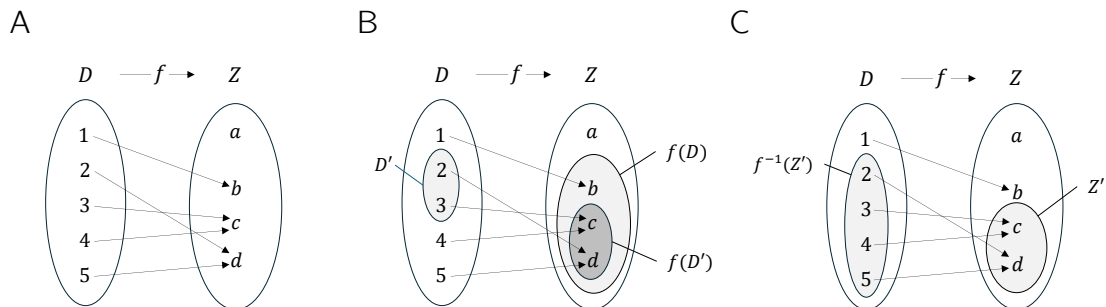


Abbildung 4.2. Bildmenge, Wertebereich, Urbildmenge, Urbild

Grundlegende Eigenschaften von Funktionen werden in folgender Definition benannt.

Definition 4.3 (Injektivität, Surjektivität, Bijektivität). $f : D \rightarrow Z, x \mapsto f(x)$ sei eine Funktion. f heißt *injektiv*, wenn es zu jedem Bild $z \in f(D)$ genau ein Urbild $x \in D$ gibt. Äquivalent gilt, dass f injektiv ist, wenn aus $x_1, x_2 \in D$ mit $x_1 \neq x_2$ folgt, dass $f(x_1) \neq f(x_2)$ ist. f heißt *surjektiv*, wenn $f(D) = Z$ gilt, wenn also jedes Element der Zielmenge Z ein Urbild in der Definitionsmenge D hat. Schließlich heißt f *bijektiv*, wenn f injektiv und surjektiv ist. Bijektive Funktionen werden auch *eineindeutige Funktionen* (engl. *one-to-one mappings*) genannt.

•

Beispiele

Wir verdeutlichen Definition 4.3 zunächst anhand dreier (Gegen)beispiele in Abbildung 4.3. Abbildung 4.3 A visualisiert dabei die *nicht-injektive* Funktion

$$f : \{1, 2, 3\} \rightarrow \{a, b\}, x \mapsto f(x) := \begin{cases} f(1) := a \\ f(2) := a \\ f(3) := b \end{cases} . \quad (4.6)$$

Die Funktion ist nicht-injektiv, weil es zum Element a in der Bildmenge von f mehr als ein Urbild in der Definitionsmenge von f gibt, nämlich die Elemente 1 und 2. Abbildung 4.3 B visualisiert die *nicht-surjektive* Funktion

$$g : \{1, 2, 3\} \rightarrow \{a, b, c, d\}, x \mapsto g(x) := \begin{cases} g(1) := a \\ g(2) := b \\ g(3) := d \end{cases} . \quad (4.7)$$

Die Funktion ist nicht surjektiv, weil das Element c in der Zielmenge von f kein Urbild in der Definitionsmenge von f hat. Abbildung 4.3 C schließlich visualisiert die bijektive Funktion

$$h : \{1, 2, 3\} \rightarrow \{a, b, c\}, x \mapsto h(x) := \begin{cases} h(1) := a \\ h(2) := b \\ h(3) := c \end{cases} . \quad (4.8)$$

Zu *jedem* Element in der Zielmenge von h gibt es *genau ein* Urbild, die Funktion ist also injektiv und surjektiv und damit bijektiv. Bijektive Abbildungen werden auch Eins-zu-Eins-Abbildungen (eng. one-to-one mappings) genannt.

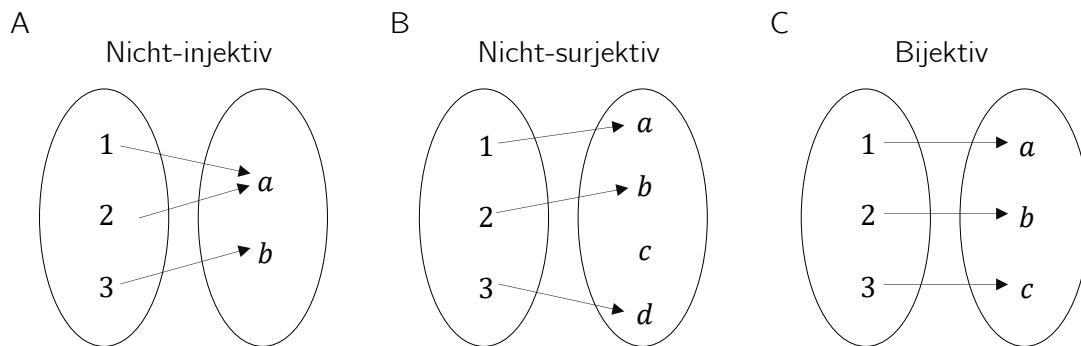


Abbildung 4.3. Injektivität, Surjektivität, Bijektivität.

Als weiteres Beispiel betrachten wir die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2 \quad (4.9)$$

Diese Funktion ist nicht injektiv, weil z.B. für $x_1 = 2 \neq -2 = x_2$ gilt, dass $f(x_1) = 2^2 = 4 = (-2)^2 = f(x_2)$. Weiterhin ist f auch nicht surjektiv, weil z.B. $-1 \in \mathbb{R}$ kein Urbild unter f hat. Schränkt man die Definitionsmenge von f allerdings auf die nicht-negativen reellen Zahlen ein, definiert man also die Funktion

$$\tilde{f} : [0, \infty[\rightarrow [0, \infty[, x \mapsto \tilde{f}(x) := x^2, \quad (4.10)$$

so ist \tilde{f} im Gegensatz zu f injektiv und surjektiv, also bijektiv.

4.2. Funktionentypen

Durch Verkettung lassen sich aus Funktionen weitere Funktionen bilden.

Definition 4.4 (Verkettung von Funktionen). Es seien $f : D \rightarrow Z$ und $g : Z \rightarrow S$ zwei Funktionen, wobei die Wertemenge von f mit der Definitionsmenge von g übereinstimmen sollen. Dann ist durch

$$g \circ f : D \rightarrow S, x \mapsto (g \circ f)(x) := g(f(x)) \quad (4.11)$$

eine Funktion definiert, die die *Verkettung von f und g* genannt wird.

•

Die Schreibweise für verkettete Funktionen ist etwas gewöhnungsbedürftig. Wichtig ist es zu erkennen, dass $g \circ f$ die verkettete Funktion und $(g \circ f)(x)$ ein Element in der Zielmenge der verketteten Funktion bezeichnen. Intuitiv wird bei der Auswertung von $(g \circ f)(x)$ zunächst die Funktion f auf x angewendet und dann die Funktion g das Element auf $f(x)$ von R angewendet. Dies ist in der funktionalen Form $g(f(x))$ festgehalten. Der Einfachheit halber benennt man die Verkettung zweier Funktionen auch oft mit einem einzelnen Buchstaben und schreibt beispielsweise, $h := g \circ f$ mit $h(x) = g(f(x))$. Leicht zur Verwirrung kann es führen, wenn Elemente in der Zielmenge von f mit y bezeichnet werden, also die Schreibweise $y = f(x)$ und $h(x) = g(y)$ genutzt wird. Allerdings ist diese Schreibweise manchmal zur notationellen Vereinfachung nötig. Wir visualisieren Definition 4.4 in Abbildung 4.4.

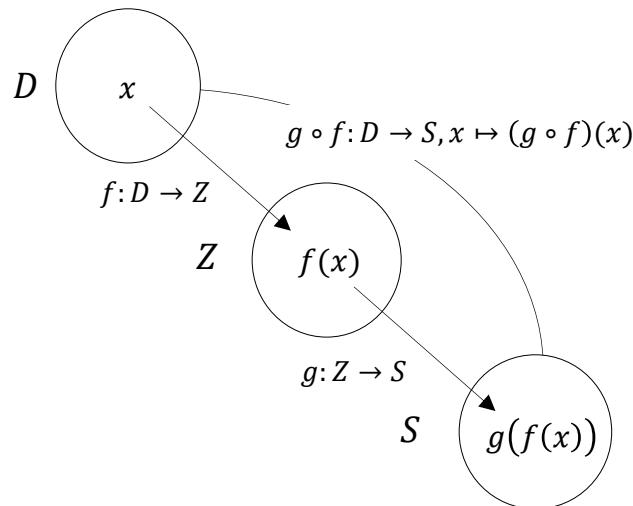


Abbildung 4.4. Verkettung von Funktionen.

Als Beispiel für die Verkettung zweier Funktionen betrachten wir

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := -x^2 \quad (4.12)$$

und

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto g(x) := \exp(x). \quad (4.13)$$

Die Verkettung von f und g ergibt sich in diesem Fall zu

$$g \circ f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto (g \circ f)(x) := g(f(x)) = \exp(-x^2). \quad (4.14)$$

Eine erste Anwendung der Verkettung von Funktionen findet sich in folgender Definition.

Definition 4.5 (Inverse Funktion). Es sei $f : D \rightarrow Z, x \mapsto f(x)$ eine bijektive Funktion. Dann heißt die Funktion f^{-1} mit

$$f^{-1} \circ f : D \rightarrow D, x \mapsto (f^{-1} \circ f)(x) := f^{-1}(f(x)) = x \quad (4.15)$$

inverse Funktion, Umkehrfunktion oder einfach *Inverse* von f .

•

Inverse Funktionen sind immer bijektiv. Dies folgt, weil f bijektiv ist und damit jedem $x \in D$ genau ein $f(x) = z \in Z$ zugeordnet wird. Damit wird aber auch jedem $z \in Z$ genau ein $x \in D$, nämlich $f^{-1}(f(x)) = x$ zugeordnet. Intuitiv macht die inverse Funktion von f den Effekt von f auf ein Element x rückgängig. Wir visualisieren Definition 4.5 in Abbildung 4.5 A.

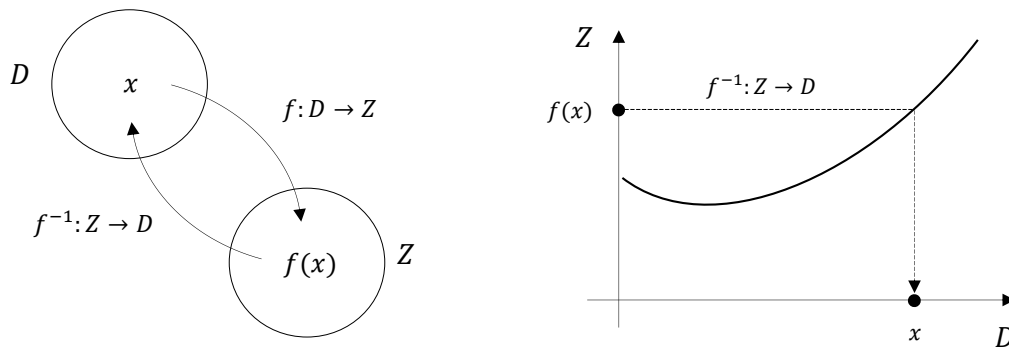


Abbildung 4.5. Inverse Funktion.

Betrachtet man konkret den Graphen einer Funktion in einem Kartesischen Koordinatensystem, so führt die Anwendung von einem Wert auf der x -Achse zu einem Wert auf der y -Achse. Die Anwendung der inversen Funktion führt dementsprechend von einem Wert auf der y -Achse zu einem Wert auf der x -Achse. Wir visualisieren dies in Abbildung 4.5 B. Betrachten wir beispielsweise die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := 2x =: y. \quad (4.16)$$

Dann ist die inverse Funktion von f gegeben durch

$$f^{-1} : \mathbb{R} \rightarrow \mathbb{R}, y \mapsto f^{-1}(y) := \frac{1}{2}y, \quad (4.17)$$

weil für jedes $x \in \mathbb{R}$ gilt, dass

$$(f^{-1} \circ f)(x) := f^{-1}(f(x)) = f^{-1}(2x) = \frac{1}{2} \cdot 2x = x. \quad (4.18)$$

Eine wichtige Klasse von Funktionen sind *lineare Abbildungen*.

Definition 4.6 (Lineare Abbildung). Eine Abbildung $f : D \rightarrow Z, x \mapsto f(x)$ heißt *lineare Abbildung*, wenn für $x, y \in D$ und einen Skalar c gelten, dass

$$f(x + y) = f(x) + f(y) \quad f(cx) = cf(x) \quad (\text{Additivität})$$

und

$$f(cx) = cf(x) \quad (\text{Homogenität})$$

Eine Abbildung, für die obige Eigenschaften nicht gelten, heißt *nicht-lineare Abbildung*.

•

Lineare Abbildungen sind oft als “gerade Linien” bekannt. Die allgemeine Definition linearer Abbildungen ist mit dieser Intuition nicht komplett kongruent. Insbesondere sind lineare Abbildungen nur solche Funktionen, die den Nullpunkt auf den Nullpunkt abbilden. Wir zeigen dazu folgendes Theorem.

Theorem 4.1 (Lineare Abbildung der Null). $f : D \rightarrow Z$ sei eine lineare Abbildung. Dann gilt

$$f(0) = 0. \quad (4.19)$$

◦

Beweis. Wir halten zunächst fest, dass mit der Additivität von f gilt, dass

$$f(0) = f(0 + 0) = f(0) + f(0). \quad (4.20)$$

Addition von $-f(0)$ auf beiden Seiten obiger Gleichung ergibt dann

$$\begin{aligned} f(0) - f(0) &= f(0) + f(0) - f(0) \\ 0 &= f(0) \end{aligned} \quad (4.21)$$

und damit ist alles gezeigt. \square

Wir wollen den Begriff der linearen Abbildung noch an zwei Beispielen verdeutlichen.

- Für $a \in \mathbb{R}$ ist die Abbildung

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax \quad (4.22)$$

eine lineare Abbildung, weil gilt, dass

$$f(x + y) = a(x + y) = ax + ay = f(x) + f(y) \quad \text{und} \quad f(cx) = acx = cax = cf(x). \quad (4.23)$$

- Für $a, b \in \mathbb{R}$ ist dagegen die Abbildung

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax + b \quad (4.24)$$

nicht-linear, weil z.B. für $a := b := 1$ gilt, dass

$$f(x + y) = 1(x + y) + 1 = x + y + 1 \neq x + 1 + y + 1 = f(x) + f(y). \quad (4.25)$$

Eine Abbildung der Form $f(x) := ax + b$ heißt *linear-affine Abbildung* oder *linear-affine Funktion*. Etwas unsauber werden Funktionen der Form $f(x) := ax + b$ auch manchmal als *lineare Funktionen* bezeichnet.

Neben den bisher diskutierten Funktionentypen gibt es noch viele weitere Klassen von Funktionen. In folgender Definition klassifizieren wir Funktionen anhand der Dimensionalität ihrer Definitions- und Zielmengen. Diese Art der Funktionsklassifikation ist oft hilfreich, um sich einen ersten Überblick über ein mathematisches Modell zu verschaffen.

Definition 4.7 (Funktionenarten). Wir unterscheiden

- *univariate reellwertige Funktionen* der Form

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x), \quad (4.26)$$

- *multivariate reellwertige Funktionen* der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n), \quad (4.27)$$

- und *multivariate vektorwertige Funktionen* der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto f(x) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}, \quad (4.28)$$

wobei $f_i, i = 1, \dots, m$ die *Komponenten(funktionen)* von f genannt werden.

•

In der Physik werden multivariate reellwertige Funktionen *Skalarfelder* und multivariate vektorwertige Funktionen *Vektorfelder* genannt. In manchen Anwendungen treten zum Beispiel auch *matrixvariante matrixwertige Funktionen* auf.

4.3. Elementare Funktionen

Als *elementare Funktionen* bezeichnen wir eine kleine Schar von univariaten reellwertigen Funktionen, die häufig als Bausteine komplexerer Funktionen auftreten. Dies sind die *Polynomfunktionen*, die *Exponentialfunktion*, die *Logarithmusfunktion* und die *Gammafunktion*. Im Folgenden geben wir die Definitionen dieser Funktionen sowie ihre wesentlichen Eigenschaften als Theoreme an und stellen ihre Graphen an. Für Beweise der Eigenschaften der hier vorgestellten Funktionen verweisen wir auf die weiterführende Literatur.

Definition 4.8 (Polynomfunktionen). Eine Funktion der Form

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \sum_{i=0}^k a_i x^i = a_0 + a_1 x^1 + a_2 x^2 + \dots + a_k x^k \quad (4.29)$$

heißt *Polynomfunktion* k -ten Grades mit Koeffizienten $a_0, a_1, \dots, a_k \in \mathbb{R}$. Typische Polynomfunktionen sind in nachfolgender Tabelle aufgelistet:

Name	Form	Koeffizienten
Konstante Funktion	$f(x) = a$	$a_0 := a, a_i := 0, i > 0$
Identitätsfunktion	$f(x) = x$	$a_0 := 0, a_1 := 1, a_i := 0, i > 1$
Linear-affine Funktion	$f(x) = ax + b$	$a_0 := b, a_1 := a, a_i := 0, i > 1$
Quadratfunktion	$f(x) = x^2$	$a_0 := 0, a_1 := 0, a_2 := 1, a_i := 0, i > 2$

•

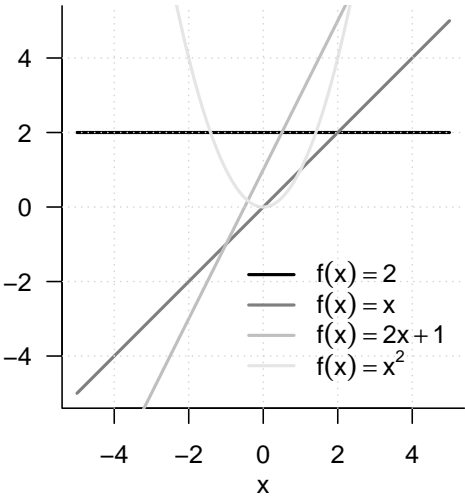


Abbildung 4.6. Ausgewählte Polynomfunktionen

Abbildung 4.6 zeigt die Graphen der in Definition 4.8 aufgelisteten Polynomfunktionen. Ein wichtiges Funktionenpaar sind die Exponentialfunktion und die Logarithmusfunktion.

Theorem 4.2 (Exponentialfunktion und ihre Eigenschaften). *Die Exponentialfunktion ist definiert als*

$$\exp : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \exp(x) := e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

(4.30)

Die Exponentialfunktion hat folgende Eigenschaften:

Eigenschaft	Bedeutung
Wertebereich	$x \in]-\infty, 0[\Rightarrow \exp(x) \in]0, 1[$ $x \in]0, \infty[\Rightarrow \exp(x) \in]1, \infty[$
Monotonie	$x < y \Rightarrow \exp(x) < \exp(y)$
Spezielle Werte	$\exp(0) = 1$ und $\exp(1) = e$

Eigenschaft	Bedeutung
Summationseigenschaft	$\exp(x + y) = \exp(x) \exp(y)$
Subtraktionseigenschaft	$\exp(x - y) = \frac{\exp(x)}{\exp(y)}$

◦

Insbesondere nimmt die Exponentialfunktion also nur positive Werte an und schneidet die y -Achse also bei $x = 0$. Die Zahl $\exp(1) := e \approx 2.71\dots$ heißt *Eulersche Zahl*. Schließlich gilt mit den speziellen Werten der Exponentialfunktion insbesondere auch

$$\exp(x) \exp(-x) = \exp(x - x) = \exp(0) = 1. \quad (4.31)$$

Theorem 4.3 (Logarithmusfunktion und ihre Eigenschaften). *Die Logarithmusfunktion ist definiert als inverse Funktion der Exponentialfunktion,*

$$\ln :]0, \infty[\rightarrow \mathbb{R}, x \mapsto \ln(x) \text{ mit } \ln(\exp(x)) = x \text{ für alle } x \in \mathbb{R}. \quad (4.32)$$

Die Logarithmusfunktion hat folgende Eigenschaften:

Eigenschaft	Bedeutung
Wertebereich	$x \in]0, 1[\Rightarrow \ln(x) \in]-\infty, 0[$ $x \in]1, \infty[\Rightarrow \ln(x) \in]0, \infty[$
Monotonie	$x < y \Rightarrow \ln(x) < \ln(y)$
Spezielle Werte	$\ln(1) = 0$ und $\ln(e) = 1$
Produkteigenschaft	$\ln(xy) = \ln(x) + \ln(y)$
Potenzeigenschaft	$\ln(x^c) = c \ln(x)$
Divisionseigenschaft	$\ln\left(\frac{1}{x}\right) = -\ln(x)$

◦

Im Gegensatz zur Exponentialfunktion nimmt die Logarithmusfunktion sowohl negative als auch positive Werte an und Logarithmusfunktion schneidet die x -Achse bei $x = 1$. Die Produkteigenschaft und die Potenzseigenschaften sind beim Rechnen mit der Logarithmusfunktion zentral. Man merkt sie sich intuitiv als “Die Logarithmusfunktion wandelt Produkte in Summen und Potenzen in Produkte um.” Die Graphen der Exponential- und Logarithmusfunktion sind in Abbildung 4.7 abgebildet.

Ein häufiger Begleiter in der Wahrscheinlichkeitstheorie ist die *Gammafunktion*.

Definition 4.9 (Gammafunktion). Die *Gammafunktion* ist definiert durch

$$\Gamma : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \Gamma(x) := \int_0^\infty \xi^{x-1} \exp(-\xi) d\xi \quad (4.33)$$

Die Gammafunktion hat folgende Eigenschaften:

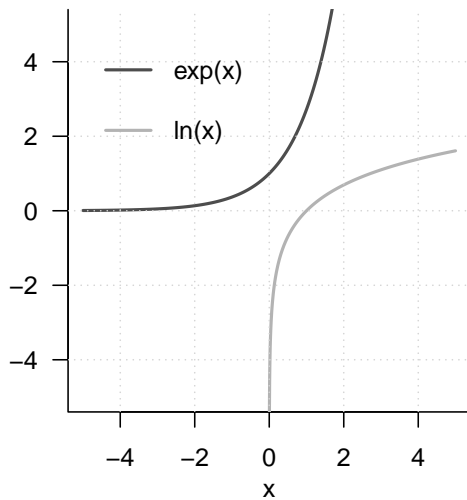


Abbildung 4.7. Exponentialfunktion und Logarithmusfunktion

Eigenschaft	Bedeutung
Spezielle Werte	$\Gamma(1) = 1$ $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$
Rekursionseigenschaft	Für $x > 0$ gilt $\Gamma(x+1) = x\Gamma(x)$

•

Ein Ausschnitt des Graphen der Gammafunktion ist in [Abbildung 4.8](#) dargestellt.

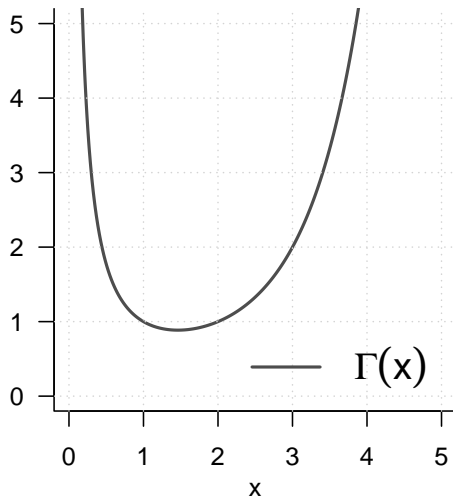


Abbildung 4.8. Gammafunktion
Probabilistische Datenwissenschaft für die Psychologie | © 2025 Dirk Ostwald CC BY 4.0

5. Folgen, Grenzwerte, Stetigkeit

Die in diesem Kapitel behandelten Themen sind in der probabilistischen Datenanalyse nicht zentral, sondern bilden Grundpfeiler der reellen Analysis. Durch die enge Verschränkung der modernen Wahrscheinlichkeitstheorie mit analytischen Ansätzen dienen sie jedoch dem Verständnis von zum Beispiel dem Zentralen Grenzwertsatz, der eine Hauptgrundlage für die weit verbreitete Normalverteilungsannahme in der probabilistischen Datenanalyse darstellt. In aller Kürze ist der Zentrale Grenzwertsatz eine Aussage über die Grenzfunktion einer Funktionenfolge, nämlich einer Folge von Zufallsvariablen. Das Wissen um das Wesen von Folgen, Funktionenfolgen und ihren Grenzwerten erlaubt also ein tieferes Verständnis wichtiger Grundannahmen der probabilistischen Datenanalyse. Weiterhin ermöglichen die in diesem Kapitel behandelten Themen zumindest einen ersten Einstieg in das Verständnis der Stetigkeit und Glattheit von Funktionen, die insbesondere in der nichtlinearen Optimierung zu Bestimmung von Parameterschätzern in probabilistischen Modellen wichtige Grundkonzepte bilden.

5.1. Folgen

Wir beginnen mit der Definition des Begriffs der *reellen Folge*.

Definition 5.1 (Reelle Folge). Eine *reelle Folge* ist eine Funktion der Form

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) \quad (5.1)$$

Die Funktionswerte $f(n)$ einer reellen Folge werden üblicherweise mit x_n bezeichnet und *Folgenglieder* genannt. Übliche Schreibweisen für Folgen sind

$$(x_1, x_2, \dots) \text{ oder } (x_n)_{n=1}^{\infty} \text{ oder } (x_n)_{n \in \mathbb{N}} \text{ oder } (x_n). \quad (5.2)$$

•

Man beachte, dass weil es unendlich viele natürliche Zahlen gibt, eine reelle Folge immer unendlich viele Folgenglieder hat. Dies sollte man sich insbesondere bei der Schreibweise (x_1, x_2, \dots) bewusst machen. Wir wollen zwei Standardbeispiele für reelle Folgen betrachten.

Beispiele für reelle Folgen

(1) Reelle Folgen der Form

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) := \left(\frac{1}{n}\right)^{\frac{p}{q}} \text{ mit } p, q \in \mathbb{N} \quad (5.3)$$

nennen wir *harmonische Folgen*. Für $p := q := 1$ hat eine harmonische Folge die Folgengliederform

$$\left(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \dots\right). \quad (5.4)$$

(2) Reelle Folgen der Form

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) := q^n \text{ mit } q \in]-1, 1[\quad (5.5)$$

werden *geometrische Folgen* genannt. Für $q := \frac{1}{2}$ hat eine geometrische Folge die Folengliederform

$$\begin{aligned} \left(\left(\frac{1}{2}\right)^1, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \dots \right) &= \left(\left(\frac{1}{2}\right)^1, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \dots \right) \\ &= \left(\frac{1^1}{2^1}, \frac{1^2}{2^2}, \frac{1^3}{2^3}, \dots \right) \\ &= \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots \right) \end{aligned} \quad (5.6)$$

Neben den reellen Folgen, die Folgen reeller Zahlen sind, kann man auch Folgen anderer mathematischer Objekte betrachten. Eine wichtige Folgenart sind die *Funktionenfolgen*.

Definition 5.2 (Funktionenfolge). Es sei ϕ eine Menge univariater reellwertiger Funktionen mit Definitionsmenge $D \subseteq \mathbb{R}$. Dann ist eine Funktionenfolge eine Funktion der Form

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n). \quad (5.7)$$

Die Funktionswerte $F(n)$ einer Funktionenfolgen werden üblicherweise mit f_n bezeichnet und *Folenglieder* genannt. Übliche Schreibweisen für Funktionenfolgen sind

$$(f_1, f_2, \dots) \text{ oder } (f_n)_{n=1}^{\infty} \text{ oder } (f_n)_{n \in \mathbb{N}} \text{ oder } (f_n). \quad (5.8)$$

•

Die Definition einer Funktionenfolge ist offenbar analog zur Definition einer reellen Folge. Der Unterschied zwischen einer reellen Folge und einer Funktionenfolge ist, dass die Folenglieder einer reellen Folge reelle Zahlen, die Folenglieder einer Funktionenfolgen dagegen univariate reellwertige Funktionen sind. Auch hier wollen wir zwei Standardbeispiel diskutieren.

Beispiele für Funktionenfolgen

(1) Wir betrachten die Menge ϕ der univariaten reellwertigen Funktionen der Form

$$\phi := \{f_n | f_n : [0, 1] \rightarrow \mathbb{R}, x \mapsto f_n(x) := x^n \text{ für } n \in \mathbb{N}\} \quad (5.9)$$

Dann definiert

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (5.10)$$

eine Funktionenfolge. Für die Funktionswerte der Folenglieder von F gilt

$$f_1(x) := x^1, f_2(x) := x^2, f_3(x) := x^3, \dots \quad (5.11)$$

(2) Wir betrachten die Menge ϕ der univariaten reellwertigen Funktionen der Form

$$\phi := \{f_n | f_n : [-a, a] \rightarrow \mathbb{R}, x \mapsto f_n(x) := \sum_{k=0}^n \frac{x^k}{k!} \text{ für } n \in \mathbb{N}\} \quad (5.12)$$

Dann definiert

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (5.13)$$

eine Funktionenfolge. Für die Funktionswerte der Folgenglieder von F gilt

$$f_1(x) := \sum_{k=0}^1 \frac{x^k}{k!}, f_2(x) := \sum_{k=0}^2 \frac{x^k}{k!}, f_3(x) := \sum_{k=0}^3 \frac{x^k}{k!}, \dots \quad (5.14)$$

5.2. Grenzwerte

Wenn man die Folgenglieder einer Folge betrachtet, kann man sich fragen, welche Werte eine Folge wohl annimmt, wenn der Folgenindex n sehr groß wird, also gegen unendlich strebt. Wenn in diesem Fall die Folgenglieder sehr ähnliche Werte annehmen (und nicht etwa auch unendlich groß werden), so ist man auf den Begriff des *Grenzwerts* für reelle Folgen bzw. der *Grenzfunktion* für Funktionenfolgen geführt.

Definition 5.3 (Grenzwert einer Folge). $x \in \mathbb{R}$ heißt Grenzwert einer reellen Folge $(x_n)_{n=1}^\infty$, wenn es zu jedem $\epsilon > 0$ ein $m \in \mathbb{N}$ gibt, so dass

$$|x_n - x| < \epsilon \text{ für alle } n \geq m. \quad (5.15)$$

Eine Folge, die einen Grenzwert besitzt, wird *konvergente Folge* genannt, eine Folge die keinen Grenzwert besitzt, wird *divergente Folge* genannt. Dafür, dass $x \in \mathbb{R}$ Grenzwert der Folge $(x_n)_{n=1}^\infty$ ist, schreibt man auch

$$\lim_{n \rightarrow \infty} x_n = x \text{ oder } x_n \rightarrow x \text{ für } n \rightarrow \infty \text{ oder } x_n \xrightarrow{n \rightarrow \infty} x. \quad (5.16)$$

•

Der Grenzwert einer Folge kann also, aber muss nicht existieren. So hat zum Beispiel die Folge

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) := n \quad (5.17)$$

keinen Grenzwert, da hier sowohl n als auch $f(n)$ unendlich groß werden. Die oben betrachteten Beispiele für reelle Folgen dagegen haben Grenzwert. Dies ist Inhalt folgender Beispiele

Beispiele

(1) Für die verallgemeinerten harmonischen Folgen gilt mit $p, q \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \right)^{\frac{p}{q}} = 0. \quad (5.18)$$

(2) Für die geometrischen Folgen gilt mit $q \in]-1, 1[$

$$\lim_{n \rightarrow \infty} q^n = 0. \quad (5.19)$$

Man nennt die harmonischen und geometrischen Folgen entsprechend auch *Nullfolgen*. Für Beweise von Gleichung 5.18 und Gleichung 5.19 verweisen wir auf die weiterführende Literatur. Tatsächlich sind diese Beweise nicht trivial und rühren an die Grundannahmen über das Wesen der reellen Zahlen. Wir visualisieren die ersten zehn Folgenglieder sowie die Grenzwerte der harmonischen Folge für $p := q := 1$ und der geometrischen Folge für $q := 1/2$ in Abbildung 5.1.

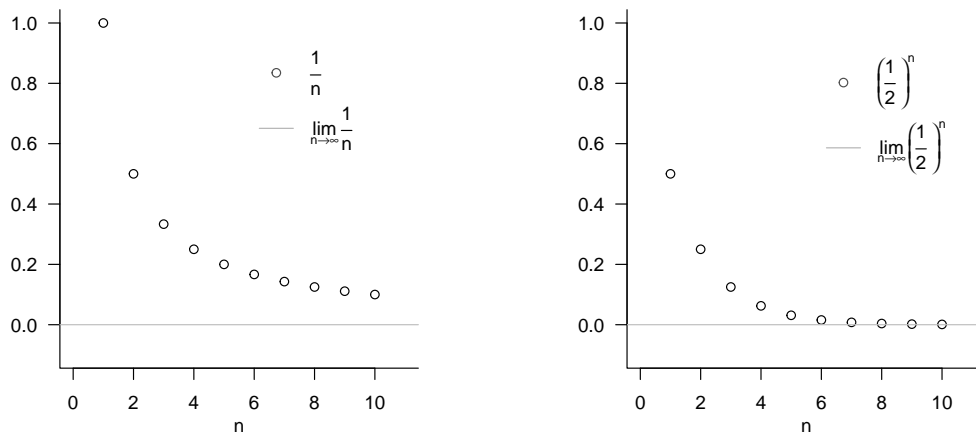


Abbildung 5.1. Beispiele für Grenzwerte reeller Folgen.

Für Funktionenfolgen ist eine Möglichkeit der Erweiterung der Begriffe der Konvergenz und des Grenzwertes folgende.

Definition 5.4 (Punktweise Konvergenz und Grenzfunktion einer Funktionenfolge). $F = (f_n)_{n \in \mathbb{N}}$ sei eine Funktionenfolge von univariaten reellwertigen Funktionen mit Definitionsbereich D . F heißt *punktweise konvergent*, wenn die reelle Folge $(f_n(x))_{n \in \mathbb{N}}$ für jedes $x \in D$ eine konvergente Folge ist, also einen Grenzwert besitzt. Die Funktion, die jedem $x \in D$ diesen Grenzwert von $(f_n(x))_{n \in \mathbb{N}}$ zuordnet, heißt dann die *Grenzfunktion* der Funktionenfolge F und hat die Form

$$f : D \rightarrow \mathbb{R}, x \mapsto f(x) := \lim_{n \rightarrow \infty} f_n(x). \quad (5.20)$$

•

Man beachte, dass die Grenzwerte von konvergenten reellen Folgen reelle Zahlen sind, die Grenzfunktionen von punktweise konvergenten Funktionenfolgen dagegen sind Funktionen. Neben der punktweisen Konvergenz von Funktionenfolgen gibt es noch den mächtigeren Begriff der *gleichmäßigen Konvergenz* von Funktionenfolgen, für den wir aber auf die weiterführende Literatur verweisen. Als Beispiel betrachten wir die Grenzfunktionen der oben diskutierten Funktionenfolgen, wobei wir für Beweise ebenfalls auf die weiterführende Literatur verweisen.

Beispiele

(1) Wir betrachten die Funktionenfolge

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (5.21)$$

mit

$$\phi := \{f_n | f_n : [0, 1] \rightarrow \mathbb{R}, x \mapsto f_n(x) := x^n \text{ für } n \in \mathbb{N}\} \quad (5.22)$$

Dann ist F punktweise konvergent mit Grenzfunktion

$$f : [0, 1] \rightarrow \mathbb{R}, x \mapsto f(x) := \begin{cases} 0, & \text{für } x \in [0, 1[\\ 1, & \text{für } x = 1 \end{cases} \quad (5.23)$$

da $f_n(x) := x^n$ für $x \in [0, 1[$ eine geometrische Folge und damit eine Nullfolge ist und $f_n(x) := x^n$ für $x = 1$ eine konstante Folge ist, für die alle Folgenglieder den Abstand 0 von 1 haben. Die Funktionenfolge F konvergiert also gegen eine Funktion, die auf dem gesamten Intervall $[0, 1]$ gleich Null ist, außer im Punkt 1. Diese Funktion hat offenbar einen Sprung.

(2) Wir betrachten die Funktionenfolge

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (5.24)$$

mit

$$\phi := \{f_n | f_n : [-a, a] \rightarrow \mathbb{R}, x \mapsto f_n(x) := \sum_{k=0}^n \frac{x^k}{k!} \text{ für } n \in \mathbb{N}\} \quad (5.25)$$

Dann ist F punktweise konvergent mit Grenzfunktion

$$f : [-a, a] \rightarrow \mathbb{R}, x \mapsto f(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!} =: \exp(x) \quad (5.26)$$

Die Funktionenfolge F konvergiert also gegen die Exponentialfunktion auf $[-a, a]$. Umgekehrt betrachtet ist die Exponentialfunktion gerade durch

$$\exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (5.27)$$

definiert.

5.3. Stetigkeit

In diesem Abschnitt versuchen wir uns dem Begriff der *Stetigkeit* einer Funktion zu nähern. Intuitiv ist eine Funktion stetig, wenn sie keine Sprünge hat oder äquivalent, wenn kleine Änderungen in ihren Argumenten stets nur zu kleinen Änderungen in ihren Funktionswerten (und damit eben keinen Sprüngen) führen. Zur Definition der Stetigkeit benötigen wir zunächst den Begriff des *Grenzwertes einer Funktion*.

Definition 5.5 (Grenzwert einer Funktion).

Für $D \subseteq \mathbb{R}$ und $Z \subseteq \mathbb{R}$ sei $f : D \rightarrow Z, x \mapsto f(x)$ eine Funktion und es seien $a, b \in \mathbb{R}$. b heißt *Grenzwert der Funktion f für x gegen a* , wenn

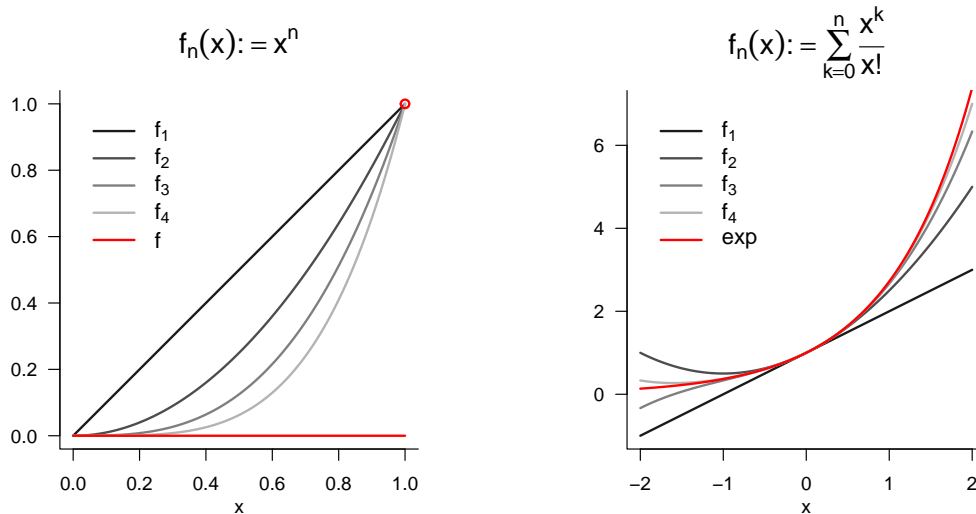


Abbildung 5.2. Beispiele für Grenzwerte von Funktionenfolgen

- (1) es eine reelle Folge $(x_n)_{n=1}^{\infty}$ mit Folgengliedern in D mit Grenzwert a gibt, also $\lim_{n \rightarrow \infty} x_n = a$ gilt, und
- (2) für jede solche Folge gilt, dass b der Grenzwert der Folge der Funktionswerte $f(x_n)$ der Folgenglieder von $(x_n)_{n=1}^{\infty}$ ist, also $\lim_{n \rightarrow \infty} f(x_n) = b$ gilt.

Wenn b Grenzwert der Funktion f für x gegen a ist, so schreibt man auch $\lim_{x \rightarrow a} f(x) = b$.

•

In [Abbildung 5.3](#) visualisieren wir den Grenzwert der Exponentialfunktion in $a = 1$ durch Darstellung von Folgengliedern $x_n \rightarrow 1$ und den entsprechenden Folgengliedern $f(x_n)$. Offenbar gilt $\lim_{x \rightarrow 1} \exp(x) = e$.

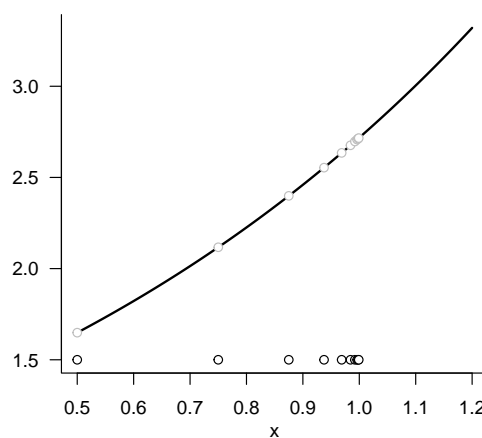


Abbildung 5.3. Beispiele für einen Grenzwert einer Funktion

Wir können nun den Begriff der Stetigkeit einer Funktion definieren.

Definition 5.6 (Stetigkeit einer Funktion). Eine Funktion $f : D \rightarrow Z$ mit $D \subseteq \mathbb{R}, Z \subseteq \mathbb{R}$ heißt *stetig in* $a \in D$, wenn

$$\lim_{x \rightarrow a} f(x) = f(a). \quad (5.28)$$

Ist f in jedem $x \in D$ stetig, so heißt f *stetig auf* D .

•

Man beachte, dass für eine in a stetige Funktion folgt, dass

$$\lim_{x \rightarrow a} f(x) = f\left(\lim_{x \rightarrow a} x\right) \quad (5.29)$$

Bei stetigen Funktion können also Grenzwertbildung und Auswertung der Funktion vertauscht werden.

6. Differentialrechnung

Die Differentialrechnung befasst sich mit der Änderung von Funktionen. Sie bildet einerseits die Grundlage für die mathematische Modellierung mithilfe von *Differentialgleichungen*, also der Beschreibung von Funktionen anhand ihrer Änderungsraten. Zum anderen bildet die Differentialrechnung die Grundlage der *Optimierung*, also des Bestimmens von Extremstellen von Funktionen. In Kapitel 6.1 führen wir zunächst den Begriff der Ableitung und mit ihm verbundene elementare Rechenregeln ein. In Kapitel 6.2 widmen wir uns dann der Frage, wie man mithilfe von Ableitungen Extremstellen von Funktionen bestimmen kann.

6.1. Definitionen und Rechenregeln

Wir beginnen mit folgender Definition.

Definition 6.1 (Differenzierbarkeit und Ableitung). Es sei $I \subseteq \mathbb{R}$ ein Intervall und

$$f : I \rightarrow \mathbb{R}, x \mapsto f(x) \quad (6.1)$$

eine univariate reellwertige Funktion. f heißt in $a \in I$ *differenzierbar*, wenn der Grenzwert

$$f'(a) := \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \quad (6.2)$$

existiert. $f'(a)$ heißt dann die *Ableitung von f an der Stelle a* . Ist f differenzierbar für alle $x \in I$, so heißt f *differenzierbar* und die Funktion

$$f' : I \rightarrow \mathbb{R}, x \mapsto f'(x) \quad (6.3)$$

heißt *Ableitung von f* .

