

Multivariate Verfahren

Dirk Ostwald

2023-11-23

Inhaltsverzeichnis

	3
I. Mathematische Grundlagen	4
1. Sprache und Logik	5
1.1. Mathematik ist eine Sprache	5
1.2. Grundbausteine mathematischer Kommunikation	6
1.3. Aussagenlogik	7
1.4. Beweistechniken	12
1.5. Selbstkontrollfragen	13
2. Mengen	14
2.1. Grundlegende Definitionen	14
2.2. Verknüpfungen von Mengen	16
2.3. Spezielle Mengen	17
2.4. Selbstkontrollfragen	20
3. Summen, Produkte, Potenzen	21
3.1. Summen	21
3.2. Produkte	23
3.3. Potenzen	24
3.4. Selbstkontrollfragen	25
4. Funktionen	26
4.1. Definition und Eigenschaften	26
4.2. Funktionentypen	28
4.3. Elementare Funktionen	31
4.4. Selbstkontrollfragen	35
5. Differentialrechnung	36
5.1. Definitionen und Rechenregeln	36
5.2. Analytische Optimierung	38
5.3. Differentialrechnung multivariater reellwertiger Funktionen	42
5.4. Selbstkontrollfragen	49
6. Folgen, Grenzwerte, Stetigkeit	50
6.1. Folgen	50
6.2. Grenzwerte	52
6.3. Stetigkeit	54
6.4. Selbstkontrollfragen	56

7. Integralrechnung	57
7.1. Unbestimmte Integrale	57
7.2. Bestimmte Integrale	59
7.3. Uneigentliche Integrale	66
7.4. Mehrdimensionale Integrale	67
7.5. Selbstkontrollfragen	68
8. Vektoren	70
8.1. Reeller Vektorraum	70
8.2. Euklidischer Vektorraum	74
8.3. Lineare Unabhängigkeit	82
8.4. Vektorraumbasen	84
8.5. Vektorkoordinatentransformation	87
8.6. Selbstkontrollfragen	90
9. Matrizen	92
9.1. Definition	92
9.2. Grundlegende Matrixoperationen	94
9.3. Matrixmultiplikation	100
9.4. Matrixinversion	104
9.5. Determinanten	108
9.6. Spezielle Matrizen	112
9.7. Literaturhinweise	115
9.8. Selbstkontrollfragen	115
10. Eigenanalyse	117
10.1. Eigenvektoren und Eigenwerte	117
10.2. Orthonormalzerlegung	123
10.3. Singulärwertzerlegung	125
10.4. Literaturhinweise	126
10.5. Selbstkontrollfragen	127
11. Zufallsvektoren	128
11.1. Definition und multivariate Verteilungen	128
11.2. Erwartungswert, Kovarianzmatrix und Korrelationsmatrix	132
11.3. Marginalverteilungen	135
11.4. Bedingte Verteilungen	137
11.5. Unabhängige Zufallsvariablen	139
11.6. Normalverteilungen	142
11.7. Literaturhinweise	152
11.8. Selbstkontrollfragen	152
II. Frequentistische Inferenz	154
12. Datendeskription	155
12.1. Datenanalyseszenarien	155
12.2. Deskriptivstatistiken	156
12.3. Literaturhinweise	161
12.4. Selbstkontrollfragen	161

13. Einstichproben-T^2-Tests	163
13.1. Anwendungsszenario	163
13.2. Modellformulierung und Modellevaluation	164
13.3. Modellevaluation	165
13.4. Anwendungsbeispiel	178
13.5. Literaturhinweise	180
13.6. Selbstkontrollfragen	180
14. Einfaktorielle Varianzanalyse	181
14.1. Anwendungsszenario	181
14.2. Modellformulierung und Modellschätzung	183
14.3. Modellevaluation	186
14.4. Anwendungsbeispiel	194
14.5. Literaturhinweise	195
14.6. Selbstkontrollfragen	195
15. Kanonische Korrelationsanalyse	197
15.1. Anwendungsszenario und Grundzüge	197
15.2. Mathematischer Hintergrund	199
15.3. Modellformulierung	202
15.4. Modellschätzung	206
15.5. Anwendungsbeispiel	209
15.6. Literaturhinweise	210
15.7. Selbstkontrollfragen	210
Referenzen	211

Teil 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), Friston et al. (2023), Ostwald et al. (2014), Blei et al. (2017)). 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 Begriffsbildungen zu arbeiten. Sie geht dabei insbesondere streng hierarchisch vor, so dass an späterer Stelle eingeführte Begrifflichkeiten oft ein gutes Verständnis der ihnen zugrundeliegenden und an früherer Stelle eingeführten Begrifflichkeiten voraussetzen.

Die Genauigkeit der mathematischen Sprache impliziert dabei 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 an 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 nachzubauen.