•

Für $h > 0$ heißt der Ausdruck

$$\frac{f(a+h) - f(a)}{h} \quad (6.4)$$

Newtonscher Differenzquotient. Wie in Abbildung 6.1 dargestellt, misst der Newtonsche Differenzquotient die Änderung $f(a+h) - f(a)$ von f auf der y -Achse pro Strecke h auf der x -Achse. Wenn also zum Beispiel $f(a)$ und $f(a+h)$ die Position eines Objektes zu einem Zeitpunkt a und zu einem späteren Zeitpunkt $a+h$ repräsentieren, dann ist $f(a+h) - f(a)$ die von diesem Objekt in der Zeit h zurückgelegte Strecke, also seine durchschnittliche Geschwindigkeit über den Zeitraum h . Für $h \rightarrow 0$ misst der Newtonsche Differenzquotient dann die instantane Änderungsrate von f in a , also im Beispiel die Geschwindigkeit des Objektes zu einem Zeitpunkt a .

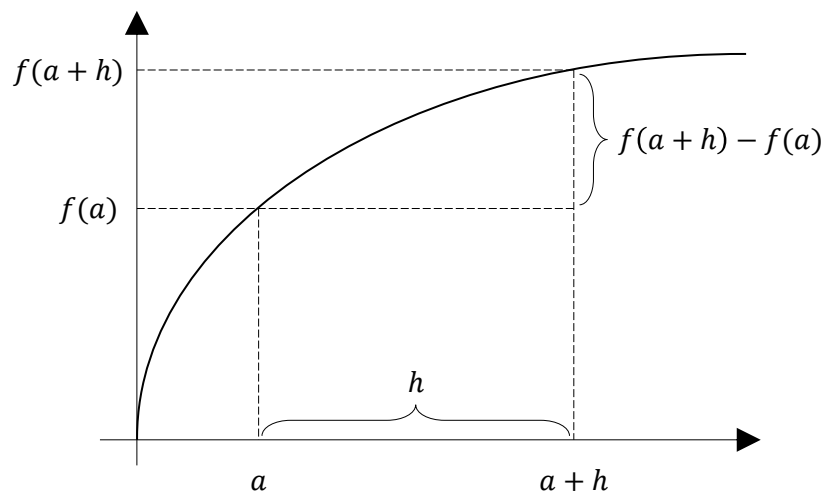


Abbildung 6.1. Newtonscher Differenzquotient

Aus mathematischer Sicht ist es wichtig, bei der Definition der Ableitung zwischen den Symbolen $f'(a)$ und f' zu unterscheiden. Wie üblich bezeichnet $f'(a)$ den Wert einer Funktion, also eine Zahl. f' dagegen bezeichnet eine Funktion, nämlich die Funktion, deren Werte als $f'(a)$ für alle $a \in \mathbb{R}$ bestimmt sind.

Es existieren in der Literatur verschiedene, historisch gewachsene Notationen für Ableitungen, welche alle das identische Konzept der Ableitung repräsentieren.

Definition 6.2 (Notation für Ableitungen univariater reellwertiger Funktionen). Es sei f eine univariate reellwertige Funktion. Äquivalente Schreibweisen für die Ableitung von f und die Ableitung von f an einer Stelle x sind

- die *Lagrange-Notation* f' und $f'(x)$,
- die *Leibniz-Notation* $\frac{df}{dx}$ und $\frac{df(x)}{dx}$,
- die *Newton-Notation* \dot{f} und $\dot{f}(x)$, sowie
- die *Euler-Notation* Df und $Df(x)$,

respektive.

•

Wir werden im Folgenden für univariate reellwertige Funktionen vor allem die Lagrange-Notation f' und $f'(x)$ als Bezeichner wählen. In Berechnungen nutzen wir auch eine adaptierte Form der Leibniz-Notation und verstehen dort die Schreibweise $\frac{d}{dx}f(x)$ als den Auftrag, die Ableitung von f zu berechnen. Die Newton-Notation wird vor allem eingesetzt, wenn das Funktionsargument die Zeit repräsentiert und dann üblicherweise mit t für *time* bezeichnet wird. $\dot{f}(t)$ bezeichnet dann die Änderungsrate von f zum Zeitpunkt t . Die Euler-Notation ist vor allem im Kontext multivariater reell- oder vektorwertiger Funktionen nützlich.

Wir setzen eine Reihe von Ableitungen elementarer Funktionen als bekannt voraus, diese sind in Tabelle 5.1 zusammengestellt. Für Beweise verweisen wir wiederum auf die weiterführende Literatur.

Tabelle 6.1. Ableitungen elementarer Funktionen

Name	Definition	Ableitung
Polynomfunktion	$f(x) := \sum_{i=0}^n a_i x^i$	$f'(x) = \sum_{i=1}^n i a_i x^{i-1}$
Konstante Funktion	$f(x) := a$	$f'(x) = 0$
Identitätsfunktion	$f(x) := x$	$f'(x) = 1$
Linear-affine Funktion	$f(x) := ax + b$	$f'(x) = a$
Quadratfunktion	$f(x) := x^2$	$f'(x) = 2x$
Exponentialfunktion	$f(x) := \exp(x)$	$f'(x) = \exp(x)$
Logarithmusfunktion	$f(x) := \ln(x)$	$f'(x) = \frac{1}{x}$

In Abbildung 6.2 visualisieren wir die Identitätsfunktion, eine linearen Funktion und die Quadratfunktion zusammen mit ihrer jeweiligen Ableitung. In Abbildung 6.3 visualisieren wir die Exponential- und Logarithmusfunktionen zusammen mit ihrer jeweiligen Ableitung.

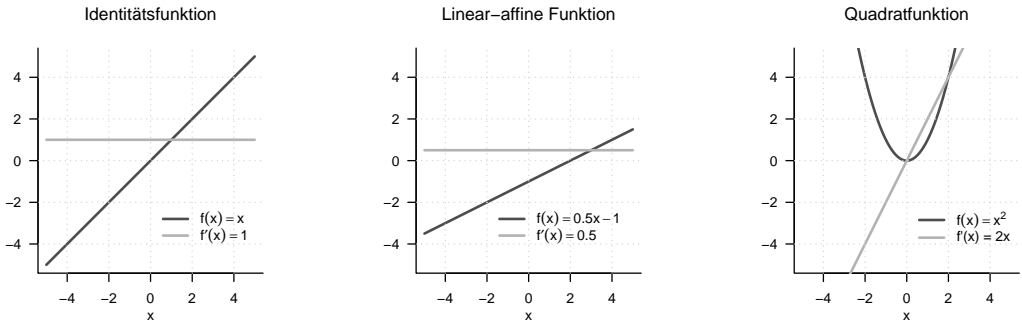


Abbildung 6.2. Ableitungen dreier elementarer Funktionen

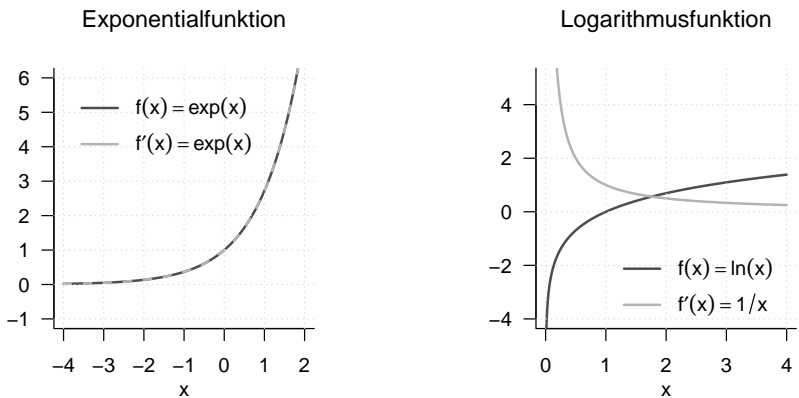


Abbildung 6.3. Ableitungen von Exponentialfunktion und Logarithmusfunktion

Basierend auf der Definition der Ableitung einer univariaten reellwertigen Funktionen lassen sich leicht weitere Ableitungen einer solchen Funktion definieren.

Definition 6.3 (Höhere Ableitungen). Es sei f eine univariate reellwertige Funktion und

$$f^{(1)} := f' \tag{6.5}$$

sei die Ableitung von f . Die k -te Ableitung von f ist rekursiv definiert durch

$$f^{(k)} := (f^{(k-1)})' \text{ für } k \geq 1, \quad (6.6)$$

unter der Annahme, dass $f^{(k-1)}$ differenzierbar ist. Insbesondere ist die *zweite Ableitung* von f definiert durch die Ableitung von f' , also

$$f'' := (f')'. \quad (6.7)$$

•

In Analogie zu oben Gesagtem schreiben wir in Berechnungen auch $\frac{d^2}{dx^2}f(x)$ für den Auftrag, die zweite Ableitung einer Funktion f zu bestimmen. Die nullte Ableitung $f^{(0)}$ von f ist f selbst. Der Tradition und Einfachheit halber schreibt man für $k < 4$ gemäß der Lagrange-Notation meist f' , f'' und f''' anstelle von $f^{(1)}$, $f^{(2)}$ und $f^{(3)}$.

Beispiel

Es sei

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2 \quad (6.8)$$

Dann gilt

$$f^{(1)}(x) = f'(x) = \frac{d}{dx}(x^2) = 2x \quad (6.9)$$

Weiterhin gelten

$$\begin{aligned} f^{(2)}(x) &= (f^{(2-1)})'(x) = (f^{(1)})'(x) = \frac{d}{dx}(2x) = 2, \\ f^{(3)}(x) &= (f^{(3-1)})'(x) = (f^{(2)})'(x) = \frac{d}{dx}(2) = 0, \\ f^{(4)}(x) &= (f^{(4-1)})'(x) = (f^{(3)})'(x) = \frac{d}{dx}(0) = 0. \end{aligned} \quad (6.10)$$

Man berechnet also durchgängig lediglich erste Ableitungen.

Zum Bestimmen der Ableitung einer Funktion sind eine Reihe von Rechenregeln hilfreich, die es erlauben, die Ableitung einer Funktion aus den Ableitungen ihrer Unterfunktionen herzuleiten. Für Beweise der in folgendem Theorem eingeführten Rechenregeln verweisen wir auf die weiterführende Literatur.

Theorem 6.1 (Rechenregeln für Ableitungen). *Für $i = 1, \dots, n$ seien g_i reellwertige univariate differenzierbare Funktionen. Dann gelten folgende Rechenregeln:*

(1) *Summenregel*

$$\text{Für } f(x) := \sum_{i=1}^n g_i(x) \text{ gilt } f'(x) = \sum_{i=1}^n g'_i(x). \quad (6.11)$$

(2) *Produktregel*

$$\text{Für } f(x) := g_1(x)g_2(x) \text{ gilt } f'(x) = g'_1(x)g_2(x) + g_1(x)g'_2(x). \quad (6.12)$$

(3) Quotientenregel

$$\text{Für } f(x) := \frac{g_1(x)}{g_2(x)} \text{ gilt } f'(x) = \frac{g_1'(x)g_2(x) - g_1(x)g_2'(x)}{g_2^2(x)}. \quad (6.13)$$

(4) Kettenregel

$$\text{Für } f(x) := g_1(g_2(x)) \text{ gilt } f'(x) = g_1'(g_2(x))g_2'(x). \quad (6.14)$$

◦

Beispiele

Für ein Beispiel zur Anwendung der Summenregel, sei

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := 4x^3 + 3x^2 \quad (6.15)$$

Dann hat f die Form

$$f(x) = \sum_{i=1}^2 g_i(x) = g_1(x) + g_2(x) \text{ mit } g_1(x) := 4x^3 \text{ und } g_2(x) := 3x^2. \quad (6.16)$$

Es gelten außerdem

$$g_1'(x) = \frac{d}{dx}(4x^3) = 12x^2 \text{ und } g_2'(x) = \frac{d}{dx}(3x^2) = 6x. \quad (6.17)$$

Also gilt mit der Summenregel

$$f'(x) = \sum_{i=1}^2 g_i'(x) = g_1'(x) + g_2'(x) = 12x^2 + 6x. \quad (6.18)$$

Für ein Beispiel zur Anwendung der Kettenregel sei

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \exp(-x^2) \quad (6.19)$$

Dann hat f die Form

$$f(x) = g_2(g_1(x)) \text{ mit } g_2(z) := \exp(z) \text{ und } g_1(x) := -x^2 \quad (6.20)$$

wobei man beachte, dass für das Funktionsargument z in g_2 im Fall von f gerade $z = g_1(x) := -x^2$ eingesetzt wird. Es gelten außerdem

$$g_1'(x) = \frac{d}{dx}(-x^2) = -2x \text{ und } g_2'(z) = \frac{d}{dz}(\exp(z)) = \exp(z) \quad (6.21)$$

Also gilt mit der Kettenregel

$$f'(x) = g_2'(g_1(x))g_1'(x) = \exp(-x^2)(-2x) = -2x \exp(-x^2). \quad (6.22)$$

Man kann sich die resultierende Ableitung der Kettenregel also merken als “Die Ableitung der äußeren Funktion an der Stelle der inneren Funktion mal die Ableitung der inneren Funktion.”

6.2. Analytische Optimierung

Eine wichtige Anwendung der Differentialrechnung ist das Bestimmen von Extremstellen von Funktionen. Dabei geht es im Kern um die Frage, für welche Werte ihrer Definitionsmenge eine Funktion ein Maximum oder ein Minimum annimmt. Bei einfachen Funktionen ist dies analytisch möglich. Die generelle Vorgehensweise dabei ist oft auch unter dem Stichwort “Kurvendiskussion” bekannt. In der Anwendung ist ein analytisches Vorgehen zur Optimierung von Funktionen meist nicht möglich und es werden Computeralgorithmen zur Bestimmung von Extremstellen genutzt. Ein Verständnis dieser Algorithmen setzt allerdings ein Verständnis der Prinzipien der analytischen Optimierung voraus. In diesem Abschnitt geben wir eine Einführung in die analytische Optimierung von univariaten reellwertigen Funktionen. Wir gehen dabei eher informell vor. Einen formaleren Zugang geben wir an späterer Stelle im Kontext der nichtlinearen Optimierung. Wir beginnen damit, die Begriffe der erwähnten Maxima und Minima von univariaten reellwertigen Funktionen zu präzisieren.

Definition 6.4 (Extremstellen und Extremwerte). Es seien $U \subseteq \mathbb{R}$ und $f : U \rightarrow \mathbb{R}$ eine univariate reellwertige Funktion. f hat an der Stelle $x_0 \in U$

- ein *lokales Minimum*, wenn es ein Intervall $I :=]a, b[$ gibt mit $x_0 \in]a, b[$ und

$$f(x_0) \leq f(x) \text{ für alle } x \in I \cap U, \quad (6.23)$$

- ein *globales Minimum*, wenn gilt, dass

$$f(x_0) \leq f(x) \text{ für alle } x \in U, \quad (6.24)$$

- ein *lokales Maximum*, wenn es ein Intervall $I :=]a, b[$ gibt mit $x_0 \in]a, b[$ und

$$f(x_0) \geq f(x) \text{ für alle } x \in I \cap U, \quad (6.25)$$

- ein *lokales Maximum*, wenn gilt, dass

$$f(x_0) \geq f(x) \text{ für alle } x \in U. \quad (6.26)$$

Der Wert $x_0 \in U$ der Definitionsmenge von f heißt entsprechend *lokale* oder *globale Minimalstelle* oder *Maximalstelle*, der Funktionswert $f(x_0) \in \mathbb{R}$ heißt entsprechend *lokales* oder *globales Minimum* oder *Maximum*. Generell heißt der Wert $x_0 \in U$ *Extremstelle* und der Funktionswert $f(x_0) \in \mathbb{R}$ *Extremwert*.

•

Extremstellen von Funktionen werden häufig mit

$$\underset{x \in I \cap U}{\operatorname{argmin}} f(x) \text{ oder } \underset{x \in I \cap U}{\operatorname{argmax}} f(x) \quad (6.27)$$

bezeichnet und Extremwerte von Funktionen werden häufig mit

$$\min_{x \in I \cap U} f(x) \text{ oder } \max_{x \in I \cap U} f(x) \quad (6.28)$$

bezeichnet.

Die analytische Optimierung von univariaten reellwertigen Funktionen basiert auf den sogenannten *notwendigen* und *hinreichenden Bedingungen für Extrema*. Erstere macht eine Aussage über das Verhalten der ersten Ableitung einer Funktion an einer Extremstelle, letztere macht eine Aussage über das Verhalten einer Funktion an einer Stelle, die bestimmten Forderungen an ihre erste und zweite Ableitung genügt.

Theorem 6.2 (Notwendige Bedingung für Extrema). *f sei eine univariate reellwertige Funktion. Dann gilt*

$$x_0 \text{ ist Extremstelle von } f \Rightarrow f'(x_0) = 0. \quad (6.29)$$

◦

Wenn x_0 eine Extremstelle von f ist, dann ist also die erste Ableitung von f in x_0 gleich null. Anstelle eines Beweises überlegen wir uns, dass zum Beispiel an eine lokaler Maximalstelle x_0 von f gilt: links von x_0 steigt f an, rechts von x_0 fällt f ab. In x_0 aber steigt f weder an, noch fällt f ab, es ist also nachvollziehbar, dass $f'(x_0) = 0$ ist.

Theorem 6.3 (Hinreichende Bedingungen für lokale Extrema). *f sei eine zweimal differenzierbare univariate reellwertige Funktion.*

- Wenn für $x_0 \in U \subseteq \mathbb{R}$

$$f'(x_0) = 0 \text{ und } f''(x_0) > 0 \quad (6.30)$$

gilt, dann hat f an der Stelle x_0 ein Minimum.

- Wenn für $x_0 \in U \subseteq \mathbb{R}$

$$f'(x_0) = 0 \text{ und } f''(x_0) < 0 \quad (6.31)$$

gilt, dann hat f an der Stelle x_0 ein Maximum.

◦

Wir verzichten wiederum auf einen Beweis und verdeutlichen uns die Bedingung an dem in Abbildung 6.4 gezeigtem Beispiel. Hier ist offenbar $x_0 = 1$ eine lokale Minimalstelle von $f(x) = (x - 1)^2$. Man erkennt: links von x_0 fällt f ab, rechts von x_0 steigt f an. In x_0 steigt f weder an, noch fällt f ab, also ist $f'(x_0) = 0$. Weiter gilt, dass links und rechts von x_0 und in x_0 die Änderung f'' von f' positiv ist: links von x_0 schwächt sich die Negativität von f' zu 0 ab und rechts von x_0 verstärkt sich die Positivität von f' .

Insbesondere die hinreichende Bedingung für das Vorliegen von Extremstellen legt folgendes *Standardverfahren* zur Bestimmung von lokalen Extremstellen nahe.

Theorem 6.4 (Standardverfahren der analytischen Optimierung). *f sei eine univariate reellwertige Funktion. Lokale Extremstellen von f können mit folgendem Standardverfahren der analytischen Optimierung identifiziert werden:*

1. Berechnen der ersten und zweiten Ableitung von f .
2. Bestimmen von Nullstellen x^* von f' durch Auflösen von $f'(x^*) = 0$ nach x^* . Die Nullstellen von f' sind dann Kandidaten für Extremstellen von f .
3. Evaluation von $f''(x^*)$: Wenn $f''(x^*) > 0$ ist, dann ist x^* lokale Minimumstelle von f ; wenn $f''(x^*) < 0$ ist, dann ist x^* lokale Maximumstelle von f ; wenn $f''(x^*) = 0$ ist, dann ist x^* keine Extremstelle von f .

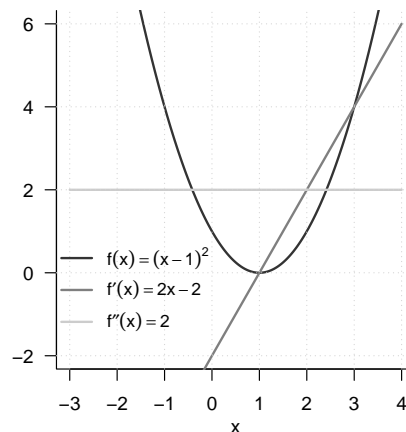


Abbildung 6.4. Analytische Optimierung von $f(x) := (x - 1)^2$

◦

Anstelle eines Beweises betrachten wir beispielhaft die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := (x - 1)^2. \quad (6.32)$$

aus Abbildung 6.4. Die erste Ableitung von f ergibt sich mit der Kettenregel zu

$$f'(x) = \frac{d}{dx} ((x - 1)^2) = 2(x - 1) \cdot \frac{d}{dx} (x - 1) = 2x - 2. \quad (6.33)$$

Die zweite Ableitung von f ergibt sich zu

$$f''(x) = \frac{d}{dx} f'(x) = \frac{d}{dx} (2x - 2) = 2 > 0 \text{ für alle } x \in \mathbb{R}. \quad (6.34)$$

Auflösen von $f'(x^*) = 0$ nach x^* ergibt

$$f'(x^*) = 0 \Leftrightarrow 2x^* - 2 = 0 \Leftrightarrow 2x^* = 2 \Leftrightarrow x^* = 1. \quad (6.35)$$

$x^* = 1$ ist folglich eine Minimalstelle von f mit zugehörigen Minimalwert $f(1) = 0$.

6.3. Differentialrechnung multivariater reellwertiger Funktionen

Wir erinnern zunächst an den Begriff der multivariaten reellwertigen Funktion.

Definition 6.5 (Multivariate reellwertige Funktion). Eine Funktion der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n) \quad (6.36)$$

heißt *multivariate reellwertige Funktion*.

•

Die Argumente multivariater reellwertiger Funktionen sind also reelle n -Tupel der Form $x := (x_1, \dots, x_n)$ während ihre Funktionswerte reelle Zahlen sind. Ein Beispiel für eine multivariate reellwertige für $n := 2$ ist

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2 \quad (6.37)$$

Wir visualisieren diese Funktion in Abbildung 6.5. Dabei zeigt die rechte Abbildung eine Darstellung mithilfe sogenannter *Isokonturen*, also Linien im Definitionsbereich der Funktion, für die die Funktion identische Werte annimmt. Die entsprechenden Werte sind für ausgewählte Isokonturen in der Abbildung vermerkt.

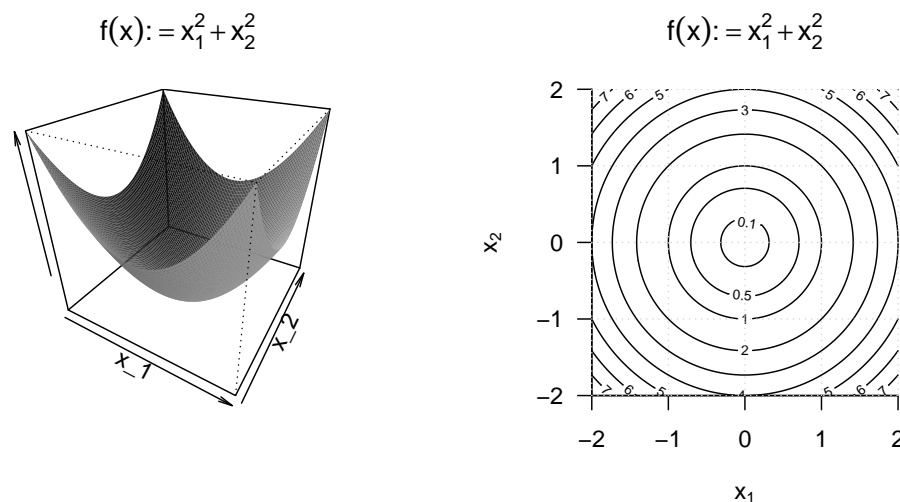


Abbildung 6.5. Visualisierungen einer bivariaten Funktion.

Wir wollen nun beginnen, die Begriffe der Differenzierbarkeit und der Ableitung univariater reellwertiger Funktionen auf den Fall multivariater reellwertiger Funktionen zu erweitern. Dazu führen wir zunächst die Begriffe der *partiellen Differenzierbarkeit* und der *partiellen Ableitung* ein.

Definition 6.6 (Partielle Differenzierbarkeit und partielle Ableitung). Es sei $D \subseteq \mathbb{R}^n$ eine Menge und

$$f : D \rightarrow \mathbb{R}, x \mapsto f(x) \quad (6.38)$$

eine multivariate reellwertige Funktion. f heißt in $a \in D$ nach x_i *partiell differenzierbar*, wenn der Grenzwert

$$\frac{\partial}{\partial x_i} f(x) := \lim_{h \rightarrow 0} \frac{f(a + h e_i) - f(a)}{h} \quad (6.39)$$

existiert. $\frac{\partial}{\partial x_i} f(a)$ heißt dann die *partielle Ableitung von f nach x_i an der Stelle a* . Wenn f für alle $x \in D$, nach x_i partiell differenzierbar ist, dann heißt f nach x_i *partiell differenzierbar* und die Funktion

$$\frac{\partial}{\partial x_i} f : D \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_i} f(x) \quad (6.40)$$

heißt *partielle Ableitung von f nach x_i* . f heißt *partiell differenzierbar in $x \in D$* , wenn f für alle $i = 1, \dots, n$ in $x \in D$ nach x_i partiell differenzierbar ist, und

f heißt *partiell differenzierbar*, wenn f für alle $i = 1, \dots, n$ in allen $x \in D$ nach x_i partiell differenzierbar ist.

•

In Definition 6.6 bezeichnet $e_i \in \mathbb{R}^n$ bezeichnet den i ten kanonischen Einheitsvektor, für den gilt, dass $e_{ij} = 1$ für $i = j$ und $e_{ij} = 0$ für $i \neq j$ mit $j = 1, \dots, n$ (vgl. Definition 8.14). In Analogie und Verallgemeinerung zum Newtonschen Differenzquotient misst der hier auftretende Differenzquotient

$$\frac{f(x + he_i) - f(x)}{h} \quad (6.41)$$

die Änderung $f(x + he_i) - f(x)$ von f pro Strecke h in Richtung e_i . Wir visualisieren die Bestandteile dieses Quotienten im Falle einer bivariaten Funktion in Abbildung 6.6.

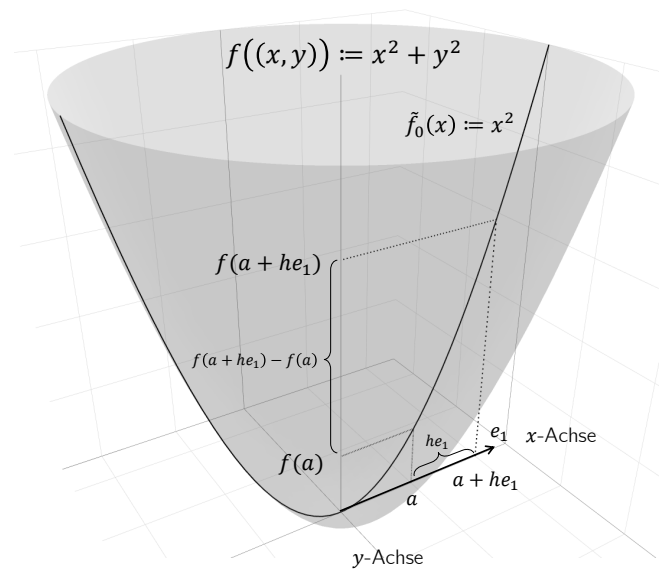


Abbildung 6.6. Partiieller Newtonscher Differenzquotient

Für $h \rightarrow 0$ misst der Differenzquotient entsprechend die *Änderungsrate* von f in x in Richtung e_i . Wie bei der Betrachtung von Ableitungen gilt, dass $\frac{\partial}{\partial x_i} f(x)$ eine Zahl, $\frac{\partial}{\partial x_i} f$ dagegen eine Funktion ist. Praktisch berechnet man $\frac{\partial}{\partial x_i} f$ als die (einfache) Ableitung

$$\frac{d}{dx_i} \tilde{f}_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n}(x_i) \quad (6.42)$$

der univariaten reellwertigen Funktion

$$\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}, x_i \mapsto \tilde{f}_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n}(x_i) := f(x_1, \dots, x_i, \dots, x_n). \quad (6.43)$$

Man betrachtet für die i te partielle Ableitung also alle x_j mit $j \neq i$ als Konstanten und ist auf das gewohnte Berechnen von Ableitungen von univariaten reellwertigen Funktionen geführt. Wir wollen das Vorgehen zum Berechnen von partiellen Ableitungen an einem ersten Beispiel verdeutlichen.

Beispiel (1)

Wir betrachten die Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \quad (6.44)$$

Weil die Definitionsmenge dieser Funktion zweidimensional ist, kann man zwei partielle Ableitungen berechnen

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) \text{ und } \frac{\partial}{\partial x_2} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x). \quad (6.45)$$

Um die erste dieser partiellen Ableitungen zu berechnen, betrachtet man die Funktion

$$f_{x_2} : \mathbb{R} \rightarrow \mathbb{R}, x_1 \mapsto f_{x_2}(x_1) := x_1^2 + x_2^2, \quad (6.46)$$

wobei x_2 hier die Rolle einer Konstanten einnimmt. Um explizit zu machen, dass x_2 kein Argument der Funktion ist, die Funktion aber weiterhin von x_2 abhängt haben wir die Subskriptnotation $f_{x_2}(x_1)$ verwendet. Um nun die partielle Ableitung zu berechnen, berechnen wir die (einfache) Ableitung von f_{x_2} ,

$$f'_{x_2}(x) = 2x_1. \quad (6.47)$$

Es ergibt sich also

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) = \frac{\partial}{\partial x_1} (x_1^2 + x_2^2) = f'_{x_2}(x) = 2x_1. \quad (6.48)$$

Analog gilt mit der entsprechenden Formulierung von f_{x_1} , dass

$$\frac{\partial}{\partial x_2} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x) = \frac{\partial}{\partial x_2} (x_1^2 + x_2^2) = f'_{x_1}(x) = 2x_2. \quad (6.49)$$

Wie bei der Ableitung einer univariaten reellwertigen Funktion ist es auch für eine multivariate reellwertige Funktion möglich, rekursiv eine höhere Ableitung zu definieren.

Definition 6.7 (Zweite partielle Ableitungen). $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion und $\frac{\partial}{\partial x_i} f$ sei die partielle Ableitung von f nach x_i . Dann ist die zweite partielle Ableitung von f nach x_i und x_j definiert als

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) := \frac{\partial}{\partial x_j} \left(\frac{\partial}{\partial x_i} f \right). \quad (6.50)$$

•

Man beachte, dass es zu jeder partiellen Ableitung $\frac{\partial}{\partial x_i} f$ für $i = 1, \dots, n$ insgesamt n zweite partiellen Ableitungen $\frac{\partial^2}{\partial x_j \partial x_i} f$ für $j = 1, \dots, n$ gibt. Die so resultierenden n^2 zweiten partiellen Ableitungen sind jedoch nicht alle verschieden. Dies ist eine wesentliche Aussage des *Satzes von Schwarz*

Theorem 6.5 (Satz von Schwarz). $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine partiell differenzierbare multivariate reellwertige Funktion. Dann gilt

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \text{ für alle } 1 \leq i, j \leq n. \quad (6.51)$$

◦

Für einen Beweis verweisen wir auf die weiterführende Literatur. Der Satz von Schwarz besagt insbesondere also auch, dass bei Bildung der zweiten partiellen Ableitungen die Reihenfolge des partiellen Ableitens irrelevant ist. Das Theorem erleichtert auf diese Weise die Berechnung von zweiten partiellen Ableitungen und hilft zudem, analytische Fehler bei der Berechnung zweiter partieller Ableitungen aufzudecken. Wir verdeutlichen dies in Fortführung obigen Beispiels.

Beispiel (1)

Wir wollen die partiellen Ableitungen zweiter Ordnung der Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \quad (6.52)$$

berechnen. Mit den Ergebnissen für die partiellen Ableitungen erster Ordnung dieser Funktion ergibt sich

$$\begin{aligned} \frac{\partial^2}{\partial x_1 x_1} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_1} (2x_1) = 2 \\ \frac{\partial^2}{\partial x_1 x_2} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_1} (2x_2) = 0 \\ \frac{\partial^2}{\partial x_2 x_1} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_2} (2x_1) = 0 \\ \frac{\partial^2}{\partial x_2 x_2} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_2} (2x_2) = 2 \end{aligned} \quad (6.53)$$

Offenbar gilt

$$\frac{\partial^2}{\partial x_1 x_2} f(x) = \frac{\partial^2}{\partial x_2 x_1} f(x). \quad (6.54)$$

Beispiel (2)

Als weiteres Beispiel wollen wir die partiellen Ableitungen erster und zweiter Ordnung der Funktion

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}. \quad (6.55)$$

berechnen. Mit den Rechenregeln für Ableitungen ergibt sich für die partiellen Ableitungen erster Ordnung

$$\begin{aligned} \frac{\partial}{\partial x_1} f(x) &= \frac{\partial}{\partial x_1} (x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}) = 2x_1 + x_2, \\ \frac{\partial}{\partial x_2} f(x) &= \frac{\partial}{\partial x_2} (x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}) = x_1 + \sqrt{x_3}, \\ \frac{\partial}{\partial x_3} f(x) &= \frac{\partial}{\partial x_3} (x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}) = \frac{x_2}{2\sqrt{x_3}}. \end{aligned} \quad (6.56)$$

Für die zweiten partiellen Ableitungen hinsichtlich x_1 ergibt sich

$$\begin{aligned} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_1} (2x_1 + x_2) = 2, \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_2} (2x_1 + x_2) = 1, \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_3} (2x_1 + x_2) = 0. \end{aligned} \quad (6.57)$$

Für die zweiten partiellen Ableitungen hinsichtlich x_2 ergibt sich

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_1} (x_1 + \sqrt{x_3}) = 1, \\ \frac{\partial^2}{\partial x_2 \partial x_2} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_2} (x_1 + \sqrt{x_3}) = 0, \\ \frac{\partial^2}{\partial x_3 \partial x_2} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_3} (x_1 + \sqrt{x_3}) = \frac{1}{2\sqrt{x_3}}.\end{aligned}\tag{6.58}$$

Beispiel (2) Für die zweiten partiellen Ableitungen hinsichtlich x_3 ergibt sich

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_1} \left(\frac{x_2}{2} \sqrt{x_3} \right) = 0, \\ \frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_2} \left(\frac{x_2}{2\sqrt{x_3}} \right) = \frac{1}{2\sqrt{x_3}}, \\ \frac{\partial^2}{\partial x_3 \partial x_3} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_3} \left(x_2 \frac{1}{2} x_3^{-\frac{1}{2}} \right) = -\frac{1}{4} x_2 x_3^{-\frac{3}{2}}.\end{aligned}\tag{6.59}$$

Weiterhin erkennt man, dass die Reihenfolge der partiellen Ableitungen irrelevant ist, denn es gilt

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial^2}{\partial x_2 \partial x_1} f(x) = 1, \\ \frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_1} f(x) = 0, \\ \frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_2} f(x) = \frac{1}{2\sqrt{x_3}}.\end{aligned}\tag{6.60}$$

Wie oben gesehen gibt es für eine multivariate reellwertige Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ insgesamt n erste partielle Ableitungen und n^2 zweite partielle Ableitungen. Diese werden im *Gradienten* und der *Hesse-Matrix* einer multivariaten reellwertigen Funktion zusammengefasst.

Definition 6.8 (Gradient). $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann ist der *Gradient* $\nabla f(x)$ von f an der Stelle $x \in \mathbb{R}^n$ definiert als

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^n.\tag{6.61}$$

•

Man beachte, dass Gradienten multivariate vektorwertige Funktionen der

$$\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n, x \mapsto \nabla f(x)\tag{6.62}$$

sind. Für $n = 1$ gilt $\nabla f(x) = f'(x)$. Eine wichtige Eigenschaft des Gradienten ist, dass $-\nabla f(x)$ die Richtung des steilsten Abstiegs von f in \mathbb{R}^n anzeigt. Diese Einsicht ist aber

nicht trivial und soll an späterer Stelle vertieft werden. Als Beispiele betrachten wir die Gradienten der oben analysierten Funktionen

Beispiel (1)

Für die in Beispiel (1) betrachtete Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ gilt

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \in \mathbb{R}^2. \quad (6.63)$$

In Abbildung 6.7 visualisieren wir ausgewählte Werte dieses Gradienten für

$$x := \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix}, \quad , \quad x := \begin{pmatrix} -0.3 \\ 0.1 \end{pmatrix}, \quad x := \begin{pmatrix} -0.5 \\ -0.4 \end{pmatrix}, \quad x := \begin{pmatrix} 0.1 \\ -1.0 \end{pmatrix}$$

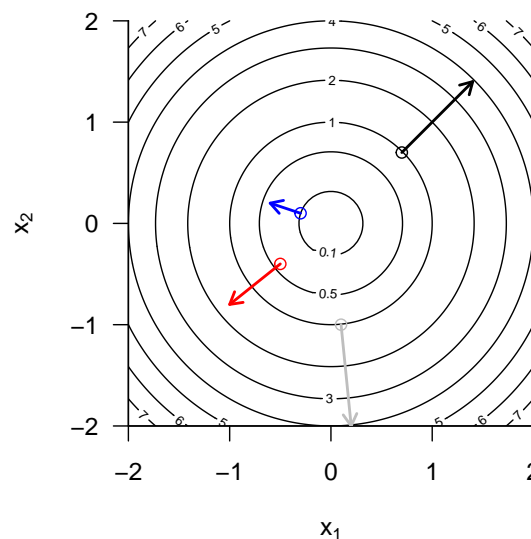


Abbildung 6.7. Exemplarische Gradientenwerte der bivariaten Funktion $f(x) = x_1^2 + x_2^2$.

Beispiel (2)

Für die in Beispiel (2) betrachtete Funktion $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ gilt

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \frac{\partial}{\partial x_3} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + \sqrt{x_3} \\ \frac{x_2}{2\sqrt{x_3}} \end{pmatrix} \in \mathbb{R}^3. \quad (6.64)$$

Schließlich widmen wir uns der Zusammenfassung der zweiten partiellen Ableitungen einer multivariaten reellwertigen Funktion in der *Hesse-Matrix*.

Definition 6.9 (Hesse-Matrix). $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann ist die *Hesse-Matrix* $\nabla^2 f(x)$ von f an der Stelle $x \in \mathbb{R}^n$ definiert als

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^{n \times n}. \quad (6.65)$$

•

Man beachte, dass Hesse-Matrizen multivariate matrixwertige Abbildungen der Form

$$\nabla^2 f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}, x \mapsto \nabla^2 f(x) \quad (6.66)$$

sind. Für $n = 1$ gilt $\nabla^2 f(x) = f''(x)$. Weiterhin folgt aus

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x) \text{ für } 1 \leq i, j \leq n \quad (6.67)$$

dass die Hesse-Matrix symmetrisch ist, dass also

$$(\nabla^2 f(x))^T = \nabla^2 f(x) \quad (6.68)$$

gilt.

Beispiel (1)

Für die in Beispiel (1) betrachtete Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ gilt

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \in \mathbb{R}^{2 \times 2} \quad (6.69)$$

Die Hesse-Matrix dieser Funktion ist also eine konstante Funktion, die nicht von x abhängt.

Beispiel (2)

Für die in Beispiel (2) betrachtete Funktion $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ gilt

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_3} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \frac{\partial^2}{\partial x_2 \partial x_3} f(x) \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(x) & \frac{\partial^2}{\partial x_3 \partial x_2} f(x) & \frac{\partial^2}{\partial x_3 \partial x_3} f(x) \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & \frac{1}{2\sqrt{3}} \\ 0 & \frac{1}{2\sqrt{3}} & -\frac{1}{4}x_2x_3^{-3/2} \end{pmatrix}. \quad (6.70)$$

Im Gegensatz zu Beispiel (1) ist die Hesse-Matrix der hier betrachteten Funktion keine konstante Funktion und ihr Wert hängt vom Wert des Funktionsarguments $x \in \mathbb{R}^3$ ab.

6.4. Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs der Ableitung $f'(a)$ einer Funktion f an einer Stelle a wieder.
2. Geben Sie die Definition des Begriffs der Ableitung f' einer Funktion f .
3. Erläutern Sie die Symbole $f'(x)$, $\dot{f}(x)$, $\frac{df(x)}{dx}$, und $\frac{d}{dx}f(x)$.
4. Geben Sie die Definition des Begriffs der zweiten Ableitung f'' einer Funktion f wieder.
5. Geben Sie die Summenregel für Ableitungen wieder.
6. Geben Sie die Produktregel für Ableitungen wieder.
7. Geben Sie die Quotientenregel für Ableitungen wieder.
8. Geben Sie die Kettenregel für Ableitungen wieder.
9. Bestimmen Sie die erste Ableitung der Funktion $f(x) := 3x^2 + \exp(-x^2) - x \ln(x)$.
10. Bestimmen Sie die erste Ableitung der Funktion $f(x) := \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2$ für $\mu \in \mathbb{R}$.
11. Geben Sie die Definition der Begriffe des globalen und lokalen Maximums/Minimums einer univariaten reellwertigen Funktion wieder.
12. Geben Sie die notwendige Bedingung für ein Extremum einer Funktion wieder.
13. Geben Sie die hinreichende Bedingung für ein lokales Extremum einer Funktion wieder.
14. Geben Sie das Standardverfahren der analytischen Optimierung wieder.
15. Bestimmen Sie einen Extremwert von $f(x) := \exp(-\frac{1}{2}(x - \mu)^2)$ für $\mu \in \mathbb{R}$.
16. Berechnen Sie die partiellen Ableitungen der Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right). \quad (6.71)$$

17. Berechnen Sie die zweiten partiellen Ableitungen obiger Funktion f .
18. Geben Sie den Satz von Schwarz wieder.
19. Geben Sie die Definition des Gradienten einer multivariaten reellwertigen Funktion wieder.
20. Geben Sie den Gradienten der Funktion in Gleichung 6.71 an und werten Sie ihn in $x = (1, 2)^T$ aus.
21. Geben Sie die Definition der Hesse-Matrix einer multivariaten reellwertigen Funktion wieder.
22. Geben Sie die Hesse-Matrix der Funktion in Gleichung 6.71 an und werten Sie sie in $x = (1, 2)^T$ aus.

7. Integralrechnung

Dieses Kapitel gibt einen Überblick über zentrale Begriffe der Integralrechnung. Das Hauptaugenmerk liegt dabei durchgängig auf der Klärung von Begrifflichkeiten, ihrer mathematischen Symbolik und der durch sie vermittelten Intuition und weniger auf der konkreten Berechnung von Integralen.

7.1. Unbestimmte Integrale

Wir beginnen mit der Definition des unbestimmten Integrals und dem Begriff der Stammfunktion.

Definition 7.1 (Unbestimmtes Integral und Stammfunktion). Für ein Intervall $I \subseteq \mathbb{R}$ sei $f : I \rightarrow \mathbb{R}$ eine univariate reellwertige Funktion. Dann heißt eine differenzierbare Funktion $F : I \rightarrow \mathbb{R}$ mit der Eigenschaft

$$F' = f \tag{7.1}$$

Stammfunktion von f . Ist F eine Stammfunktion von f , dann heißt

$$\int f(x) dx := F + c \text{ mit } c \in \mathbb{R} \tag{7.2}$$

unbestimmtes Integral der Funktion f . Das unbestimmte Integral einer Funktion bezeichnet damit die Menge aller Stammfunktionen einer Funktion.

•

Obige Definition besagt, dass die Ableitung der Stammfunktion einer Funktion f eben f ist. Das unbestimmte Integral einer Funktion f ist darüber hinaus die Menge *aller* durch Addition verschiedener Konstanten $c \in \mathbb{R}$ gegebenen Stammfunktionen von f . Eine solche Konstante $c \in \mathbb{R}$ heißt auch *Integrationskonstante*; es gilt natürlich $\frac{d}{dx}c = 0$. Das Symbol $\int f(x) dx$ ist als $F + c$ definiert. $f(x)$ wird in diesem Ausdruck *Integrand* genannt. \int und dx haben keine eigentliche Bedeutung, sondern sind reine Symbole.