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 ständige Exposition mit mathematischen Ausdrucksweisen, sicherlich auch zunächst einmal das Auswendiglernen von Begriffen und das aktive Lesen 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 mathematischer Kommunikation

In diesem Abschnitt 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 auch oft als *Sätze* bezeichnet. In der angewandten, datenanalytischen Mathematik sind Theoreme oft 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 zum Beispiel 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 entsprechende mathematische Gebäude abzukommen. Beweise enden in diesem Text immer mit dem Symbol \square .

Neben den oben vorgestellten Begriffen gibt es mit *Axiomen*, *Lemmata*, *Korollaren* und *Vermutungen* noch weitere typische Grundbausteine mathematischer Texte. Wir werden diesen Begriff 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 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 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 einfachem System nähern, das es erlaubt, Beziehungen zwischen mathematischen Aussagen herzustellen und zu formalisieren. Im Folgenden spielt die Aussagenlogik zum Beispiel in der Definition von Mengenoperationen, bei Optimierungsbedingungen von Funktionen und in vielen Beweisen eine tragende Rolle. In der mathematischen Anwendung ist Aussagenlogik die Grundlage der Booleschen Logik der Programmierung. In der mathematischen Psychologie ist die Aussagenlogik zum Beispiel 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.

A	$\neg A$
w	f
f	w

Tabellen dieser Form nennt man *Wahrheitstafeln*. Sie sind ein beliebtes Hilfsmittel in der Aussagenlogik. 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.

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 nun 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

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 ”.

•

Man mag \Rightarrow auch als “daraus folgt” lesen. Die Definition der Implikation impliziert folgende Wahrheitstafel.

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

Ein 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. Betrachtet man obige Wahrheitstafel unter diesem Gesichtspunkt, 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 (dass also gilt, wenn A wahr ist, dass

$A \Rightarrow B$ 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. Aus einer wahren Voraussetzung folgt also nur bei wahrer Implikation eine wahre Konklusion. Insbesondere genügt die Definition der Implikation damit der Forderung "Aus Falschem folgt beliebiges (ex falso sequitur quodlibet)". Aus falschen Aussagen kann man also mithilfe der Implikation nichts richtiges folgern.

Im Kontext der Implikation ergeben sich die Begriffe der *hinreichenden* und der *notwendigen Aussagen (Bedingungen)*. Diese sind definiert wie folgt: wenn $A \Rightarrow B$ wahr ist, sagt man " A ist *hinreichend* für B " und " B ist *notwendig* für A ". Diese Sprachregelung erklärt sich 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

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 folgendes Theorem.

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.

□

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 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*.

1.4. Beweistechniken

Im letzten Abschnitt wollen wir mit den Begriffen der *direkten* und *indirekten Beweise* sowie des *Beweises durch Widerspruch* kurz drei Beweistechniken skizzieren, von denen vor allem die erste in diesem Text immer wieder zur Begründung von Theoremen herangezogen wird. Dabei haben typische Theoreme die Form $A \Rightarrow B$ für Aussagen A und B .

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.

Um diese Techniken an einem Beispiel zu erläutern, erinnern wir kurz an folgende *Äquivalenzumformungen von Gleichungen*:

- Addition oder Subtraktion einer Zahl auf beiden Seiten der Gleichung, zum Beispiel

$$2x + 4 = 10 \Leftrightarrow 2x = 6, \quad (1.3)$$

- Multiplikation mit einer oder Division durch eine von Null verschiedene Zahl auf beiden Seiten der Gleichung, zum Beispiel

$$2x = 6 \Leftrightarrow x = 3, \quad (1.4)$$

- Anwendung einer injektiven Funktion auf beiden Seiten der Gleichung, zum Beispiel

$$\exp(x) = 2 \Leftrightarrow x = \ln(2), \quad (1.5)$$

sowie an folgende elementaren *Äquivalenzumformungen von Ungleichungen*:

- Addition oder Subtraktion einer Zahl auf beiden Seiten der Ungleichung, zum Beispiel

$$-2x + 4 \geq 10 \Leftrightarrow -2x \geq 6, \quad (1.6)$$

- Multiplikation mit einer Zahl oder Division durch eine von Null verschiedene Zahl auf beiden Seiten der Ungleichung, wobei die Multiplikation oder Division mit einer negativen Zahl die Umkehrung der Ungleichung impliziert, zum Beispiel

$$-2x \geq 6 \Leftrightarrow x \leq -3, \quad (1.7)$$

- Anwendung monotoner Funktionen auf beiden Seiten der Ungleichung

$$\exp(x) \geq 2 \Leftrightarrow x \geq \ln(2). \quad (1.8)$$

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

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

□

1.5. 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 den Begriff des direkten Beweises.
14. Erläutern Sie den Begriff des indirekten Beweises.
15. Erläutern Sie den Begriff des Beweises durch Widerspruch.
16. Beweisen Sie mit einem direkten Beweis, dass gilt

$$x^2 + px + q = 0 \Leftrightarrow x = -\frac{p}{2} - \sqrt{\left(\frac{p}{2}\right)^2 - q} \vee x = -\frac{p}{2} + \sqrt{\left(\frac{p}{2}\right)^2 - q}. \quad (1.14)$$

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 unsere 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, das 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). A und B seien zwei Mengen.

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

$$A \subseteq B \quad (2.2)$$

und nennt A *Untermenge* von B und B *Obermenge* von A .

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

$$A \subset B. \quad (2.3)$$

- Zwei Mengen A und B heißen *gleich*, wenn für jedes Element $a \in A$ gilt, dass auch $a \in B$, und wenn für jedes Element $b \in B$ gilt, dass auch $b \in A$. Sind die Mengen A und B gleich, so schreibt man

$$A = B. \quad (2.4)$$

•

Betrachten wir zum Beispiel die Mengen $A := \{1\}$, $B := \{1, 2\}$, und $C := \{1, 2\}$. Dann gilt mit obigen Definitionen, dass $A \subset B$, weil $1 \in A$ und $1 \in B$, aber $2 \in B$ und $2 \notin A$. Weiterhin gilt, dass $B \subseteq C$, weil $1 \in B$ und $1 \in C$ sowie $2 \in B$ und $2 \in C$ und es kein Element von C gibt, welches nicht in B ist. Ebenso gilt $C \subseteq B$, weil $1 \in C$ und $1 \in B$ sowie $2 \in C$ und $2 \in B$ und es kein Element von B gibt, welches nicht in C ist. Schließlich gilt sogar $B = C$, weil für jedes Element $b \in B$ gilt, dass auch $b \in C$, und gleichzeitig für jedes Element $c \in C$ gilt, dass auch $c \in B$.

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 $A := \{1, 2, 3\}$, $B = \{a, b, c, d\}$ und $C := \{\}$. Dann gelten $|A| = 3$, $|B| = 4$ und $|C| = 0$.

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 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 ein-, zwei-, als auch dreielementige Teilmengen und es gilt

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

Hinsichtlich der Kardinalitäten einer Menge und ihrer Potenzmenge kann man beweisen, dass aus $|M| = n$ mit $n > 0$ folgt, dass die Kardinalität der Potenzmenge $|\mathcal{P}(M)| = 2^n$ ist. In den obigen Beispielen haben wir die Fälle $|M_1| = 1$ und somit $|\mathcal{P}(M_1)| = 2^1 = 2$, $|M_2| = 2$ und somit $|\mathcal{P}(M_2)| = 2^2 = 4$ und schließlich $|M_3| = 3$ und somit $|\mathcal{P}(M_3)| = 2^3 = 8$, wovon man sich durch Nachzählen schnell überzeugt.

2.2. Verknüpfungen von Mengen

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 wie immer im inklusiven Sinne 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 *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.

•

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 $5 \notin \{2, 3\}$.

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

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

$$M = \cup_{i=1}^n N_i \wedge N_i \cap N_j = \emptyset \text{ für } i = 1, \dots, n, j = 1, \dots, n, i \neq j. \quad (2.13)$$

Dann heißt 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 Möglichkeiten eine gegebene Menge zu partitionieren. Betrachten wir zum Beispiel 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 mögliche Partitionen von M .

2.3. Spezielle Mengen

In der Naturwissenschaft versucht man, Phänomene der Welt mit Zahlen zu beschreiben. Je nach Phänomen bieten sich dazu *diskrete* oder *kontinuierliche* Zahlenmengen an. Die Mathematik stellt dazu unter anderem die in folgender Definition gegebenen Zahlenmengen 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, wie oben definiert, durch Brüche ganzer Zahlen ausdrücken lassen. Dies sind alle ganzen Zahlen sowie die negativen und positiven Dezimalzahlen wie z.B. $-\frac{9}{10} = -0.9$, $\frac{1}{3} = 1.3\bar{3}$, 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 kann man beweisen (Cantor (1892)), 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

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

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 kann. Sie gehören auch nicht zu den in obiger Definition gegebenen Zahlenmengen, es gilt also sowohl $\infty \notin \mathbb{R}$ als auch $-\infty \notin \mathbb{R}$.

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. Wir vertiefen die Theorie komplexer Zahlen an dieser Stelle nicht.

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

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*

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

•

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 $\mathbb{R} =] - \infty, \infty[$.

Oft möchte man mehrere 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 auch *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

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

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

$$M^2 := M \times M. \tag{2.19}$$

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

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

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

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

•

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

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* 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

$$\mathbb{R}^n := \prod_{i=1}^n \mathbb{R} := \{x := (x_1, \dots, x_n) | x_i \in \mathbb{R}\} \tag{2.22}$$

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

$$x := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \tag{2.23}$$

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

•

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

$$x = \begin{pmatrix} 0.16 \\ 1.76 \\ 0.23 \\ 7.11 \end{pmatrix}. \quad (2.24)$$

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

3.1. Summen

Diese Einheit führt einige Schreibweisen für die Grundrechenarten ein.