Für die in vorherigen Abschnitten eingeführten elementaren Funktionen ergeben sich die in Tabelle 7.1 aufgelisteten Stammfunktionen. Man überzeugt sich davon durch Ableiten der jeweiligen Stammfunktion mithilfe der Rechenregeln der Differentialrechnung. Die uneigentlichen Integrale dieser elementaren Funktionen ergeben sich dann direkt aus diesen Stammfunktionen durch Addition einer Integrationskonstanten.

Tabelle 7.1. Stammfunktionen elementarer Funktionen

Name	Definition	Stammfunktion
Polynomfunktion	$f(x) := \sum_{i=0}^n a_i x^i$	$F(x) = \sum_{i=0}^n \frac{a_i}{i+1} x^{i+1}$
Konstante Funktion	$f(x) := a$	$F(x) = ax$
Identitätsfunktion	$f(x) := x$	$F(x) = \frac{1}{2}x^2$

Name	Definition	Stammfunktion
Linear-affine Funktion	$f(x) := ax + b$	$F(x) = \frac{1}{2}ax^2 + bx$
Quadratfunktion	$f(x) := x^2$	$F(x) = \frac{1}{3}x^3$
Exponentialfunktion	$f(x) := \exp(x)$	$F(x) = \exp(x)$
Logarithmusfunktion	$f(x) := \ln(x)$	$F(x) = x \ln x - x$

Die in nachfolgendem Theorem zusammengestellten Rechenregeln sind oft hilfreich, um Stammfunktionen von Funktionen zu bestimmen, die sich aus Funktionen mit bekannten Stammfunktionen zusammensetzen.

Theorem 7.1 (Rechenregeln für Stammfunktionen). *f und g seien univariate reellwertige Funktion, die Stammfunktionen besitzen, und g sei invertierbar. Dann gelten folgende Rechenregeln für die Bestimmung von Stammfunktionen*

(1) *Summenregel*

$$\int af(x) + bg(x) dx = a \int f(x) dx + b \int g(x) dx \text{ für } a, b \in \mathbb{R} \quad (7.3)$$

(2) *Partielle Integration*

$$\int f'(x)g(x) dx = f(x)g(x) - \int f(x)g'(x) dx \quad (7.4)$$

(3) *Substitutionsregel*

$$\int f(g(x))g'(x) dx = \int f(t) dt \text{ mit } t = g(x) \quad (7.5)$$

◦

Beweis. Für einen Beweis der Summenregel verweisen wir auf die weiterführende Literatur. Die Rechenregel der partiellen Integration ergibt sich durch Integration der Produktregel der Differentiation. Wir erinnern uns, dass gilt

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x). \quad (7.6)$$

Integration beider Seiten der Gleichung und Berücksichtigung der Summenregel für Stammfunktionen ergibt dann

$$\begin{aligned} \int (f(x)g(x))' dx &= \int f'(x)g(x) + f(x)g'(x) dx \\ &\Leftrightarrow f(x)g(x) = \int f'(x)g(x) dx + \int f(x)g'(x) dx \\ &\Leftrightarrow \int f'(x)g(x) dx = f(x)g(x) - \int f(x)g'(x) dx. \end{aligned} \quad (7.7)$$

Die Substitutionsregel ergibt sich für $F' = f$ durch Anwendung der Kettenregel der Differentiation auf die verkettete Funktion $F(g)$. Speziell gilt zunächst

$$(F(g(x)))' = F'(g(x))g'(x) = f(g(x))g'(x). \quad (7.8)$$

Integration beider Seiten der Gleichung

$$(F(g(x)))' = f(g(x))g'(x) \quad (7.9)$$

ergibt dann

$$\begin{aligned} \int (F(g(x)))' dx &= \int f(g(x))g'(x) dx \\ &\Leftrightarrow F(g(x)) + c = \int f(g(x))g'(x) dx \\ &\Leftrightarrow \int f(g(x))g'(x) dx = \int f(t) dt \text{ mit } t := g(x). \end{aligned} \quad (7.10)$$

Dabei ist die rechte Seite der letzten obigen Gleichung zu verstehen als $F(g(x)) + c$, also als Stammfunktion von f evaluiert an der Stelle $t := g(x)$. Das dt ist nicht durch $dg(x)$ zu ersetzen, sondern rein notationeller Natur. \square

Unbestimmte Integrale nehmen in der Lösung von Differentialgleichungen einen zentralen Platz ein. Naheliegender ist aber zunächst die Anwendung unbestimmter Integrale im Kontext der Auswertung *bestimmter Integrale*, wie im nächsten Abschnitt eingeführt.

7.2. Bestimmte Integrale

Anschaulich entspricht ein bestimmtes Integral der vorzeichenbehafteten und auf ein Intervall $[a, b]$ beschränkten Fläche zwischen dem Graphen einer Funktion f und der x -Achse (vgl. Abbildung 7.1). *Vorzeichenbehaftet* heißt dabei, dass Flächen zwischen der x -Achse und positiven Werten von f positiv zur Fläche beitragen, Flächen zwischen der x und negativen Werten von f dagegen negativ. So ergeben sich zum Beispiel der Wert des in Abbildung 7.1 A gezeigten bestimmten Integral zu 0.68, der Wert des in Abbildung 7.1 B gezeigten bestimmten Integrals zu 0.95 (die eingezeichnete Fläche ist offensichtlich größer als in Abbildung 7.1 A) und der Wert des in Abbildung 7.1 C gezeigten bestimmten Integrals zu 0 (die eingezeichneten positiven und negativen Flächen gleichen sich genau aus). Letzteres Beispiel legt auch die Interpretation des Integrals als Durchschnittswert einer Funktion f über einem Intervall $[a, b]$ nahe.

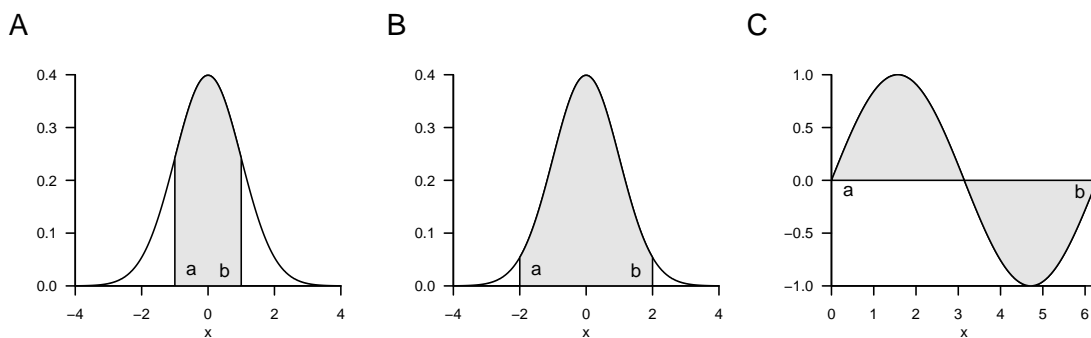


Abbildung 7.1. Beispiele bestimmter Integrale

Um den Begriff des *bestimmten Integrals* im Sinne des *Riemannschen Integrals* einführen zu können, müssen wir zunächst etwas Vorarbeit leisten. Wir beginnen damit, einen Begriff für die Aufteilung eines Intervalls in kleinere Abschnitte einzuführen.

Definition 7.2 (Zerlegung eines Intervalls und Feinheit). Es sei $[a, b] \subset \mathbb{R}$ ein Intervall und $x_0, x_1, x_2, \dots, x_n \in [a, b]$ eine Menge von Punkten mit

$$a =: x_0 < x_1 < x_2 < \dots < x_n := b \quad (7.11)$$

und

$$\Delta x_i := x_i - x_{i-1} \text{ für } i = 1, \dots, n. \quad (7.12)$$

Dann heißt die Menge

$$Z := \{[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]\} \quad (7.13)$$

der durch $x_0, x_1, x_2, \dots, x_n$ definierten Teilintervalle von $[a, b]$ eine *Zerlegung von $[a, b]$* . Weiterhin heißt

$$Z_{\max} := \max_{i \in n} \Delta x_i, \quad (7.14)$$

also die größte der Teilintervalllängen Δx_i , die *Feinheit von Z* .

•

Anschaulich ist Δx_i die Breite der Rechtecke in Abbildung 7.2, wie wir in der Folge sehen werden. Mithilfe der Begriffe der Zerlegung eines Intervalls können wir nun den Begriff der *Riemannschen Summen* einführen.

Definition 7.3 (Riemannsche Summen). $f : [a, b] \rightarrow \mathbb{R}$ sei eine beschränkte Funktion auf $[a, b]$, d.h. $|f(x)| < c$ für $0 < c < \infty$ und alle $x \in [a, b]$, Z sei eine Zerlegung von $[a, b]$ mit Teilintervalllängen Δx_i für $i = 1, \dots, n$. Weiterhin sei ξ_i für $i = 1, \dots, n$ ein beliebiger Punkt im Teilintervall $[x_{i-1}, x_i]$ der Zerlegung Z . Dann heißt

$$R(Z) := \sum_{i=1}^n f(\xi_i) \Delta x_i \quad (7.15)$$

Riemannsche Summe von f auf $[a, b]$ bezüglich der Zerlegung Z .

•

Wählt man zum Beispiel in der Riemannschen Summe in jedem Teilintervall das Maximum von f , so ergibt sich die sogenannte *Riemannsche Obersumme*,

$$R_o(Z) := \sum_{i=1}^n \left(\max_{[x_{i-1}, x_i]} f(\xi_i) \right) \Delta x_i. \quad (7.16)$$

Wählt man dagegen in jedem Teilintervall dagegen das Minimum von f , so ergibt sich dies sogenannte *Riemannsche Untersumme*.

$$R_u(Z) := \sum_{i=1}^n \left(\min_{[x_{i-1}, x_i]} f(\xi_i) \right) \Delta x_i. \quad (7.17)$$

Abbildung 7.2 verdeutlicht die Definition dieser Riemannschen Summen: die dunkelgrauen Rechtecke haben jeweils die Fläche $[x_{i-1}, x_i] \cdot \min_{[x_{i-1}, x_i]} f(\xi)$ und bilden damit die Summenterme in der Riemannschen Untersumme

$$R_u(Z) := \sum_{i=1}^4 \left(\min_{[x_{i-1}, x_i]} f(\xi_i) \right) \cdot \Delta x_i. \quad (7.18)$$

Die vertikale Kombination aus dunkelgrauen und hellgrauen Rechtecken hat jeweils die Fläche $[x_{i-1}, x_i] \cdot \max_{[x_{i-1}, x_i]} f(\xi)$ und bilden damit die Summenterme in der Riemannschen Obersumme

$$R_o(Z) := \sum_{i=1}^4 \left(\max_{[x_{i-1}, x_i]} f(\xi_i) \right) \cdot \Delta x_i. \quad (7.19)$$

Stellt man sich nun vor, dass man Δx_i für alle $i = 1, \dots, n$ gegen Null gehen lässt, verkleinert man die Feinheit der Zerlegung Z also immer weiter, so werden sich die Werte von $\min_{[x_{i-1}, x_i]} f(\xi_i)$ und $\max_{[x_{i-1}, x_i]} f(\xi_i)$ und damit auch die Werte von $R_u(Z)$ und $R_o(Z)$ immer weiter annähern. Diesen Grenzprozess macht man sich in der Definition des Riemannschen Integrals zunutze.

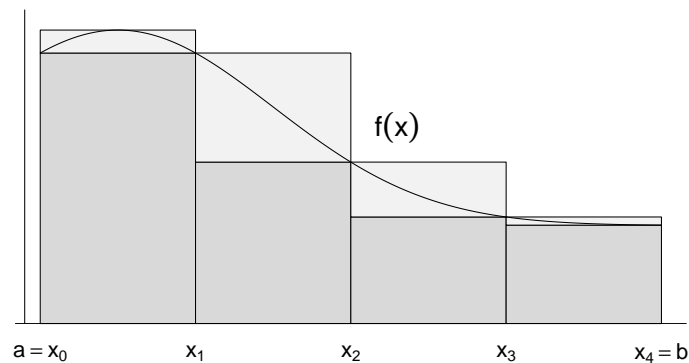


Abbildung 7.2. Riemannsche Summen

Definition 7.4 (Bestimmtes Riemannsches Integral). $f : [a, b] \rightarrow \mathbb{R}$ sei eine beschränkte reellwertige Funktion auf $[a, b]$. Weiterhin sei für Z_k mit $k = 1, 2, 3, \dots$ eine Folge von Zerlegungen von $[a, b]$ mit zugehörigen Feinheit $Z_{\max, k}$. Wenn für jede Folge von Zerlegungen Z_1, Z_2, \dots mit $|Z_{\max, k}| \rightarrow 0$ für $k \rightarrow \infty$ und für beliebig gewählte Punkte ξ_{ki} mit $i = 1, \dots, n$ im Teilintervall $[x_{k,i-1}, x_{k,i}]$ der Zerlegung Z_k gilt, dass die Folge der zugehörigen Riemannschen Summen $R(Z_1), R(Z_2), \dots$ gegen den gleichen Grenzwert strebt, dann heißt f auf $[a, b]$ *integrierbar*. Der entsprechende Grenzwert der Folge von Riemannschen Summen wird *bestimmtes Riemannsches Integral* genannt und mit

$$\int_a^b f(x) dx := \lim_{k \rightarrow \infty} R(Z_k) \text{ für } |Z_{\max, k}| \rightarrow 0 \quad (7.20)$$

bezeichnet. Die Werte a und b bezeichnet man in diesem Kontext als *untere* und *obere* Integrationsgrenzen, respektive, $f(x)$ als *Integrand* und x als *Integrationsvariable*.

•

Die Riemannsche Integrierbarkeit einer Funktion und der Wert eines bestimmten Riemannschen Integrals sind also im Sinne einer Grenzwertbildung definiert. Die Theorie der Riemannschen Integrale lässt sich allerdings um die Hauptsätze der Differential- und Integralrechnung erweitern, so dass zur konkreten Berechnung eines bestimmten Integrals die Bildung von Zerlegungen und die Bestimmung eines Grenzwertes nur selten nötig ist. Der Einfachheit halber verzichten wir in der Folge auf die Bezeichnungen *Riemannsche* und sprechen einfach von *bestimmten Integralen*.

Ein erster Schritt zur Vereinfachung der Berechnung von bestimmten Integralen ist das Feststellen folgender Rechenregeln, für deren Beweis wir auf die weiterführende Literatur verweisen.

Theorem 7.2 (Rechenregeln für bestimmte Integrale). *Es seien f und g integrierbare Funktionen auf $[a, b]$. Dann gelten folgende Rechenregeln.*

(1) *Linearität.* Für $c_1, c_2 \in \mathbb{R}$ gilt

$$\int_a^b (c_1 f(x) + c_2 g(x)) dx = c_1 \int_a^b f(x) dx + c_2 \int_a^b g(x) dx. \quad (7.21)$$

(2) *Additivität. Für $a < c < b$ gilt*

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (7.22)$$

(3) *Vorzeichenwechsel bei Umkehrung der Integralgrenzen*

$$\int_a^b f(x) dx = - \int_b^a f(x) dx. \quad (7.23)$$

(4) *Unabhängigkeit von der Wahl der Integrationsvariable*

$$\int_a^b f(x) dx = \int_a^b f(y) dy. \quad (7.24)$$

(5) *Unabhängigkeit des Integrals von Art des Intervalls. Es gilt*

$$\int_a^b f(x) dx = \int_{[a,b[} f(x) dx = \int_{[a,b]} f(x) dx = \int_{]a,b]} f(x) dx = \int_{]a,b]} f(x) dx. \quad (7.25)$$

wobei \int_I das bestimmte Integral von f auf dem Intervall $I \subseteq \mathbb{R}$ bezeichnet.

◦

Eine graphische Darstellung der Rechenregel der Additivität findet sich in Abbildung 7.3. Die Summe der durch die bestimmten Integrale gegebenen Flächen $\int_a^c f(x) dx$ und $\int_c^b f(x) dx$ mit $a < c < b$ ergibt sich dabei zur Fläche von $\int_a^b f(x) dx$.

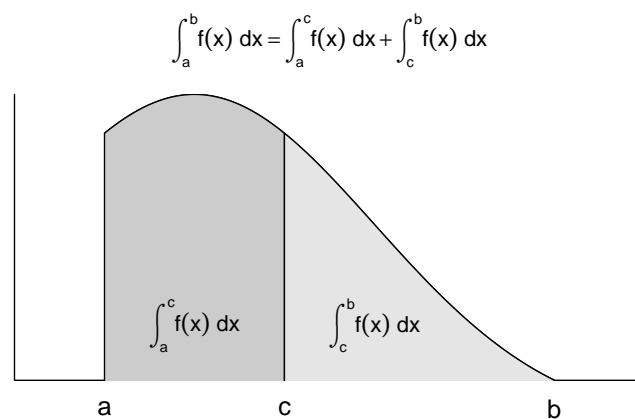


Abbildung 7.3. Additivität bestimmter Integrale

Die in der Nachfolge vermerkten Hauptsätze der Differential- und Integralrechnung schließlich, ermöglichen es, bestimmte Integrale einer Funktion f direkt mithilfe der Stammfunktion F von f zu berechnen.

Theorem 7.3 (Erster Hauptsatz der Differential- und Integralrechnung). *Ist $f : I \rightarrow \mathbb{R}$ eine auf dem Intervall $I \subset \mathbb{R}$ stetige Funktion, dann ist die Funktion*

$$F : I \rightarrow \mathbb{R}, x \mapsto F(x) := \int_a^x f(t) dt \text{ mit } x, a \in I \quad (7.26)$$

eine Stammfunktion von f .

◦

Beweis. Wir betrachten den Differenzquotienten

$$\frac{1}{h}(F(x+h) - F(x)) \quad (7.27)$$

Mit der Definition $F(x) := \int_a^x f(t) dt$ und der Additivität des bestimmten Integrals gilt dann

$$\frac{1}{h}(F(x+h) - F(x)) = \frac{1}{h} \left(\int_a^{x+h} f(t) dt - \int_a^x f(t) dt \right) = \frac{1}{h} \int_x^{x+h} f(t) dt \quad (7.28)$$

Mit dem Mittelwertsatz der Integralrechnung gibt es also ein $\xi \in]x, x+h[$, so dass

$$\frac{1}{h}(F(x+h) - F(x)) = f(\xi) \quad (7.29)$$

Grenzwertbildung ergibt dann

$$\lim_{h \rightarrow 0} \frac{1}{h}(F(x+h) - F(x)) = \lim_{h \rightarrow 0} f(\xi) \text{ für } \xi \in]x, x+h[\Leftrightarrow F'(x) = f(x). \quad (7.30)$$

□

Für den Beweis des Ersten Hauptsatzes der Differential- und Integralrechnung benötigen wir offenbar den Mittelwertsatz der Integralrechnung, welchen wir hier ohne Beweis wiedergeben und in Abbildung 7.4 veranschaulichen.

Theorem 7.4 (Mittelwertsatz der Integralrechnung). *Für eine stetige Funktion $f : [a, b] \rightarrow \mathbb{R}$ existiert ein $\xi \in]a, b[$ mit*

$$\int_a^b f(x) dx = f(\xi)(b-a) \quad (7.31)$$

◦

Der Mittelwertsatz der Integralrechnung garantiert die Existenz eines $\xi \in [a, b]$, so dass das bestimmte Integral $\int_a^b f(x) dx$ gleich dem Produkt aus der “Rechteckhöhe” $f(\xi)$ und und der “Rechteckbreite” $(b-a)$ ist. In Abbildung 7.4 liegt dieses ξ genau mittig zwischen a und b . Dass die sich so ergebene grau eingefärbte Rechteckfläche gleich $\int_a^b f(x) dx$ ist, ergibt sich aus der visuell zumindest nachvollziehbaren Tatsache, dass die Flächen zwischen $f(x)$ und $f(\xi)$ im Intervall $[a, \xi]$ und zwischen $f(\xi)$ und $f(x)$ im Intervall $[\xi, b]$ den gleichen Betrag haben, erstere aber mit einem negativen Vorzeichen behaftet ist. Der Mittelwertsatz der Integralrechnung garantiert im Allgemeinen aber nur die Existenz eines $\xi \in [a, b]$ mit der diskutierten Eigenschaft, gibt aber keine Formel zu Bestimmung von ξ an.

Der Zweite Hauptsatz der Differential- und Integralrechnung schließlich besagt, wie man mithilfe der Stammfunktion ein bestimmtes Integral berechnet.

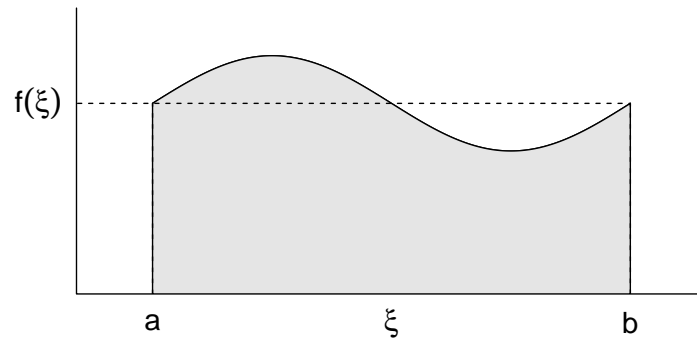


Abbildung 7.4. Zum Mittelwertsatz der Integralrechnung

Theorem 7.5 (Zweiter Hauptsatz der Differential- und Integralrechnung). *Ist F eine Stammfunktion einer stetigen Funktion $f : I \rightarrow \mathbb{R}$ auf einem Intervall I , so gilt für $a, b \in I$ mit $a \leq b$*

$$\int_a^b f(x) dx = F(b) - F(a) =: F(x)|_a^b \quad (7.32)$$

◦

Beweis. Mit den Rechenregeln für bestimmte Integrale und dem ersten Hauptsatz der Differential- und Integralrechnung ergibt sich

$$F(b) - F(a) = \int_a^b f(t) dt - \int_a^a f(t) dt = \int_a^b f(x) dx \quad (7.33)$$

□

Wir wollen den Zweiten Hauptsatz der Differential- und Integralrechnung in drei Beispielen anwenden (vgl. Abbildung 7.5).

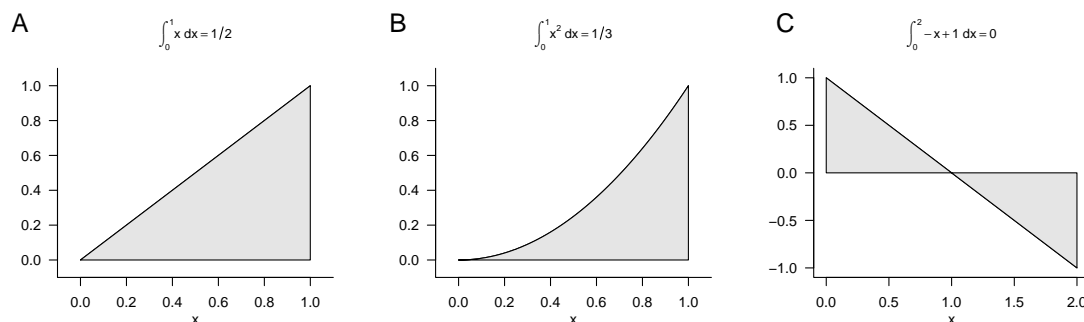


Abbildung 7.5. Beispiele zum Zweiten Hauptsatz der Differential- und Integralrechnung

Beispiel (1)

Wir betrachten die Identitätsfunktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x \quad (7.34)$$

und wollen das bestimmte Integral dieser Funktion auf dem Intervall $[0, 1]$, also

$$\int_0^1 f(x) dx = \int_0^1 x dx \quad (7.35)$$

berechnen. Dazu erinnern wir uns, dass eine Stammfunktion von f durch

$$F : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto F(x) := \frac{1}{2}x^2 \quad (7.36)$$

gegeben ist, weil

$$F'(x) = \frac{d}{dx} \left(\frac{1}{2}x^2 \right) = 2 \cdot \frac{1}{2}x^{2-1} = x. \quad (7.37)$$

Einsetzen in den Zweiten Hauptsatz der Differential- und Integralrechnung ergibt dann sofort

$$\int_0^1 x dx = \frac{1}{2}1^2 - \frac{1}{2}0^2 = \frac{1}{2}. \quad (7.38)$$

Dieses Ergebnis ist mit der Intuition, die sich anhand der grauen Fläche in Abbildung 7.5 A, ergibt kongruent.

Beispiel (2)

Als nächstes betrachten wir die Quadratfunktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2 \quad (7.39)$$

und wollen das bestimmte Integral auch dieser Funktion auf dem Intervall $[0, 1]$, also

$$\int_0^1 f(x) dx = \int_0^1 x^2 dx \quad (7.40)$$

berechnen. Dazu erinnern wir uns, dass eine Stammfunktion von f durch

$$F : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto F(x) := \frac{1}{3}x^3 \quad (7.41)$$

gegeben ist, weil

$$F'(x) = \frac{d}{dx} \left(\frac{1}{3}x^3 \right) = 3 \cdot \frac{1}{3}x^{3-1} = x^2. \quad (7.42)$$

Einsetzen in den Zweiten Hauptsatz der Differential- und Integralrechnung ergibt dann sofort

$$\int_0^1 x^2 dx = \frac{1}{3}1^3 - \frac{1}{3}0^3 = \frac{1}{3}. \quad (7.43)$$

Dieses Ergebnis ist mit der Intuition, die sich aus dem Vergleich der grauen Flächen in Abbildung 7.5 A und Abbildung 7.5 B ergibt, kongruent.

Beispiel (3)

Schließlich betrachten wir die linear-affine Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := -x + 1 \quad (7.44)$$

und wollen das bestimmte Integral auch dieser Funktion auf dem Intervall $[0, 2]$, also

$$\int_0^2 f(x) dx = \int_0^2 -x + 1 dx \quad (7.45)$$

berechnen. Dazu erinnern wir uns, dass eine Stammfunktion der linearen Funktion mit $a = -1$ und $b = 1$ (vgl. Tabelle 7.1) durch

$$F : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto F(x) := -\frac{1}{2}x^2 + x \quad (7.46)$$

gegeben ist, weil

$$F'(x) = \frac{d}{dx} \left(-\frac{1}{2}x^2 + x \right) = -2 \cdot \frac{1}{2}x^{2-1} + 1 \cdot x^{1-1} = -x + 1. \quad (7.47)$$

Einsetzen in den Zweiten Hauptsatz der Differential- und Integralrechnung ergibt dann sofort

$$\int_0^2 -x + 1 dx = \left(-\frac{1}{2}2^2 + 2 \right) - \left(-\frac{1}{2}0^2 + 0 \right) = -2 + 2 - 0 = 0. \quad (7.48)$$

Dieses Ergebnis ist mit der Intuition kongruent, dass sich die “positive” und die “negative” graue Fläche in Abbildung 7.5 C ausgleichen, kongruent.

7.3. Uneigentliche Integrale

Uneigentliche Integrale sind bestimmte Integrale bei denen mindestens eine Integrationsgrenze keine reelle Zahl ist, sondern $-\infty$ oder ∞ . Wir beleuchten die Natur uneigentlicher Integrale mit folgender Definition und einem Beispiel.

Definition 7.5 (Uneigentliche Integrale). $f : \mathbb{R} \rightarrow \mathbb{R}$ sei eine univariate reellwertige Funktion. Mit den Definitionen

$$\int_{-\infty}^b f(x) dx := \lim_{a \rightarrow -\infty} \int_a^b f(x) dx \text{ und } \int_a^{\infty} f(x) dx := \lim_{b \rightarrow \infty} \int_a^b f(x) dx \quad (7.49)$$

und der Additivität von Integralen

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^b f(x) dx + \int_b^{\infty} f(x) dx \quad (7.50)$$

wird der Begriff des bestimmten Integrals auf die unbeschränkten Integrationsintervalle $]-\infty, b]$, $[a, \infty[$ und $]-\infty, \infty[$ erweitert. Integrale mit unbeschränkten Integrationsintervallen heißen *uneigentliche Integrale*. Wenn die entsprechenden Grenzwerte existieren, sagt man, dass die uneigentlichen Integrale *konvergieren*.

•

Als Beispiel betrachten wir das uneigentliche Integral der Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) \frac{1}{x^2} \quad (7.51)$$

auf dem Intervall $[1, \infty[$, also

$$\int_1^\infty \frac{1}{x^2} dx. \quad (7.52)$$

Nach den Festlegungen in der Definition uneigentlicher Integrale gilt

$$\int_1^\infty \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2} dx. \quad (7.53)$$

Mit der Stammfunktion $F(x) = -x^{-1}$ von $f(x) = x^{-2}$ ergibt sich für das bestimmte Integral in obiger Gleichung

$$\int_1^b \frac{1}{x^2} dx = F(b) - F(1) = -\frac{1}{b} - \left(-\frac{1}{1}\right) = -\frac{1}{b} + 1. \quad (7.54)$$

Es ergibt sich also

$$\int_1^\infty \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \left(-\frac{1}{b} + 1\right) = -\lim_{b \rightarrow \infty} \frac{1}{b} + \lim_{b \rightarrow \infty} 1 = 0 + 1 = 1. \quad (7.55)$$

7.4. Mehrdimensionale Integrale

Bisher haben wir nur Integrale univariater reellwertiger Funktionen betrachtet. Der Integralbegriff lässt sich auch auf multivariate reellwertige Funktionen erweitern. Allerdings ist dann der Integrationsbereich der Funktion nicht notwendigerweise so einfach zu beschreiben wie ein Intervall; insbesondere sind zum Beispiel schon im zweidimensionalen arbiträr geformte zweidimensionale Integrationsbereiche möglich. Wir wollen hier nun den einfachsten Fall eines Hyperrechtecks betrachten. In diesem Fall können wir mehrdimensionale bestimmte Integrale wie folgt definieren.

Definition 7.6 (Mehrdimensionale Integrale). $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann heißen Integrale der Form

$$\int_{[a_1, b_1] \times \dots \times [a_n, b_n]} f(x) dx = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (7.56)$$

mehrdimensionale bestimmte Integrale auf Hyperrechtecken. Weiterhin heißen Integrale der Form

$$\int_{\mathbb{R}^n} f(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (7.57)$$

mehrdimensionale uneigentliche Integrale.

•

Wie schon erwähnt kann man multivariate reellwertige Funktion nicht nur auf Hyperrechtecken, sondern im Prinzip auf beliebigen Hyperflächen integrieren. Dies kann sich jedoch oft schwierig gestalten.

Als Beispiel betrachten wir das zweidimensionale bestimmte Integral der Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, (x_1, x_2) \mapsto f(x_1, x_2) := x_1^2 + 4x_2 \quad (7.58)$$

auf dem Rechteck $[0, 1] \times [0, 1]$. Der *Satz von Fubini* der Theorie mehrdimensionaler Integrale besagt, dass man mehrdimensionale Integrale in beliebiger Koordinatenfolge auswerten kann. Es gilt also zum Beispiel, dass

$$\int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x_1, x_2) dx_2 \right) dx_1 = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right) dx_2. \quad (7.59)$$

In diesem Sinne betrachten wir für das Beispiel

$$\int_0^1 \int_0^1 x_1^2 + 4x_2 dx_1 dx_2 = \int_0^1 \left(\int_0^1 x_1^2 + 4x_2 dx_1 \right) dx_2 \quad (7.60)$$

also zunächst das innere Integral. x_2 nimmt dabei die Rolle einer Konstanten ein. Eine Stammfunktion von $g(x_1) := x_1^2 + 4x_2$ ist $G(x_1) = \frac{1}{3}x_1^3 + 4x_2x_1$, wie man sich durch Ableiten von G überzeugt. Es ergibt sich also für das innere Integral

$$\begin{aligned} \int_0^1 x_1^2 + 4x_2 dx_1 &= G(1) - G(0) \\ &= \frac{1}{3} \cdot 1^3 + 4x_2 \cdot 1 - \frac{1}{3} \cdot 0^3 - 4x_2 \cdot 0 \\ &= \frac{1}{3} + 4x_2. \end{aligned} \quad (7.61)$$

Betrachten des äußeren Integrals ergibt dann mit der Stammfunktion

$$H(x_2) = \frac{1}{3}x_2 + 2x_2^2 \quad (7.62)$$

von

$$h(x_2) := \frac{1}{3} + 4x_2, \quad (7.63)$$

dass

$$\begin{aligned} \int_0^1 \int_0^1 x_1^2 + 4x_2 dx_1 dx_2 &= \int_0^1 \frac{1}{3} + 4x_2 dx_2 \\ &= H(1) - H(0) \\ &= \frac{1}{3} \cdot 1 + 4 \cdot 1^2 - \frac{1}{3} \cdot 0 + 4 \cdot 0^2 \\ &= \frac{13}{3}. \end{aligned} \quad (7.64)$$

Folgendes Theorem ist in der Analyse von probabilistischen Modellen oft nützlich, da es die mehrfache Integration einer faktorisierten Funktion auf das Produkt einfacher Integrationen auf die Funktionsfaktoren zurückführt.

Theorem 7.6 (Vertauschung von Integration und Multiplikation bei Produktfunktionen). Für $i = 1, \dots, n$ seien $f_i : \mathbb{R} \rightarrow \mathbb{R}$ integrierbare univariate reellwertige Funktionen. Dann gilt

$$\int \left(\prod_{i=1}^n f_i(x_i) \right) dx_1 \dots dx_n = \prod_{i=1}^n \left(\int f_i(x_i) dx_i \right) \quad (7.65)$$

◦

Beweis. Das Theorem folgt mit den Linearitätseigenschaften von Integralen. Wir zeigen das Theorem durch Induktion nach n .

Induktionsanfang. Es sei $n = 2$. Dann gilt

$$\begin{aligned}
 \int \left(\prod_{i=1}^2 f_i(x_i) \right) dx_1 dx_2 &= \int \left(\int f_1(x_1) f_2(x_2) dx_1 \right) dx_2 \\
 &= \int f_2(x_2) \left(\int f_1(x_1) dx_1 \right) dx_2 \\
 &= \left(\int f_1(x_1) dx_1 \right) \left(\int f_2(x_2) dx_2 \right) \\
 &= \prod_{i=1}^2 \left(\int f_i(x_i) dx_i \right)
 \end{aligned} \tag{7.66}$$

Induktionsschritt. Es gelte Gleichung 7.65 für $n - 1$, beispielsweise für $n - 1 = 2$. Dann gilt

$$\begin{aligned}
 \int \left(\prod_{i=1}^n f_i(x_i) \right) dx_1 \dots dx_n &= \int \left(\int \dots \left(\int f_1(x_1) \dots f_{n-1}(x_{n-1}) f_n(x_n) dx_1 \right) \dots dx_{n-1} \right) dx_n \\
 &= \int f_n(x_n) \left(\int \dots \left(\int f_1(x_1) \dots f_{n-1}(x_{n-1}) dx_1 \right) \dots dx_{n-1} \right) dx_n \\
 &= \int f_n(x_n) \left(\int \left(\prod_{i=1}^{n-1} f_i(x_i) \right) dx_{1:n-1} \right) dx_n \\
 &= \left(\int f_n(x_n) dx_n \right) \left(\int \left(\prod_{i=1}^{n-1} f_i(x_i) \right) dx_{1:n-1} \right) \\
 &= \left(\int f_n(x_n) dx_n \right) \left(\prod_{i=1}^{n-1} \left(\int f_i(x_i) dx_i \right) \right) \\
 &= \prod_{i=1}^n \left(\int f_i(x_i) dx_i \right),
 \end{aligned} \tag{7.67}$$

wobei wir die Annahme, dass Gleichung 7.65 für $n - 1$ in der vorletzten Gleichung genutzt haben. □

7.5. Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs der Stammfunktion wieder.
2. Geben Sie die Definition des Begriffs des unbestimmten Integrals wieder.
3. Erläutern Sie die intuitive Bedeutung des Begriff des Riemannschen Integrals.
4. Geben Sie den ersten Hauptsatz der Differential- und Integralrechnung wieder.
5. Geben Sie den zweiten Hauptsatz der Differential- und Integralrechnung wieder.
6. Erläutern Sie den Begriff des uneigentlichen Integrals.
7. Erläutern Sie den Begriff des mehrdimensionalen Integrals.

8. Vektoren

In der naturwissenschaftlichen Modellbildung betrachtet man häufig Phänomene, die sich durch das Vorliegen mehrerer quantitativer Merkmale auszeichnen. So ist zum Beispiel die Position eines Objektes im dreidimensionalen Raum durch drei Koordinaten hinsichtlich der drei Achsen eines Kartesischen Koordinatensystems festgelegt. Analog mag der Gesundheitszustand einer Person durch das Vorliegen dreier Messwerte, beispielsweise eines Selbstauskunftscore, eines Biomarkers und einer Expert:inneneinschätzung charakterisiert sein. Zum modellieren und analysieren solcher mehrdimensionalen quantitativen Phänomene stellt die Mathematik mit dem reellen Vektorraum ein vielseitig einsetzbares Hilfsmittel bereit. In diesem Kapitel wollen wir zunächst den Begriff des reellen Vektorraums und das grundlegende Rechnen mit Vektoren einführen (Kapitel 8.1). Eine Vektorraumstruktur, die sich stark an der dreidimensionalen räumlichen Intuition orientiert bietet dann der Euklidische Vektorraum (Kapitel 8.2). Mithilfe der Vektorrechnung können alle Vektoren eines Vektorraums aus einer kleinen Schar ausgezeichneter Vektoren gebildet werden. Die diesem Prinzip zugrundeliegenden Konzepte diskutieren wir in Kapitel 8.3 und Kapitel 8.4.

8.1. Reeller Vektorraum

Wir beginnen mit der allgemeinen Definition eines Vektorraums, die grundlegende Regeln zum Rechnen mit Vektoren festlegt.

Definition 8.1 (Vektorraum). Es seien V eine nichtleere Menge und S eine Menge von Skalaren. Weiterhin sei eine Abbildung

$$+ : V \times V \rightarrow V, (v_1, v_2) \mapsto +(v_1, v_2) =: v_1 + v_2, \quad (8.1)$$

genannt *Vektoraddition*, definiert. Schließlich sei eine Abbildung

$$\cdot : S \times V \rightarrow V, (s, v) \mapsto \cdot(s, v) =: sv, \quad (8.2)$$

genannt *Skalarmultiplikation* definiert. Dann wird das Tupel $(V, S, +, \cdot)$ genau dann *Vektorraum* genannt, wenn für beliebige Elemente $v, w, u \in V$ und $a, b \in S$ folgende Bedingungen gelten:

(1) *Kommutativität der Vektoraddition.*

$$v + w = w + v.$$

(2) *Assoziativität der Vektoraddition.*

$$(v + w) + u = v + (w + u)$$

(3) *Existenz eines neutralen Elements der Vektoraddition.*

Es gibt einen Vektor $0 \in V$ mit $v + 0 = 0 + v = v$.

(4) *Existenz inverser Elemente der Vektoraddition*

Für alle Vektoren $v \in V$ gibt es einen Vektor $-v \in V$ mit $v + (-v) = 0$.

(5) *Existenz eines neutralen Elements der Skalarmultiplikation.*

Es gibt einen Skalar $1 \in S$ mit $1 \cdot v = v$.

(6) *Assoziativität der Skalarmultiplikation.*

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

(7) *Distributivität hinsichtlich der Vektoraddition.*

$$a \cdot (v + w) = a \cdot v + a \cdot w.$$

(8) *Distributivität hinsichtlich der Skalaraddition.*

$$(a + b) \cdot v = a \cdot v + b \cdot v.$$

•

Es fällt auf, dass Definition 8.1 zwar festlegt, wie mit Vektoren gerechnet werden soll, jedoch keine Aussage darüber macht, was ein Vektor, über ein Element einer Menge hinaus, eigentlich ist. Dies ist der Tatsache geschuldet, dass es verschiedenste mathematische Objekte gibt, für die Vektorraumstrukturen definiert werden können. Beispiele dafür sind die Menge der reellen m -Tupel, die Menge der Matrizen, die Menge der Polynome, die Menge der Lösungen eines linearen Gleichungssystems, die Menge der reellen Folgen, die Menge der stetigen Funktionen u.v.a.m.

Wir sind hier zunächst nur am Vektorraum der Menge reellen m -Tupel interessiert. Wir erinnern dazu daran, dass wir die reellen m -Tupel mit

$$\mathbb{R}^m := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \mid x_i \in \mathbb{R} \text{ für alle } 1 \leq i \leq m \right\} \quad (8.3)$$

bezeichnen und \mathbb{R}^m als “ \mathbb{R} hoch m ” aussprechen. Die Elemente $x \in \mathbb{R}^m$ nennen wir *reelle Vektoren* oder auch einfach *Vektoren*. Wir wollen nun der Definition eines Vektorraums die Menge \mathbb{R}^m zugrunde legen. Dazu definieren wir zunächst die Vektoraddition für Elemente von \mathbb{R}^m und die Skalarmultiplikation für Elemente von \mathbb{R} und \mathbb{R}^m

Definition 8.2 (Vektoraddition und Skalarmultiplikation in \mathbb{R}^m). Für alle $x, y \in \mathbb{R}^m$ und $a \in \mathbb{R}$ sei die *Vektoraddition* durch

$$+ : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m, (x, y) \mapsto x + y = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_m + y_m \end{pmatrix} \quad (8.4)$$

und die *Skalarmultiplikation* durch

$$\cdot : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^m, (a, x) \mapsto ax = a \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} := \begin{pmatrix} ax_1 \\ \vdots \\ ax_m \end{pmatrix} \quad (8.5)$$

definiert.

•

Es ergibt sich dann folgendes Resultat.

Theorem 8.1 (Reeller Vektorraum). $(\mathbb{R}^m, +, \cdot)$ mit den Rechenregeln der Addition und Multiplikation in \mathbb{R} einen Vektorraum.

◦

Für einen Beweis, auf den wir hier verzichten wollen, muss man die Bedingungen (1) bis (8) aus Definition 8.1 für die hier betrachtete Menge und die hier festgelegten Formen der Vektoraddition und der Skalarmultiplikation nachweisen. Diese ergeben sich aber leicht aus den Rechenregeln von Addition und Multiplikation in \mathbb{R} und der Tatsache, dass Vektoraddition und Skalarmultiplikation für Elemente von \mathbb{R}^m in Definition 8.2 komponentenweise definiert wurden. Wir definieren damit den Begriff des *reellen Vektorraums*.

Definition 8.3 (Reeller Vektorraum). Für \mathbb{R}^m seien $+$ und \cdot die in Definition 8.2 definierte Vektoraddition und Skalarmultiplikation. Dann nennen wir auf Grundlage von Theorem 8.1 den Vektorraum $(\mathbb{R}^m, +, \cdot)$ den *reellen Vektorraum*.

•

Auf Grundlage von Definition 8.3 wollen wir uns nun das Rechnen mit reellen Vektoren anhand einiger Beispiele verdeutlichen.