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 die sinnvolle Benutzung des Summenzeichens ist es essentiell, dass mit mithilfe des Subskripts und des Superskripts Anfang und Ende der Summation festgelegt werden. Die genaue Bezeichnung des Laufindexes 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 natürlicher Zahlen*. Es gilt

$$\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5 = 15. \quad (3.5)$$

- *Summation gerader natürlicher Zahlen.* Es gilt

$$\sum_{i=1}^5 2i = 2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 + 2 \cdot 4 + 2 \cdot 5 = 2 + 4 + 6 + 8 + 10 = 30. \quad (3.6)$$

- *Summation ungerader natürlicher Zahlen.* Es gilt

$$\sum_{i=1}^5 (2i-1) = 2 \cdot 1 - 1 + 2 \cdot 2 - 1 + 2 \cdot 3 - 1 + 2 \cdot 4 - 1 + 2 \cdot 5 - 1 = 1 + 3 + 5 + 7 + 9 = 25. \quad (3.7)$$

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

Theorem 3.1 (Rechenregeln für Summen).

- (1) *Summen gleicher Summanden*

$$\sum_{i=1}^n x = nx \quad (3.8)$$

- (2) *Assoziativität bei Summen gleicher Länge*

$$\sum_{i=1}^n x_i + \sum_{i=1}^n y_i = \sum_{i=1}^n (x_i + y_i) \quad (3.9)$$

- (3) *Distributivität bei Multiplikation mit einer Konstante*

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad (3.10)$$

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

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

- (5) *Umindizierung*

$$\sum_{i=0}^n x_i = \sum_{j=m}^{n+m} x_{j-m} \quad (3.12)$$

◦

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

$$\begin{aligned} \sum_{i=1}^n x_i + \sum_{i=1}^n y_i &= x_1 + x_2 + \cdots + x_n + y_1 + y_2 + \cdots + y_n \\ &= x_1 + y_1 + x_2 + y_2 + \cdots + x_n + y_n \\ &= \sum_{i=1}^n (x_i + y_i). \end{aligned} \quad (3.13)$$

Hinsichtlich letzterer gilt

$$\begin{aligned}\sum_{i=1}^n ax_i &= ax_1 + ax_2 + \dots + ax_n \\ &= a(x_1 + x_2 + \dots + x_n) \\ &= a \sum_{i=1}^n x_i.\end{aligned}\tag{3.14}$$

□

Als Beispiel für die Anwendung einer Rechenregel 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 obiger Rechenregel (3) 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 von Rechenregel (3) mit $a = 1/n$, dass

$$\frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n \frac{x_i}{n}.\tag{3.15}$$

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

$$\frac{1}{4} \sum_{i=1}^4 x_i = \frac{1}{4}(1 + 4 + 2 + 1) = \frac{8}{4} = 2 = \frac{8}{4} = \frac{1}{4} + \frac{4}{4} + \frac{2}{4} + \frac{1}{4} = \sum_{i=1}^4 \frac{x_i}{4}.\tag{3.16}$$

3.2. Produkte

Eine analoge Schreibweise zum Summenzeichen bietet das Produktzeichen für Produkte.

Definition 3.2 (Produktzeichen). Es bezeichnet

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n.\tag{3.17}$$

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

$$\prod_{i=1}^n x_i = \prod_{j=1}^n x_j.\tag{3.18}$$

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

$$\prod_{j \in J} x_j := x_0 \cdot x_1 \cdot x_2. \quad (3.19)$$

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

$$a^0 := 1 \text{ und } a^{n+1} := a^n \cdot a. \quad (3.20)$$

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

$$a^{-n} := (a^n)^{-1} := \frac{1}{a^n}. \quad (3.21)$$

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 3.2 (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:

$$a^n a^m = a^{n+m} \quad (3.22)$$

$$(a^n)^m = a^{nm} \quad (3.23)$$

$$(ab)^n = a^n b^n \quad (3.24)$$

◦

Wir verzichten auf einen Beweis. Beispielsweise gelten also

$$2^2 \cdot 2^3 = (2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) = 2^5 = 2^{2+3}, \quad (3.25)$$

$$(3^2)^3 = (3 \cdot 3)^3 = (3 \cdot 3) \cdot (3 \cdot 3) \cdot (3 \cdot 3) = 3^6 = 3^{2 \cdot 3}, \quad (3.26)$$

und

$$(2 \cdot 4)^2 = (2 \cdot 4) \cdot (2 \cdot 4) = (2 \cdot 2) \cdot (4 \cdot 4) = 2^2 \cdot 4^2. \quad (3.27)$$