Beispiele

(1) Für

$$x := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^4 \text{ und } y := \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^4$$

gilt

$$x + y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+2 \\ 2+1 \\ 3+0 \\ 4+1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 5 \end{pmatrix} \in \mathbb{R}^4.$$

In **R** implementiert dieses Beispiel wie folgt

```
x = matrix(c(1,2,3,4), nrow = 4) # Vektordefinition
y = matrix(c(2,1,0,1), nrow = 4) # Vektordefinition
x + y # Vektoraddition
```

```
[,1]
[1,] 3
[2,] 3
[3,] 3
[4,] 5
```

(2) Für

$$x := \begin{pmatrix} 2 \\ 3 \end{pmatrix} \in \mathbb{R}^2 \text{ und } y := \begin{pmatrix} 1 \\ 3 \end{pmatrix} \in \mathbb{R}^2$$

gilt

$$x - y = \begin{pmatrix} 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2-1 \\ 3-3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{R}^2.$$

In **R** implementiert man dieses Beispiel wie folgt

```
x = matrix(c(2,3), nrow = 2) # Vektordefinition
y = matrix(c(1,3), nrow = 2) # Vektordefinition
x - y                         # Vektorsubtraktion
```

```
[,1]
[1,] 1
[2,] 0
```

(3) Für

$$x := \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \in \mathbb{R}^3 \text{ und } a := 3 \in \mathbb{R}$$

gilt

$$ax = 3 \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 \\ 3 \cdot 1 \\ 3 \cdot 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \\ 9 \end{pmatrix} \in \mathbb{R}^3.$$

In **R** implementiert man dieses Beispiel wie folgt

```
x = matrix(c(2,1,3), nrow = 3) # Vektordefinition
a = 3                          # Skalardefinition
a*x                            # Skalarmultiplikation
```

```
[,1]
[1,] 6
[2,] 3
[3,] 9
```

Für $m \in \{1, 2, 3\}$ kann man sich reelle Vektoren und das Rechnen mit ihnen visuell veranschaulichen. Für $m > 3$, wenn also zum Beispiel für eine Person mehr als drei quantitative Merkmale zu ihrem Gesundheitszustand vorliegen, was in der Anwendung regelmäßig der Fall ist, ist dies nicht möglich. Trotzdem mag die visuelle Intuition für $m \leq 3$ einen Einstieg in das Verständnis von Vektorräumen erleichtern. Wir fokussieren hier auf den Fall $m := 2$. In diesem Fall liegen die betrachteten reellen Vektoren in der zweidimensionalen Ebene und werden üblicherweise als Punkte oder Pfeile visualisiert (Abbildung 8.1).

Abbildung 8.2 visualisiert die Vektoraddition

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}. \quad (8.6)$$

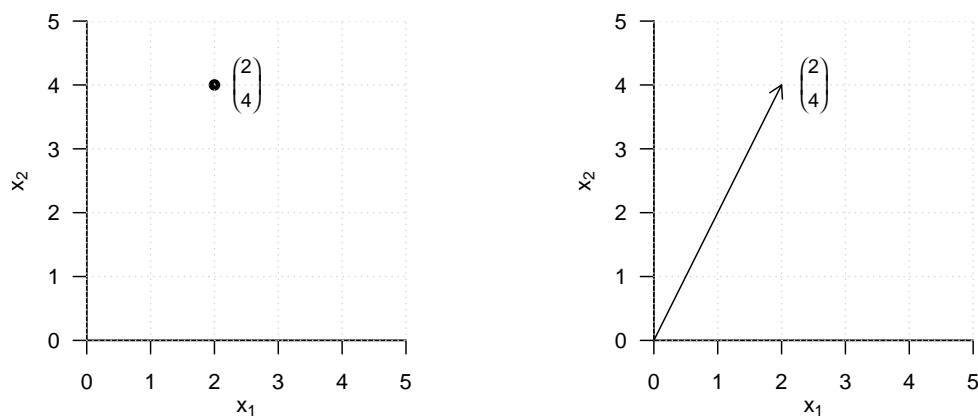


Abbildung 8.1. Visualisierung von Vektoren in \mathbb{R}^2

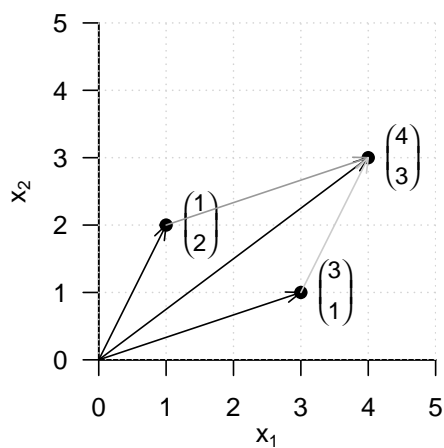


Abbildung 8.2. Vektoraddition in \mathbb{R}^2

Der Summenvektor entspricht dabei der Diagonale des von den beiden Summanden aufgespannten Parallelogramms.

Abbildung 8.3 visualisiert die Vektorsubtraktion

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} -3 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad (8.7)$$

Der resultierende Vektor entspricht dabei der Diagonale des von dem ersten Vektors und dem entgegengesetzten Vektor des zweiten Vektors aufgespannten Parallelogramms.

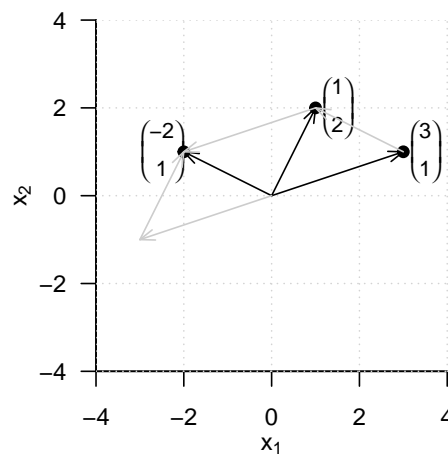


Abbildung 8.3. Vektorsubtraktion in \mathbb{R}^2

Abbildung 8.4 schließlich visualisiert die Skalarmultiplikation

$$3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \quad (8.8)$$

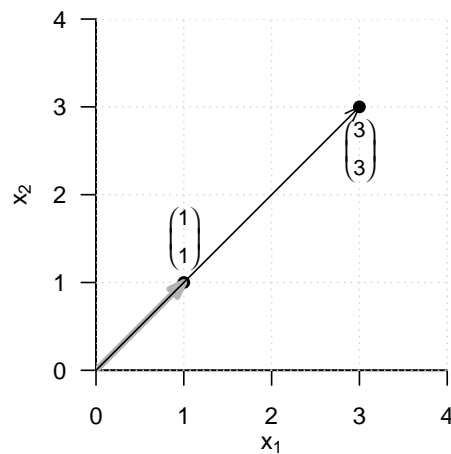
Die Multiplikation eines Vektors mit einem Skalar ändert dabei immer nur seine Länge, nicht jedoch seine Richtung.

8.2. Euklidischer Vektorraum

Der reelle Vektorraum kann durch Definition des *Skalarprodukts* im Sinne eines *Euklidischen Vektorraums* mit räumlich-geometrischer Intuition versehen werden. Diese ermöglicht es insbesondere, Begriffe wie die *Länge eines Vektors*, den *Abstand zwischen zwei Vektoren*, und nicht zuletzt den *Winkel zwischen zwei Vektoren* zu definieren und zu berechnen. Wir führen zunächst das *Skalarprodukt* ein.

Definition 8.4 (Skalarprodukt auf \mathbb{R}^m). Das *Skalarprodukt auf \mathbb{R}^m* ist definiert als die Abbildung

$$\langle \rangle : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}, (x, y) \mapsto \langle (x, y) \rangle := \langle x, y \rangle := \sum_{i=1}^m x_i y_i. \quad (8.9)$$

Abbildung 8.4. Skalarmultiplikation in \mathbb{R}^2

•

Das Skalarprodukt heißt Skalarprodukt, weil es einen Skalar ergibt, nicht etwa, weil mit Skalaren multipliziert wird. Das Skalarprodukt steht in enger Beziehung zum Matrixprodukt, wie wir an späterer Stelle sehen werden. Wir betrachten zunächst ein Beispiel und seine Implementation in **R**.

Beispiel

Es seien

$$x := \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ und } y := \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \quad (8.10)$$

Dann ergibt sich

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3 = 1 \cdot 2 + 2 \cdot 0 + 3 \cdot 1 = 2 + 0 + 3 = 5. \quad (8.11)$$

In **R** gibt es verschiedene Möglichkeiten, ein Skalarprodukt auszuwerten. Wir führen zwei von ihnen für das gegebene Beispiel untenstehend auf.

```
# Vektordefinitionen
x = matrix(c(1,2,3), nrow = 3)
y = matrix(c(2,0,1), nrow = 3)

# Skalarprodukt mithilfe von R's komponentenweiser Multiplikation und sum() Funktion
sum(x*y)
```

```
[1] 5
```

```
# Skalarprodukt mithilfe von R's Matrixtransposition und -multiplikation
t(x) %*% y
```

```
 [,1]
[1,] 5
```

Mithilfe des Skalarprodukts kann der Begriff des reellen Vektorraums zum Begriff des *reellen kanonischen Euklidischen Vektorraums* erweitert werden.

Definition 8.5 (Euklidischer Vektorraum). Das Tupel $((\mathbb{R}^m, +, \cdot), \langle \rangle)$ aus dem reellen Vektorraum $(\mathbb{R}^m, +, \cdot)$ und dem Skalarprodukt $\langle \rangle$ auf \mathbb{R}^m heißt *reeller kanonischer Euklidischer Vektorraum*.

•

Generell heißt jedes Tupel aus einem Vektorraum und einem Skalarprodukt “Euklidischer Vektorraum”. Informell sprechen wir aber oft auch einfach von \mathbb{R}^m als “Euklidischer Vektorraum” und insbesondere bei $((\mathbb{R}^m, +, \cdot), \langle \rangle)$ vom “Euklidischen Vektorraum”. Ein Euklidischer Vektorraum ist ein Vektorraum mit geometrischer Struktur, die durch das Skalarprodukt induziert wird. Insbesondere bekommen im Euklidischen Vektorraum nun die geometrischen Begriffe von *Länge*, *Abstand* und *Winkel* eine Bedeutung. Wir definieren sie wie folgt.

Definition 8.6. $((\mathbb{R}^m, +, \cdot), \langle \rangle)$ sei der Euklidische Vektorraum.

(1) Die *Länge* eines Vektors $x \in \mathbb{R}^m$ ist definiert als

$$\|x\| := \sqrt{\langle x, x \rangle}. \quad (8.12)$$

(2) Der *Abstand* zweier Vektoren $x, y \in \mathbb{R}^m$ ist definiert als

$$d(x, y) := \|x - y\|. \quad (8.13)$$

(3) Der *Winkel* α zwischen zwei Vektoren $x, y \in \mathbb{R}^m$ mit $x, y \neq 0$ ist definiert durch

$$0 \leq \alpha \leq \pi \text{ und } \cos \alpha := \frac{\langle x, y \rangle}{\|x\| \|y\|} \quad (8.14)$$

•

Die Länge $\|x\|$ eines Vektors $x \in \mathbb{R}^m$ heißt auch *Euklidische Norm von x* oder ℓ_2 -Norm von x oder einfach *Norm von x*. Sie wird häufig auch mit $\|x\|_2$ bezeichnet. Wir betrachten drei Beispiele für die Bestimmung der Länge eines Vektors und ihre entsprechende **R** Implementation. Wir veranschaulichen diese Beispiele in Abbildung 8.5.

Beispiel (1)

$$\left\| \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\| = \sqrt{\left\langle \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\rangle} = \sqrt{2^2 + 0^2} = \sqrt{4} = 2.00 \quad (8.15)$$

```
norm(matrix(c(2,0),nrow = 2), type = "2")           # Vektorlänge = 1_2 Norm
```

[1] 2

Beispiel (2)

$$\left\| \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\| = \sqrt{\left\langle \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\rangle} = \sqrt{2^2 + 2^2} = \sqrt{8} \approx 2.83 \quad (8.16)$$

```
norm(matrix(c(2,2),nrow = 2), type = "2") # Vektorlänge = 1_2 Norm
```

```
[1] 2.828427
```

Beispiel (3)

$$\left\| \begin{pmatrix} 2 \\ 4 \end{pmatrix} \right\| = \sqrt{\left\langle \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix} \right\rangle} = \sqrt{2^2 + 4^2} = \sqrt{20} \approx 4.47 \quad (8.17)$$

```
norm(matrix(c(2,4),nrow = 2), type = "2") # Vektorlänge = 1_2 Norm
```

```
[1] 4.472136
```

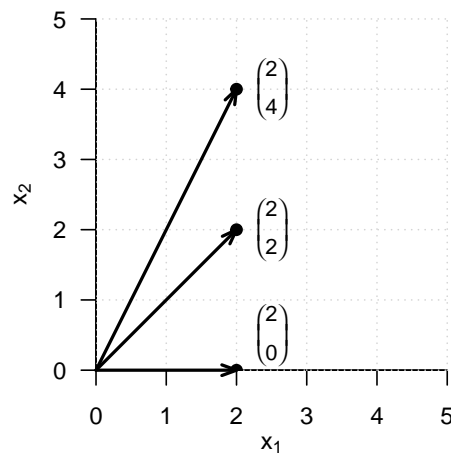


Abbildung 8.5. Vektorlänge in \mathbb{R}^2

Für den Abstand $d(x, y)$ zweier Vektoren $x, y \in \mathbb{R}^m$ halten wir ohne Beweis fest, dass er zum einen nicht-negativ und symmetrisch ist, also dass

$$d(x, y) \geq 0, d(x, x) = 0 \text{ und } d(x, y) = d(y, x) \quad (8.18)$$

gelten. Zudem erfüllt $d(x, y)$ die sogenannte *Dreiecksungleichung*, die besagt, dass die direkte Wegstrecke zwischen zwei Punkten im Raum immer kürzer ist als eine indirekte Wegstrecke über einen dritten Punkt,

$$d(x, y) \leq d(x, z) + d(z, y). \quad (8.19)$$

Damit erfüllt $d(x, y)$ wichtige Aspekte der räumlichen Anschauung. Wir geben zwei Beispiele für die Bestimmung von Abständen von Vektoren in \mathbb{R}^2 , die wir in [Abbildung 8.6](#) visualisieren.

Beispiel (1)

$$d\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right) = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\| = \left\| \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\| = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} \approx 1.41 \quad (8.20)$$

```
norm(matrix(c(1,1),nrow = 2) - matrix(c(2,2),nrow = 2), type = "2")
```

```
[1] 1.414214
```

Beispiel (2)

$$d\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \end{pmatrix}\right) = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 4 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} -3 \\ 0 \end{pmatrix} \right\| = \sqrt{(-3)^2 + 0^2} = \sqrt{9} = 3 \quad (8.21)$$

```
norm(matrix(c(1,1),nrow = 2) - matrix(c(4,1),nrow = 2), type = "2")
```

```
[1] 3
```

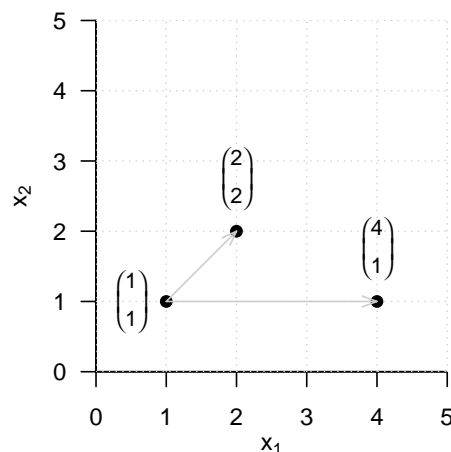


Abbildung 8.6. Vektorabstände in \mathbb{R}^2

Schließlich halten wir fest, dass für die Berechnung des Winkels zwischen zwei Vektoren anhand obiger Definition gilt, dass die Kosinusfunktion \cos auf $[0, \pi]$ bijektiv, also invertierbar mit der Umkehrfunktion \arccos , der Arkuskosinusfunktion, ist. Auch für den Begriff des Winkels wollen wir zwei Beispiele betrachten. Man beachte dabei insbesondere, dass die Definition 8.6 den Winkel in Radians angibt. Für eine Angabe in Grad ist eine entsprechende Umrechnung erforderlich.

Beispiel (1)

$$\arccos\left(\frac{\left\langle \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \end{pmatrix} \right\rangle}{\left\| \begin{pmatrix} 3 \\ 0 \end{pmatrix} \right\| \left\| \begin{pmatrix} 3 \\ 3 \end{pmatrix} \right\|}\right) = \arccos\left(\frac{3 \cdot 3 + 3 \cdot 0}{\sqrt{3^2 + 0^2} \cdot \sqrt{3^2 + 3^2}}\right) = \arccos\left(\frac{9}{3 \cdot \sqrt{18}}\right) = \frac{\pi}{4} \approx 0.785 \quad (8.22)$$

Die Umrechnung in Grad ergibt dann

$$0.785 \cdot \frac{180^\circ}{\pi} = 45^\circ \quad (8.23)$$

In **R** implementiert man dies wie folgt.

```
x = matrix(c(3,0), nrow = 2) # Vektor 1
y = matrix(c(3,3), nrow = 2) # Vektor 2
w = acos(sum(x*y)/(sqrt(sum(x*x))*sqrt(sum(y*y)))) * 180/pi # Winkel in Grad
print(w)
```

[1] 45

Beispiel (2)

$$\alpha = \arccos \left(\frac{\left\langle \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\rangle}{\left\| \begin{pmatrix} 3 \\ 0 \end{pmatrix} \right\| \left\| \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\|} \right) = \arccos \left(\frac{3 \cdot 0 + 0 \cdot 3}{\sqrt{3^2 + 0^2} \cdot \sqrt{0^2 + 3^2}} \right) = \arccos \left(\frac{0}{3 \cdot 3} \right) = \frac{\pi}{2} \approx 1.57 \quad (8.24)$$

Die Umrechnung in Grad ergibt dann

$$\frac{\pi}{2} \cdot \frac{180^\circ}{\pi} = 90^\circ \quad (8.25)$$

Die entsprechende **R** Implementation lautet wie folgt.

```
x = matrix(c(3,0), nrow = 2) # Vektor 1
y = matrix(c(0,3), nrow = 2) # Vektor 2
w = acos(sum(x*y)/(sqrt(sum(x*x))*sqrt(sum(y*y)))) * 180/pi # Winkel in Grad
print(w)
```

[1] 90

Die Tatsache, dass zwei Vektoren einen rechten Winkel bilden können, also gewissermaßen maximal nicht-parallel sein können, ist ein wichtiges geometrisches Prinzip und wird deshalb mit folgender Definition speziell ausgezeichnet.

Definition 8.7 (Orthogonalität und Orthonormalität von Vektoren). $((\mathbb{R}^m, +, \cdot), \langle \rangle)$ sei der Euklidische Vektorraum.

(1) Zwei Vektoren $x, y \in \mathbb{R}^m$ heißen *orthogonal*, wenn gilt, dass

$$\langle x, y \rangle = 0 \quad (8.26)$$

(2) Zwei Vektoren $x, y \in \mathbb{R}^m$ heißen *orthonormal*, wenn gilt, dass

$$\langle x, y \rangle = 0 \text{ und } \|x\| = \|y\| = 1. \quad (8.27)$$

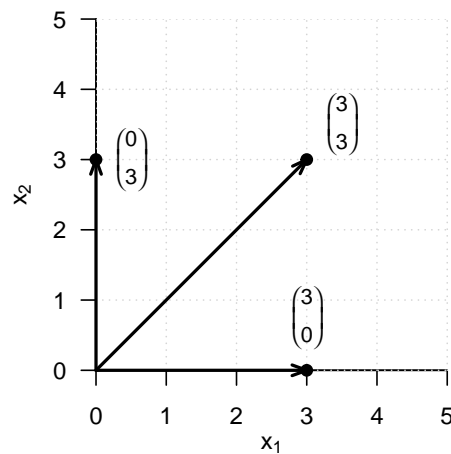
•

Für orthogonale und orthonormale Vektoren gilt also insbesondere auch

$$\cos \alpha = \frac{\langle x, y \rangle}{\|x\| \|y\|} = \frac{0}{\|x\| \|y\|} = 0, \quad (8.28)$$

also

$$\alpha = \frac{\pi}{2} = 90^\circ. \quad (8.29)$$

Abbildung 8.7. Winkel in \mathbb{R}^2

8.3. Lineare Unabhängigkeit

In diesem Abschnitt führen wir den Begriff der *linearen Unabhängigkeit* von Vektoren ein. Wir definieren dazu zunächst den Begriff der *Linearkombination* von Vektoren.

Definition 8.8 (Linearkombination). $\{v_1, v_2, \dots, v_k\}$ sei eine Menge von k Vektoren eines Vektorraums V und a_1, a_2, \dots, a_k seien Skalare. Dann ist die *Linearkombination* der Vektoren in $\{v_1, v_2, \dots, v_k\}$ mit den *Koeffizienten* a_1, a_2, \dots, a_k definiert als der Vektor

$$w := \sum_{i=1}^k a_i v_i \in V. \quad (8.30)$$

•

Beispiel

Es seien

$$v_1 := \begin{pmatrix} 2 \\ 1 \end{pmatrix}, v_2 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_3 := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ und } a_1 := 2, a_2 := 3, a_3 := 0. \quad (8.31)$$

Dann ergibt sich die Linearkombination von v_1, v_2, v_3 mit den Koeffizienten a_1, a_2, a_3 zu

$$\begin{aligned} w &= a_1 v_1 + a_2 v_2 + a_3 v_3 \\ &= 2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 7 \\ 5 \end{pmatrix}. \end{aligned} \quad (8.32)$$

Basierend auf dem Begriff der Linearkombination kann man nun den Begriff der *Linearen Unabhängigkeit* von Vektoren definieren.

Definition 8.9 (Lineare Unabhängigkeit). V sei ein Vektorraum. Eine Menge $W := \{w_1, w_2, \dots, w_k\}$ von Vektoren in V heißt *linear unabhängig*, wenn die einzige Repräsentation des Nullelements $0 \in V$ durch eine Linearkombination der $w \in W$ die sogenannte *triviale Repräsentation*

$$0 = a_1 w_1 + a_2 w_2 + \dots + a_k w_k \text{ mit } a_1 = a_2 = \dots = a_k = 0 \quad (8.33)$$

ist. Wenn die Menge W nicht linear unabhängig ist, dann heißt sie *linear abhängig*.

•

Um zu prüfen, ob eine gegebene Menge von Vektoren linear abhängig oder unabhängig ist, muss man prinzipiell für jede mögliche Linearkombination der gegebenen Vektoren, ob sie Null ist. Theorem 8.2 und Theorem 8.3 zeigen, wie dies für zwei bzw. endliche viele Vektoren auch mit weniger Aufwand gelingen kann.

Theorem 8.2 (Lineare Abhängigkeit von zwei Vektoren). V sei ein Vektorraum. Zwei Vektoren $v_1, v_2 \in V$ sind *linear abhängig*, wenn einer der Vektoren ein skalares Vielfaches des anderen Vektors ist.

◦

Beweis. v_1 sei ein skalares Vielfaches von v_2 , also

$$v_1 = \lambda v_2 \text{ mit } \lambda \neq 0. \quad (8.34)$$

Dann gilt

$$v_1 - \lambda v_2 = 0. \quad (8.35)$$

Dies aber entspricht der Linearkombination

$$a_1 v_1 + a_2 v_2 = 0 \quad (8.36)$$

mit $a_1 = 1 \neq 0$ und $a_2 = -\lambda \neq 0$. Es gibt also eine Linearkombination des Nullelementes, die nicht die triviale Repräsentation ist, und damit sind v_1 und v_2 nicht linear unabhängig. \square

Theorem 8.3 (Lineare Abhängigkeit einer Menge von Vektoren). V sei ein Vektorraum und $w_1, \dots, w_k \in V$ sei eine Menge von Vektoren in V . Wenn einer der Vektoren w_i mit $i = 1, \dots, k$ eine Linearkombination der anderen Vektoren ist, dann ist die Menge der Vektoren *linear abhängig*.

◦

Beweis. Die Vektoren w_1, \dots, w_k sind genau dann linear abhängig, wenn gilt, dass $\sum_{i=1}^k a_i w_i = 0$ mit mindestens einem $a_i \neq 0$. Es sei also zum Beispiel $a_j \neq 0$. Dann gilt

$$0 = \sum_{i=1}^k a_i w_i = \sum_{i=1, i \neq j}^k a_i w_i + a_j w_j \quad (8.37)$$

Also folgt

$$a_j w_j = - \sum_{i=1, i \neq j}^k a_i w_i \quad (8.38)$$

und damit

$$w_j = -a_j^{-1} \sum_{i=1, i \neq j}^k a_i w_i = - \sum_{i=1, i \neq j}^k (a_j^{-1} a_i) w_i \quad (8.39)$$

Also ist w_j eine Linearkombination der w_i für $i = 1, \dots, k$ mit $i \neq j$. \square

8.4. Vektorraumbasen

In diesem Abschnitt wollen wir den Begriff der *Vektorraumbasis* einführen. Eine Basis eines Vektorraums ist eine Untermenge von Vektoren des Vektorraums, die zur Darstellung aller Vektoren des Vektorraums genutzt werden kann. Im Sinne der linearen Kombination von Vektoren enthält also eine Vektorraumbasis alle nötige Information zur Konstruktion des entsprechenden Vektorraums. Allerdings ist eine Vektorraumbasis in der Regel nicht eindeutig und die viele Vektorräume haben in der Tat unendlich viele Basen. Die folgenden Definition sagt zunächst aus, wie aus einer beschränkten Anzahl von Vektoren mithilfe von Linearkombinationen unendlich viele Vektoren gebildet werden können.

Definition 8.10 (Lineare Hülle und Aufspannen). V sei ein Vektorraum und es sei $W := \{w_1, \dots, w_k\} \subset V$. Dann ist die *lineare Hülle* von W definiert als die Menge aller Linearkombinationen der Elemente von W ,

$$\text{Span}(W) := \left\{ \sum_{i=1}^k a_i w_i \mid a_1, \dots, a_k \text{ sind skalare Koeffizienten} \right\} \quad (8.40)$$

Man sagt, dass eine Menge von Vektoren $W \subseteq V$ *einen Vektorraum V aufspannt*, wenn jedes $v \in V$ als eine Linearkombination von Vektoren in W geschrieben werden kann.

•

Wir definieren nun den Begriff der *Basis* eines Vektorraums.

Definition 8.11 (Basis). V sei ein Vektorraum und es sei $B \subseteq V$. B heißt eine *Basis* von V , wenn

- (1) die Vektoren in B linear unabhängig sind und
- (2) die Vektoren in B den Vektorraum V aufspannen.

•

Basen von Vektorräumen haben folgende wichtige Eigenschaften.

Theorem 8.4 (Eigenschaften von Basen).

- (1) *Alle Basen eines Vektorraums beinhalten die gleiche Anzahl von Vektoren.*
- (2) *Jede Menge von m linear unabhängigen Vektoren ist Basis eines m -dimensionalen Vektorraums.*

◦

Für einen Beweis dieses sehr tiefen Theorems verweisen wir auf die weiterführende Literatur. Die mit obigem Theorem benannte eindeutige Anzahl der Vektoren einer Basis eines Vektorraums heißt die *Dimension des Vektorraums*. Da es in der Regel unendlich viele Mengen von m linear unabhängigen Vektoren in einem Vektorraum gibt haben Vektorräume in der Regel unendlich viele Basen.

Betrachtet man nun einen einzelnen Vektor in einem Vektorraum, so kann man sich fragen, wie man diesen mithilfe einer Vektorraumbasis darstellen kann. Dies führt auf folgende Begriffsbildungen.

Definition 8.12 (Basisdarstellung und Koordinaten). $B := \{b_1, \dots, b_m\}$ sei eine Basis eines m -dimensionalen Vektorraumes V und es sei $v \in V$. Dann heißt die Linearkombination

$$v = \sum_{i=1}^m c_i b_i \quad (8.41)$$

die *Darstellung von v bezüglich der Basis B* und die Koeffizienten c_1, \dots, c_m heißen die *Koordinaten von v bezüglich der Basis B* .

•

Bei fester Basis sind auch die Koordinaten eines Vektors bezüglich dieser Basis fest und eindeutig. Dies ist die Aussage folgenden Theorems.

Theorem 8.5 (Eindeutigkeit der Basisdarstellung). *Die Basisdarstellung eines $v \in V$ bezüglich einer Basis B ist eindeutig.*

◦

Beweis. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass der Vektorraum von Dimension m ist. Nehmen wir an, dass zwei Darstellungen von v bezüglich der Basis B existieren, also dass

$$\begin{aligned} v &= a_1 b_1 + \dots + a_m b_m \\ v &= c_1 b_1 + \dots + c_m b_m \end{aligned} \quad (8.42)$$

Subtraktion der unteren von der oberen Gleichung ergibt

$$0 = (a_1 - c_1)b_1 + \dots + (a_m - c_m)b_m \quad (8.43)$$

Weil die b_1, \dots, b_m linear unabhängig sind, gilt aber, dass $(a_i - c_i) = 0$ für alle $i = 1, \dots, m$ und somit sind die beiden Darstellungen von v bezüglich der Basis B identisch. \square

Zum Abschluss dieses Abschnitts wollen wir eine spezielle Basis des reellen Vektorraums betrachten.

Definition 8.13 (Orthonormalbasis von \mathbb{R}^m). Eine Menge von m Vektoren $v_1, \dots, v_m \in \mathbb{R}^m$ heißt *Orthonormalbasis* von \mathbb{R}^m , wenn v_1, \dots, v_m jeweils die Länge 1 haben und wechselseitig orthogonal sind, also wenn

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases} \quad (8.44)$$

•

Wir wollen zunächst ein Beispiel für eine Orthonormalbasis betrachten.

Beispiel (1)

Es ist

$$B_1 := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \quad (8.45)$$

eine Orthonormalbasis von \mathbb{R}^2 , denn B_1 besteht aus zwei Vektoren und es gelten

$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle = 1 \cdot 1 + 0 \cdot 0 = 1 + 0 = 1 \quad (8.46)$$

sowie

$$\left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 0 \cdot 0 + 1 \cdot 1 = 0 + 1 = 1 \quad (8.47)$$

und

$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 1 \cdot 0 + 0 \cdot 1 = 0 + 0 = 0 \quad (8.48)$$

Für allgemeine reelle Vektorräume werden Basen der Form von B_1 mit dem Begriff der *kanonischen Basis* speziell ausgezeichnet.

Definition 8.14 (Kanonische Basis und kanonische Einheitsvektoren). Die Orthonormalbasis

$$B := \{e_1, \dots, e_m | e_{i_j} = 1 \text{ für } i = j \text{ und } e_{i_j} = 0 \text{ für } i \neq j\} \subset \mathbb{R}^m \quad (8.49)$$

heißt die *kanonische Basis* von \mathbb{R}^m und die e_{i_j} heißen *kanonische Einheitsvektoren*.

•

B_1 aus Beispiel (1) ist also die kanonische Basis von \mathbb{R}^2 .

Die kanonische Basis von \mathbb{R}^3 ist

$$B := \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}. \quad (8.50)$$

Allerdings gibt es auch nicht kanonische Orthonormalbasen. Dazu betrachten wir ein weiteres Beispiel

Beispiel (2)

Es ist auch

$$B_2 := \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\} \quad (8.51)$$

eine Orthonormalbasis von \mathbb{R}^2 , denn B_2 besteht aus zwei Vektoren und es gelten

$$\left\langle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\rangle = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1, \quad (8.52)$$

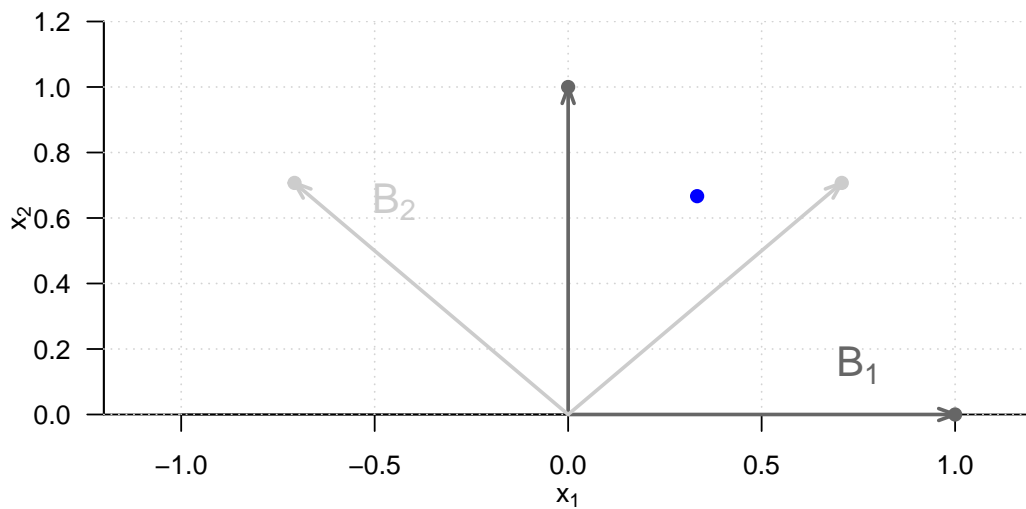
sowie

$$\left\langle \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\rangle = \left(-\frac{1}{\sqrt{2}}\right) \cdot \left(-\frac{1}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1 \quad (8.53)$$

und

$$\left\langle \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\rangle = -\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = -\frac{1}{2} + \frac{1}{2} = 0 \quad (8.54)$$

Wir visualisieren die beiden Orthonormalbasen B_1 und B_2 von \mathbb{R}^2 in Abbildung 8.8.

Abbildung 8.8. Zwei Basen von \mathbb{R}^2

8.5. Selbstkontrollfragen

1. Geben Sie die Definition eines Vektorraums wieder.
2. Geben Sie die Definition des reellen Vektorraums wieder.
3. Es seien

$$x := \begin{pmatrix} 2 \\ 1 \end{pmatrix}, y := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ und } a := 2. \quad (8.55)$$

Berechnen Sie

$$v = a(x + y) \text{ und } w = \frac{1}{a}(y - x) \quad (8.56)$$

4. Geben Sie die Definition des Skalarproduktes auf \mathbb{R}^m wieder.
5. Für

$$x := \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, y := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, z := \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \quad (8.57)$$

berechnen Sie

$$\langle x, y \rangle, \langle x, z \rangle, \langle y, z \rangle \quad (8.58)$$

6. Geben Sie die Definition des Euklidischen Vektorraums wieder.
7. Geben Sie die Definition der Länge eines Vektors im Euklidischen Vektorraum wieder.
8. Berechnen Sie die Längen der Vektoren x, y, z aus Gleichung 8.57.
9. Geben Sie Definition des Abstands zweier Vektoren im Euklidischen Vektorraum wieder.
10. Berechnen Sie $d(x, y), d(x, z)$ und $d(y, z)$ für x, y, z aus Gleichung 8.57.
11. Geben Sie die Definition des Winkels zwischen zwei Vektoren im Euklidischen Vektorraum wieder.
12. Berechnen Sie die Winkel zwischen den Vektoren x und y , x und z , sowie y und z aus Gleichung 8.57.
13. Geben Sie die Definitionen der Orthogonalität und Orthonormalität von Vektoren wieder.
14. Geben Sie die Definition der Linearkombination von Vektoren wieder.
15. Geben Sie die Definition der linearen Unabhängigkeit von Vektoren wieder.
16. Woran kann man erkennen, ob zwei reelle Vektoren linear abhängig sind oder nicht?
17. Geben Sie die Definition der linearen Hülle einer Menge von Vektoren wieder.
18. Geben Sie die Definition der Basis eines Vektorraums wieder.
19. Geben Sie das Theorem zu den Eigenschaften von Vektorraumbasen wieder.
20. Geben Sie die Definition der Basisdarstellung eines Vektors wieder.

21. Geben Sie die Definition einer Orthonormalbasis von \mathbb{R}^m wieder.
22. Geben Sie die Definition der kanonischen Basis von \mathbb{R}^m wieder.

9. Matrizen

Matrizen sind die Worte der Sprache der modernen Datenanalyse. Ein Verständnis moderner datenanalytischer Verfahren und ihrer Implementation ist ohne ein Grundverständnis des Matrixbegriffs und ein Wissen um die grundlegenden Matrixoperationen nicht möglich. Matrizen können dabei sehr unterschiedliche Rollen spielen. So können Matrizen zum Beispiel Daten, experimentelle Designs und Modellparameter repräsentieren. Im Kontext der Linearen Algebra dienen Matrizen zur Repräsentation linearer Abbildungen und von Vektorräumen, hier werden Vektoren dann als spezielle Matrizen aufgefasst.

In diesem Kapitel geben wir eine Einführung zum Umgang mit Matrizen, wobei wir auf abstrakte Begrifflichkeiten der Linearen Algebra im Wesentlichen verzichten. Wir führen zunächst den Matrixbegriff ein und diskutieren dann mit der Matrixaddition, Matrixsubtraktion, Skalarmultiplikation und der Matrixtransposition erste grundlegende Matrixoperationen (Kapitel 9.1 und Kapitel 9.2). Wir führen dann die zentralen Begriffe der Matrixmultiplikation und der Matrixinversion ein (Kapitel 9.3 und Kapitel 9.4). Mit der Matrixdeterminante diskutieren wir dann in Kapitel 9.5 eine erste Maßzahl zur Beschreibung von Matrizen. Wir schließen in Kapitel 9.6 mit einer Übersicht zu besonders häufig auftretenden Matrizen.

9.1. Definition

Wir beginnen mit der Definition einer Matrix.

Definition 9.1. Eine Matrix ist eine rechteckige Anordnung von Zahlen, die wie folgt bezeichnet wird

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}. \quad (9.1)$$

•

Matrizen bestehen aus *Zeilen* (*rows*) und *Spalten* (*columns*). Die Matrixeinträge a_{ij} werden mit einem *Zeilenindex* i und einem *Spaltenindex* j indiziert. Zum Beispiel gilt für

$$A := \begin{pmatrix} 2 & 7 & 5 & 2 \\ 8 & 2 & 5 & 6 \\ 6 & 4 & 0 & 9 \\ 9 & 2 & 1 & 2 \end{pmatrix}, \quad (9.2)$$

dass $a_{32} = 4$. Die *Größe* oder *Dimension* einer Matrix ergibt sich aus der Anzahl ihrer Zeilen $n \in \mathbb{N}$ und Spalten $m \in \mathbb{N}$. Matrizen mit $n = m$ heißen *quadratische Matrizen*.

In der Folge benötigen wir nur Matrizen mit reellen Einträgen, also $a_{ij} \in \mathbb{R}$ für alle $i = 1, \dots, n$ und $j = 1, \dots, m$. Wir nennen die Matrizen mit reellen Einträge *reelle Matrizen* und bezeichnen die Menge der reellen Matrizen mit n Zeilen und m Spalten mit $\mathbb{R}^{n \times m}$. An dem Ausdruck

$$A \in \mathbb{R}^{n \times m} \quad (9.3)$$

können wir also ablesen, dass A eine reelle Matrix mit n Zeilen und m Spalten ist. Wir identifizieren dabei die Menge $\mathbb{R}^{1 \times 1}$ mit der Menge \mathbb{R} , die Menge $\mathbb{R}^{n \times 1}$ mit der Menge \mathbb{R}^n . Reelle Matrizen mit einer Spalte und n Zeilen entsprechen also n -dimensionalen reellen Vektoren und reelle Matrizen mit einer Spalte und einer Zeile entsprechen reellen Zahlen.

Definition von Matrizen in R

In **R** werden Matrizen definiert, indem **R** Vektoren mithilfe der `matrix()` Funktion in die Repräsentation einer mathematischen Matrix transformiert werden. Die Einträge eines **R** Vektors werden dabei anhand der spezifizierten Zeilenanzahl `nrow` anhand ihrer Gesamtanzahl auf die Matrix verteilt. Wollen wir beispielsweise die Matrix

$$A := \begin{pmatrix} 2 & 3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \quad (9.4)$$

in **R** definieren, so ergibt sich

```
# Spaltenweise Definition von A (R default)
A = matrix(c(2,1,3,6,0,5), nrow = 2)
print(A)
```

```
      [,1] [,2] [,3]
[1,]    2    3    0
[2,]    1    6    5
```

R folgt hier per default einer sogenannten *column-major-order*, das heißt, die Elemente des **R** Vektors `c(2,1,3,6,0,5)` werden der Reihe nach von oben nach unten in die Spalten der Matrix von links nach rechts überführt. Einen etwas klareren Zusammenhang zwischen dem visuellen Layout des **R** Codes und der resultierenden Matrix erhält man, indem man den **R** Vektor mithilfe von Zeilenumbrüchen anhand des intendierten Matrixlayouts formatiert und dann die *column-major-order* mithilfe des Arguments `byrow = TRUE` zu einer *row-major-order* umstellt. Es wird dann zunächst die erste Zeile der Matrix von links nach rechts mit den Elementen des **R** Vektors gefüllt wird und dann die zweite Zeile usw. bis alle Elemente des Vektors auf die Matrix verteilt sind.

```
# Reihenweise Definition von A (R default)
A = matrix(c(2,3,0,
             1,6,5),
           nrow = 2,
           byrow = TRUE)
print(A)
```

```
      [,1] [,2] [,3]
[1,]    2    3    0
[2,]    1    6    5
```

```
# Zeilenweise Definition von B
B = matrix(c(4,1,0,
             -4,2,0),
           nrow = 2,
           byrow = TRUE)
print(B)
```

```
      [,1] [,2] [,3]
[1,]    4    1    0
[2,]   -4    2    0
```

9.2. Grundlegende Matrixoperationen

Man kann mit Matrizen rechnen. Dabei sind folgende Matrixoperationen grundlegend:

- Die Addition von Matrizen gleicher Größe, genannt *Matrixaddition*,
- die Subtraktion von Matrizen gleicher Größe, genannt *Matrixsubtraktion*,
- die Multiplikation einer Matrix mit einem Skalar, genannt *Skalarmultiplikation*,
- das Vertauschen der Zeilen- und Spalten einer Matrix, genannt *Matrixtransposition*.

Wir führen diese Operationen in der Folge in Operatorform, also als Funktionen ein. Dies dient insbesondere dazu, bei jeder Operation mit Hilfe ihrer Definitionsmenge zu betonen, von welcher Art die Objekte der jeweiligen Operation sind und mithilfe ihrer Bildmenge zu betonen, von welcher Art das Resultat der jeweiligen Operation ist.

9.2.1. Matrixaddition

Definition 9.2. Es seien $A, B \in \mathbb{R}^{n \times m}$. Dann ist die *Addition* von A und B definiert als die Abbildung

$$+ : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto +(A, B) := A + B \quad (9.5)$$

mit

$$\begin{aligned} A + B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\ &:= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}. \end{aligned} \quad (9.6)$$

•

Die Definition der Matrixaddition legt insbesondere fest, dass nur Matrizen gleicher Größe addiert werden können und dass die Operation der Matrixaddition elementweise definiert ist.

Beispiel

Es seien $A, B \in \mathbb{R}^{2 \times 3}$ definiert als

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix}. \quad (9.7)$$

Da A und B gleich groß sind, können wir sie addieren

$$\begin{aligned} C = A + B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2+4 & -3+1 & 0+0 \\ 1-4 & 6+2 & 5+0 \end{pmatrix} \\ &= \begin{pmatrix} 6 & -2 & 0 \\ -3 & 8 & 5 \end{pmatrix}. \end{aligned} \quad (9.8)$$

In **R** führt man obige Rechnung wie folgt aus.

```
# Definition
A = matrix(c(2, -3, 0,
             1, 6, 5),
           nrow = 2,
           byrow = TRUE)
B = matrix(c(4, 1, 0,
             -4, 2, 0),
           nrow = 2,
           byrow = TRUE)

# Addition
C = A + B
print(C)
```

```
      [,1] [,2] [,3]
[1,]    6  -2    0
[2,]   -3    8    5
```

9.2.2. Matrixsubtraktion

Die Subtraktion von Matrizen gleicher Größe ist analog zur Addition definiert.