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 Zahl r , so dass

$$r^n = a. \quad (3.28)$$

•

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

Theorem 3.3 (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*

$$r = a^{\frac{1}{n}} \quad (3.29)$$

◦

Beweis. Es gilt

$$\left(a^{\frac{1}{n}}\right)^n = a^{\frac{1}{n}} \cdot a^{\frac{1}{n}} \cdot \dots \cdot a^{\frac{1}{n}} = a^{\sum_{i=1}^n \frac{1}{n}} = a^1 = a. \quad (3.30)$$

Also gilt mit der Definition der *n*-ten Wurzel, dass $r = a^{\frac{1}{n}}$.

□

Das Rechnen mit Quadratwurzeln wird durch die Potenzschreibweise $\sqrt{x} = x^{\frac{1}{2}}$ sehr erleichtert. Zum Beispiel gilt

$$\frac{2\pi}{\sqrt{2\pi}} = \frac{2\pi}{(2\pi)^{\frac{1}{2}}} = (2\pi)^1 \cdot (2\pi)^{-\frac{1}{2}} = (2\pi)^{1-\frac{1}{2}} = (2\pi)^{\frac{1}{2}} = \sqrt{2\pi}. \quad (3.31)$$

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 bilden zusammen mit den Mengen 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. Funktionen werden äquivalent auch als Abbildungen bezeichnet.

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 Zielmenge Z zuordnet. D wird dabei *Definitionsmenge* von f und Z wird *Zielmenge* 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.

Funktionen sind immer *eindeutig*, in dem Sinne dass sie jedem $x \in D$ bei jeder Anwendung der Funktion immer dasselbe $f(x) \in Z$ zuordnen. Funktionen setzen dabei Elemente von Mengen miteinander in Beziehung. Die Mengen dieser Elemente erhalten spezielle Bezeichnungen.

Definition 4.2 (Bildmenge und Urbildmenge). 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

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

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

•

Man beachte, dass der Wertebereich $f(D)$ von f und die Zielmenge Z von f sind nicht notwendigerweise identisch sein müssen. Grundlegende Eigenschaften von Funktionen werden in folgender Definition festgelegt.

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.

•

Abbildung 4.1 verdeutlicht diese Definitionen anhand dreier (Gegen)beispiele.

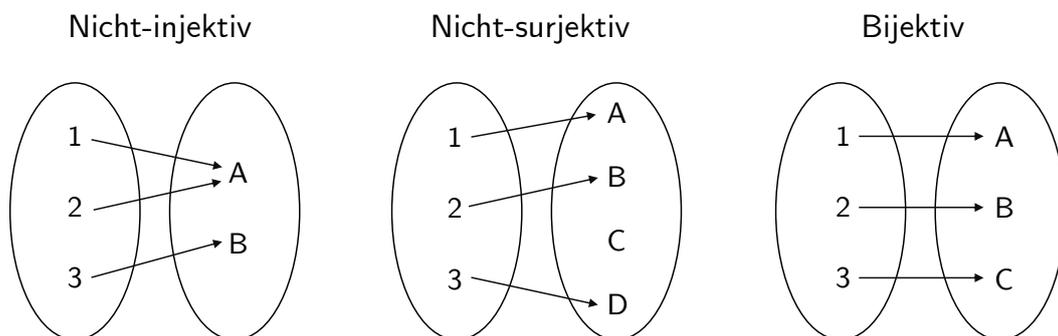


Abbildung 4.1. Injektivität, Surjektivität, Bijektivität.

Abbildung 4.1 A visualisiert die *nicht-injektive* Funktion

$$f : \{1, 2, 3\} \rightarrow \{A, B\}, x \mapsto f(x) := \begin{cases} f(1) & := A \\ f(2) & := A \\ f(3) & := B \end{cases} \tag{4.5}$$

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 1 und 2.

Abbildung 4.1 B visualisiert die nicht-surjektive Funktion

$$g : \{1, 2, 3\} \rightarrow \{A, B, C, D\}, x \mapsto g(x) := \begin{cases} g(1) & := A \\ g(2) & := B \\ g(3) & := D \end{cases} \tag{4.6}$$

Die Funktion ist nicht surjektiv, weil das Element D in der Zielmenge von f kein Urbild in der Definitionsmenge von f hat. Abbildung 4.1 C schließlich visualisiert die bijektive Funktion

$$h : \{1, 2, 3\} \rightarrow \{A, B, C\}, x \mapsto h(x) := \begin{cases} h(1) & := A \\ h(2) & := B \\ h(3) & := C \end{cases} \quad (4.7)$$

Zu *jedem* Element in der Zielmenge von h gibt es *genau ein* Urbild, die Funktion ist also injektiv und surjektiv und damit bijektiv.

Als weiteres Beispiel betrachten wir die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2 \quad (4.8)$$

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

$$\tilde{f} : [0, \infty[\rightarrow [0, \infty[, x \mapsto \tilde{f}(x) := x^2, \quad (4.9)$$

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

$$g \circ f : D \rightarrow S, x \mapsto (g \circ f)(x) := g(f(x)) \quad (4.10)$$

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.

Als Beispiel für die Verkettung zweier Funktionen betrachten wir

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := -x^2 \quad (4.11)$$

und

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto g(x) := \exp(x). \tag{4.12}$$

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

$$g \circ f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto (g \circ f)(x) := g(f(x)) = \exp(-x^2). \tag{4.13}$$

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

$$f^{-1} \circ f : D \rightarrow D, x \mapsto (f^{-1} \circ f)(x) := f^{-1}(f(x)) = x \tag{4.14}$$

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. Betrachtet man 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. Betrachten wir zum Beispiel die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := 2x =: y. \tag{4.15}$$

Dann ist die inverse Funktion von f gegeben durch

$$f^{-1} : \mathbb{R} \rightarrow \mathbb{R}, y \mapsto f^{-1}(y) := \frac{1}{2}y, \tag{4.16}$$

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

$$(f^{-1} \circ f)(x) := f^{-1}(f(x)) = f^{-1}(2x) = \frac{1}{2} \cdot 2x = x. \tag{4.17}$$

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

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

und

$$f(cx) = cf(x) \tag{Homogenität}$$

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*

$$f(0) = 0. \tag{4.18}$$

◦

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

$$f(0) = f(0 + 0) = f(0) + f(0). \tag{4.19}$$

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

$$\begin{aligned} f(0) - f(0) &= f(0) + f(0) - f(0) \\ 0 &= f(0) \end{aligned} \tag{4.20}$$

und damit ist alles gezeigt.

□

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

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

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax \tag{4.21}$$

eine lineare Abbildung, weil gilt, dass

$$f(x+y) = a(x+y) = ax+ay = f(x)+f(y) \text{ und } f(cx) = acx = cax = cf(x). \tag{4.22}$$

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

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax + b \tag{4.23}$$

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

$$f(x+y) = 1(x+y) + 1 = x+y+1 \neq x+1+y+1 = f(x)+f(y). \tag{4.24}$$

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

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x), \quad (4.25)$$

- *multivariate reellwertige Funktionen* der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n), \quad (4.26)$$

- und *multivariate vektorwertige Funktionen* der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto f(x) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}, \quad (4.27)$$

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 wesentliche Eigenschaften dieser Funktionen und 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

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \sum_{i=0}^k a_i x^i = a_0 + a_1 x^1 + a_2 x^2 + \dots + a_k x^k \quad (4.28)$$

heißt *Polynomfunktion* k -ten Grades mit *Koeffizienten* $a_0, a_1, \dots, a_k \in \mathbb{R}$.

•

Einige ausgewählte Polynomfunktionen sind in Tabelle 4.1 aufgelistet, Abbildung 4.2 zeigt die entsprechende Graphen.

Tabelle 4.1. Ausgewählte Polynomfunktionen

Name	Funktionale 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	$f(x) = x^2$	$a_0 := 0, a_1 := 0, a_2 := 1, a_i := 0, i > 2$

Ein wichtiges Funktionenpaar sind die Exponentialfunktion und die Logarithmusfunktion. Die Graphen der Exponential- und Logarithmusfunktion sind in Abbildung 4.3 abgebildet.

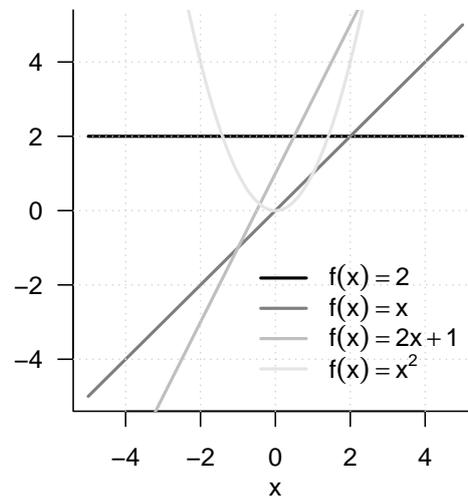


Abbildung 4.2. Ausgewählte Polynomfunktionen

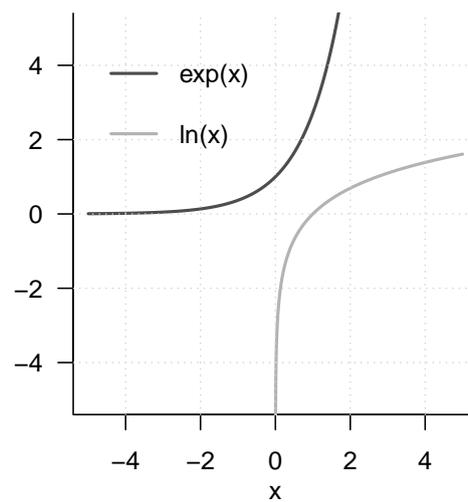


Abbildung 4.3. Exponentialfunktion und Logarithmusfunktion

Definition 4.9 (Exponentialfunktion). Die *Exponentialfunktion* ist definiert als

$$\exp : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \exp(x) := e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \quad (4.29)$$

•

Die Exponentialfunktion hat unter anderem folgende Eigenschaften.