Definition 9.3 (Matrixsubtraktion). Es seien $A, B \in \mathbb{R}^{n \times m}$. Dann ist die *Subtraktion* von A und B definiert als die Abbildung

$$- : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto -(A, B) := A - B \quad (9.9)$$

mit

$$\begin{aligned} A - B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\ &:= \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \cdots & a_{1m} - b_{1m} \\ a_{21} - b_{21} & a_{22} - b_{22} & \cdots & a_{2m} - b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} - b_{n1} & a_{n2} - b_{n2} & \cdots & a_{nm} - b_{nm} \end{pmatrix}. \end{aligned} \quad (9.10)$$

•

Wie bei der Matrixaddition legt die Definition der Matrixsubtraktion fest, dass nur Matrizen gleicher Größe voneinander subtrahiert werden können und dass die Subtraktion zweier gleich großer Matrizen elementweise definiert ist.

Beispiel

Wir können die im Beispiel zur Matrixaddition definierten Matrizen A und B auch voneinander subtrahieren,

$$\begin{aligned} D = A - B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} - \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2-4 & -3-1 & 0-0 \\ 1+4 & 6-2 & 5-0 \end{pmatrix} \\ &= \begin{pmatrix} -2 & -4 & 0 \\ 5 & 4 & 5 \end{pmatrix}. \end{aligned} \quad (9.11)$$

In **R** führt man diese Rechnung wie folgt aus.

```
# Subtraktion
D = A - B
print(D)
```

```
      [,1] [,2] [,3]
[1,]   -2   -4    0
[2,]    5    4    5
```

9.2.3. Skalarmultiplikation

Die *Skalarmultiplikation* einer Matrix bezeichnet die Multiplikation eines Skalars mit einer Matrix.

Definition 9.4 (Skalarmultiplikation). Es sei $c \in \mathbb{R}$ ein Skalar und $A \in \mathbb{R}^{n \times m}$. Dann ist die *Skalarmultiplikation* von c und A definiert als die Abbildung

$$\cdot : \mathbb{R} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (c, A) \mapsto \cdot(c, A) := cA \quad (9.12)$$

mit

$$cA = c \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}. \quad (9.13)$$

•

Die Skalarmultiplikation ist mit dieser Definition also elementweise definiert.

Beispiel

Es seien $c := -3$ und $A \in \mathbb{R}^{4 \times 3}$ definiert als

$$A := \begin{pmatrix} 3 & 1 & 1 \\ 5 & 2 & 5 \\ 2 & 7 & 1 \\ 3 & 4 & 2 \end{pmatrix}. \quad (9.14)$$

Dann ergibt sich

$$B := cA = -3 \begin{pmatrix} 3 & 1 & 1 \\ 5 & 2 & 5 \\ 2 & 7 & 1 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} -3 \cdot 3 & -3 \cdot 1 & -3 \cdot 1 \\ -3 \cdot 5 & -3 \cdot 2 & -3 \cdot 5 \\ -3 \cdot 2 & -3 \cdot 7 & -3 \cdot 1 \\ -3 \cdot 3 & -3 \cdot 4 & -3 \cdot 2 \end{pmatrix} = \begin{pmatrix} -9 & -3 & -3 \\ -15 & -6 & -15 \\ -6 & -21 & -3 \\ -9 & -12 & -6 \end{pmatrix}. \quad (9.15)$$

In **R** führt man diese Skalarmultiplikation aus wie folgt.

```
# Definitionen
A = matrix(c(3,1,1,
             5,2,5,
             2,7,1,
             3,4,2),
           nrow = 4,
           byrow = TRUE)

c = -3

# Skalarmultiplikation
B = c*A
print(B)
```

	[,1]	[,2]	[,3]
[1,]	-9	-3	-3
[2,]	-15	-6	-15
[3,]	-6	-21	-3
[4,]	-9	-12	-6

Mithilfe der Definition von Matrixaddition und Skalarmultiplikation ist es möglich, einen Vektorraum zu definieren, dessen Elemente die reellen Matrizen sind. Insbesondere legt diese Definition auch die Rechenregeln beim Umgang mit Matrixaddition und Skalarmultiplikation fest.

Theorem 9.1 (Vektorraum der reellwertigen Matrizen). *Das Tripel $(\mathbb{R}^{n \times m}, +, \cdot)$ mit der oben definierten Matrixaddition und Skalarmultiplikation ist ein Vektorraum. Insbesondere gelten damit für $A, B, C \in \mathbb{R}^{n \times m}$ und $r, s, t \in \mathbb{R}$ folgende Rechenregeln:*

- (1) *Kommutativität der Addition: $A + B = B + A$.*
- (2) *Assoziativität der Addition: $(A + B) + C = A + (B + C)$.*
- (3) *Existenz eines neutralen Elements der Addition: $\exists 0 \in \mathbb{R}^{n \times m}$ mit $A + 0 = 0 + A = A$.*
- (4) *Existenz inverser Elemente der Addition: $\forall A \exists -A \in \mathbb{R}^{n \times m}$ mit $A + (-A) = 0$.*
- (5) *Existenz eines neutralen Elements der Skalarmultiplikation: $\exists 1 \in \mathbb{R}$ mit $1 \cdot A = A$.*
- (6) *Assoziativität der Skalarmultiplikation: $r \cdot (s \cdot t) = (r \cdot s) \cdot t$.*
- (7) *Distributivität hinsichtlich der Matrixaddition: $r \cdot (A + B) = r \cdot A + r \cdot B$.*
- (8) *Distributivität hinsichtlich der Skalaraddition: $(r + s) \cdot A = r \cdot A + s \cdot A$.*

◦

Wir verzichten auf einen Beweis, der sich mit einigem Notationsaufwand direkt aus dem elementweisen Charakter von Matrixaddition und Skalarmultiplikation sowie den aus dem Umgang mit den reellen Zahlen bekannten Rechenregeln ergibt. Das im Theorem erwähnte neutrale Element der Addition wird *Nullmatrix* genannt, wir werden dazu später eine allgemeine Notation einführen. Die inversen Elemente der Addition sind durch

$$-A := (-a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m} \quad (9.16)$$

gegeben und erlauben es, die Matrixsubtraktion als Spezialfall der Matrixaddition zu betrachten.

9.2.4. Matrixtransposition

Eine weitere häufig auftretende grundlegende Matrixoperation ist das Vertauschen der Zeilen- und Spaltenanordnung einer Matrix, genannt *Matrixtransposition*.

Definition 9.5 (Matrixtransposition). Es sei $A \in \mathbb{R}^{n \times m}$. Dann ist die *Transposition* von A definiert als die Abbildung

$$\cdot^T : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{m \times n}, A \mapsto \cdot^T(A) := A^T \quad (9.17)$$

mit

$$A^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}^T := \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}. \quad (9.18)$$

•

Für $A \in \mathbb{R}^{n \times m}$ gilt damit also immer $A^T \in \mathbb{R}^{m \times n}$. Weiterhin gelten folgende Rechenregeln der Matrixtransposition, wie man sich an Beispielen klar macht:

(1) Für $A \in \mathbb{R}^{1 \times 1}$ gilt

$$A^T = A. \quad (9.19)$$

(2) Es gilt

$$(A^T)^T = A. \quad (9.20)$$

(3) Es gilt

$$(a_{ii})_{1 \leq i \leq \min(n,m)} = (a_{ii})_{1 \leq i \leq \min(n,m)}^T. \quad (9.21)$$

Letztere Eigenschaft der Transposition besagt, dass die Elemente auf der Hauptdiagonalen einer Matrix bei Transposition unberührt bleiben.

Beispiel

Es sei $A \in \mathbb{R}^{2 \times 3}$ definiert durch

$$A := \begin{pmatrix} 2 & 3 & 0 \\ 1 & 6 & 5 \end{pmatrix}, \quad (9.22)$$

Dann gilt $A^T \in \mathbb{R}^{3 \times 2}$ und speziell

$$A^T := \begin{pmatrix} 2 & 1 \\ 3 & 6 \\ 0 & 5 \end{pmatrix}. \quad (9.23)$$

Weiterhin gilt offenbar $\min(m, n) = 2$ und folglich

$$(a_{11}) = (a_{11})^T \text{ und } (a_{22}) = (a_{22})^T. \quad (9.24)$$

In **R** führt man die Transposition einer Matrix wie folgt durch.

```
# Definition
A = matrix(c(2,3,0,
             1,6,5),
           nrow = 2,
           byrow = TRUE)
print(A)
```

```
      [,1] [,2] [,3]
[1,]    2    3    0
[2,]    1    6    5
```

```
# Transposition
AT = t(A)
print(AT)
```

```
      [,1] [,2]
[1,]    2    1
[2,]    3    6
[3,]    0    5
```

Schließlich gelten in der Verbindung mit der Matrixaddition, Matrixsubtraktion und der Skalarmultiplikation folgende Rechenregeln, wie man sich an Beispielen klar macht:

(1) Für $A, B \in \mathbb{R}^{n \times m}$ gilt

$$(A + B)^T = A^T + B^T. \quad (9.25)$$

(2) Für $A, B \in \mathbb{R}^{n \times m}$ gilt

$$(A - B)^T = A^T - B^T. \quad (9.26)$$

(3) Für $c \in \mathbb{R}$ und $A \in \mathbb{R}^{n \times m}$ gilt

$$(cA)^T = cA^T. \quad (9.27)$$

9.3. Matrixmultiplikation

Die Matrixmultiplikation ist die zentrale Operation beim Rechnen mit Matrizen. Sie ist definiert wie folgt.

Definition 9.6 (Matrixmultiplikation). Es seien $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times k}$. Dann ist die *Matrixmultiplikation* von A und B definiert als die Abbildung

$$\cdot : \mathbb{R}^{n \times m} \times \mathbb{R}^{m \times k} \rightarrow \mathbb{R}^{n \times k}, (A, B) \mapsto \cdot(A, B) := AB \quad (9.28)$$

mit

$$\begin{aligned} AB &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{pmatrix} \\ &:= \begin{pmatrix} \sum_{i=1}^m a_{1i}b_{i1} & \sum_{i=1}^m a_{1i}b_{i2} & \cdots & \sum_{i=1}^m a_{1i}b_{ik} \\ \sum_{i=1}^m a_{2i}b_{i1} & \sum_{i=1}^m a_{2i}b_{i2} & \cdots & \sum_{i=1}^m a_{2i}b_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ni}b_{i1} & \sum_{i=1}^m a_{ni}b_{i2} & \cdots & \sum_{i=1}^m a_{ni}b_{ik} \end{pmatrix} \\ &= \left(\sum_{i=1}^m a_{ji}b_{il} \right)_{1 \leq j \leq n, 1 \leq l \leq k} \end{aligned} \quad (9.29)$$

•

Das Matrixprodukt AB ist also nur dann definiert, wenn A genau so viele Spalten hat wie B Zeilen hat. Informell gilt für die beteiligten Matrixgrößen dabei die Merkregel

$$(n \times m)(m \times k) = (n \times k). \quad (9.30)$$

Der Eintrag $(AB)_{ij}$ in AB entspricht der Summe der multiplizierten i ten Zeile von A und j ten Spalte von B . Zum Berechnen von $(AB)_{ij}$ geht man für $i = 1, \dots, n$ und $j = 1, \dots, k$ also in Gedanken wie folgt vor:

- (1) Man legt die Transposition der i ten Zeile von A über die j te Spalte von B .
- (2) Weil A genau m Spalten hat und B genau m Zeilen hat, gibt es dann zu jedem Element der Zeile aus A ein korrespondierendes Element in der Spalte von B .
- (3) Man multipliziert die korrespondierenden Elemente miteinander.
- (4) Die Summe dieser Produkte ist dann der Eintrag mit Index ij in AB .

Beispiel

$A \in \mathbb{R}^{2 \times 3}$ und $B \in \mathbb{R}^{3 \times 2}$ seien definiert als

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix}. \quad (9.31)$$

Wir wollen $C := AB$ und $D := BA$ berechnen. Mit $n = 2, m = 3$ und $k = 2$ wissen wir schon, dass $C \in \mathbb{R}^{2 \times 2}$ und $D \in \mathbb{R}^{3 \times 3}$, weil

$$(2 \times 3)(3 \times 2) = (2 \times 2) \quad (9.32)$$

und

$$(3 \times 2)(2 \times 3) = (3 \times 3). \quad (9.33)$$

Es gilt hier also sicher $AB \neq BA$. Für C ergibt sich dann

$$\begin{aligned} C &= AB \\ &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 2 \cdot 4 + (-3) \cdot (-1) + 0 \cdot 1 & 2 \cdot 2 + (-3) \cdot 0 + 0 \cdot 3 \\ 1 \cdot 4 + 6 \cdot (-1) + 5 \cdot 1 & 1 \cdot 2 + 6 \cdot 0 + 5 \cdot 3 \end{pmatrix} \\ &= \begin{pmatrix} 8 + 3 + 0 & 4 + 0 + 0 \\ 4 - 6 + 5 & 2 + 0 + 15 \end{pmatrix} \\ &= \begin{pmatrix} 11 & 4 \\ 3 & 17 \end{pmatrix}. \end{aligned} \quad (9.34)$$

In **R** nutzt man für die Matrixmultiplikation den `%%` Operator.

```
# Definitionen
A = matrix(c(2,-3,0,
            1, 6,5),
           nrow = 2,
           byrow = TRUE)
B = matrix(c( 4,2,
            -1,0,
            1,3),
           nrow = 3,
           byrow = TRUE)

# Matrixmultiplikation
C = A %% B
print(C)
```

```
      [,1] [,2]
[1,]   11    4
[2,]    3   17
```

Für D ergibt sich weiterhin

$$\begin{aligned} D &= BA \\ &= \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 4 \cdot 2 + 2 \cdot 1 & 4 \cdot (-3) + 2 \cdot 6 & 4 \cdot 0 + 2 \cdot 5 \\ (-1) \cdot 2 + 0 \cdot 1 & (-1) \cdot (-3) + 0 \cdot 6 & (-1) \cdot 0 + 0 \cdot 5 \\ 1 \cdot 2 + 3 \cdot 1 & 1 \cdot (-3) + 3 \cdot 6 & 1 \cdot 0 + 3 \cdot 5 \end{pmatrix} \\ &= \begin{pmatrix} 8 + 2 & -12 + 12 & 0 + 5 \\ -2 + 0 & 3 + 0 & 0 + 0 \\ 2 + 3 & -3 + 18 & 0 + 15 \end{pmatrix} \\ &= \begin{pmatrix} 10 & 0 & 10 \\ -2 & 3 & 0 \\ 5 & 15 & 15 \end{pmatrix} \end{aligned} \quad (9.35)$$

In **R** überprüft man diese Rechnung wie folgt.

```
# Definitionen
A = matrix(c(2,-3,0,
            1, 6,5),
          nrow = 2,
          byrow = TRUE)
B = matrix(c( 4,2,
            -1,0,
            1,3),
          nrow = 3,
          byrow = TRUE)

# Matrixmultiplikation
D = B %*% A
print(D)
```

```
      [,1] [,2] [,3]
[1,]   10    0   10
[2,]   -2    3    0
[3,]    5   15   15
```

Ist allerdings eine Matrixmultiplikation aufgrund nicht-adäquater Matrizengrößen nicht definiert, so lässt sich diese auch nicht numerisch auswerten.

```
# Beispiel für eine undefinierte Matrixmultiplikation
E = t(A) %*% B      # (3 x 2)(3 x 2)
```

```
Error in t(A) %*% B: non-conformable arguments
```

Folgendes Theorem, das wir nicht beweisen wollen, stellt den Bezug zwischen dem Skalarprodukt zweier Vektoren und der Multiplikation zweier Matrizen her. Dieser ergibt sich im Wesentlichen durch die Identifikation von \mathbb{R}^n und $\mathbb{R}^{n \times 1}$ und der Tatsache, dass nach Definition der Eintrag $(AB)_{ij}$ im Produkt von $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times k}$ dem Vektorskalarprodukt der i ten Spalte von A^T und der j ten Spalte von B entspricht.

Theorem 9.2 (Matrixmultiplikation und Vektorskalarprodukt). *Es seien $x, y \in \mathbb{R}^n$. Dann gilt*

$$\langle x, y \rangle = x^T y. \quad (9.36)$$

Weiterhin seien für $A \in \mathbb{R}^{n \times m}$ für $i = 1, \dots, n$

$$\bar{a}_i := (a_{ji})_{1 \leq j \leq m} \in \mathbb{R}^m \quad (9.37)$$

die Spalten von A^T und für $B \in \mathbb{R}^{m \times k}$ für $i = 1, \dots, k$

$$\bar{b}_j := (b_{ij})_{1 \leq i \leq m} \in \mathbb{R}^m \quad (9.38)$$

die Spalten von B , also

$$A^T = (\bar{a}_1 \quad \bar{a}_2 \quad \dots \quad \bar{a}_n) \in \mathbb{R}^{m \times n} \text{ und } B = (\bar{b}_1 \quad \bar{b}_2 \quad \dots \quad \bar{b}_k) \in \mathbb{R}^{m \times k}. \quad (9.39)$$

Dann gilt

$$AB = (\langle \bar{a}_i, \bar{b}_j \rangle)_{1 \leq i \leq n, 1 \leq j \leq k}. \quad (9.40)$$

◦

9.3.1. Rechenregeln der Matrixmultiplikation

Im Folgenden stellen wir einige grundlegende Rechenregeln der Matrixmultiplikation, insbesondere auch in Kombination mit anderen Matrixoperationen zusammen.

Für Beweise der folgenden zwei Theoreme zur Assoziativität und Distributivität, die sich im Wesentlichen mit den entsprechenden Rechenregeln für Summen und Produkte der reellen Zahlen ergeben, verweisen wir auf die weiterführende Literatur.

Theorem 9.3 (Assoziativität). *Es seien $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{m \times k}$, $C \in \mathbb{R}^{k \times p}$ und $c \in \mathbb{R}$. Dann gelten*

(1) *Die Multiplikation von Matrizen ist assoziativ, es gilt*

$$A(BC) = (AB)C. \quad (9.41)$$

(2) *Die Kombination von Matrizenmultiplikation und Skalarmultiplikation ist assoziativ,*

$$c(AB) = (cA)B = A(cB). \quad (9.42)$$

◦

Die Assoziativität von Matrizenmultiplikation und Skalarmultiplikation erkennt man leicht bei Betrachtung des j, l ten Elements von $c(AB)$, $(cA)B$ und $A(cB)$ anhand von

$$c \left(\sum_{i=1}^m a_{ji} b_{il} \right) = \sum_{i=1}^m (ca_{ji}) b_{il} = \sum_{i=1}^m a_{ji} (cb_{il}). \quad (9.43)$$

Theorem 9.4 (Distributivität). *Es seien $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times p}$. Dann gelten*

$$(A + B)C = AC + BC \quad (9.44)$$

und

$$C^T(A + B) = C^T A + C^T B \quad (9.45)$$

◦

Im Gegensatz zur Kommutativität der Multiplikation reeller Zahlen ist die Matrixmultiplikation im Allgemeinen nicht kommutativ.

Theorem 9.5 (Nichtkommutativität). *Es seien $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times p}$. Dann gilt im Allgemeinen*

$$AB \neq BA. \quad (9.46)$$

◦

Beweis. Im Fall $p \neq n$ ist BA nicht definiert, wir betrachten also nur den Fall $p = n$. Wir zeigen durch Angabe eines Gegenbeispiels mit $A, B \in \mathbb{R}^{2 \times 2}$, dass im Allgemeinen $AB = BA$ nicht gilt. Es seien

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ und } B := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (9.47)$$

Dann gilt

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = BA. \quad (9.48)$$

□

Theorem 9.6 (Kombination von Matrixmultiplikation und Transposition). *Es seien $A \in \mathbb{R}^{m \times n}$ und $B \in \mathbb{R}^{n \times k}$. Dann gilt*

$$(AB)^T = B^T A^T. \quad (9.49)$$

◦

Beweis. Ein Beweis ergibt sich wie folgt

$$\begin{aligned} (AB)^T &= \left(\left(\sum_{i=1}^m a_{ji} b_{il} \right)_{1 \leq j \leq n, 1 \leq l \leq k} \right)^T \\ &= \left(\sum_{i=1}^m a_{ij} b_{li} \right)_{1 \leq i \leq k, 1 \leq j \leq n} \\ &= \left(\sum_{i=1}^m b_{li} a_{ij} \right)_{1 \leq j \leq k, 1 \leq l \leq n} \\ &= B^T A^T. \end{aligned} \quad (9.50)$$

□

9.4. Matrixinversion

Um den Begriff der inversen Matrix zu motivieren, betrachten wir zunächst das Problem des *Lösens eines linearen Gleichungssystems*. Dazu seien $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$ und $b \in \mathbb{R}^n$ und es gelte

$$Ax = b. \quad (9.51)$$

A und b seien als bekannt vorausgesetzt, x sei unbekannt. Konkret seien beispielsweise

$$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ und } b := \begin{pmatrix} 5 \\ 11 \end{pmatrix}. \quad (9.52)$$

Dann liegt folgendes lineares Gleichungssystem mit zwei Gleichungen und zwei Unbekannten vor:

$$Ax = b \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \Leftrightarrow \begin{array}{rcl} 1x_1 + 2x_2 & = & 5 \\ 3x_1 + 4x_2 & = & 11 \end{array} \quad (9.53)$$

Ziel des Lösens von linearen Gleichungssystemen ist bekanntlich, herauszufinden, für welche x das Gleichungssystem erfüllt ist. Um in diesem Kontext den Begriff der inversen Matrix von A einzuführen, vereinfachen wir die Situation weiter. Wir nehmen an, dass $A = a$ eine 1×1 Matrix, also ein Skalar, sei und ebenso x und b , dass wir also für $a, x, b \in \mathbb{R}$ die Gleichung

$$ax = b \quad (9.54)$$

haben. Um diese Gleichung nach x aufzulösen würde man natürlich beide Seiten der Gleichung mit dem *multiplikativem Inversen* von a multiplizieren, wobei das *multiplikative Inverse* von a den Wert bezeichnet, der mit a multipliziert 1 ergibt. Dieser ist bekanntlich durch

$$a^{-1} = \frac{1}{a} \quad (9.55)$$

gegeben. Dann würde gelten

$$ax = b \Leftrightarrow a^{-1}ax = a^{-1}b \Leftrightarrow 1 \cdot x = a^{-1}b \Leftrightarrow x = \frac{b}{a}. \quad (9.56)$$

Ganz konkret etwa

$$2x = 6 \Leftrightarrow 2^{-1}2x = 2^{-1}6 \Leftrightarrow \frac{1}{2}2x = \frac{1}{2}6 \Leftrightarrow x = 3. \quad (9.57)$$

Analog zu dem Fall, dass die Matrizen in $Ax = b$ allesamt Skalare sind, möchte man im Fall eines linearen Gleichungssystems beide Seiten der Gleichung mit dem *multiplikativen Inversen* A^{-1} von A multiplizieren können, sodass eine Gleichung der Form

$$A^{-1}A = "1". \quad (9.58)$$

resultiert. Dann hätte man nämlich

$$Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b. \quad (9.59)$$

Diese intuitive Idee des multiplikativen Inversen einer Matrix A wird im Folgenden unter dem Begriff der *inversen Matrix* formalisiert. Dazu benötigen wir zunächst den Begriff der *Einheitsmatrix*.

Definition 9.7 (Einheitsmatrix). Die Matrix

$$I_n := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \quad (9.60)$$

mit $a_{ij} = 1$ für $i = j$ und $a_{ij} = 0$ für $i \neq j$ heißt *n-dimensionale Einheitsmatrix*.

•

In **R** wird I_n mit dem Befehl `diag(n)` erzeugt. Die Einheitsmatrix ist für die Matrixmultiplikation das Analog zur 1 bei der Multiplikation reeller Zahlen. Das ist die Aussage folgenden Theorems.

Theorem 9.7 (Neutrales Element der Matrixmultiplikation). I_n ist das neutrale Element der Matrixmultiplikation, das heißt es gilt für $A \in \mathbb{R}^{n \times m}$, dass

$$I_n A = A \text{ und } A I_m = A. \quad (9.61)$$

◦

Beweis. Es sei $B = (b_{ij}) = I_n A \in \mathbb{R}^{n \times m}$. Dann gilt für alle $1 \leq i \leq n$ und alle $1 \leq j \leq m$

$$b_{ij} = 0 \cdot a_{1j} + 0 \cdot a_{2j} + \cdots + 0 \cdot a_{i-1,j} + 1 \cdot a_{ij} + \cdots + 0 \cdot a_{i+1,j} + 0 \cdot a_{nj} = a_{ij}. \quad (9.62)$$

Analog zeigt man dies für $A I_m$. □

Mit dem Begriff der Einheitsmatrix können wir jetzt die Begriffe der inversen Matrix und der invertierbaren Matrix definieren:

Definition 9.8 (Invertierbare Matrix und inverse Matrix). Eine quadratische Matrix $A \in \mathbb{R}^{n \times n}$ heißt *invertierbar*, wenn es eine quadratische Matrix $A^{-1} \in \mathbb{R}^{n \times n}$ gibt, so dass

$$A^{-1}A = AA^{-1} = I_n \quad (9.63)$$

ist. Die Matrix A^{-1} heißt die *inverse Matrix von A*.

•

Man beachte, dass sich die Begriffe der inversen Matrix und der Invertierbarkeit *nur* auf quadratische Matrizen beziehen. Insbesondere können quadratische Matrizen invertierbar sein, müssen es aber nicht sein (lineare Gleichungssysteme können also Lösungen haben, müssen es aber nicht). Nicht invertierbare Matrizen nennt man auch *singuläre* Matrizen, invertierbare Matrizen manchmal auch *nicht-singuläre* Matrizen. Schließlich beachte man, dass Definition 9.8 lediglich aussagt, was eine inverse Matrix ist, aber nicht wie man sie berechnet.

Beispiel für eine invertierbare Matrix

Die Matrix

$$A := \begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix} \quad (9.64)$$

ist invertierbar und ihre inverse Matrix ist gegeben durch

$$A^{-1} = \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix}, \quad (9.65)$$

denn

$$\begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix} \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 2.0 & 1.0 \\ 3.0 & 4.0 \end{pmatrix}, \quad (9.66)$$

wovon man sich durch Nachrechnen überzeugt.

Beispiel für eine nicht-invertierbare Matrix

Die Matrix

$$B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (9.67)$$

ist nicht invertierbar, denn wäre B invertierbar, dann gäbe es

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (9.68)$$

mit

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (9.69)$$

Das würde aber bedeuten, dass $0 = 1$ in \mathbb{R} und das ist ein Widerspruch. Also kann B nicht invertierbar sein.

Zum Berechnen inverser Matrizen

2×2 bis etwa 5×5 Matrizen kann man prinzipiell per Hand invertieren, dazu stellt die Lineare Algebra verschiedene Verfahren bereit. Wir wollen hier auf eine Einführung in die Matrizeninvertierung per Hand verzichten, da in der Anwendung Matrizen standardmäßig

numerisch invertiert werden. Die numerische Matrixinversion ist dann auch ein großes Feld der Forschung zur Numerischen Mathematik, die eine Vielzahl von Algorithmen zu diesem Zweck bereitstellt. In **R** werden Matrizen per default mit der Funktion `solve()`, in Anlehnung an das Lösen linearer Gleichungssysteme, invertiert. Für das obige Beispiel einer invertierbaren Matrix ergibt sich dabei folgender **R** Code.

```
# Definition
A = matrix(c(2,1,
             3,4),
           nrow = 2,
           byrow = TRUE)

# Berechnen von A^{-1}
print(solve(A))
```

```
      [,1] [,2]
[1,]  0.8 -0.2
[2,] -0.6  0.4
```

```
# Überprüfen der Eigenschaften einer inversen Matrix
print(solve(A) %*% A)
```

```
      [,1] [,2]
[1,] 1.000000e+00 0
[2,] -1.110223e-16 1
```

```
# Bei der umgekehrten Berechnung ergeben sich kleine Rundungsfehler
print(A %*% solve(A))
```

```
      [,1] [,2]
[1,]  1  0
[2,]  0  1
```

Nicht-invertierbare Matrizen sind dabei natürlich auch numerisch nicht-invertierbar, wie folgende Fehlermeldung in **R** bezüglich obigen Beispiels einer nicht-invertierbaren Matrix demonstriert.

```
B = matrix(c(1,0,
             0,0),
           nrow = 2,
           byrow = 2)
solve(B)
```

```
Error in solve.default(B): Lapack routine dgesv: system is exactly singular: U[2,2] = 0
```

9.5. Determinanten

Die Determinante ist eine vielseitig einsetzbare Maßzahl einer quadratischen Matrix. Für das Verständnis der *Eigenanalyse* und der *Matrixzerlegung* ist der Begriff der Determinante im Kontext des *charakteristischen Polynoms* grundlegend.

Allgemein ist eine Determinante eine nichtlineare Abbildung der Form

$$|\cdot| : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}, A \mapsto |A|, \quad (9.70)$$

das heißt, eine Determinante ordnet einer quadratischen Matrix A die reelle Zahl $|A|$ zu. Die Zahl $|A|$ wird dabei rekursiv anhand folgender Definition bestimmt.

Definition 9.9 (Determinante). Für $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ mit $n > 1$ sei $A_{ij} \in \mathbb{R}^{(n-1) \times (n-1)}$ die Matrix, die aus A durch Entfernen der i ten Zeile und der j ten Spalte entsteht. Dann heißt die Zahl

$$|A| := a_{11} \quad \text{für } n = 1 \quad (9.71)$$

$$|A| := \sum_{j=1}^n a_{1j}(-1)^{1+j} \det(A_{1j}) \quad \text{für } n > 1 \quad (9.72)$$

die *Determinante* von A .

•

Die Definition führt die Bestimmung der Determinante einer quadratischen Matrix also sukzessive durch Streichen von Zeilen und Spalten auf die Determinante einer 1×1 Matrix zurück, die durch ihr einziges Element gegeben ist. Für

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \quad (9.73)$$

ergeben sich dabei zum Beispiel folgende Matrizen der Form $A_{ij} \in \mathbb{R}^{(3-1) \times (3-1)}$:

$$A_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, A_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, A_{21} = \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}, A_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}. \quad (9.74)$$

Für die Berechnung der Determinanten von zwei- und dreidimensionalen quadratischen Matrizen gibt es direkte, nicht-rekursive Rechenregeln, die in folgendem Theorem festgehalten sind.

Theorem 9.8 (Determinanten von zwei- und dreidimensionalen Matrizen).

Es sei $A = (a_{ij})_{1 \leq i, j \leq 2} \in \mathbb{R}^{2 \times 2}$. Dann gilt

$$|A| = a_{11}a_{22} - a_{12}a_{21}. \quad (9.75)$$

Es sei $A = (a_{ij})_{1 \leq i, j \leq 3} \in \mathbb{R}^{3 \times 3}$. Dann gilt

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \quad (9.76)$$

◦

Beweis. Für $A \in \mathbb{R}^{2 \times 2}$ gilt nach Definition

$$\begin{aligned} |A| &= \sum_{j=1}^2 a_{1j}(-1)^{1+j} |A_{1j}| \\ &= a_{11}(-1)^{1+1} |A_{11}| + a_{12}(-1)^{1+2} |A_{12}| \\ &= a_{11}|(a_{22})| - a_{12}|(a_{21})| \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned} \quad (9.77)$$

Für $A \in \mathbb{R}^{3 \times 3}$ gilt nach Definition und mit der Formel für Determinanten von 2×2 Matrizen

$$\begin{aligned}
 |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\
 &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| + a_{13}(-1)^{1+3}|A_{13}| \\
 &= a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| \\
 &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
 &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\
 &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\
 &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}.
 \end{aligned} \tag{9.78}$$

□

Für die Bestimmung der Determinanten von 2×2 und 3×3 Matrizen gilt somit die sogenannte *Sarrusche Merkregel*:

“Summe der Produkte auf den Diagonalen minus Summe der Produkte auf den Gegendiagonalen.”

Dabei bezieht sich die Merkregeln bei 3×3 Matrizen auf das Schema

$$\left(\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \right). \tag{9.79}$$

Beispiele für Determinanten von 2×2 und 3×3 Matrizen

Es seien

$$A := \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ und } C := \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \tag{9.80}$$

Dann ergeben sich

$$|A| = 2 \cdot 4 - 1 \cdot 3 = 8 - 3 = 5 \tag{9.81}$$

und

$$|B| = 1 \cdot 0 - 0 \cdot 0 = 0 - 0 = 0 \tag{9.82}$$

und

$$|C| = 2 \cdot 1 \cdot 3 + 0 \cdot 0 \cdot 0 + 0 \cdot 0 \cdot 0 - 0 \cdot 0 \cdot 3 - 0 \cdot 0 \cdot 0 - 0 \cdot 1 \cdot 0 = 2 \cdot 1 \cdot 3 = 6. \tag{9.83}$$

In **R** rechnet man dies mithilfe der `det()` Funktion wie folgt nach.

```
# Matrixdefinition und Determinantenberechnung
A = matrix(c(2,1,
             3,4),
           nrow = 2,
           byrow = TRUE)
det(A)
```

```
[1] 5
```

```
# Matrixdefinition und Determinantenberechnung
B = matrix(c(1,0,
             0,0),
           nrow = 2,
           byrow = TRUE)
det(B)
```

```
[1] 0
```

```
# Matrixdefinition und Determinantenberechnung
C = matrix(c(2,0,0,
             0,1,0,
             0,0,3),
           nrow = 3,
           byrow = TRUE)
det(C)
```

[1] 6

Für Determinanten bestehen zahlreiche Rechenregeln im Zusammenspiel mit Matrixmultiplikation und Matrixinversion. Ohne Beweis stellen wir diese in folgendem Theorem zusammen.

Theorem 9.9 (Rechenregeln für Determinanten).

(Determinantenmultiplikationssatz). Für $A, B \in \mathbb{R}^{n \times n}$ gilt

$$|AB| = |A||B|. \quad (9.84)$$

(Transposition). Für $A \in \mathbb{R}^{n \times n}$ gilt

$$|A| = |A^T|. \quad (9.85)$$

(Inversion). Für eine invertierbare Matrix $A \in \mathbb{R}^{n \times n}$ gilt

$$|A^{-1}| = \frac{1}{|A|}. \quad (9.86)$$

(Dreiecksmatrizen). Für Matrizen $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ mit $a_{ij} = 0$ für $i > j$ oder $a_{ij} = 0$ für $j > i$ gilt

$$|A| = \prod_{i=1}^n a_{ii}. \quad (9.87)$$

◦

Folgendes sehr tiefgehendes Theorem, welches wir nicht vollständig beweisen wollen, gibt eine Möglichkeit an, anhand der Determinante einer quadratischen Matrix zu bestimmen, ob sie invertierbar ist.

Theorem 9.10. $A \in \mathbb{R}^{n \times n}$ ist dann und nur dann invertierbar, wenn gilt, dass $|A| \neq 0$. Es gilt also

$$A \text{ ist invertierbar} \Leftrightarrow |A| \neq 0 \text{ und } A \text{ ist nicht invertierbar} \Leftrightarrow |A| = 0. \quad (9.88)$$

◦

Beweis. Wir deuten einen Beweis lediglich an und zeigen, dass aus der Invertierbarkeit von A folgt, dass $|A|$ nicht gleich Null sein kann. Nehmen wir also an, dass A invertierbar ist. Dann gibt es eine Matrix B mit $AB = I_n$ und mit dem Determinantenmultiplikationssatz folgt

$$|AB| = |A||B| = |I_n| = 1. \quad (9.89)$$

Also kann $|A| = 0$ nicht gelten, denn sonst wäre $0 = 1$. □

Visuelle Intuition

Der abstrakte Begriff der Determinante einer quadratischen Matrix kann mithilfe des Vektorraumbegriffs etwas veranschaulicht werden. Dazu seien $a_1, \dots, a_n \in \mathbb{R}^n$ die Spalten von $A \in \mathbb{R}^{n \times n}$. Dann gilt (wie wir nicht beweisen wollen), dass $|A|$ dem signierten Volumen des von $a_1, \dots, a_n \in \mathbb{R}^n$ aufgespannten Parallelotops entspricht. Um dies visuell zu veranschaulichen betrachten wir die Matrizen

$$A_1 = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, A_3 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \quad (9.90)$$

mit den jeweiligen Determinanten

$$|A_1| = 3 \cdot 2 - 1 \cdot 1 = 5, \quad |A_2| = 2 \cdot 2 - 0 \cdot 0 = 4, \quad |A_3| = 2 \cdot 2 - 2 \cdot 2 = 0. \quad (9.91)$$

Abbildung 9.1 visualisiert die entsprechende Intuition.

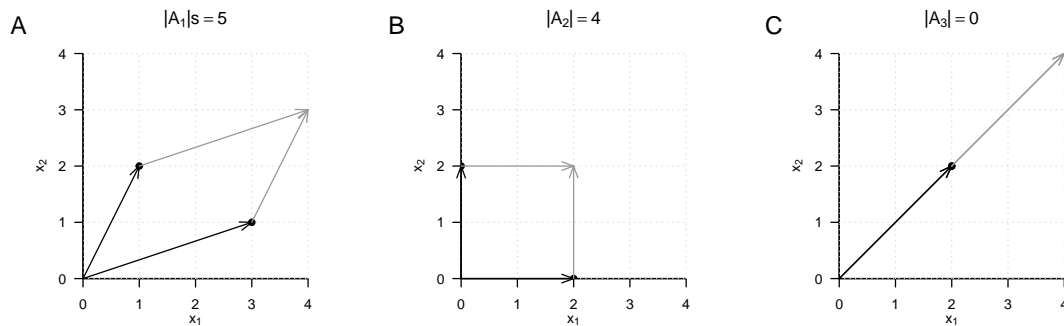


Abbildung 9.1. Determinanten als Parallelotopvolumina.

9.6. Spezielle Matrizen

In dieser Sektion stellen wir einige häufig auftretende Typen von Matrizen und ihre Eigenschaften zusammen. Zum Beweis der allermeisten Eigenschaften verweisen wir dabei auf die weiterführende Literatur.

9.6.1. Einheitsmatrizen

Die Einheitsmatrix und die Einheitsvektoren haben wir bereits kennengelernt. Wir fassen sie hier noch einmal in einer gemeinsamen Definition zusammen.

Definition 9.10 (Einheitsmatrix und Einheitsvektoren). Wir bezeichnen die *Einheitsmatrix* mit

$$I_n := (i_{jk})_{1 \leq j \leq n, 1 \leq k \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ für } j = k \text{ und } i_{jk} = 0 \text{ für } j \neq k. \quad (9.92)$$

Wir bezeichnen die *Einheitsvektoren* $e_i, i = 1, \dots, n$ mit

$$e_i := (e_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{ij} = 1 \text{ für } i = j \text{ und } e_{ij} = 0 \text{ für } i \neq j. \quad (9.93)$$

•

Die Einheitsmatrix I_n besteht nur aus Nullen und Diagonalelementen gleich Eins, die Einheitsvektoren bestehen nur aus Nullen und einer Eins in der jeweils indizierten Komponente. Es gilt

$$I_n = (e_1 \quad \cdots \quad e_n) \in \mathbb{R}^{n \times n} \quad (9.94)$$

Für $n = 3$ gilt also zum Beispiel

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ und } e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (9.95)$$

Weiterhin gelten für die Einheitsvektoren bekanntlich für $1 \leq i, j \leq n$

$$e_i^T e_j = 0 \text{ für } i \neq j, e_i^T e_i = 1 \text{ und } e_i^T v = v^T e_i = v_i \text{ für } v \in \mathbb{R}^n. \quad (9.96)$$

9.6.2. Einsmatrizen und Nullmatrizen

Definition 9.11 (Nullmatrizen, Nullvektoren, Einsmatrizen, Einsvektoren). Wir bezeichnen *Nullmatrizen* und *Nullvektoren* mit

$$0_{nm} := (0)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n. \quad (9.97)$$

Wir bezeichnen *Einsmatrizen* und *Einsvektoren* mit

$$1_{nm} := (1)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 1_n := (1)_{1 \leq i \leq n} \in \mathbb{R}^n. \quad (9.98)$$

•

0_{nm} und 0_n bestehen also nur aus Nullen und 1_{nm} und 1_n bestehen nur aus Einsen. Es gilt also beispielsweise

$$0_{32} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, 0_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, 1_{32} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \text{ und } 1_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad (9.99)$$

Weiterhin gelten zum Beispiel

$$0_n 0_n^T = 0_{nn} \text{ und } 1_n 1_n^T = 1_{nn}, \quad (9.100)$$

wovon man sich durch Nachrechnen überzeugt.

9.6.3. Diagonalmatrizen

Definition 9.12 (Diagonalmatrix). Eine Matrix $D \in \mathbb{R}^{n \times m}$ heißt *Diagonalmatrix*, wenn $d_{ij} = 0$ für $1 \leq i \leq n, 1 \leq j \leq m$ mit $i \neq j$.

•

Eine quadratische Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ mit den Diagonalelementen $d_1, \dots, d_n \in \mathbb{R}$ schreibt man auch als

$$D = \text{diag}(d_1, \dots, d_n). \quad (9.101)$$

Zum Beispiel gelten

$$D := \text{diag}(1, 2, 3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad (9.102)$$

und für $\sigma^2 \in \mathbb{R}$

$$\Sigma = \text{diag}(\sigma^2, \sigma^2, \sigma^2) = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_3. \quad (9.103)$$

In folgendem Theorem stellen wir einige wichtige Eigenschaften von quadratischen Diagonalmatrizen zusammen.

Theorem 9.11 (Eigenschaften quadratischer Diagonalmatrizen).

(Determinante.) $D := \text{diag}(d_1, \dots, d_n) \in \mathbb{R}^{n \times n}$ sei eine quadratische Diagonalmatrix. Dann gilt

$$|D| = \prod_{i=1}^n d_i. \quad (9.104)$$

◦

9.6.4. Symmetrische Matrizen

Symmetrische Matrizen sind quadratische Matrizen, die bei Transposition unverändert bleiben:

Definition 9.13. Eine Matrix $S \in \mathbb{R}^{n \times n}$ heißt *symmetrisch*, wenn $S^T = S$.

•

Ein Beispiel für eine symmetrische Matrix ist

$$S := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}. \quad (9.105)$$

In folgendem Theorem stellen wir einige wichtige Eigenschaften symmetrischer Matrizen zusammen.

Theorem 9.12 (Eigenschaften symmetrischer Matrizen).

(Summation.) $S_1 \in \mathbb{R}^{n \times n}$ und $S_2 \in \mathbb{R}^{n \times n}$ seien symmetrische Matrizen. Dann gilt

$$S_1 + S_2 = (S_1 + S_2)^T. \quad (9.106)$$

(Inverse.) S sei eine invertierbare symmetrische Matrix und S^{-1} ihre Inverse. Dann ist auch S^{-1} eine symmetrische Matrix, das heißt es gilt

$$(S^{-1})^T = S^{-1}. \quad (9.107)$$

◦

9.6.5. Orthogonale Matrizen

Definition 9.14. Eine Matrix $Q \in \mathbb{R}^{n \times n}$ heißt *orthogonal*, wenn $Q^T Q = I_n$.