Wertebereich der Exponentialfunktion

- $x \in]-\infty, 0[\Rightarrow \exp(x) \in]0, 1[$
- $x \in]0, \infty[\Rightarrow \exp(x) \in]1, \infty[$

Insbesondere nimmt die Exponentialfunktion also nur positive Werte an.

Monotonieeigenschaft der Exponentialfunktion

- $x < y \Rightarrow \exp(x) < \exp(y)$

Spezielle Werte der Exponentialfunktion

- $\exp(0) = 1$
- $\exp(1) = e \approx 2.71$

Die Logarithmusfunktion schneidet die y -Achse also bei 0. Die Zahl e heißt *Eulersche Zahl*.

Summationseigenschaft und Subtraktionseigenschaft der Exponentialfunktion

- $\exp(x + y) = \exp(x) \exp(y)$
- $\exp(x - y) = \frac{\exp(x)}{\exp(y)}$

Mit den speziellen Werten der Exponentialfunktion gilt dann insbesondere auch

$$\exp(x) \exp(-x) = \exp(x - x) = \exp(0) = 1. \quad (4.30)$$

Definition 4.10 (Logarithmusfunktion). Die *Logarithmusfunktion* ist definiert als inverse Funktion der Exponentialfunktion,

$$\ln :]0, \infty[\rightarrow \mathbb{R}, x \mapsto \ln(x) \text{ mit } \ln(\exp(x)) = x \text{ für alle } x \in \mathbb{R}. \quad (4.31)$$

•

Die Logarithmusfunktion hat unter anderem folgende Eigenschaften.

Wertebereich der Logarithmusfunktion

- $x \in]0, 1[\Rightarrow \ln(x) \in]-\infty, 0[$
- $x \in]1, \infty[\Rightarrow \ln(x) \in]0, \infty[$

Die Logarithmusfunktion nimmt also sowohl negative als auch positive Werte an.

Monotonie der Logarithmusfunktion

- $x < y \Rightarrow \ln(x) < \ln(y)$

Spezielle Werte der Logarithmusfunktion

- $\ln(1) = 0$ und $\ln(e) = 1$.

Die Logarithmusfunktion schneidet die x -Achse also bei 1.

Produkteigenschaft, Potenz eigenschaft und Divisionseigenschaft der Logarithmusfunktion

- $\ln(xy) = \ln(x) + \ln(y)$
- $\ln(x^c) = c \ln(x)$
- $\ln\left(\frac{1}{x}\right) = -\ln(x)$

Letztere Eigenschaft sind beim Rechnen Logarithmusfunktionen zentral. Man merkt sie sich intuitiv als “Die Logarithmusfunktion wandelt Produkte in Summen und Potenzen in Produkte um.”

Ein häufiger Begleiter in der Wahrscheinlichkeitstheorie ist die *Gammafunktion*. Ein Ausschnitt des Graphen der Gammafunktion ist in Abbildung 4.4 dargestellt.

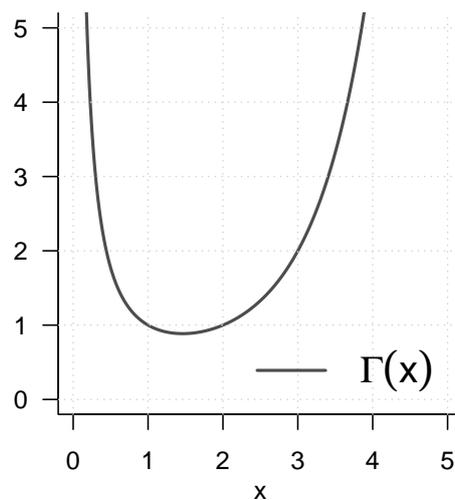


Abbildung 4.4. Gammafunktion

Definition 4.11 (Gammafunktion). Die *Gammafunktion* ist definiert durch

$$\Gamma : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \Gamma(x) := \int_0^{\infty} \xi^{x-1} \exp(-\xi) d\xi \quad (4.32)$$

•

Die Gammafunktion hat folgende Eigenschaften:

Spezielle Werte der Gammafunktion

- $\Gamma(1) = 1$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$.