•

Die Spalten einer orthogonalen Matrix sind also paarweise orthogonal, es gilt für

$$Q = (q_1 \quad \cdots \quad q_n) \text{ mit } q_i \in \mathbb{R}^n \text{ für } 1 \leq i \leq n, \quad (9.108)$$

dass

$$q_i^T q_j = 0 \text{ für } i \neq j \text{ und } q_i^T q_j = 1 \text{ für } i = j \text{ mit } 1 \leq i, j \leq n. \quad (9.109)$$

Theorem 9.13 (Eigenschaften orthogonaler Matrizen). $Q \in \mathbb{R}^{n \times n}$ sei eine orthogonale Matrix. Dann gelten folgende Eigenschaften von Q .

(Inverse.) Die Inverse von Q ist Q^T , es gilt

$$Q^{-1} = Q^T. \quad (9.110)$$

(Transposition) Die Zeilen von Q sind orthonormal, es gilt

$$Q Q^T = I_n \quad (9.111)$$

◦

Beweis. (Inverse) Unter der Annahme, dass Q^{-1} existiert, gilt

$$Q^T Q = I_n \Leftrightarrow Q^T Q Q^{-1} = I_n Q^{-1} \Leftrightarrow Q^{-1} = Q^T. \quad (9.112)$$

(Transposition) Es gilt

$$Q^T Q = I_n \Leftrightarrow Q Q^T Q = Q I_n \Leftrightarrow Q Q^T Q Q^T = Q Q^T \Leftrightarrow Q Q^T = I_n. \quad (9.113)$$

□

9.6.6. Positiv-definite Matrizen

Positiv-definite Matrizen sind für die probabilistische Modellbildung unter Verwendung multivariater Normalverteilungen zentral.

Definition 9.15. Eine quadratische Matrix $C \in \mathbb{R}^{n \times n}$ heißt positiv-definit (p.d.), wenn

- C eine symmetrische Matrix ist und
- für alle $x \in \mathbb{R}^n, x \neq 0_n$ gilt, dass $x^T C x > 0$ ist.

•

In folgendem Theorem stellen wir einige wichtige Eigenschaften positiv-definiten Matrizen zusammen.

Theorem 9.14 (Eigenschaften positiv-definiten Matrizen).

(Inverse.) $C \in \mathbb{R}^{n \times n}$ sei eine positiv-definite Matrix. Dann gilt, dass C^{-1} existiert und ebenfalls positiv-definit ist.

◦

9.7. Literaturhinweise

Searle (1982) gibt eine umfassende Einführung in die Matrixtheorie vor dem Hintergrund der probabilistischen Datenanalyse, Strang (2009) gibt eine umfassende Einführung in die Matrixtheorie im Kontext der linearen Algebra. In ihrer modernen Inkarnation tauchen Matrizen als algebraische Objekte wohl zunächst in den Arbeiten von Arthur Cayley (1821-1895) auf, siehe zum Beispiel Cayley (1858).

9.8. Selbstkontrollfragen

1. Geben Sie die Definition einer Matrix wieder.
2. Nennen Sie sechs Matrixoperationen.
3. Geben Sie die Definitionen der Matrixaddition und der Matrixsubtraktion wieder.
4. Geben Sie die Definition der Skalarmultiplikation für Matrizen wieder.
5. Geben Sie die Definition der Matrixtransposition wieder.
6. Es seien

$$A := \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, B := \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix} \text{ und } c := 2. \quad (9.114)$$

Berechnen Sie

$$D := c(A - B^T) \text{ und } E := (cA)^T + B. \quad (9.115)$$

7. Geben Sie die Definition der Matrixmultiplikation wieder.
8. Es seien $A \in \mathbb{R}^{3 \times 2}$, $B \in \mathbb{R}^{2 \times 4}$ und $C \in \mathbb{R}^{3 \times 4}$. Prüfen Sie, ob folgende Matrixprodukte definiert sind, und wenn ja, geben Sie die Größe der resultierenden Matrix an:

$$ABC, ABC^T, A^T C B^T, BAC. \quad (9.116)$$

9. Es seien

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 2 & 0 \end{pmatrix} B := \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 0 & 0 \end{pmatrix} \text{ und } C := \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}. \quad (9.117)$$

Berechnen Sie die Matrixprodukte

$$AB, B^T A^T, (B^T A^T)^T, AC. \quad (9.118)$$

10. Definieren Sie die Begriff der inversen Matrix und der Invertierbarkeit einer Matrix.
11. Geben Sie die Formel für die Determinante von $A := (A_{ij})_{1 \leq i, j \leq 2} \in \mathbb{R}^2$ wieder.
12. Geben Sie die Formel für die Determinante von $A := (A_{ij})_{1 \leq i, j \leq 3} \in \mathbb{R}^3$ wieder.
13. Berechnen Sie die Determinanten von

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} B := \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} \text{ und } C := \text{diag}(1, 2, 3). \quad (9.119)$$

14. Geben Sie die Definitionen von Einheitsmatrix und Einheitsvektoren wieder.
15. Geben Sie die Definitionen von Nullmatrizen und Einsmatrizen wieder.
16. Geben Sie die Definition einer symmetrischen Matrix wieder.
17. Geben Sie die Definition einer Diagonalmatrix wieder.
18. Geben Sie die Definition einer positiv-definiten Matrix wieder.

9.9. Eigenanalyse

Mit der *Eigenanalyse einer quadratischen Matrix*, der *Orthonormalzerlegung einer symmetrischen Matrix* und der *Singulärwertzerlegung einer beliebigen Matrix* behandeln wir in diesem Abschnitt drei eng zusammenhängende Konzepte der Matrixtheorie, die in vielen Gebieten der datenanalytischen Anwendung zentrale Rollen spielen.

Allerdings erschließt sich die Bedeutung dieser Konzepte dann vor allem im jeweiligen Anwendungskontext, so dass dieser Abschnitt notwendigerweise etwas abstrakt anmuten mag.

9.10. Eigenvektoren und Eigenwerte

Unter der *Eigenanalyse* einer quadratischen Matrix versteht man das bestimmen ihrer *Eigenvektoren* und *Eigenwerte*. Diese sind für eine quadratische Matrix wie folgt definiert.

Definition 9.16 (Eigenvektor und Eigenwert). $A \in \mathbb{R}^{m \times m}$ sei eine quadratische Matrix. Dann heißt jeder vom Nullvektor 0_m verschiedene Vektor $v \in \mathbb{R}^m$, für den mit einem Skalar $\lambda \in \mathbb{R}$ gilt, dass

$$Av = \lambda v \quad (9.120)$$

ist, ein *Eigenvektor* von A und λ heißt dann ein *Eigenwert* von A .

•

Nach Definition hat also jeder Eigenvektor einen zugehörigen Eigenwert, allerdings können die Eigenwerte verschiedener Eigenvektoren durchaus identisch sein. Intuitiv bedeutet die Definition von Eigenvektor und Eigenwert, dass ein Eigenvektor einer Matrix durch Multiplikation mit eben dieser Matrix in seiner Länge, nicht aber in seiner Richtung, verändert wird. Der zugehörige Eigenwert des Eigenvektors entspricht dem Faktor der Längenänderung. Allerdings ist die Zuordnung von Eigenvektoren und Eigenwerten nicht eindeutig, wie folgendes Theorem zeigt.

Theorem 9.15 (Multiplikativität von Eigenvektoren). $A \in \mathbb{R}^{m \times m}$ sei eine quadratische Matrix. Wenn $v \in \mathbb{R}^m$ Eigenvektor von A mit Eigenwert $\lambda \in \mathbb{R}$ ist, dann ist für $c \in \mathbb{R}$ auch $cv \in \mathbb{R}^m$ Eigenvektor von A und zwar wiederum mit Eigenwert $\lambda \in \mathbb{R}$.

◦

Beweis. Es gilt

$$Av = \lambda v \Leftrightarrow cAv = c\lambda v \Leftrightarrow A(cv) = \lambda(cv). \quad (9.121)$$

Also ist cv ein Eigenvektor von A mit Eigenwert λ . □

Um nun die Uneindeutigkeit in der Definition des zu einem Eigenwert zugeordneten Eigenvektors aufzulösen, nutzen wir die Konvention, nur diejenigen Vektoren also Eigenvektoren zu einem Eigenwert λ zu betrachten, die die Länge 1 haben, für die also gilt, dass

$$\|v\| = 1. \quad (9.122)$$

Sollten wir also einen Eigenvektor v zu einem Eigenwert λ einer Matrix A finden, der nicht von der Länge 1 ist, so können wir ihn immer mit $\|v\|^{-1}$ multiplizieren. Der resultierende Vektor $v' = v/\|v\|$ hat dann die Länge 1 und ist nach Theorem 9.15 ebenso ein Eigenvektor von A zum Eigenwert λ . Bevor wir uns der Bestimmung von Eigenwerten und Eigenvektoren widmen, wollen wir die Konzepte von Eigenwert und Eigenvektor für den Fall einer 2×2 Matrix an einem Beispiel veranschaulichen

Beispiel

Es sei

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (9.123)$$

Dann ist der Vektor der Länge 1

$$v := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (9.124)$$

ein Eigenvektor von A zum Eigenwert $\lambda = 3$, da gilt, dass

$$\begin{aligned} Av &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 3 \\ 3 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= 3 \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \\ &= \lambda v. \end{aligned} \quad (9.125)$$

Inspektion von Abbildung 9.2 zeigt dementsprechend, dass für die hier definierte Matrix A die Vektoren v und Av in die gleiche Richtung zeigen, dass aber Av um den Faktor λ länger ist als v .

Der Vektor

$$w := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (9.126)$$

dagegen hat zwar die Länge 1, ist aber im Gegensatz zu v kein Eigenvektor von A , da es im Falle von

$$Aw = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad (9.127)$$

keinen Skalar λ geben kann, der mit Null, dem zweiten Eintrag von w , multipliziert einen Wert ungleich Null ergeben kann. Inspektion von Abbildung 9.2 zeigt dementsprechend, dass der aus der Multiplikation von w mit A resultierende Vektor in eine andere Richtung zeigt als w .

Bestimmung von Eigenwerten und Eigenvektoren

Folgendes Theorem besagt, wie die Eigenwerte und Eigenvektoren einer quadratischen Matrix berechnet werden können.

Theorem 9.16 (Bestimmung von Eigenwerten und Eigenvektoren). *$A \in \mathbb{R}^{m \times m}$ sei eine quadratische Matrix. Dann ergeben sich die Eigenwerte von A als die Nullstellen des charakteristischen Polynoms*

$$\chi_A(\lambda) := |A - \lambda I_m| \quad (9.129)$$

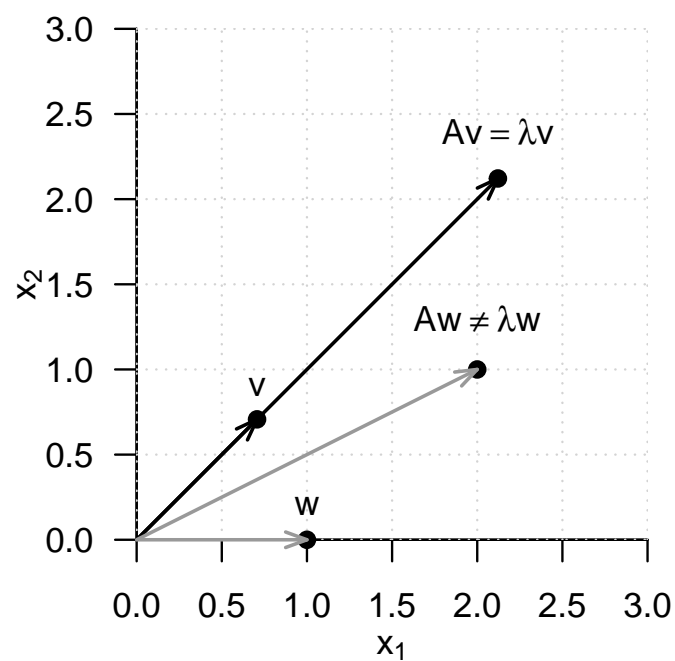


Abbildung 9.2. Eigenvektor einer 2×2 Matrix. Für die Matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (9.128)$$

ist v ein Eigenvektor, w jedoch nicht

von A . Weiterhin seien $\lambda_i^*, i = 1, 2, \dots$ die auf diese Weise bestimmten Eigenwerte von A . Die entsprechenden Eigenvektoren $v_i, i = 1, 2, \dots$ von A können dann durch Lösen der linearen Gleichungssysteme

$$(A - \lambda_i^* I_m) v_i = 0_m \text{ für } i = 1, 2, \dots \quad (9.130)$$

bestimmt werden.

◦

Beweis. (1) Bestimmen von Eigenwerten

Wir halten zunächst fest, dass mit der Definition von Eigenvektoren und Eigenwerten gilt, dass

$$Av = \lambda v \Leftrightarrow Av - \lambda v = 0_m \Leftrightarrow (A - \lambda I_m)v = 0_m. \quad (9.131)$$

Für den Eigenwert λ wird der Eigenvektor v also durch Multiplikation mit $(A - \lambda I_m)$ auf den Nullvektor 0_m abgebildet. Weil aber per Definition $v \neq 0_m$ gilt, ist die Matrix $(A - \lambda I_m)$ somit nicht invertierbar: sowohl der Nullvektor als auch v werden durch A auf 0_m abgebildet, die Abbildung

$$f: \mathbb{R}^m \rightarrow \mathbb{R}^m, x \mapsto (A - \lambda I_m)x \quad (9.132)$$

ist also nicht bijektiv, und $(A - \lambda I_m)^{-1}$ kann nicht existieren. Die Tatsache, dass $(A - \lambda I_m)$ nicht invertierbar ist, ist aber äquivalent dazu, dass die Determinante von $(A - \lambda I_m)$ gleich Null ist. Also ist

$$\chi_A(\lambda) = |A - \lambda I_m| = 0 \quad (9.133)$$

eine notwendige und hinreichende Bedingung dafür, dass λ ein Eigenwert von A ist.

(2) Bestimmen von Eigenvektoren

Es sei λ_i^* ein Eigenwert von A . Dann gilt mit den obigen Überlegungen, dass Auflösen von

$$(A - \lambda_i^* I_m) v_i^* = 0_m \quad (9.134)$$

nach v_i^* einen Eigenvektor zum Eigenwert λ_i^* ergibt. \square

Allgemein müssen zur Bestimmung von Eigenwerten und Eigenvektoren also Polynomnullstellen bestimmt und lineare Gleichungssysteme gelöst werden. Dies kann für kleine Matrizen mit $m \leq 4$ durchaus manuell geschehen. Die in der Anwendung auftretenden Matrizen sind jedoch meist weitaus größer, so dass zur Eigenanalyse numerische Verfahren der Nullstellenbestimmung und des Lösen linearer Gleichungssysteme eingesetzt werden, die zum Beispiel in Funktionen wie **R's** `eigen()`, **SciPy's** `linalg.eig()` oder **Julia's** `eigvals()` und `eigvecs()` genutzt werden. Für Details zu diesen Verfahren verweisen wir auf die weiterführende Literatur, zum Beispiel Burden et al. (2016) und Richter & Wick (2017). Wir wollen Theorem 9.16 hier lediglich anhand eines Beispiels illustrieren.

Beispiel

Dazu sei wiederum

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (9.135)$$

Wir wollen zunächst die Eigenwerte von A berechnen. Nach Theorem 9.16 sind dies die Nullstellen des charakteristischen Polynoms von A . Wir berechnen also zunächst das charakteristische Polynom von A durch

$$\chi_A(\lambda) = \left| \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = \left| \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} \right| = (2-\lambda)^2 - 1. \quad (9.136)$$

Mithilfe der pq-Formel zur Lösung quadratischer Gleichungen findet man dann

$$(2 - \lambda_{1/2}^*)^2 - 1 = 0 \Leftrightarrow \lambda_1^* = 3 \text{ oder } \lambda_2^* = 1. \quad (9.137)$$

Die Eigenwerte von A sind also $\lambda_1 = 3$ und $\lambda_2 = 1$. Die zugehörigen Eigenvektoren ergeben sich dann für $i = 1, 2$ durch Lösen des linearen Gleichungssystems

$$(A - \lambda_i I_2)v_i = 0_2. \quad (9.138)$$

Speziell ergibt sich hier, dass für $\lambda_1 = 3$ aus

$$(A - 3I_2)v_1 = 0_2 \Leftrightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} v_{1_1} \\ v_{1_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (9.139)$$

folgt, dass

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (9.140)$$

ein Eigenvektor zum Eigenwert λ_1 ist und dass für $\lambda_2 = 1$ aus

$$(A - 1I_2)v_2 = 0_2 \Leftrightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} v_{2_1} \\ v_{2_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (9.141)$$

folgt, dass

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad (9.142)$$

ein Eigenvektor zum Eigenwert $\lambda_2 = 1$ ist. Weiterhin gelten hier offenbar

$$v_1^T v_2 = 0 \text{ und } \|v_1\| = \|v_2\| = 1. \quad (9.143)$$

Folgender **R** Code demonstriert die Bestimmung der Eigenwerte und Eigenvektoren der hier betrachteten Matrix mithilfe der `eigen()` Funktion.

```
# Matrixdefinition
A = matrix(c(2,1,
             1,2),
           nrow = 2,
           byrow = TRUE)

# Eigenanalyse
eigen(A)

eigen() decomposition
$values
[1] 3 1

$vectors
      [,1]      [,2]
[1,] 0.7071068 -0.7071068
[2,] 0.7071068  0.7071068
```

Zum Abschluss dieses Abschnittes betrachten wir zwei technische Theoreme, die Aussagen zum Zusammenhang spezieller Matrixprodukte und ihrer Eigenwerte und Eigenvektoren machen. Wir benötigen dieses Theoreme im Kontext der Kanonischen Korrelationsanalyse (`?@sec-kanonische-korrelationsanalyse`).

Theorem 9.17 (Eigenwerte und Eigenvektoren von Matrixprodukten). *Für $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times n}$ sind die Eigenwerte von $AB \in \mathbb{R}^{n \times n}$ und $BA \in \mathbb{R}^{m \times m}$ gleich. Weiterhin gilt, dass für einen Eigenvektor v zu einem von Null verschiedenen Eigenwert λ von AB $w := Bv$ ein Eigenvektor von BA zum Eigenwert λ ist.*

◦

Für einen Beweis verweisen wir auf Mardia et al. (1979), S. 468. Wir demonstrieren die Aussage dieses Theorems anhand untenstehenden **R** Codes.

```
A = matrix(1:6, nrow = 2, byrow = T) # Matrix A \in \mathbb{R}^{2 \times 3}
B = matrix(1:6, ncol = 2, byrow = T) # Matrix B \in \mathbb{R}^{3 \times 2}
EAB = eigen(A %*% B) # Eigenanalyse von AB \in \mathbb{R}^{2 \times 2}
EBA = eigen(B %*% A) # Eigenanalyse von BA \in \mathbb{R}^{3 \times 3}
w = B %*% EAB$eigenvalues[,1] # Eigenvektor von BA
cat("Eigenwerte von AB :", EAB$values[1:2],
    "\nEigenwerte von BA :", EBA$values[1:2],
    "\nBAw mit w = Bv :", B %*% A %*% w,
    "\nlw mit w = Bv :", EBA$values[1] * w)
```

```
Eigenwerte von AB : 85.57934 0.4206623
Eigenwerte von BA : 85.57934 0.4206623
BAw mit w = Bv : -191.1333 -416.7586 -642.3839
lw mit w = Bv : -191.1333 -416.7586 -642.3839
```

Theorem 9.18. Für $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{p \times n}$, $a \in \mathbb{R}^m$ und $b \in \mathbb{R}^p$ gilt, dass der einzige von Null verschiedene Eigenwert von $Aab^T B \in \mathbb{R}^{n \times n}$ gleich $b^T B A a$ mit zugehörigem Eigenvektor Aa ist.

◦

Für einen Beweis verweisen wir auf Mardia et al. (1979), S. 468. Wir demonstrieren die Aussage dieses Theorems anhand untenstehenden **R** Codes.

```
A = matrix(1:6, nrow = 2, byrow = T) # Matrix A \in \mathbb{R}^{2 \times 3}
B = matrix(1:8, ncol = 2, byrow = T) # Matrix B \in \mathbb{R}^{4 \times 2}
a = matrix(1:3, nrow = 3, byrow = T) # Vektor a \in \mathbb{R}^{3 \times 1}
b = matrix(1:4, nrow = 4, byrow = T) # Vektor b \in \mathbb{R}^{4 \times 1}
EAabTB = eigen(A %*% a %*% t(b) %*% B) # Eigenanalyse von Aab^T B \in \mathbb{R}^{4 \times 4}
cat("Eigenwerte von AabTB :", EAabTB$values,
    "\nbTBaAa :", t(b) %*% B %*% A %*% a,
    "\nAa :", A %*% a,
    "\n(AabTB)Aa :", (A %*% a %*% t(b) %*% B) %*% A %*% a, # Mv
    "\n(bTBaAa)Aa :", as.vector((t(b) %*% B %*% A %*% a) * (A %*% a)) # = \lambda v
```

```
Eigenwerte von AabTB : 2620 0
bTBaAa : 2620
Aa : 14 32
(AabTB)Aa : 36680 83840
(bTBaAa)Aa : 36680 83840
```

9.11. Orthonormalzerlegung

Mit dem Begriff der *Zerlegung* einer Matrix wird das Aufspalten einer gegebenen Matrix in das Matrixprodukt mehrerer Matrizen bezeichnet. Verschiedenste Matrixzerlegungen spielen in vielen mathematischen Anwendungen eine wichtige Rolle, für einen Überblick siehe beispielsweise Golub & Van Loan (2013). In diesem Abschnitt führen wir mit der *Orthonormalzerlegung einer symmetrischen Matrix* eine spezielle Matrixzerlegung ein, die direkt auf der Eigenanalyse aufbaut. Wir halten zunächst folgendes grundlegendes Theorem zu den Eigenwerten und Eigenvektoren symmetrischer Matrizen fest.

Theorem 9.19 (Eigenwerte und Eigenvektoren symmetrischer Matrizen).

$S \in \mathbb{R}^{m \times m}$ sei eine symmetrische Matrix. Dann gelten

- (1) Die Eigenwerte von S sind reell.
- (2) Die Eigenvektoren zu je zwei verschiedenen Eigenwerten von S sind orthogonal.

◦

Beweis. Wir setzen die Tatsache, dass eine symmetrische Matrix m reelle Eigenwerte hat, als gegeben voraus und zeigen lediglich, dass die Eigenvektoren zu je zwei verschiedenen Eigenwerten einer symmetrischen Matrix orthogonal sind. Ohne Beschränkung der Allgemeinheit seien also $\lambda_i, \lambda_j \in \mathbb{R}$ mit $1 \leq i, j \leq m$ und $\lambda_i \neq \lambda_j$ zwei verschiedenen Eigenwerte von S mit zugehörigen Eigenvektoren q_i und q_j , respektive. Dann ergibt sich wie unten gezeigt, dass

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \quad (9.144)$$

Mit $q_i \neq 0_m, q_j \neq 0_m$ und $\lambda_i \neq \lambda_j$ folgt damit $q_i^T q_j = 0$, weil es keine andere Zahl c als die Null gibt, für die bei $a, b \in \mathbb{R}$ und $a \neq b$ gilt, dass

$$ac = bc. \quad (9.145)$$

Um abschließend

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \quad (9.146)$$

zu zeigen, halten wir zunächst fest, dass

$$Sq_i = \lambda_i q_i \Leftrightarrow (Sq_i)^T = (\lambda_i q_i)^T \Leftrightarrow q_i^T S^T = q_i^T \lambda_i^T \Leftrightarrow q_i^T S = q_i^T \lambda_i \Leftrightarrow q_i^T Sq_j = \lambda_i q_i^T q_j \quad (9.147)$$

und

$$Sq_j = \lambda_j q_j \Leftrightarrow q_j^T S = q_j^T \lambda_j \Leftrightarrow q_j^T Sq_i = \lambda_j q_j^T q_i \Leftrightarrow (q_j^T Sq_i)^T = (\lambda_j q_j^T q_i)^T \Leftrightarrow q_i^T Sq_j = \lambda_j q_i^T q_j \quad (9.148)$$

gelten. Sowohl $\lambda_i q_i^T q_j$ als auch $\lambda_j q_i^T q_j$ sind also mit $q_i^T Sq_j$ und damit auch miteinander identisch. \square

Offenbar haben wir nur Aussage (2) von Theorem 9.19 bewiesen. Ein vollständiger Beweis des Theorems findet sich zum Beispiel bei Strang (2009). Wir merken außerdem an, dass, weil wir nach Konvention Eigenvektoren der Länge 1 betrachten, die in Theorem 9.19 angesprochenen orthogonalen Eigenvektoren insbesondere auch orthonormal sind. Mithilfe von Theorem 9.19 können wir nun die Orthonormalzerlegung einer symmetrischen Matrix formulieren und ihre Existenz beweisen.

Theorem 9.20 (Orthonormalzerlegung einer symmetrischen Matrix). *$S \in \mathbb{R}^{m \times m}$ sei eine symmetrische Matrix mit m verschiedenen Eigenwerten. Dann kann S geschrieben werden als*

$$S = Q\Lambda Q^T, \quad (9.149)$$

wobei $Q \in \mathbb{R}^{m \times m}$ eine orthogonale Matrix ist und $\Lambda \in \mathbb{R}^{m \times m}$ eine Diagonalmatrix ist.

◦

Beweis. Es seien $\lambda_1 > \lambda_2 > \dots > \lambda_m$ die der Größe nach geordneten Eigenwerte von S und q_1, \dots, q_m die zugehörigen orthonormalen Eigenvektoren. Mit

$$Q := (q_1 \quad q_2 \quad \dots \quad q_m) \in \mathbb{R}^{m \times m} \text{ und } \Lambda := \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^{m \times m}, \quad (9.150)$$

folgt dann mit den Definitionen von Eigenwerten und Eigenvektoren zunächst, dass

$$Sq_i = \lambda_i q_i \text{ für } i = 1, \dots, m \Leftrightarrow SQ = Q\Lambda. \quad (9.151)$$

Rechtseitige Multiplikation mit Q^T ergibt dann mit $QQ^T = I_m$, dass

$$SQQ^T = Q\Lambda Q^T \Leftrightarrow SI_m = Q\Lambda Q^T \Leftrightarrow S = Q\Lambda Q^T. \quad (9.152)$$

 \square

Man nennt das Aufspalten von S in das Matrixprodukt $Q\Lambda Q^T$ aufgrund der Diagonalität von Λ auch eine *Diagonalisierung von S* . Wie im Beweis gezeigt, wählt man zur Darstellung von S in Diagonaldarstellung für die Diagonalelemente von Λ die der Größe nach geordneten Eigenwerte von S und für die Spalten von Q die jeweils zugehörigen Eigenvektoren von S . Wir verdeutlichen dies an einem Beispiel.

Beispiel

Für die symmetrische Matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (9.153)$$

mit den oben bestimmten Eigenwerten $\lambda_1 = 3$ und $\lambda_2 = 1$ sowie den zugehörigen orthonormalen Eigenvektoren

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad (9.154)$$

seien

$$Q := (v_1 \ v_2) \text{ und } \Lambda = \text{diag}(\lambda_1, \lambda_2). \quad (9.155)$$

Dann ergibt sich offenbar

$$\begin{aligned} Q\Lambda Q^T &= (v_1 \ v_2) \text{diag}(\lambda_1, \lambda_2) (v_1 \ v_2)^T \\ &= \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right) \\ &= \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 3 & -1 \\ 3 & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right) \\ &= \frac{1}{2} \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \\ &= A \end{aligned}$$

und wir haben Theorem 9.20 für dieses Beispiel verifiziert.

Symmetrische Quadratwurzel einer Matrix

Die Definition der Orthonormalzerlegung einer symmetrischen Matrix erlaubt es, den Begriff der *symmetrischen Quadratwurzel einer Matrix* einzuführen.

Definition 9.17 (Symmetrische Quadratwurzel einer Matrix). $S \in \mathbb{R}^{m \times m}$ sei eine invertierbare symmetrische Matrix mit positiven Eigenwerten. Dann sind für $r \in \mathbb{N}^0$ und $s \in \mathbb{N}$ die rationalen Potenzen von S mit der orthonormalen Matrix $Q \in \mathbb{R}^{m \times m}$ der Eigenvektoren von S und der Diagonalmatrix $\Lambda = \text{diag}(\lambda_i) \in \mathbb{R}^{m \times m}$ der zugehörigen Eigenwerte $\lambda_1, \dots, \lambda_m$ von S definiert als

$$S^{r/s} = Q\Lambda^{r/s}Q^T \text{ mit } \Lambda^{r/s} = \text{diag}(\lambda_i^{r/s}). \quad (9.156)$$

Der Spezialfall $r := 1, s := 2$ wird als *symmetrische Quadratwurzel von S* bezeichnet und hat die Form

$$S^{1/2} = Q\Lambda^{1/2}Q^T \text{ mit } \Lambda^{1/2} = \text{diag}(\lambda_i^{1/2}). \quad (9.157)$$

•

Wir halten fest, dass mit Definition 9.17 offenbar gilt, dass

$$(S^{1/2})^2 = Q\Lambda^{1/2}Q^TQ\Lambda^{1/2}Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = Q\Lambda Q^T = S. \quad (9.158)$$

Weiterhin gilt, dass

$$(S^{-1/2})^2 = Q\Lambda^{-1/2}Q^TQ\Lambda^{-1/2}Q^T = Q\Lambda^{-1/2}\Lambda^{-1/2}Q^T = Q\Lambda^{-1}Q^T = S^{-1}. \quad (9.159)$$

Schließlich gilt, dass

$$\begin{aligned} S^{-1/2}SS^{-1/2} &= Q\Lambda^{-1/2}Q^TQ\Lambda Q^TQ\Lambda^{-1/2}Q^T \\ &= Q\Lambda^{-1/2}\Lambda\Lambda^{-1/2}Q^T \\ &= Q\Lambda^{-1}Q^T \\ &= I_m \end{aligned} \quad (9.160)$$

9.12. Singularwertzerlegung

Eine vielseitig einsetzbare Matrixzerlegung einer beliebigen Matrix ist die *Singularwertzerlegung*. Wir sind an dieser Stelle lediglich an dem Zusammenhang von Singularwertzerlegung und Eigenanalyse interessiert und verweisen für eine ausführliche Diskussion der Singularwertzerlegung auf die weiterführende Literatur, beispielsweise Strang (2009). Der Begriff der Singularwertzerlegung ist wie folgt definiert.

Definition 9.18 (Singularwertzerlegung). $Y \in \mathbb{R}^{m \times n}$ sei eine Matrix. Dann heißt die Zerlegung

$$Y = USV^T, \quad (9.161)$$

wobei $U \in \mathbb{R}^{m \times m}$ eine orthogonale Matrix ist, $S \in \mathbb{R}^{m \times n}$ eine Diagonalmatrix ist und $V \in \mathbb{R}^{n \times n}$ eine orthogonale Matrix ist, *Singularwertzerlegung* von Y . Die Diagonalelemente von S heißen die *Singularwerte* von Y .

•

Singularwertzerlegungen werden auf Englisch *singular value decompositions* genannt und entsprechend mit *SVD* abgekürzt. Wir verzichten auf eine Diskussion der Berechnung einer Singularwertzerlegung und weisen lediglich daraufhin, dass Singularwertzerlegungen zum Beispiel in **R** mit der Funktion `svd()`, in **SciPy** mit `scipy.linalg.svd()` und in **Julia** mit `svd()` berechnet werden können. Folgendes Theorem beschreibt den Zusammenhang zwischen Singularwertzerlegung und Eigenanalyse und wird an vielen Stellen eingesetzt.

Theorem 9.21 (Singularwertzerlegung und Eigenanalyse).

$Y \in \mathbb{R}^{m \times n}$ sei eine Matrix und

$$Y = USV^T \quad (9.162)$$

sei ihre Singularwertzerlegung. Dann gilt:

- Die Spalten von U sind die Eigenvektoren von YY^T ,

- die Spalten von V sind die Eigenvektoren von $Y^T Y$ und
- die Singulärwerte sind die Quadratwurzeln der zugehörigen Eigenwerte.

◦

Beweis. Wir halten zunächst fest, dass mit

$$(YY^T)^T = YY^T \text{ und } (Y^T Y)^T = Y^T Y, \quad (9.163)$$

YY^T und $Y^T Y$ symmetrische Matrizen sind und somit Orthonormalzerlegungen haben. Wir halten weiterhin fest, dass mit $V^T V = I_n$, $U^T U = I_m$ und $S^T = S$ gilt, dass

$$YY^T = USV^T (USV^T)^T = USV^T V S^T U^T = USSU^T =: U\Lambda U^T \quad (9.164)$$

und

$$Y^T Y = (USV^T)^T USV^T = V S^T U^T U S^T V^T =: V\Lambda V^T \quad (9.165)$$

ist, wobei wir $\Lambda := SS$ definiert haben. Weil das Produkt von Diagonalmatrizen wieder eine Diagonalmatrix ist, ist Λ eine Diagonalmatrix und per Definition sind U und V orthogonale Matrizen. Wir haben also YY^T und $Y^T Y$ in Form der Orthonormalzerlegungen

$$YY^T = U\Lambda U^T \text{ und } Y^T Y = V\Lambda V^T \quad (9.166)$$

geschrieben, wobei für die Diagonalelemente von Λ gilt, dass sie die quadrierten Werte der Diagonalelemente von S sind. \square

9.13. Literaturhinweise

Die in diesem Kapitel behandelten Konzepte werden ausführlich zum Beispiel in Searle (1982) und Strang (2009) behandelt. Die Verwendung des Präfix *Eigen-* für die beschriebenen Vektoren und Skalare in Bezug zu einer Matrix beginnt offenbar Hilbert (1904) im Kontext der Analyse von Integralgleichungen und hat sich auch im Englischen durchgesetzt.

9.14. Selbstkontrollfragen

1. Geben Sie die Definition eines Eigenvektors einer quadratischen Matrix wieder.
2. Geben Sie die Definition eines Eigenwerts einer quadratischen Matrix wieder.
3. Geben Sie das Theorem zur Bestimmung von Eigenwerten und Eigenvektoren wieder.
4. Geben Sie das Theorem zu den Eigenwerten und Eigenvektoren symmetrischer Matrizen wieder.
5. Geben Sie das Theorem zur Orthonormalzerlegung einer symmetrischen Matrix wieder.
6. Geben Sie die Definition der symmetrischen Quadratwurzel einer Matrix wieder.
7. Geben Sie die Definition einer Singulärwertzerlegung wieder.
8. Geben Sie das Theorem zum Zusammenhang von Singulärwertzerlegung und Eigenanalyse wieder.

10. Deskriptivstatistiken

Ziel der Auswertung von Deskriptivstatistiken ist es, Datensätze möglichst effizient der menschlichen Betrachtung zugänglich zu machen. Sobald die Anzahl der Datenpunkte eines Datensatzes eine gewisse Anzahl überschreitet, sind reine Zahlendarstellungen für die menschliche Kognition meist eher schwer zu erfassen. Hier setzt die Deskriptivstatistik an und bietet neben graphischen Darstellungen größerer Datensätze auch Maßzahlen an, die im Sinne der Datenreduktion bestimmte charakteristische Aspekte eines Datensatzes, wie seinen “durchschnittlichen” Wert oder seine Variabilität beschreiben. In diesem Abschnitt wollen wir mit den univariaten Deskriptivstatistiken erste Verfahren und Maße kennenlernen, die es erlauben, Datensätze, in denen pro Studieneinheit eine Zahl vorliegt, zusammenzufassen. Wir betrachten dabei insbesondere *Häufigkeitsverteilungen* und *Histogramme*, *Verteilungsfunktionen* und *Quantile*, sowie *Maße der zentralen Tendenz*, die Aussagen über “durchschnittliche” Datenwerte erlauben und *Maße der Datenvariabilität*, die Aussagen über die Streuung der Datenwerte erlauben.

In Abgrenzung zur probabilistischen Modellierung im Rahmen der Frequentistischen und Bayesianischen Inferenz basiert die deskriptive Statistik nicht auf der Wahrscheinlichkeitstheorie, kann also losgelöst von dieser betrachtet werden. Allerdings ergeben viele Aspekte der Deskriptivstatistik letztlich nur vor dem Hintergrund wahrscheinlichkeitstheoretischer Betrachtungen Sinn und umgekehrt sind viele Begriffe der Wahrscheinlichkeitstheorie durch deskriptivstatistische Konzepte motiviert. Wenn also der vorliegende Abschnitt ohne einen wahrscheinlichkeitstheoretischen Hintergrund auskommt, so erschließt sich seine volle Bedeutung sicherlich erst im Kontext der Begriffe der Wahrscheinlichkeitstheorie.

Wir fokussieren in diesem Abschnitt auf Deskriptivstatistiken zur Beschreibung eines Datensatzes der Form

$$y := (y_1, \dots, y_n) \text{ mit } y_i \in \mathbb{R} \text{ für } i = 1, \dots, n. \quad (10.1)$$

Dabei ist es zunächst unerheblich, ob ein solcher Datensatz in Form eines Spaltenvektors $y \in \mathbb{R}^{n \times 1}$ oder eines Zeilenvektors $y \in \mathbb{R}^{1 \times n}$ vorliegt. Weiterhin fokussieren wir auf die praktische Auswertung der betrachteten Deskriptivstatistiken, geben also viele Beispiele für ihre Berechnung mit **R**.

Anwendungsbeispiel

Als Anwendungsbeispiel betrachten wir durchgängig einen Datensatz zur evidenzbasierten Evaluation von Psychotherapie bei Depression. Dazu nehmen wir an, dass $n = 100$ Patient:innen zufällig auf eine Treatment- und eine Kontrollstudienbedingung aufgeteilt wurden und zu Beginn der jeweiligen Intervention mithilfe des BDI-II Fragebogens zu ihrer Depressionsschwere befragt wurden. Die nach den Interventionen erhobenen BDI-II Werte wollen wir in diesem Kapitel nicht betrachten. Wir nehmen also an, dass wir einen Datensatz wie durch folgenden **R** Code in Tabelle 10.1 dargestellt vorliegen haben.

```
D = read.csv("./_data/111-deskriptive-statistik.csv")
knitr::kable(D[,1:2], digits = 2, align = "c")
```

```
# Laden des Datensatzes
# RMarkdowntabellenoutput
```

Tabelle 10.1. BDI-II Werte (PRE) von $n = 100$ Patient:innen in einer Treatment (T) und einer Kontrollstudiengruppe (C) vor Beginn einer psychologischen Intervention

GROUP	PRE
T	17
T	20
T	16
T	18
T	21
T	17
T	17
T	17
T	18
T	18
T	20
T	17
T	16
T	18
T	16
T	18
T	17
T	14
T	18
T	18
T	20
T	20
T	21
T	19
T	19
T	17
T	20
T	21
T	17
T	17
T	19
T	20
T	22
T	20
T	21
T	20
T	16
T	15
T	15
T	18
T	21
T	17
T	18
T	21
T	17
T	19
T	16
T	17
T	17
T	18
C	22
C	19
C	21
C	18
C	19
C	17
C	20
C	19
C	19
C	19
C	17
C	20
C	18
C	20
C	18
C	18
C	15
C	18
C	17
C	19
C	19
C	19
C	20
C	19
C	20
C	19
C	21
C	19
C	19
C	18
C	20
C	16
C	21
C	19

Tabelle 10.1. BDI-II Werte (PRE) von $n = 100$ Patient:innen in einer Treatment (T) und einer Kontrollstudiengruppe (C) vor Beginn einer psychologischen Intervention

GROUP	PRE
C	20
C	18
C	23
C	18
C	19
C	18
C	21
C	17
C	20
C	21
C	21
C	20
C	18
C	18
C	19
C	19

10.1. Häufigkeitsverteilungen

Wir beginnen mit den Begriffen der *absoluten* und der *relativen Häufigkeitsverteilung*. Für Datensätze mit Werten, in denen der gleiche Wert mehrfach auftritt, können Häufigkeitsverteilungen einen ersten Eindruck von der Beschaffenheit des Datensatzes liefern.

Definition 10.1 (Absolute und relative Häufigkeitsverteilungen). $y := (y_1, \dots, y_n)$ sei ein Datensatz und $W := \{w_1, \dots, w_k\}$ mit $k \leq n$ seien die im Datensatz vorkommenden Werte. Dann heißt die Funktion

$$h : W \rightarrow \mathbb{N}, w \mapsto h(w) := \text{Anzahl der } y_i \text{ aus } y \text{ mit } y_i = w \quad (10.2)$$

die *absolute Häufigkeitsverteilung* der Werte von y und die Funktion

$$r : W \rightarrow [0, 1], w \mapsto r(w) := \frac{h(w)}{n} \quad (10.3)$$

heißt die *relative Häufigkeitsverteilung* der Werte von y .

•

Häufigkeitsverteilungen können in **R** durch Kombination der `table()` und `barplot()` Funktion ausgewertet und dargestellt werden. Folgender **R** Code erzeugt zunächst die absolute und relative Häufigkeitsverteilungen der Pre-Interventions-BDI-II Daten des Beispieldatensatzes.

```
D = read.csv("./data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Double vector der PRE Werte
n = length(y) # Anzahl der Datenwerte (100)
H = as.data.frame(table(y)) # absolute Häufigkeitsverteilung (dataframe)
names(H) = c("w", "ah") # Spaltenbenennung
H$rh = H$ah/n # relative Häufigkeitsverteilung
print(H, digits = 1) # Ausgabe
```

```
  w ah  rh
1 14  1 0.01
2 15  3 0.03
3 16  6 0.06
4 17 17 0.17
5 18 21 0.21
6 19 20 0.20
7 20 17 0.17
8 21 12 0.12
9 22  2 0.02
10 23  1 0.01
```

An der Ausgabe des Dataframes H können wir ablesen, dass im betrachteten Datensatz 10 verschiedene Werte w zwischen 14 und 23 vorkommen, wobei zum Beispiel die Werte 18 und 22 jeweils 21 bzw. 2 mal auftreten. Sowohl die absolute als auch die relative Häufigkeitsverteilung zeigen, dass Werte um 18 im vorliegenden Datensatz gehäuft auftreten, Extremwerte wie 13 oder 23 dagegen eher selten. Man beachte, dass es sich bei der relativen Häufigkeitsverteilung lediglich um eine skalierte Form der absoluten Häufigkeitsverteilung handelt. Qualitativ, also hinsichtlich des gehäuft oder selteneren Vorkommens bestimmter Werte, unterscheiden sich beide Häufigkeitsverteilungen nicht. Dies wird insbesondere auch deutlich, wenn man wie in Abbildung 10.1 beide Häufigkeitsverteilungen visualisiert. Folgender **R** nutzt zu diesem Zweck die Funktion `barplot()`. Die Höhe des über einem Wert w aufgetragenen Balkens entspricht dabei den Funktionswerten $h(w)$ bzw. $r(w)$. Die visuelle Einsicht entspricht natürlich der Inspektion des Dataframes H . Beide Häufigkeitsverteilungen zeigen, dass Werte um 18 im vorliegenden Datensatz gehäuft auftreten, wohingegen niedrigere oder höhere Werte wie 13 bzw. 23 dagegen eher selten sind.

```
pdf(
  file      = "_figures/111-haeufigkeitsverteilungen.pdf",
  width     = 7,
  height    = 4)
par(
  mfcol     = c(1,2),
  las       = 1,
  cex       = 0.7,
  cex.main  = 1.3)
h           = H$ah
r           = H$rh
names(h)    = H$w
names(r)    = H$w
barplot(h,
  col       = "gray90",
  xlab      = "w",
  ylab      = "h(w)",
  ylim     = c(0,25),
  main      = "Absolute Häufigkeitsverteilung")
barplot(r,
  col       = "gray90",
  xlab      = "w",
  ylab      = "r(w)",
  ylim     = c(0,.25),
  main      = "Relative Häufigkeitsverteilung")
dev.off()
```

PDF Kopiefunktion
Dateiname
Breite (inch)
Höhe (inch)
Abbildungsparameter
Zeilen und Spaltenlayout
x-Tick Orientierung
Annotationsvergrößerung
Titelvergrößerung
h(w) Werte
r(w) Werte
barplot braucht w Werte als names
barplot braucht w Werte als names
Balkendiagramm
absolute Haeufigkeiten
Balkenfarbe
x Achsenbeschriftung
y Achsenbeschriftung
y Achsengrenzen
Titel
Balkendiagramm
absolute Haeufigkeiten
Balkenfarbe
x Achsenbeschriftung
y Achsenbeschriftung
y Achsengrenzen
Titel
Beenden der Abbildungsbearbeitung

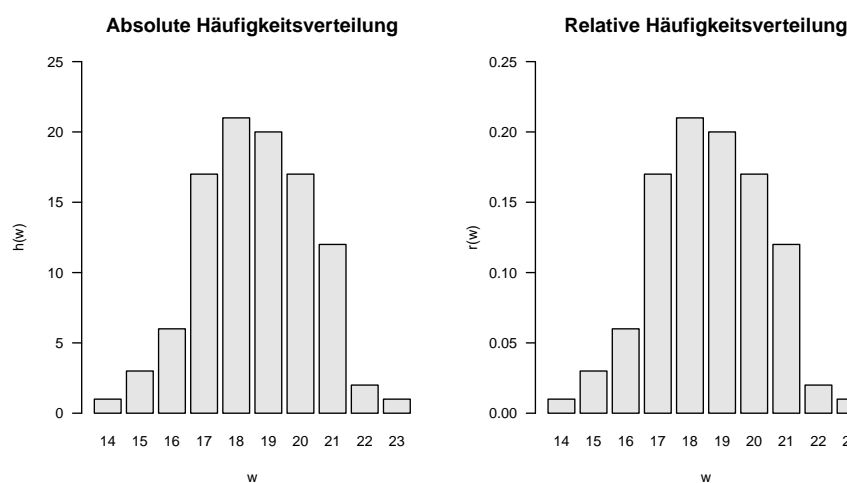


Abbildung 10.1. Absolute und relative Häufigkeitsverteilungen der Pre-Interventions-BDI-II-Werte .

10.2. Histogramme

Wenn in einem Datensatz keine Werte mehrfach vorkommen, nehmen die absoluten und relativen Häufigkeitsverteilungen für jeden dieser Werte den Wert 1 bzw. $1/n$ an und bieten dann keinen Mehrwert über die visuelle Inspektion des Datensatzes hinaus. Die in diesem Abschnitt vorgestellten Histogramme bieten jedoch eine Möglichkeit, auch in diesem Fall visuelle Darstellungen von Häufigkeitsverteilungen zu generieren. Dabei werden verschiedene Werte, die in einem wohldefinierten Sinne “nah beieinander” liegen, zu Klassen zusammengefasst und die absoluten oder relativen Häufigkeiten dieser Klassen bestimmt. Da bei Datensätzen, in denen die gleichen Werte mehrfach vorkommen, diese Werte der Natur nach in der gleichen Klasse liegen, können Histogramme auch in dem in Kapitel 10.1 betrachteten Fall sinnvoll angewendet werden. Histogramme verallgemeinern also die im vorherigen Abschnitt betrachteten Häufigkeitsverteilungen bei Datensätzen mit mehrfach auftretenden Werten.

Wir geben zunächst eine Definition des Histogrammbegriffs. Man beachte, dass in dieser Definition der Vorgang der Klassenbildung, die Bestimmung ihrer absoluten und relativen Häufigkeiten, sowie die Visualisierung dieser Häufigkeiten eng verzahnt sind. Dies ist der Konvention geschuldet, dass der Begriff des Histogramms sowohl eine bestimmte Form der Datenanalyse als auch ihre Visualisierung beschreibt.

Definition 10.2 (Histogramm). Ein *Histogramm* ist ein Diagramm, in dem zu einem Datensatz $y = (y_1, \dots, y_n)$ mit Werten $W := \{w_1, \dots, w_m\}$ mit $m \leq n$ für $j = 1, \dots, k$ über benachbarten Intervallen $[b_{j-1}, b_j[$ mit $b_0 := \min W$ und $b_k := \max W$ Rechtecke mit der Breite $d_j := b_j - b_{j-1}$ und der Höhe $h(w)$ oder $r(w)$ für $w \in [b_{j-1}, b_j[$ abgebildet sind. Dabei werden die Intervalle $[b_{j-1}, b_j[$ Klassen (engl. *bins*) genannt.

•

Natürlich ist das visuelle Erscheinungsbild eines Histogramms stark davon abhängig, wie genau die Werte eines Datensatzes zu Klassen zusammengefasst werden. Der entscheidende Parameter dafür ist nach Definition 10.2 die Anzahl der Klassen k , die zwischen dem Minimalwert b_0 und dem Maximalwert b_k des Datensatzes gewählt werden. Im Folgenden stellen wir einige konventionelle Werte für die Anzahl der Klassen k vor und berechnen und visualisieren sie am Beispiel des Beispieldatensatzes. Dabei gehen wir durchgängig davon aus, dass die benachbarten Intervalle $[b_{j-1}, b_j[$ mit $j = 1, \dots, k$ für die Anzahl von Klassen k mit konstanter Breite $d = b_j - b_{j-1}$ gewählt werden.

Zur Berechnung und Visualisierung von Histogrammen nutzen wir die **R** Funktion `hist()`. Bei dieser werden die Klassen $[b_{j-1}, b_j[$ für $j = 1, \dots, k$ als das Argument `breaks` festgelegt. Speziell ist `breaks` der **R** Vector `c(b_0, b_1, \dots, b_k)` mit Länge $k + 1$. Um den Effekt verschiedener Wahlen der Anzahl der Klassen zu veranschaulichen, lesen wir zunächst den Datensatz ein und bestimmen seine minimalen und maximalen Werte b_0 und b_k .

```
# Datensatz
D = read.csv("../data/111-deskriptive-statistik.csv")
y = D$PRE
b_0 = min(y)
b_k = max(y)

# Laden des Datensatzes
# Pre-Interventions-BDI-II-Daten
# b_0
# b_k
```

```
Minimum b_0      : 14
Maximum b_k      : 23
```

Eine erste Möglichkeit der Wahl der Klassen k besteht darin, möglichst genau eine gewünschte konstante Breite d zu erzielen. Hierzu setzt man

$$k := \lceil (b_k - b_0)d \rceil. \quad (10.4)$$

Man beachte, dass die Anwendung der Aufrundungsfunktion impliziert, dass die so resultierende Klassenbreite höchstens dem gewünschten Wert entspricht, aber auch kleiner sein kann. Folgender **R** Code verdeutlicht diesen Effekt.

```
# Histogramm mit gewünschter Klassenbreite d = 2
d      = 2                                # gewünschte Klassenbreite
b_0    = min(y)                          # b_0
b_k    = max(y)                          # b_k
k      = ceiling((b_k - b_0)/d)          # Anzahl der Klassen
b      = seq(b_0, b_k, length.out = k + 1) # Klassen [b_{j-1}, b_j[ (breaks)
```

```
Gewünschte Klassenbreite d : 2
Klassenanzahl k           : 5
Intervalle [b_{j-1}, b_j] : 14 15.8 17.6 19.4 21.2 23
Erzielte Klassenbreite d : 1.8
```

Die Wahl einer gewünschten Klassenbreite von $d = 1.5$ dagegen führt für den vorliegenden Datensatz auf eine resultierende Klassenbreite, die dem Wunsch entspricht.

```
# Histogramm mit gewünschter Klassenbreite d = 1.5
d      = 1.5                              # gewünschte Klassenbreite
k      = ceiling((b_k - b_0)/d)          # Anzahl der Klassen
b      = seq(b_0, b_k, length.out = k + 1) # Klassen [b_{j-1}, b_j[ (breaks)
```

```
Gewünschte Klassenbreite d : 1.5
Klassenanzahl k           : 6
Intervalle [b_{j-1}, b_j] : 14 15.5 17 18.5 20 21.5 23
Erzielte Klassenbreite d : 1.5
```

Eine weitere Möglichkeit ist es, die Anzahl der Klassen abhängig von der Anzahl der Datenpunkte im Datensatz zu machen. Die Excelstandardwahl ist dabei

$$k := \lceil \sqrt{n} \rceil. \quad (10.5)$$

Folgender **R** Code implementiert diese Wahl für den den vorliegenden Datensatz.

```
# Histogramm mit Excelstandard
n      = length(y)                        # Anzahl Datenpunkte
k      = ceiling(sqrt(n))                 # Anzahl der Klassen
b      = seq(b_0, b_k, length.out = k + 1) # Klassen [b_{j-1}, b_j[ (breaks)
```

```
Anzahl Datenpunkte n      : 100
Klassenanzahl k           : 10
Intervalle [b_{j-1}, b_j] : 14 14.9 15.8 16.7 17.6 18.5 19.4 20.3 21.2 22.1 23
Erzielte Klassenbreite d : 0.9
```

Ähnlich gelagert wie der Excelstandard ist die von Sturges (1926) vorgeschlagene Klassenanzahl, welche die Darstellung unter einer impliziten Normalverteilungsannahme optimiert und durch

$$k := \lceil \log_2 n + 1 \rceil \quad (10.6)$$

gegeben ist. Im vorliegenden Fall resultiert hier:

```
# Klassenanzahl nach Sturges (1926)
n      = length(y)                        # Anzahl Datenpunkte
k      = ceiling(log2(n)+1)               # Anzahl der Klassen
b      = seq(b_0, b_k, length.out = k + 1) # Klassen [b_{j-1}, b_j[ (breaks)
```



```
Anzahl Datenpunkte n      : 100
Klassenanzahl k          : 8
Intervalle [bj-1, bj]    : 14 15.125 16.25 17.375 18.5 19.625 20.75 21.875 23
Erzielte Klassenbreite d : 1.125
```

Eine weitere Möglichkeit ist es, Deskriptivstatistiken des Datensatzes in die Wahl der Klassenanzahl einfließen zu lassen. Als Maß für die Streuung der Daten im Datensatz nutzt die Klassenanzahl nach Scott (1979) die empirische Standardabweichung (cf. Kapitel 10.6), um durch das Histogramm eine Dichteschätzung bei Normalverteilung zu realisieren. Wir wollen den nichtparametrischen Hintergrund dieses Vorgehens hier nicht vertiefen, sondern halten lediglich fest, dass die Wahl der Klassenanzahl nach Scott (1979) durch Festlegung der Klassenbreite zu

$$d := 3.49S/\sqrt[3]{n} \quad (10.7)$$

und die anschließende Wahl der Klassenanzahl durch

$$k := \lceil (b_k - b_0)d \rceil \quad (10.8)$$

gegeben ist, wobei S die Standardabweichung des Datensatzes bezeichnet. Im vorliegenden Fall resultieren hier:

```
# Klassenanzahl nach Scott (1979)
n      = length(y)
S      = sd(y)
d      = ceiling(3.49*S/(n^(1/3)))
k      = ceiling((b_k - b_0)/d)
b      = seq(b_0, b_k, length.out = k + 1)

# Anzahl Datenwerte
# Stichprobenstandardabweichung
# Klassenbreite
# Anzahl der Klassen
# Klassen [b_{j-1}, b_j] (breaks)
```

```
Anzahl Datenpunkte n      : 100
Standardabweichung        : 1.740167
Klassenbreite nach Scott d : 2
Klassenanzahl k          : 5
Intervalle [bj-1, bj]    : 14 15.8 17.6 19.4 21.2 23
Erzielte Klassenbreite d : 1.8
```

Per Default nutzt die in **R** implementierte `hist()` Funktion eine Modifikation der von Sturges (1926) vorgeschlagenen Klassenanzahl und bietet eine Reihe von weiteren Möglichkeiten zur Wahl der Klassen. Einen Überblick dazu liefert `?hist`. Folgender **R** Code visualisiert die absoluten Häufigkeiten der Werte des Beispieldatensatzes anhand der oben spezifizierten Klassen.

```
# Abbildungsparameter
pdf(
  file      = "../figures/111-histogramme.pdf",
  width     = 6,
  height    = 9)
par(
  mfcol     = c(3,2),
  las       = 1,
  cex       = .7,
  cex.main  = .9)
x_min      = 12
x_max      = 25
y_min      = 0
y_max      = 45
label      = c("", "", "Excel", "Sturges", "Scott")

# Histogramme mit vordefinierter Klassenanzahl und Klassenbreite
for(i in 1:5){
  hist(
    y,
    breaks = bs[[i]],
    xlim   = c(x_min, x_max),
    ylim   = c(y_min, y_max),
    ylab    = "Häufigkeit",
    xlab    = "BDI",
    main    = sprintf(paste(label[i], "k = %.0f, d = %.2f"), ks[i], ds[i]))
}

# R Default Histogramm

# PDF Kopiefunktion
# Dateiname
# Breite (inch)
# Höhe (inch)
# Abbildungsparameter
# Zeilen und Spaltenlayout
# x-Tick Orientierung
# Annotationsvergrößerung
# Titelvergrößerung
# x-Achsenlimit
# y-Achsenlimit
# x-Achsenlimit
# y-Achsenlimit
# y-Achsenlimit
# Titel

# Histogramm
# Datensatz
# breaks Argument
# x-Achsenlimits
# y-Achsenlimits
# y-Achsenbezeichnung
# x-Achsenbezeichnung
# Titel
```

```

hist(                                     # Histogramm
y,                                       # Datensatz
xlim = c(x_min, x_max),               # x-Achsenlimits
ylim = c(y_min, y_max),               # y-Achsenlimits
ylab = "Häufigkeit",                  # y-Achsenbezeichnung
xlab = "",                             # x-Achsenbezeichnung
main = "R Default")                  # Titel
mtext(LETTERS[6], adj=0, line=2, cex = 1.2, at = 8) # Subplot Buchs
dev.off()                             # Beenden der Abbildungsbearbeitung

```

10.3. Empirische Verteilungsfunktionen

Neben der Frage, wie häufig in einem Datensatz bestimmte Werte oder Klassen von Werten vorkommen, kann es interessant sein, zu untersuchen, wie häufig Werte vorkommen, die kleiner als ein bestimmter Wert sind. Obwohl die Antwort auf diese Frage natürlich in den absoluten und relativen Häufigkeiten bzw. den entsprechenden Histogrammen implizit ist, ist es manchmal von Vorteil, diese Information direkt verfügbar zu haben. Die in diesem Abschnitt eingeführten Begriffe der *kumulativen absoluten und relativen Häufigkeitsverteilung* sowie der *empirischen Verteilungsfunktion* bezeichnen die entsprechenden Analoga zu den in Kapitel 10.1 und Kapitel 10.2 diskutierten Begriffen. Wir beginnen mit der Definition der *kumulativen absoluten* und *kumulativen relativen Häufigkeitsverteilungen*.

Definition 10.3 (Kumulative absolute und relative Häufigkeitsverteilungen). $y = (y_1, \dots, y_n)$ sei ein Datensatz, $W := \{w_1, \dots, w_k\}$ mit $k \leq n$ die im Datensatz vorkommenden Zahlenwerte und h und r die absoluten und relativen Häufigkeitsverteilungen der Werte von y , respektive. Dann heißt die Funktion

$$H : W \rightarrow \mathbb{N}, w \mapsto H(w) := \sum_{w' \leq w} h(w') \quad (10.9)$$

die *kumulative absolute Häufigkeitsverteilung* von y und die Funktion

$$R : W \rightarrow [0, 1], w \mapsto R(w) := \sum_{w' \leq w} r(w') \quad (10.10)$$

die *kumulative relative Häufigkeitsverteilung* der Zahlwerte von y .

•

Im Sinne der Definition Definition 10.1 gilt für die kumulativen Häufigkeitsverteilungen also

$$H(w) = \text{Anzahl der } y_i \text{ aus } y \text{ mit } y_i \leq w \quad (10.11)$$

und

$$R(w) = \text{Anzahl der } y_i \text{ aus } y \text{ mit } y_i \leq w \text{ geteilt durch } n. \quad (10.12)$$

In **R** erlaubt die Funktion `cumsum()` die Berechnung kumulativer Summen und ermöglicht damit die Auswertung kumulativer Häufigkeitsverteilungen, wie folgender **R** Code anhand des Beispieldatensatzes demonstriert.

```

D = read.csv("./_data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE                                             # Pre-Interventions-BDI-II-Daten
n = length(y)                                         # Anzahl der Datenwerte
H = as.data.frame(table(y))                          # absolute Häufigkeitsverteilung als Dataframe
names(H) = c("w", "h")                              # Spaltenbenennung
H$h = cumsum(H$h)                                    # kumulative absolute Häufigkeitsverteilung
H$r = H$h/n                                           # relative Häufigkeitsverteilung
H$R = cumsum(H$r)                                    # kumulative relative Häufigkeitsverteilung
print(H, digits = 1)

```

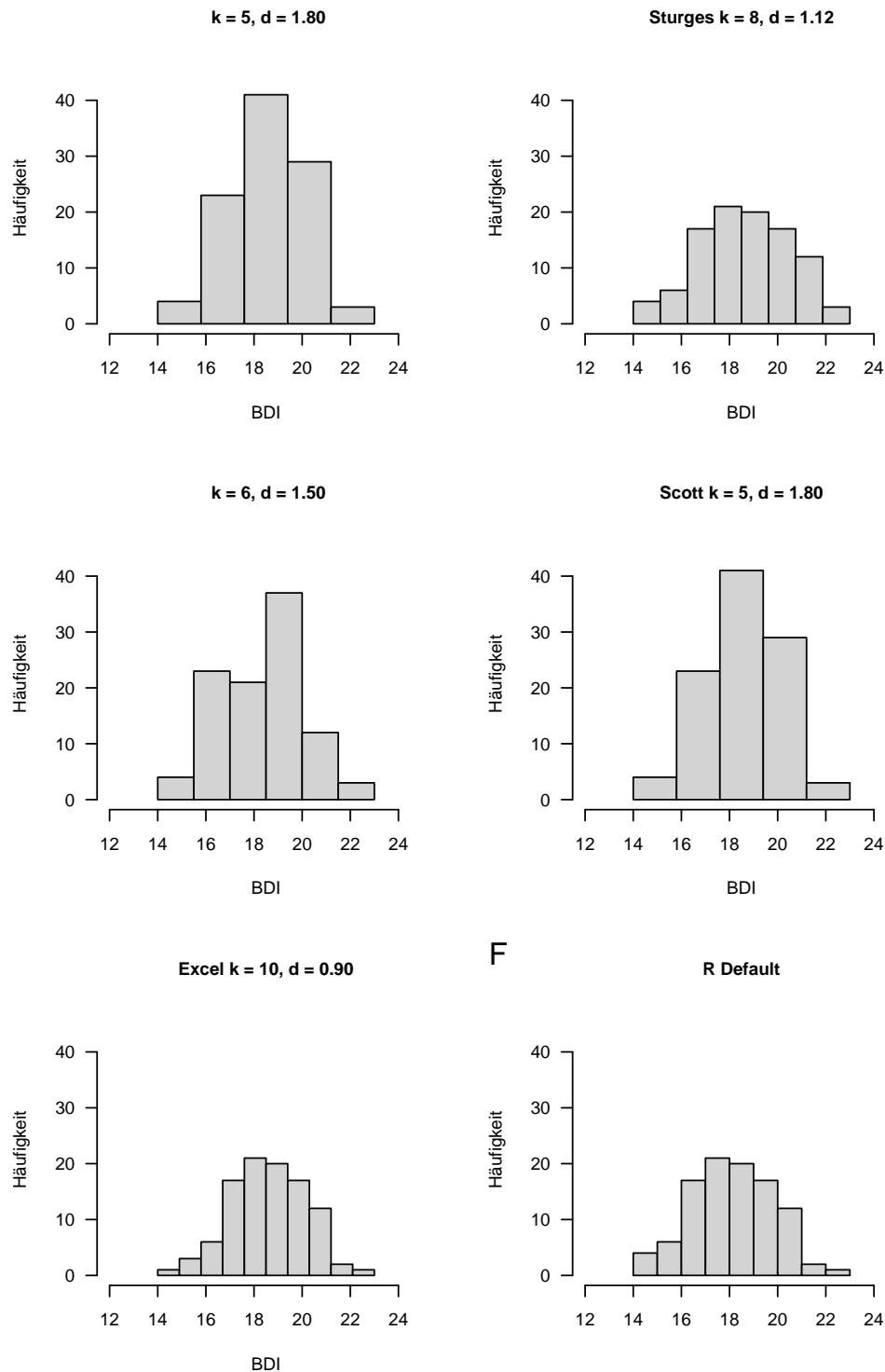


Abbildung 10.2. Histogramme der absoluten Häufigkeiten der Pre-Interventions-BDI-II-Werte. Man beachte, dass der qualitative Eindruck, dass im Datensatz Werte um 18 gehäuft auftreten und niedrigere und höhere Werte seltener sind, über alle Wahlen der Klassenanzahl konstant ist. Das genaue quantitative Erscheinungsbild ist dagegen von der exakten Wahl der Klassenanzahl und Klassenbreite abhängig.

	w	h	H	r	R
1	14	1	1	0.01	0.01
2	15	3	4	0.03	0.04
3	16	6	10	0.06	0.10
4	17	17	27	0.17	0.27
5	18	21	48	0.21	0.48
6	19	20	68	0.20	0.68
7	20	17	85	0.17	0.85
8	21	12	97	0.12	0.97
9	22	2	99	0.02	0.99
10	23	1	100	0.01	1.00

Man beachte dabei insbesondere die Gegenüberstellung von h und H sowie von r und R , die die kumulativen Summen in Definition 10.3 verdeutlichen. Folgender **R** Code ermöglicht die Darstellung der so bestimmten kumulativen Häufigkeitsverteilungen wie in Abbildung 10.3 gezeigt. Offensichtlich ist die absolute und die relative Häufigkeit von Werten kleiner als der kleinste Wert im Datensatz Null. Die absolute und relative Häufigkeit von Werten kleiner als der größte Wert im Datensatz ist dagegen n bzw. $n/n = 1$.

```
pdf(
  file      = "../figures/111-kumulative-haeufigkeitsverteilungen.pdf",
  width     = 10,
  height    = 5)
H          = H$H
R          = H$R
names(Hw)  = H$w
names(Rw)  = H$w
par(
  mfcol     = c(1,2),
  las      = 1,
  cex      = 1)
barplot(
  H,
  col      = "gray90",
  xlab     = "w",
  ylab     = "H(w)",
  ylim     = c(0,110),
  las      = 1,
  main     = "Absolute kumulative Häufigkeit PRE")
barplot(
  R,
  col      = "gray90",
  xlab     = "w",
  ylab     = "R(w)",
  ylim     = c(0,1),
  las      = 1,
  main     = "Relative kumulative Häufigkeit PRE")
dev.off()
```

PDF Kopiefunktion
Dateiname
Breite (inch)
Höhe (inch)
H(w) Werte
R(w) Werte
barplot braucht w Werte als names
barplot braucht w Werte als names
Abbildungsparameter
Zeilen und Spaltenlayout
x-Tick Orientierung
Annotationsvergrößerung
Balkendiagramm
H(w) Werte
Balkenfarbe
x-Achsenbeschriftung
y-Achsenbeschriftung
y-Achsenlimits
Achsenticklabelorientierung
Titel
Balkendiagramm
R(w) Werte
Balkenfarbe
x-Achsenbeschriftung
y-Achsenbeschriftung
y-Achsenlimits
Achsenticklabelorientierung
Titel
Beenden der Abbildungsbearbeitung

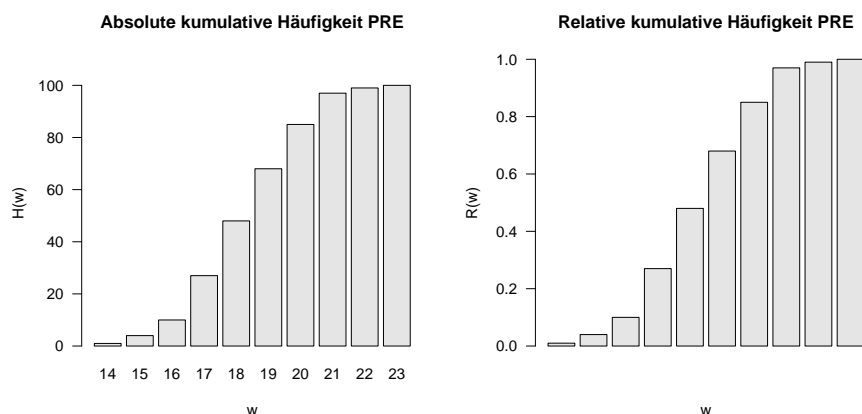


Abbildung 10.3. Absolute und relative kumulative Häufigkeitsverteilungen der Pre-Interventions-BDI-II-Werte .

Das kumulative Analogon zu einem Histogramm ist die *empirische Verteilungsfunktion*. Im Unterschied zu einem Histogramm ist für die Bestimmung einer empirischen

Verteilungsfunktion allerdings keine Definition von Klassen erforderlich. An die Stelle der Intervalleinteilung der reellen Zahlen zwischen dem Minimum und dem Maximum eines Datensatz treten bei der empirischen Verteilungsfunktion einfach die reellen Zahlen, wie folgende Definition verdeutlicht.

Definition 10.4 (Empirische Verteilungsfunktion). $y = (y_1, \dots, y_n)$ sei ein Datensatz. Dann heißt die Funktion

$$F : \mathbb{R} \rightarrow [0, 1], x \mapsto F(x) := \frac{\text{Anzahl der } y_i \text{ aus } y \text{ mit } y_i \leq x}{n} \quad (10.13)$$

die empirische Verteilungsfunktion (EVF) von y .

•

Die empirische Verteilungsfunktion wird manchmal auch die *empirische kumulative Verteilungsfunktion* genannt. Allgemein beschränkt man sich bei der empirischen Verteilungsfunktion wie in Definition 10.4 auf die Angabe der relativen Häufigkeiten. Natürlich kommen in Datensätzen nicht alle reellen Zahlen vor. Dies impliziert, dass die relative Häufigkeit, die reellen Zahlen zugeordnet wird, die zwischen zwei benachbarten Werten des Datensatzes liegen, derjenigen des kleineren der beiden Werte entspricht. Typischerweise sind empirische Verteilungsfunktionen also Treppenfunktionen. In **R** kann die Funktion `ecdf()` zur Evaluation und Visualisierung der empirischen Verteilungsfunktion eingesetzt werden, wie folgender **R** Code anhand des Beispieldatensatzes demonstriert und in Abbildung 10.4 visualisiert ist. Man liest aus Abbildung 10.4 zum Beispiel ab, dass die relative Häufigkeit von Werten kleiner als 17.9 etwa 0.3 ist. Die gleiche relative Häufigkeit haben Werte kleiner als 17.8.

```
D = read.csv("../data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Pre-Interventions-BDI-II-Daten
evf = ecdf(y) # Evaluation der EVF
pdf( # PDF Kopiefunktion
  file = "../figures/111-ecdf.pdf", # Dateiname
  width = 8, # Breite (inch)
  height = 5) # Höhe (inch)
plot( # plot() weiß mit ecdf object umzugehen
  evf, # ecdf Objekt
  xlab = "BDI", # x-Achsenbeschriftung
  ylab = "Relative Häufigkeit", # y-Achsenbeschriftung
  main = "Empirische Verteilungsfunktion") # Titel
dev.off() # Beenden der Abbildungsbearbeitung
```

10.4. Quantile und Boxplots

Die empirische Verteilungsfunktion gibt eine zumindest graphisch dargestellte Antwort auf die Frage, wie groß der relative Anteil der Werte eines Datensatzes ist, die kleiner als ein gegebener Wert sind. Die in diesem Abschnitt eingeführten Quantile können als Inversion dieser Datensatzinterrogation verstanden werden. Speziell gibt man bei Quantilen einen relativen Anteil $0 \leq p \leq 1$ der Werte eines Datensatzes vor und fragt, welcher Wert des Datensatzes so beschaffen ist, dass eben $p \cdot 100\%$ der Werte des Datensatzes kleiner als oder gleich diesem Wert sind. Wir nutzen folgende Definition des Begriffs des *p-Quantils*.

Definition 10.5 (*p*-Quantil). $y = (y_1, \dots, y_n)$ sei ein Datensatz und

$$y_s = (y_{(1)}, y_{(2)}, \dots, y_{(n)}) \text{ mit } \min_{1 \leq i \leq n} y_i = y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)} = \max_{1 \leq i \leq n} y_i \quad (10.14)$$

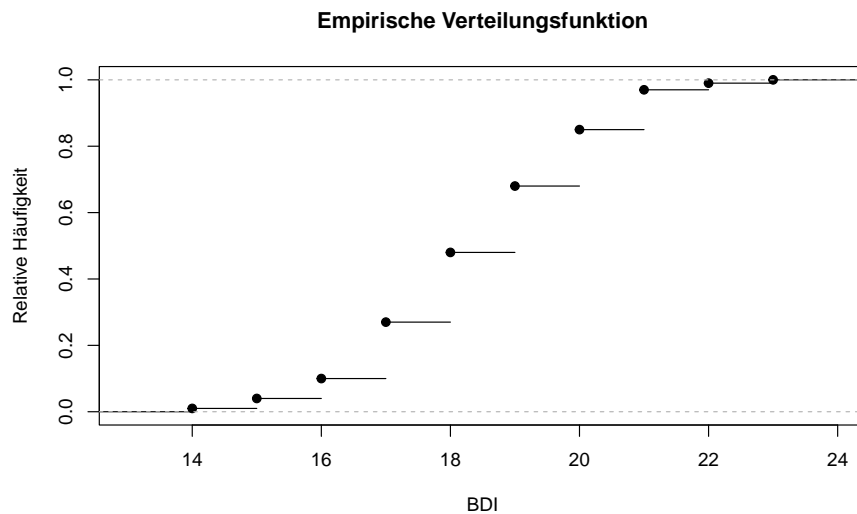


Abbildung 10.4. Empirische Verteilungsfunktion der PRE Werte.

sei der zugehörige aufsteigend sortierte Datensatz. Weiterhin bezeichne $\lfloor \cdot \rfloor$ die Abrundungsfunktion. Dann heißt für ein $p \in [0, 1]$ die Zahl

$$y_p := \begin{cases} y_{(\lfloor np+1 \rfloor)} & \text{falls } np \neq \mathbb{N} \\ \frac{1}{2} (y_{(np)} + y_{(np+1)}) & \text{falls } np \in \mathbb{N} \end{cases} \quad (10.15)$$

das p -Quantil des Datensatzes y .

•

Man beachte, dass das p -Quantil y_p nach Definition 10.5 entweder ein Wert des Datensatzes oder ein Wert zwischen zwei Werten des Datensatzes ist. Nach Konstruktion von y_p gilt dabei insbesondere, dass mindestens $p \cdot 100\%$ aller Werte des Datensatzes kleiner oder gleich y_p und mindestens $(1 - p) \cdot 100\%$ aller Werte des Datensatzes größer als y_p sind. Das p -Quantil teilt den geordneten Datensatz y_s also im Verhältnis p zu $(1 - p)$ auf. Je nach Wahl von p erhalten die p -Quantile spezielle Bezeichnungen:

- $y_{0.25}, y_{0.50}, y_{0.75}$ heißen *unteres Quartil*, *Median* und *oberes Quartil*, respektive,
- $y_{j \cdot 0.10}$ für $j = 1, \dots, 9$ heißen *Dezile* und
- $y_{j \cdot 0.01}$ für $j = 1, \dots, 99$ heißen *Percentile*.

Um den Begriff des p -Quantils zu verdeutlichen, betrachten wir ein an Henze (2018), Kapitel 5 orientiertes Beispiel. Dazu seien zunächst der in untenstehender Tabelle dargestellte Datensatz und seine aufsteigend sortierte Version gegeben.

i	1	2	3	4	5	6	7	8	9	10
y_i	8.5	1.5	75	4.5	6.0	3.0	3.0	2.5	6.0	9.0
$y_{(i)}$	1.5	2.5	3.0	3.0	4.5	6.0	6.0	8.5	9.0	75

Wir wollen als erstes das untere Quartil, also das 0.25-Quantil, dieses Datensatzes bestimmen. Es ist offenbar $n = 10$ und wir haben $p := 0.25$ gewählt. Basierend auf

Definition 10.5 finden wir, dass

$$np = 10 \cdot 0.25 = 2.5 \notin \mathbb{N}. \quad (10.16)$$

Also gilt nach Definition 10.5, dass

$$y_{0.25} = x_{(\lfloor 2.5+1 \rfloor)} = x_{(3)} = 3.0. \quad (10.17)$$

Das untere Quartil des betrachteten Datensatzes ist also der Wert 3.0 des Datensatzes.

Als zweites wollen wir das 0.80-Quantil des Datensatzes bestimmen. Es ist offenbar weiterhin $n = 10$ und wir haben nun $p := 0.80$ gewählt. Wir finden in diesem Fall, dass

$$np = 10 \cdot 0.80 = 8 \in \mathbb{N}. \quad (10.18)$$

Also folgt nach Definition 10.5

$$y_{0.80} = \frac{1}{2} (y_{(8)} + y_{(8+1)}) = \frac{1}{2} (y_{(8)} + y_{(9)}) = \frac{8.5 + 9.0}{2} = 8.75. \quad (10.19)$$

Das $y_{0.80}$ -Quantil des betrachteten Datensatzes ist mit 8.75 also der Wert zwischen den Werten 8.5 und 9 des Datensatzes. In **R** wertet man die so bestimmten Quantile wie folgt aus.

```
y = c(8.5,1.5,12,4.5,6.0,3.0,3.0,2.5,6.0,9.0) # Beispieldaten
n = length(y) # Anzahl Datenwerte
y_s = sort(y) # sortierter Datensatz
p = 0.25 # np \notin \mathbb{N}
y_p = y_s[floor(n*p + 1)] # 0.25 Quantil
cat("0.25 Quantil: ", round(y_p,2)) # Ausgabe
```

0.25 Quantil: 3

```
p = 0.80 # np \in \mathbb{N}
y_p = (1/2)*(y_s[n*p] + y_s[n*p + 1]) # 0.80 Quantil
cat("0.80 Quantil: ", round(y_p,2)) # Ausgabe
```

0.80 Quantil: 8.75

Mit `quantile()` stellt **R** auch eine Funktion bereit, die die Quantilbestimmung eines Datensatzes automatisiert. Dabei ist allerdings zu beachten, dass die in Definition 10.5 gegebene Definition des Quantilsbegriffs nur eine von mindestens neun verschiedenen Quantildefinitionen ist und die genauen Werte sich von Definition zu Definition unterscheiden können. Die spezielle Quantildefinition, die von `quantile()` genutzt wird, wird über das Funktionsargument `type` ausgewählt. Auf theoretischer Ebene geben Hyndman & Fan (1996) dazu einen Überblick, auf Implementationsebene `?quantile`. Wir betrachten exemplarisch die Bestimmung des 0.25-Quantils Pre-Interventions BDI Beispieldatensatzes anhand von `type = 8`.

```
D = read.csv("../data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Pre-Interventions-BDI-II-Daten
y_p = quantile(y, 0.25, type = 8) # 0.25 Quantil, Definition 1
```

0.25 Quantil (Type = 8): 17

Ein *Boxplot* schließlich visualisiert eine Quantil-basierte Zusammenfassung eines Datensatzes, wobei typischerweise das Minimum und das Maximum eines Datensatzes sowie das untere Quartil, der Median und das obere Quartil visualisiert werden. In **R** erstellt man einen Boxplot mithilfe der `boxplot()` Funktion, wie folgender **R** Code anhand des Pre-Interventions-BDI-II-Datensatzes demonstriert.

```

D      = read.csv("./_data/111-deskriptive-statistik.csv")      # Laden des Datensatzes
pdf(    # PDF Kopiefunktion
  file  = "./_figures/111-boxplot.pdf",      # Dateiname
  width = 8,                                # Breite (inch)
  height = 5)                                # Höhe (inch)
boxplot(D$PRE,                                # Boxplot
  horizontal = T,                             # Datensatz
  range = 0,                                  # horizontale Darstellung
  xlab = "BDI",                               # Whiskers bis zu min y und max y
  main = "Boxplot",                           # x Achsenbeschriftung
  dev.off())                                  # Titel
                                              # Beenden der Abbildungsbearbeitung

```

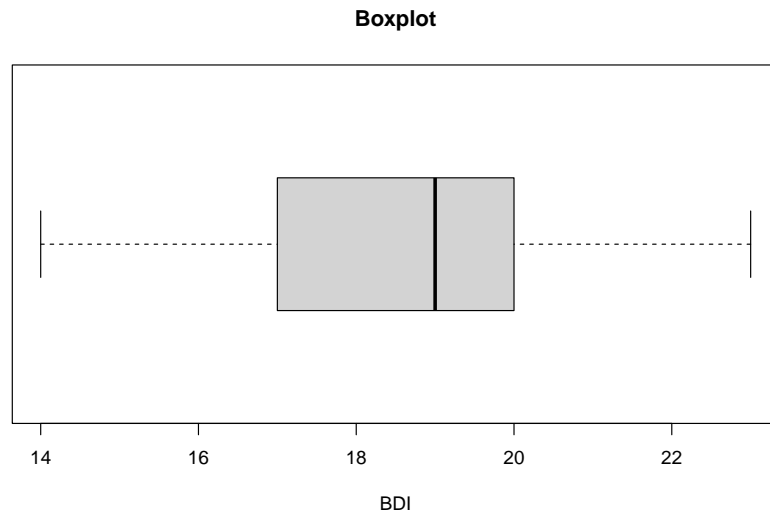


Abbildung 10.5. Boxplot der Pre-Interventions-BDI-Werte des Beispieldatensatzes.

In [Abbildung 10.5](#) werden $\min y$ und $\max y$ als sogenannte “Whiskerendpunkte” dargestellt, $y_{0.25}$ und $y_{0.75}$ bilden die unteren und oberen Grenze der zentralen grauen Box und der Median $y_{0.50}$ wird als Strich in der zentralen grauen Box abgebildet. Auch hier gilt, dass es sehr viele Boxplotvariationen gibt, vgl. McGill et al. (1978). Eine genaue Erläuterung der graphisch dargestellten Datensatzeigenschaften ist also folglich immer nötig. Schließlich sei angemerkt, dass die Darstellung von Datensätzen mithilfe von Boxplots in den letzten zwanzig Jahren etwas aus der Mode gekommen ist und man sie nur noch sehr selten in aktuellen wissenschaftlichen Publikationen findet.

10.5. Maße der Zentralen Tendenz

In diesem Abschnitt betrachten wir einige grundlegende Kennzahlen von Datensätzen, die für die deskriptive Statistik zentral sind und Aussagen zu dem “durchschnittlichen” Wert eines Datensatzes machen. Wir beginnen mit der Definition des Mittelwerts.

10.5.1. Mittelwert

Definition 10.6 (Mittelwert). $y = (y_1, \dots, y_n)$ sei ein Datensatz. Dann heißt

$$\bar{y} := \frac{1}{n} \sum_{i=1}^n y_i \quad (10.20)$$

der *Mittelwert* von y .

•

Der Mittelwert eines Datensatzes ist also die Summe der Datenwerte geteilt durch die Anzahl der Datenwerte. Der hier definierte Mittelwert wird auch als *arithmetisches Mittel* oder *Durchschnitt* bezeichnet. Im Kontext der Frequentistischen Inferenz werden wir den Mittelwert vor dem Hintergrund des Modells der Zufallsstichprobe als *Stichprobenmittel* bezeichnen. Insbesondere erlauben die Modelle der Frequentistischen Inferenz auch, die funktionale Form des Mittelwertes tiefer zu begründen, wie wir an gegebener Stelle sehen werden. Mithilfe der Summenfunktion berechnet man den Mittelwert eines Datensatzes in **R** wie folgt.

```
D = read.csv("../data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Pre-Interventions-BDI-II-Daten
n = length(y) # Anzahl der Werte
y_bar = (1/n)*sum(y) # Mittelwertberechnung
```

Mittelwert der Pre-BDI-Interventions Daten 18.61

Mit der `mean()` Funktion stellt **R** natürlich auch eine Funktion zur automatisierten Berechnung des Mittelwerts bereit.

```
cat("Mittelwert der PRE-BDI-II-Daten ", mean(y)) # Ausgabe
```

Mittelwert der PRE-BDI-II-Daten 18.61

In folgendem Theorem halten wir zunächst zwei Eigenschaften des Mittelwerts eines Datensatzes fest, die seine Einordnung als Maß der zentralen Tendenz eines Datensatzes plausibel machen.

Theorem 10.1 (Zentralitätseigenschaften des Mittelwerts). $y = (y_1, \dots, y_n)$ sei ein Datensatz und \bar{y} sei der Mittelwert von y . Dann gelten

(1) Die Summe der Abweichungen vom Mittelwert ist Null,

$$\sum_{i=1}^n (y_i - \bar{y}) = 0. \quad (10.21)$$

(2) Die absoluten Summen negativer und positiver Abweichungen vom Mittelwert sind gleich, d.h. wenn $j = 1, \dots, n_j$ die Datenpunktindizes mit $(y_j - \bar{y}) < 0$ und $k = 1, \dots, n_k$ die Datenpunktindizes mit $(y_k - \bar{y}) \geq 0$ bezeichnen, dann gilt mit $n_j + n_k$, dass

$$\left| \sum_{j=1}^{n_j} (y_j - \bar{y}) \right| = \left| \sum_{k=1}^{n_k} (y_k - \bar{y}) \right|. \quad (10.22)$$

◦

Beweis. (1) Es gilt

$$\begin{aligned}
 \sum_{i=1}^n (y_i - \bar{y}) &= \sum_{i=1}^n y_i - \sum_{i=1}^n \bar{y} \\
 &= \sum_{i=1}^n y_i - n\bar{y} \\
 &= \sum_{i=1}^n y_i - \frac{n}{n} \sum_{i=1}^n y_i \\
 &= \sum_{i=1}^n y_i - \sum_{i=1}^n y_i \\
 &= 0.
 \end{aligned} \tag{10.23}$$

(2) Seien $j = 1, \dots, n_j$ die Indizes mit $(y_j - \bar{y}) < 0$ und $k = 1, \dots, n_k$ die Indizes mit $(y_k - \bar{y}) \geq 0$, so dass $n = n_j + n_k$. Dann gilt

$$\begin{aligned}
 \sum_{i=1}^n (y_i - \bar{y}) &= 0 \\
 \Leftrightarrow \sum_{j=1}^{n_j} (y_j - \bar{y}) + \sum_{k=1}^{n_k} (y_k - \bar{y}) &= 0 \\
 \Leftrightarrow \sum_{k=1}^{n_k} (y_k - \bar{y}) &= -\sum_{j=1}^{n_j} (y_j - \bar{y}) \\
 \Leftrightarrow \left| \sum_{j=1}^{n_j} (y_j - \bar{y}) \right| &= \left| \sum_{k=1}^{n_k} (y_k - \bar{y}) \right|.
 \end{aligned} \tag{10.24}$$

□

Nach Theorem 10.1 liegt der Mittelwert eines Datensatzes also gerade so, dass sich in Summe die Abweichungen der Datenpunkte von diesem Wert zu Null ausgleichen und dass sich positive und negative Abweichungen vom Mittelwert die Waage halten. In **R** kann man Theorem 10.1 am Beispieldatensatz wie folgt verifizieren.

```

# Datensatz
D = read.csv("../data/111-deskriptive-statistik.csv")
y = D$PRE
s = sum(y - mean(y))
s_1 = sum(y[y <= mean(y)] - mean(y))
s_2 = sum(y[y > mean(y)] - mean(y))

# Laden des Datensatzes
# Pre-Interventions-BDI-II-Daten
# Summe der Abweichungen vom Mittelwert
# Summe aller negativen Abweichungen
# Summe aller positiven Abweichungen

```

```

Summe der Abweichungen vom Mittelwert : 0
Summen der positiven und negativen Abweichungen : -71.28 71.28

```

Theorem 10.2 (Summationsseigenschaften des Mittelwerts). *Gegeben seien zwei Datensätze $y = (y_1, \dots, y_n)$ und $z = (z_1, \dots, z_n)$ gleichen Umfangs und es sei durch*

$$y + z := (y_1 + z_1, \dots, y_n + z_n) \tag{10.25}$$

die Summe dieser Datensätze definiert. Weiterhin seien \bar{y} und \bar{z} die Mittelwerte der Datensätze y und z , respektive. Dann gilt

$$\overline{y + z} = \bar{y} + \bar{z} \tag{10.26}$$

◦

Beweis. Es gilt

$$\begin{aligned}
 \overline{y+z} &:= \frac{1}{n} \sum_{i=1}^n (y_i + z_i) \\
 &= \frac{1}{n} \sum_{i=1}^n y_i + \frac{1}{n} \sum_{i=1}^n z_i \\
 &=: \bar{y} + \bar{z}.
 \end{aligned} \tag{10.27}$$

□

Betrachtet man die Summe zweier Datensätze, so ist es für die Bestimmung ihres Mittelwerts also unerheblich, ob man zunächst die Datensätze aufaddiert und dann den Mittelwert des summierten Datensatzes bestimmt oder ob man für jeden der beiden Datensätze den Mittelwert bestimmt und diese aufaddiert. Wir verifizieren dieses Ergebnis exemplarisch mithilfe der Pre- und Post-Interventions BDI-II-Daten des Beispieldatensatzes in **R** wie folgt.

```

D      = read.csv("../data/111-deskriptive-statistik.csv")      # Laden des Datensatzes
y      = D$PRE                                                  # Pre-Interventions-BDI-II-Daten
y_bar  = mean(y)                                                # Mittelwert der Pre-Interventions-BDI-II-Daten
z      = D$POS                                                  # Post-Interventions BDI-II-Daten
z_bar  = mean(z)                                                # Mittelwert der Post-Interventions BDI-II-Daten
yz_bar_1 = mean(y + z)                                         # Mittelwert der summierten Pre- und Post-Daten
yz_bar_2 = y_bar + z_bar                                       # Summe der Mittelwerte der Pre- und Post-Date

```

```

Mittelwert von z      : 31.68
Summe der Mittelwerte von x und y : 31.68

```

Theorem 10.3 (Mittelwert bei linear-affiner Transformation). $y = (y_1, \dots, y_n)$ sei ein Datensatz und \bar{y} sei der Mittelwert von y . Weiterhin sei für $a, b \in \mathbb{R}$ mit

$$ay + b := (ay_1 + b, \dots, ay_n + b) \tag{10.28}$$

ein linear-affin transformierter Datensatz von y definiert. Dann gilt, dass

$$\overline{ay + b} = a\bar{y} + b \tag{10.29}$$

◦

Beweis. Es gilt

$$\begin{aligned}
 \overline{ay + b} &:= \frac{1}{n} \sum_{i=1}^n (ay_i + b) \\
 &= \sum_{i=1}^n \left(\frac{1}{n} ay_i + \frac{1}{n} b \right) \\
 &= \left(\sum_{i=1}^n \frac{1}{n} ay_i \right) + \left(\sum_{i=1}^n \frac{1}{n} b \right) \\
 &= a \left(\frac{1}{n} \sum_{i=1}^n y_i \right) + \frac{1}{n} \sum_{i=1}^n b \\
 &= a\bar{y} + \frac{1}{n} nb \\
 &= a\bar{y} + b.
 \end{aligned} \tag{10.30}$$

□

Eine linear-affine Transformation eines Datensatz transformiert den Mittelwert des Datensatzes also linear-affin. Dies kann man sich zunutze machen, wenn man sich überlegt, wie der Mittelwert lauten sollte, wenn man einen Datensatz neu skaliert, also z.B. von Gramm in Kilogramm umrechnet. Natürlich kann man zur Sicherheit aber auch einfach den Mittelwert des neu skalierten Datensatzes bestimmen. Wir demonstrieren diese Eigenschaft exemplarisch durch Reskalierung des Pre-Interventions-BDI-II-Datensatzes mit dem Wert $a = 2$ unter gleichzeitiger Addition der Konstante $b := 5$.

```
D = read.csv("../data/111-deskriptive-statistik.csv")      # Laden des Datensatzes
y = D$PRE                                                  # Pre-Interventions-BDI-II-Daten
y_bar = mean(y)                                            # Mittelwert der Pre-Interventions-BDI-II-Daten
a = 2                                                      # Multiplikationskonstante
b = 5                                                      # Additionskonstante
z = a*y + b                                                # linear-affine Transformation der PRE Daten
z_bar = mean(z)                                            # Mittelwert der transformierten PRE Daten
ay_bar_b = a*y_bar + b                                    # Transformation des PRE Daten Mittelwerts
```

```
Mittelwert des transformierten Datensatzes : 42.22
Transformation des Mittelwertes           : 42.22
```

10.5.2. Median

In den oben definierten Mittelwert geht jeder Wert eines Datensatzes summativ mit gleichem Gewicht $1/n$ ein. Dies betrifft insbesondere auch solche Werte, die im Vergleich zu den übrigen Werten des Datensatzes sehr untypisch sind. Eine Möglichkeit, die “Mitte” eines Datensatzes unabhängig von solchen untypischen Werten zu bestimmen, besteht in der Bestimmung seines Medians, den wir nachfolgend definieren. Wir haben den Median bereits in Abschnitt Kapitel 10.4 als das 0.5-Quantil kennengelernt, mit dem er identisch ist.

Definition 10.7 (Median). $y = (y_1, \dots, y_n)$ sei ein Datensatz und $y_s = (y_{(1)}, \dots, y_{(n)})$ der zugehörige aufsteigend sortierte Datensatz. Dann ist der Median von y definiert als

$$\tilde{y} := \begin{cases} y_{((n+1)/2)} & \text{falls } n \text{ ungerade,} \\ \frac{1}{2} (y_{(n/2)} + y_{(n/2+1)}) & \text{falls } n \text{ gerade.} \end{cases} \quad (10.31)$$

•

Definition 10.7 impliziert, dass mindestens 50% aller Datenwerte y_i des Datensatzes kleiner oder gleich \tilde{y} sind und dass gleichzeitig mindestens 50% aller Datenwerte y_i des Datensatzes größer oder gleich \tilde{y} sind. Anstelle eines formalen Beweises dieser Aussagen verweisen wir auf Abbildung 10.6. Wie in der Definition des Medians unterscheiden wir dort die Fälle, dass die Anzahl der Datenpunkte ungerade ist (hier $n = 5$, Abbildung 10.6 A) oder gerade ist ($n = 6$, Abbildung 10.6 B). Im Fall der ungeraden Anzahl an Datenpunkten ist der Median der Datenwert mit der Indexzahl, die in der Mitte der Indexzahlen des sortierten Datensatzes liegt, im Beispiel Abbildung 10.6 A also der Wert $y_{(3)}$. Damit sind im vorliegenden Fall $3/5$, also 60% der Datenwerte (speziell $y_{(1)}$, $y_{(2)}$ und $y_{(3)}$) kleiner oder gleich dem Median, gleichzeitig aber auch $3/5$, also 60% der Datenwerte (speziell $y_{(3)}$, $y_{(4)}$ und $y_{(5)}$) größer oder gleich dem Median. Im Fall einer geraden Anzahl von Datenpunkten ist die Lage des Medians etwas intuitiver: Der Median liegt in der Mitte der beiden mittleren Werte des sortierten Datensatzes (in Abbildung 10.6 B speziell in der Mitte der Werte $y_{(3)}$ und $y_{(4)}$). Damit sind dann $3/6$, also 50% der Datenwerte (speziell

A

Beispiel für n ungerade

$$n := 5 \Rightarrow \left(\frac{5+1}{2}\right) = (3) \Rightarrow \tilde{y} := y_{(3)}$$

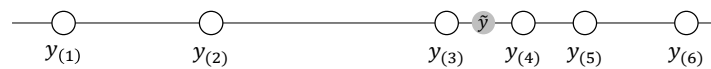


$$y_{(1)}, y_{(2)}, y_{(3)} \leq \tilde{y} \leq y_{(3)}, y_{(4)}, y_{(5)}$$

B

Beispiel für n gerade

$$n := 6 \Rightarrow \left(\frac{6}{2}\right) = (3), \left(\frac{6}{2} + 1\right) = (4) \Rightarrow \tilde{y} := \frac{1}{2}(y_{(3)} + y_{(4)})$$



$$y_{(1)}, y_{(2)}, y_{(3)} < \tilde{y} < y_{(4)}, y_{(5)}, y_{(6)}$$

Abbildung 10.6. Bestimmung und Eigenschaften des Medians.

$y_{(1)}$, $y_{(2)}$ und $y_{(3)}$ kleiner als der Median und $3/6$, also 50% der Datenwerte (speziell $y_{(4)}$, $y_{(5)}$ und $y_{(6)}$) größer als der Median.

Folgender **R** Code bestimmt den Median des Beispieldatensatzes anhand von Definition 10.7

```
D = read.csv("./_data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Pre-Interventions-BDI-II-Daten
n = length(y) # Anzahl der Werte
y_s = sort(y) # aufsteigend sortierter Vektor
if(n %% 2 == 1){ # n ungerade, n mod 2 == 1
  y_tilde = y_s[(n+1)/2] # Medianformel, Fall n ungerade
} else { # n gerade, n mod 2 == 0
  y_tilde = (y_s[n/2] + y_s[n/2 + 1])/2 # Medianformel, Fall n gerade
```

Median der PRE-BDI-II-Daten: 19

Natürlich stellt **R** mit `median()` auch eine Funktion zur direkten Bestimmung, wie folgender **R** Code demonstriert.

```
cat("Median der PRE-BDI-II-Daten: ", median(y)) # Ausgabe
```

Median der PRE-BDI-II-Daten: 19

Folgende Simulation zeigt beispielhaft, dass der Median als Maß der Mitte eines Datensatzes weniger anfällig für Ausreißer ist als der Mittelwert.

```
D = read.csv("./_data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Pre-Interventions-BDI-II-Daten
```

Mittelwert und Median für Datensatz ohne Ausreißer: 18.61 19

```
z      = y                                # neuer simulierter Datensatz
z[1]   = 10000                           # ... mit einem Extremwert
```

Mittelwert und Median für Datensatz mit Ausreißer: 118.44 19

Obwohl der Median also als robusteres Maß für die zentrale Tendenz eines Datensatzes als der Mittelwert erscheint, ist die Bestimmung von Mittelwerten sicherlich das üblichere Vorgehen. Dies ist einerseits dadurch begründet, dass sich der Mittelwert aufgrund seiner mathematisch weniger komplexen Definition einfacher im analytischen Kontext zur Gestaltung von probabilistischen Modellen eignet. Andererseits ist das Auftreten von Extremwerten üblicherweise ein Anzeichen von Datenerhebungsfehlern (z.B. sollte eine BDI-II Wert von 10.000 schlicht nicht vorkommen) oder großen Unterschieden zwischen den betrachteten experimentellen Einheiten, die durch visuelle Inspektion des Datensatzes meist vor der Bestimmung von Maßen der zentralen Tendenz auffallen. Es gibt allerdings ein ganzes Unterfeld der Statistik, dass sich mit - im weitesten Sinne - Ausreißer unabhängigen Maßen beschäftigt, vgl. z.B. Huber (1981).

10.5.3. Modalwert

Wenn in einem Datensatz manche Werte wiederholt auftreten, ist mit dem *Modalwert* als dem in einem Datensatz am häufigsten vorkommenden Wert ein drittes Maß der zentralen Tendenz gegeben. Wir nutzen folgende Definition.

Definition 10.8 (Modalwert). $y := (y_1, \dots, y_n)$ mit $y_i \in \mathbb{R}$ sei ein Datensatz, $W := \{w_1, \dots, w_k\}$ mit $k \leq n$ seien die im Datensatz vorkommenden verschiedenen Zahlenwerte und $h : W \rightarrow \mathbb{N}$ sei die absolute Häufigkeitsverteilung der Zahlwerte von y . Dann ist der *Modalwert* (oder *Modus*) von y definiert als

$$\operatorname{argmax}_{w \in W} h(w), \quad (10.32)$$

also als der im Datensatz am häufigsten vorkommende Wert.

•

In **R** kann man den Modalwert eines entsprechenden Datensatzes wie folgt bestimmen.

```
D      = read.csv("./_data/111-deskriptive-statistik.csv")      # Laden des Datensatzes
y      = D$PRE                                                  # Pre-Interventions-BDI-II-Daten
n      = length(y)                                              # Anzahl der Datenwerte (100)
H      = as.data.frame(table(y))                                # absolute Häufigkeitsverteilung (dataframe)
names(H) = c("a", "w")                                         # konsistente Benennung
mode   = H$w[which.max(H$w)]                                    # Modalwert

cat("Modalwert der PRE-BDI-II-Daten: ", as.numeric(as.vector(mode))) # Ausgabe als numeric vector, nicht factor
```

Modalwert der PRE-BDI-II-Daten: 21

10.5.4. Datenverteilungsformen und Maße zentraler Tendenz

Wie sinnvoll es ist, mit dem Mittelwert, dem Median oder dem Modalwert eine Angabe über die zentrale Tendenz eines Datensatzes zu geben, hängt stark von der Beschaffenheit des Datensatzes ab. Hat man beispielsweise einen Datensatz hochaufgelöster Daten vorliegen, in dem kein Datenwert mehrfach auftaucht, eignet sich der Modalwert sicherlich nicht, um eine quantitative Aussage über die zentrale Tendenz des Datensatzes zu machen. Man sollte die Angabe eines Maßes der zentralen Tendenz also immer im Kontext der Gesamtbeschaffenheit des Datensatzes sowie, wenn zutreffend, des inferenzstatistischen Modells, das man zur Interrogation des Datensatzes nutzt, betrachten. In Abbildung 10.7 geben wir vier Beispiele, wie sich die oben betrachteten Maße der zentralen Tendenz in verschiedenen Szenarien der Datenwertverteilung verhalten können. Diese vier Szenarien sind natürlich nicht erschöpfend und jeder Datensatz ist in Hinblick auf ein sinnvolles Maß der zentralen Tendenz kritisch zu betrachten.

Abbildung 10.7 A visualisiert die Lage von Mittelwert, Median und Modalwert in einem Datensatz, der sich symmetrisch um einen Wert von etwa $y = 50$ verteilt und für den Werte nah bei diesem zentralen Wert häufig vorkommen und Werte fern von diesem zentralen Wert eher selten. Wie wir später sehen, entspricht dies den typischen Charakteristika normalverteilter Datenwerte. In diesem Fall liegen Mittelwert, Median und Modalwert recht nahe beieinander und jedes dieser Maße beschreibt in der Tat die zentrale Tendenz des Datensatzes.

Abbildung 10.7 B visualisiert die Lage von Mittelwert, Median und Modalwert in einem Datensatz, in dem die Datenwerte um zwei Werte von etwa $y = 25$ und $y = 75$ gehäuft auftreten und jeweils in positive und negative Richtung von diesen Werten entfernt eher selten auftreten. So liegen zum Beispiel um $y = 50$ recht wenige Datenpunkte dieses Datensatzes. Mittelwert und Median nehmen nun aber gerade Werte um $y = 50$ an. Als solche tragen sie bei diesem Datensatz also keine Bedeutung darüber, welche Werte besonders häufig vorkommen, sondern lediglich darüber, wo sich der mittlere “physikalische Schwerpunkt” des Datensatzes befindet. Der Modalwert hingegen liegt, aufgrund des etwas häufigeren Auftretens von Werten um $y = 75$, bei dem größeren der beiden Datenhäufungspunkte. Insgesamt betrachtet beschreibt keines der drei Maße bei einem solchen Datensatz, den man auch *bimodal* nennt, die zentrale Tendenz der Datenwerte wirklich zufriedenstellend.

Abbildung 10.7 C visualisiert die Lage von Mittelwert, Median und Modalwert in einem Datensatz, in dem die Datenwerte in der betrachteten Breite etwa durchgängig gleichhäufig auftreten. Hier liegt der Modalwert relativ arbiträr bei etwa $y = 15$, da sich dort eine minimale Häufung von Datenwerten zeigt. Mittelwert und Median liegen wiederum im Sinne des “physikalischen Schwerpunkts” des Datensatzes bei etwa $y = 50$, auch wenn die Werte dieses Datensatzes dort keinerlei Tendenz haben, häufiger als andersorts aufzutreten. Wirklich zufriedenstellend ist auch hier die Beschreibung der zentralen Tendenz des Datensatzes durch die betrachteten Maße nicht. Zusammen betrachtet kann man anhand von Abbildung 10.7 B und Abbildung 10.7 C vielleicht festhalten, dass wenn ein Datensatz keine klare zentrale Tendenz hat, die Maße der zentralen Tendenz diese auch nicht abbilden können.

Abbildung 10.7 D schließlich zeigt die Lage von Mittelwert, Median, und Modalwert für einen Datensatz, der schiefsymmetrisch um Werte von etwa $y = 0.5$ verteilt ist. Die Daten dieses Datensatzes nehmen nur positive Werte an, viele Datenwerte liegen in der Tat um

$y = 0.5$, aber es kommen auch höhere Datenwerte vor. Wir werden an späterer Stelle sehen, dass dieses Erscheinungsbild für quadrierte normalverteilte Datenwerte typisch ist. Im psychologischen Kontext haben insbesondere Reaktionszeiten oft eine ähnliche Verteilung, da sie nicht negativ sein können, in den meisten Fällen etwa im Bereich von 0.3 bis 0.6 Sekunden liegen, aber in manchen Fällen durch Unaufmerksamkeit oder Stimulusvariabilität auch sehr viel länger sein können. In diesem Fall beschreibt der Median die tatsächliche zentrale Tendenz der Datenwerte etwas besser als der Mittelwert, da dieser durch die selten vorkommenden Ausreißerwerte etwas zu höheren Werten verschoben ist.

Insgesamt ist die Angabe eines oder mehrerer Maße der zentralen Tendenz also immer kritisch vor dem Hintergrund der Verteilung der Werte eines Datensatzes zu betrachten und qualitativ einzuordnen.

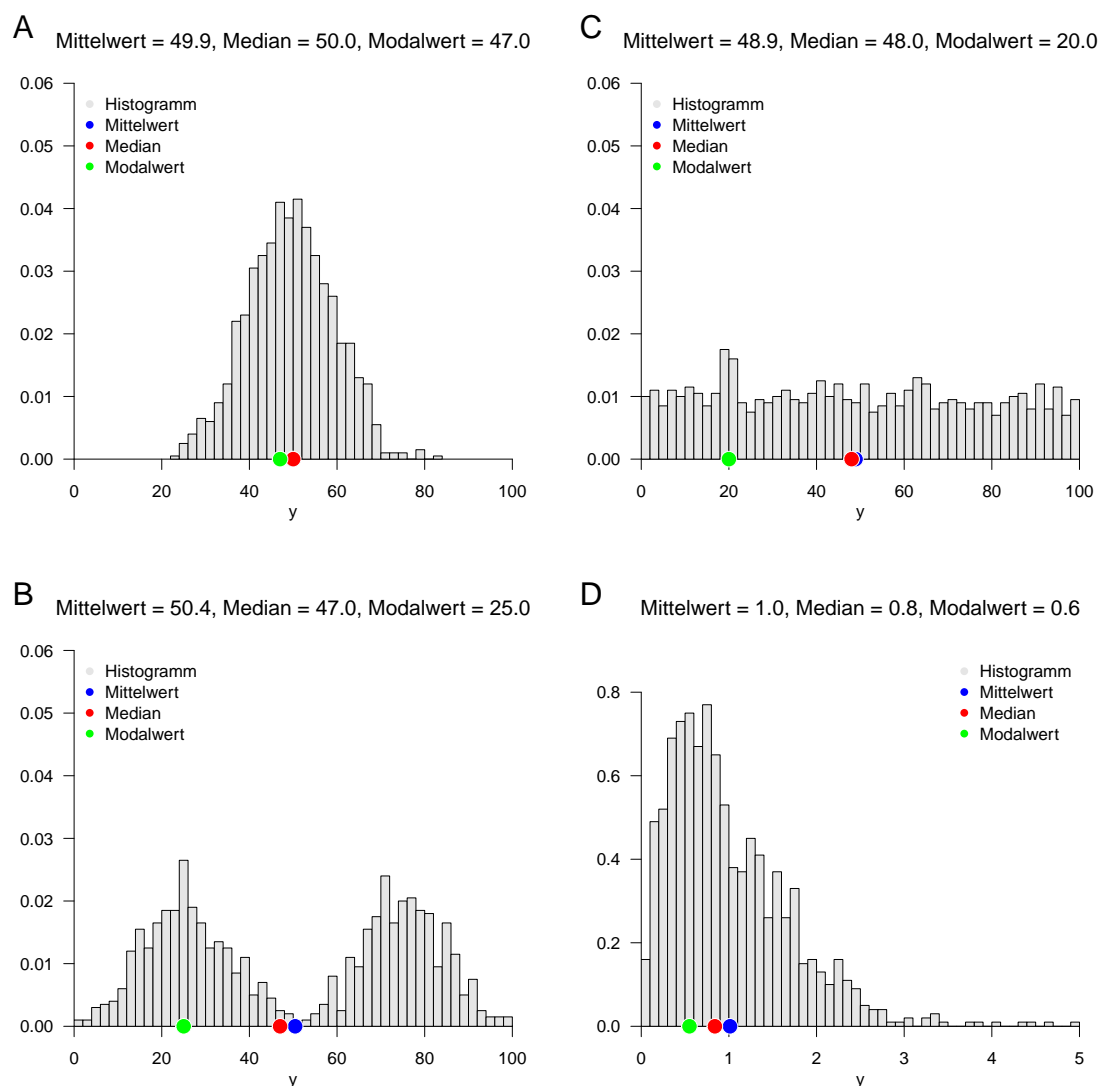


Abbildung 10.7. Visuelle Intuition zu Maßen zentraler Tendenz bei verschiedenen Datenverteilungsformen

10.6. Maße der Variabilität

Neben der Frage, welchen “durchschnittlichen Wert” ein Datensatz annimmt, ist es oft von Interesse zu quantifizieren, wie sehr die Werte im Datensatz streuen oder anders ausgedrückt, wie “variabel” sie sind. Als quantitative Maße für die Variabilität eines Datensatzes betrachten wir hier die *Spannbreite*, die *empirische Varianz* und die *empirische Standardabweichung*. Wie auch für den Mittelwert erschließt sich die volle Bedeutung insbesondere letztere Maße vor allem im Kontext inferenzstatistischer Verfahren.

10.6.1. Spannbreite

Die Spannbreite eines Datensatzes ist die Differenz zwischen seinem größten und seinem kleinsten Wert.

Definition 10.9 (Spannbreite). $y = (y_1, \dots, y_n)$ sei ein Datensatz. Dann ist die *Spannbreite* von y_1, \dots, y_n definiert als

$$r := \max(y_1, \dots, y_n) - \min(y_1, \dots, y_n). \quad (10.33)$$

•

In **R** bestimmt man die Spannbreite entweder durch Evaluation des Maximums und des Minimums des Datensatzes mithilfe der `max()` und `min()` Funktionen oder direkt mithilfe der `range()` Funktion. Folgender **R** Code demonstriert die Anwendung der `max()` und `min()` Funktionen.

```
D      = read.csv("../data/111-deskriptive-statistik.csv")      # Laden des Datensatzes
y      = D$PRE                                                  # Pre-Interventions-BDI-II-Daten
y_max  = max(y)                                                  # Maximum der PRE Werte
y_min  = min(y)                                                  # Minimum der PRE Werte
r      = y_max - y_min                                           # Spannbreite
```

Spannbreite der PRE-BDI-II-Daten: 9

Folgender **R** Code demonstriert die Anwendung der `range()` Funktion.

```
MinMax = range(y)                                                # "automatische" Berechnung von min(x), max(x)
r      = MinMax[2] - MinMax[1]                                    # Spannbreite
```

Spannbreite der PRE-BDI-II-Daten: 9

10.6.2. Empirische Varianz

Ein typisches statistisches Maß für die Variabilität von Datensätzen ist die standardisierte Summe der quadrierten Differenzen zwischen den einzelnen Datensatzwerten und ihrem Mittelwert. Die quadrierten Differenzen zwischen den einzelnen Datensatzwerten und ihrem Mittelwert werden auch als *Abweichungsquadrate* bezeichnet. Je nach Form der Standardisierung, also der Bildung eines Durchschnitts der Abweichungsquadrate, unterscheidet man eine *unkorrigierte empirische Varianz* und eine *empirische Varianz*, wie folgende Definition festsetzt.

Definition 10.10 (Unkorrigierte empirische Varianz und empirische Varianz). $y = (y_1, \dots, y_n)$ sei ein Datensatz mit $n > 1$ und \bar{y} sei sein Mittelwert. Dann ist die *unkorrigierte empirische Varianz* von y definiert als

$$\tilde{s}^2 := \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (10.34)$$

und die *empirische Varianz* von y ist definiert als

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (10.35)$$

•

Wir haben in Theorem 10.1 gesehen, dass die Summe der Differenzen zwischen den Datenwerten und ihrem Mittelwert immer gleich Null ist. Die Quadrierung der Differenzen

$$y_i - \bar{y} \text{ für } i = 1, \dots, n \quad (10.36)$$

in Definition 10.10 führt nun gerade dazu, dass sich negative und positive Differenzen von Datenpunkten nicht ausgleichen und, je nach Höhe der positiven und negativen Differenzen, mit

$$\sum_{i=1}^n (y_i - \bar{y})^2 \quad (10.37)$$

ein höheres oder geringeres Maß für die Gesamtabweichung der Datenwerte von ihrem Mittelwert bestimmt werden kann. Dabei ist die Summe der Abweichungsquadrate gerade dann Null, wenn alle Datenwerte identisch und damit auch identisch mit ihrem Mittelwert sind. Alternativ wäre zum Beispiel auch die Bestimmung der Absolutbeträge $|y_i - \bar{y}|$ denkbar, allerdings erfüllt das so entstehende Maß der Variabilität andere Eigenschaften als das Maß der quadrierten Abweichungen in Definition 10.10, weshalb dieses, auch in Hinblick auf seine Nähe zu probabilistischen Normalverteilungsmodellen in der Regel bevorzugt wird. Die unkorrigierte empirische Varianz \tilde{s}^2 und die (korrigierte) empirische Varianz s^2 unterscheiden sich dann hinsichtlich der Form der Durchschnittsbildung der quadrierten Abweichungsquadrate. Für \tilde{s}^2 wird die Summe der Abweichungsquadrate durch n , für s^2 durch $n-1$ geteilt. \tilde{s}^2 ist also immer etwas kleiner als s^2 . Allerdings wird dieser Unterschied für großes n eher gering ausfallen, man denke beispielsweise an den numerischen Unterschied zwischen $\frac{1}{3}$ und $\frac{1}{2}$ bei kleinem n sowie zwischen $\frac{1}{1001}$ und $\frac{1}{1000}$ bei großem n . Man beachte, dass die empirische Varianz für $n=1$ nicht definiert ist, die unkorrigierte empirische Varianz jedoch auch in diesem Spezialfall prinzipiell bestimmt werden kann. Der Begriff *unkorrigiert* (und implizit *korrigiert*) kann sich an dieser Stelle noch nicht erschließen, sondern ergibt nur vor dem Hintergrund eines Frequentistischen Inferenzmodells Sinn und wird in [?@sec-parameterschaetzung](#) erläutert werden.

Wir fassen die obige Diskussion in folgendem Theorem zusammen

Theorem 10.4 (Verhältnis von unkorrigierter empirischer Varianz und empirischer Varianz). $y = (y_1, \dots, y_n)$ sei ein Datensatz und \tilde{s}^2 und s^2 seien seine *unkorrigierte empirische Varianz* und *empirische Varianz*, respektive. Dann gelten

(1) (Umrechnung)

$$\tilde{s}^2 = \frac{n-1}{n} s^2 \text{ und } s^2 = \frac{n}{n-1} \tilde{s}^2. \quad (10.38)$$

(2) (Ungleichung)

$$0 \leq \tilde{s}^2 \leq s^2. \quad (10.39)$$

◦

Beweis. (1) Nach Definition 10.10 gilt zum einen

$$\begin{aligned} \tilde{s}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ \Leftrightarrow \frac{n-1}{n-1} \tilde{s}^2 &= \frac{n-1}{n-1} \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ \Leftrightarrow \tilde{s}^2 &= \frac{n-1}{n} \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ \Leftrightarrow \tilde{s}^2 &= \frac{n-1}{n} s^2 \end{aligned} \quad (10.40)$$

und zum anderen

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ \Leftrightarrow \frac{n}{n} s^2 &= \frac{n}{n} \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ \Leftrightarrow s^2 &= \frac{n}{n-1} \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ \Leftrightarrow s^2 &= \frac{n}{n-1} \tilde{s}^2. \end{aligned} \quad (10.41)$$

(2) Mit der Nichtnegativität der Quadratfunktion sowie $\frac{1}{n} > 0$ und $\frac{1}{n-1} > 0$ für $n > 1$ ergeben sich $\tilde{s}^2 \geq 0$ und $\tilde{s}^2 > 0$ direkt. Schließlich gilt für $\sum_{i=1}^n (y_i - \bar{y})^2 \neq 0$

$$\frac{1}{n} < \frac{1}{n-1} \Leftrightarrow \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 < \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \Leftrightarrow \tilde{s}^2 < s^2. \quad (10.42)$$

und für $\sum_{i=1}^n (y_i - \bar{y})^2 = 0$ offenbar $\tilde{s}^2 = s^2$. Insgesamt gilt also

$$0 \leq \tilde{s}^2 \leq s^2. \quad (10.43)$$

□

Folgender **R** Code demonstriert die Bestimmung der unkorrigierten empirischen Varianz und der empirischen Varianz anhand der in Definition 10.10 festgelegten Formeln.

```
D = read.csv("./_data/111-deskriptive-statistik.csv") # Laden des Datensatzes
y = D$PRE # Pre-Interventions-BDI-II-Daten
n = length(y) # Anzahl der Werte
S2_tilde = (1/n)*sum((y - mean(y))^2) # unkorrigierte empirische Varianz
S2 = (1/(n-1))*sum((y - mean(y))^2) # empirische Varianz
```

```
Unkorrigierte empirische Varianz der PRE-BDI-II-Daten : 2.998
Empirische Varianz der PRE-BDI-II-Daten : 3.028
```

Die **R** Funktion `var()` bestimmt per default die empirische Varianz. Mithilfe von Aussage (1) in Theorem 10.4 kann basierend auf dieser Funktion auch die unkorrigierte empirische Varianz bestimmt werden.

```

s2      = var(y)                                # empirische Varianz
s2_tilde = ((n-1)/n)*var(y)                    # unkorrigierte empirische Varianz

```

```

Unkorrigierte empirische Varianz der PRE-BDI-II-Daten: 2.998
Empirische Varianz der PRE-BDI-II-Daten              : 3.028

```

Folgendes Theorem besagt, wie sich die empirische Varianz bei linear-affiner Transformation eines Datensatzes verhält.

Theorem 10.5 (Varianz bei linear-affinen Transformationen). $y = (y_1, \dots, y_n)$ sei ein Datensatz mit empirischer Varianz S_y^2 und $z = (ay_1 + b, \dots, ay_n + b)$ sei der mit $a, b \in \mathbb{R}$ linear-affin transformierte Datensatz mit empirischer Varianz S_z^2 . Dann gilt

$$s_z^2 = a^2 s_y^2. \quad (10.44)$$

◦

Beweis.

$$\begin{aligned}
 s_z^2 &= \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n (ay_i + b - (a\bar{y} + b))^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n (ay_i + b - a\bar{y} - b)^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n (a(y_i - \bar{y}))^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n a^2 (y_i - \bar{y})^2 \\
 &= a^2 \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= a^2 S_y^2.
 \end{aligned} \quad (10.45)$$

□

Beispielsweise ändert also das Umrechnen eines Datensatzes von Meter in Zentimeter die Varianz des Ausgangsdatsatzes um den Faktor 100^2 . Wir reproduzieren Theorem 10.5 beispielhaft mithilfe folgenden **R** Codes.

```

y      = D$PRE                                # Pre-Interventions-BDI-II-Daten
s2y    = var(y)                                # Empirische Varianz von y_1,...,y_n
a      = 2                                    # Multiplikationskonstante
b      = 5                                    # Additionskonstante
z      = a*y + b                              # z_i = az_i + b
s2z    = var(z)                                # Empirische Varianz von z_1,...,z_n
cat("Empirische Varianz von z_1,...,z_n durch direkte Berechnung : ", round(s2z,3)) # Ausgabe

```

```

Empirische Varianz von z_1,...,z_n durch direkte Berechnung : 12.113

```

```

s2z    = a^2*s2y                                # Empirische Varianz von z_1,...,z_n
cat("Empirische Varianz von z_1,...,z_n nach Theorem      : ", round(s2z,3)) # Ausgabe

```

```

Empirische Varianz von z_1,...,z_n nach Theorem          : 12.113

```

Folgendes Theorem zeigt eine Möglichkeit auf, die unkorrigierte empirische Varianz eines Datensatzes allein durch die Bestimmung von Mittelwerten zu bestimmen. Das Theorem findet seine analytische Entsprechung im Verschiebungssatz der Varianz, **Thm-Varianzverschiebungssatz**.

Theorem 10.6 (Verschiebungssatz zur unkorrigierten empirischen Varianz). $y = (y_1, \dots, y_n)$ sei ein Datensatz, $y^2 := (y_1^2, \dots, y_n^2)$ sei sein elementweises Quadrat und \bar{y} und $\bar{y^2}$ seien die respektiven Mittelwerte. Dann gilt

$$\tilde{s}^2 = \bar{y^2} - \bar{y}^2 \quad (10.46)$$

◦

Beweis.

$$\begin{aligned} \tilde{s}^2 &:= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2) \\ &= \frac{1}{n} \sum_{i=1}^n y_i^2 - 2\bar{y} \frac{1}{n} \sum_{i=1}^n y_i + \frac{1}{n} \sum_{i=1}^n \bar{y}^2 \\ &= \bar{y^2} - 2\bar{y}\bar{y} + \frac{1}{n} n\bar{y}^2 \\ &= \bar{y^2} - 2\bar{y}^2 + \bar{y}^2 \\ &= \bar{y^2} - \bar{y}^2. \end{aligned} \quad (10.47)$$

□

Wir reproduzieren **Thm-verschiebungssatz-der-empirischen-varianz** beispielhaft mithilfe folgenden **R** Codes.

```
y          = D$PRE                                     # Pre-Interventions-BDI-II-Daten
s2_tilde   = mean(y^2) - (mean(y))^2                   # \bar{y^2} - \bar{y}^2
cat("Korrigierte empirische Varianz nach Transformation : ", round(s2z,3)) # Ausgabe
```

Korrigierte empirische Varianz nach Transformation : 12.113

Man beachte, dass der Verschiebungssatz der unkorrigierten empirischen Varianz für die empirische Varianz nicht zutrifft, da in diesem Fall der Multiplikationsfaktor $\frac{1}{n-1}$ nicht auf die entsprechenden Mittelwerte führt. Insbesondere haben wir oben schon gesehen, dass für den betreffenden Beispieldatensatz gilt, dass $s^2 = 3.028$.

10.6.3. Empirische Standardabweichung

Die empirische Varianz ergibt sich wie oben gesehen als durchschnittliches Abweichungsquadrat eines Datensatzwerts von seinem Mittelwert. Ist die Einheit des Datensatzes dabei zum Beispiel in einer Reaktionsszeitaufgabe die Millisekunde, so ergibt sich durch Differenzbildung und anschließende Quadrierung die Einheit Millisekunde², eine nicht sehr intuitive Einheit. Um ein Streuungsmaß zu erhalten, dessen Einheit der Einheit des Ausgangsdatsatzes entspricht, wird auf die empirische Varianz häufig noch die Quadratwurzelfunktion angewendet. Dies führt auf den Begriff der *empirischen Standardabweichung*.

Definition 10.11 (Empirische Standardabweichung). $y = (y_1, \dots, y_n)$ sei ein Datensatz. Die *empirische Standardabweichung* von y ist definiert als

$$s := \sqrt{s^2} \quad (10.48)$$

und die *unkorrigierte empirische Stichprobenstandardabweichung* von y ist definiert als

$$\tilde{s} := \sqrt{\tilde{s}^2}. \quad (10.49)$$

•

In Analogie zu den Eigenschaften der empirischen Varianz und ihrer unkorrigierten Version kann auch die unkorrigierte empirischen Standardabweichung leicht in die empirischen Standardabweichung umgerechnet werden und umgekehrt, es gelten hier offenbar

$$\tilde{s} = \sqrt{\frac{n-1}{n}} s \text{ und } s = \sqrt{\frac{n}{n-1}} \tilde{s}. \quad (10.50)$$

Folgender **R** Code demonstriert die Bestimmung der unkorrigierten empirischen Varianz und der empirischen Varianz anhand der in Definition 10.11 festgelegten Formeln.

```
y = D$PRE # Pre-Interventions-BDI-II-Daten
n = length(y) # Anzahl der Werte
s = sqrt((1/(n-1))*sum((y - mean(y))^2)) # Standardabweichung
```

Empirische Standardabweichung : 1.74

Die **R** Funktion `sd()` bestimmt per default die empirische Varianz.

```
s = sd(y)
```

Empirische Standardabweichung : 1.74

Transformiert man einen Datensatz linear-affin, so geht die Multiplikationskonstante der Transformation lediglich im Sinne ihres Betrages in die Transformation der empirischen Standardabweichung ein. Dies ist die Aussage des folgenden Theorems.

Theorem 10.7 (Empirische Standardabweichung bei linear-affinen Transformationen). $y = (y_1, \dots, y_n)$ sei ein Datensatz mit empirischer Standardabweichung s_y und $z = (az_1 + b, \dots, az_n + b)$ sei der mit $a, b \in \mathbb{R}$ linear-affin transformierte Datensatz mit empirischer Standardabweichung s_z . Dann gilt

$$s_z = |a|s_y. \quad (10.51)$$

◦

Beweis.

$$\begin{aligned}
 S_y &:= \left(\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right)^{1/2} \\
 &= \left(\frac{1}{n-1} \sum_{i=1}^n (ay_i + b - (a\bar{x} + b))^2 \right)^{1/2} \\
 &= \left(\frac{1}{n-1} \sum_{i=1}^n (a(y_i - \bar{x}))^2 \right)^{1/2} \\
 &= \left(\frac{1}{n-1} \sum_{i=1}^n a^2 (y_i - \bar{x})^2 \right)^{1/2} \\
 &= (a^2)^{1/2} \left(\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{x})^2 \right)^{1/2}.
 \end{aligned} \tag{10.52}$$

Also gilt $S_z = aS_y$, wenn $a \geq 0$ und $S_z = -aS_y$, wenn $a < 0$. Dies aber entspricht $S_z = |a|S_y$. \square

Wir reproduzieren Theorem 10.7 beispielhaft mithilfe folgenden **R** Codes einmal für den Fall $a \geq 0$ und einmal für den Fall $a < 0$

```
# a >= 0
D = read.csv("../data/111-deskriptive-statistik.csv")
y = D$PRE
a = 2
b = 5
z = a*y + b
sz = sd(z)

# Laden des Datensatzes
# Pre-Interventions-BDI-II-Daten
# Multiplikationskonstante
# Additionskonstante
# y_i = ax_i + b
# Korrigierte empirische Standardabweichung von y
```

Empirische Standardabweichung von z_1,...z_n durch direkte Berechnung : 3.48

```
sz = a*sd(y)

# Korrigierte empirische Standardabweichung von y
```

Empirische Standardabweichung von z_1,...z_n nach Theorem : 3.48

```
# a < 0
D = read.csv("../data/111-deskriptive-statistik.csv")
y = D$PRE
a = -3
b = 10
z = a*y + b
sz = sd(z)

# Laden des Datensatzes
# Pre-Interventions-BDI-II-Daten
# Multiplikationskonstante
# Additionskonstante
# y_i = ax_i + b
# Korrigierte empirische Standardabweichung von y
```

Empirische Standardabweichung nach Transformation : 5.221

```
sz = -a*sd(y)

# Korrigierte empirische Standardabweichung von y
```

Empirische Standardabweichung nach Transformation : 5.221

Referenzen

- Arens, T., Hettlich, F., Karpfinger, C., Kockelkorn, U., Lichtenegger, K., & Stachel, H. (2018). *Mathematik*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-56741-8>
- Burden, R. L., Faires, J. D., & Burden, A. M. (2016). *Numerical Analysis* (Tenth edition). Cengage Learning.
- Caley, A. (1858). A Memoir on the Theory of Matrices. *Philosophical Transactions of the Royal Society of London*, 148, 17–37. <https://doi.org/10.1098/rstl.1858.0002>
- Cantor, G. (1892). Über Eine Eigenschaft Des Inbegriffes Aller Reellen Algebraischen Zahlen. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 1.
- Cantor, G. (1895). Beiträge Zur Begründung Der Transfiniten Mengenlehre. *Mathematische Annalen*, 46(4), 481–512. <https://doi.org/10.1007/BF02124929>
- Friston, K. (2005). A Theory of Cortical Responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Golub, G. H., & Van Loan, C. F. (2013). *Matrix Computations* (Fourth edition). The Johns Hopkins University Press.
- Henze, N. (2018). *Stochastik für Einsteiger*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-22044-0>
- Hilbert, D. (1904). Grundzüge Einer Allgemeinen Theorie Der Linearen Integralgleichungen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*. https://doi.org/10.1007/978-3-322-84410-1_1
- Huber, P. J. (1981). *Robust Statistics*. Wiley.
- Hyndman, R. J., & Fan, Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician*, 50(4), 361. <https://doi.org/10.2307/2684934>
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
- McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, 32(1), 12. <https://doi.org/10.2307/2683468>
- Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica*. Royal Society.
- Richter, T., & Wick, T. (2017). *Einführung in die Numerische Mathematik: Begriffe, Konzepte und zahlreiche Anwendungsbeispiele*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-54178-4>
- Scott, D. W. (1979). *On Optimal and Data-Based Histograms*. 6.
- Searle, S. (1982). *Matrix Algebra Useful for Statistics*. Wiley-Interscience.
- Strang, G. (2009). *Introduction to Linear Algebra*. Cambridge University Press.
- Sturges, H. A. (1926). The Choice of a Class Interval. *Journal of the American Statistical Association*, 21(153), 65–66. <https://doi.org/10.1080/01621459.1926.10502161>
- Unger, L. (2000). *Grundkurs Mathematik*.