Rekursionseigenschaft der Gammafunktion

- Für $x > 0$ gilt $\Gamma(x+1) = x\Gamma(x)$

4.4. Selbstkontrollfragen

1. Geben Sie die Definition einer Funktion wieder.
2. Geben Sie die Definition der Begriffe Bildmenge, Wertebereich, und Urbildmenge wieder.
3. Geben Sie die Definitionen der Begriffe Surjektivität, Injektivität, und Bijektivität wieder.
4. Erläutern Sie, warum $f: \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2$ weder injektiv noch surjektiv ist.
5. Erläutern Sie, warum $f: [0, \infty[\rightarrow [0, \infty[, x \mapsto f(x) := x^2$ bijektiv ist.
6. Geben Sie die Definition der Verkettung von Funktionen wieder.
7. Geben Sie die Definition des Begriffs der inversen Funktion wieder.
8. Geben Sie die inverse Funktion von x^2 auf $[0, \infty[$ an.
9. Geben Sie die Definition des Begriffs der linearen Abbildung wieder.
10. Geben Sie die Definitionen der Begriffe der univariat-reellwertigen, multivariat-reellwertigen und multivariat-vektorwertigen Funktion wieder.
11. Skizzieren Sie die Identitätsfunktion und die konstante Funktion für $a := 1$.
12. Skizzieren Sie die linear-affine Funktion $f(x) = ax + b$ für $a = 2$ und $b = 3$.
13. Skizzieren Sie die Funktionen $f(x) := (x-1)^2$ und $g(x) := (x+3)^2$.
14. Skizzieren Sie die Exponential- und Logarithmusfunktionen.
15. Geben Sie die Summations- und Subtraktionseigenschaften der Exponentialfunktion an.
16. Geben Sie die Produkt-, Potenz- und Divisionseigenschaften der Logarithmusfunktion an.

5. 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 5.1 führen wir zunächst den Begriff der Ableitung und mit ihm verbundene elementare Rechenregeln ein. In Kapitel 5.2 widmen wir uns dann der Frage, wie man mithilfe von Ableitungen Extremstellen von Funktionen bestimmen kann.

5.1. Definitionen und Rechenregeln

Wir beginnen mit folgender Definition.

Definition 5.1 (Differenzierbarkeit und Ableitung). Es sei $I \subseteq \mathbb{R}$ ein Intervall und

$$f : I \rightarrow \mathbb{R}, x \mapsto f(x) \quad (5.1)$$

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

$$f'(a) := \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \quad (5.2)$$

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

$$f' : I \rightarrow \mathbb{R}, x \mapsto f'(x) \quad (5.3)$$

heißt *Ableitung von f* .

•

Für $h > 0$ heißt der Ausdruck

$$\frac{f(a+h) - f(a)}{h} \quad (5.4)$$

Newtonscher Differenzquotient. Der Newtonsche Differenzquotient misst die Änderung $f(a+h) - f(a)$ von f 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 die instantane Änderungsrate von f in a , also im Beispiel die Geschwindigkeit des Objektes zu einem Zeitpunkt a .

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 5.2 (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.

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

Definition 5.3 (Höhere Ableitungen). Es sei f eine univariate reellwertige Funktion und

$$f^{(1)} := f' \tag{5.5}$$

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

$$f^{(k)} := (f^{(k-1)})' \text{ für } k \geq 0, \tag{5.6}$$

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

$$f'' := (f')' \tag{5.7}$$

•

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 $f^{(3)}$.

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

(1) *Summenregel*

$$\text{Für } f(x) := \sum_{i=1}^n g_i(x) \text{ gilt } f'(x) = \sum_{i=1}^n g_i'(x). \quad (5.8)$$

(2) *Produktregel*

$$\text{Für } f(x) := g_1(x)g_2(x) \text{ gilt } f'(x) = g_1'(x)g_2(x) + g_1(x)g_2'(x). \quad (5.9)$$

(3) *Quotientenregel*

$$\text{Für } f(x) := \frac{g_1(x)}{g_2(x)} \text{ gilt } f'(x) = \frac{g_1'(x)g_2(x) - g_1(x)g_2'(x)}{g_2^2(x)}. \quad (5.10)$$

(4) *Kettenregel*

$$\text{Für } f(x) := g_1(g_2(x)) \text{ gilt } f'(x) = g_1'(g_2(x))g_2'(x). \quad (5.11)$$

◦

Erste Beispiele für die Anwendung obiger Rechenregeln lernen wir im Abschnitt Kapitel 5.2 kennen. 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 5.1. 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 5.1 visualisieren wir die Identitätsfunktion, eine lineare Funktion und die Quadratfunktion zusammen mit ihrer jeweiligen Ableitung. In Abbildung 5.2 visualisieren wir die Exponential- und Logarithmusfunktionen zusammen mit ihrer jeweiligen Ableitung.

5.2. 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

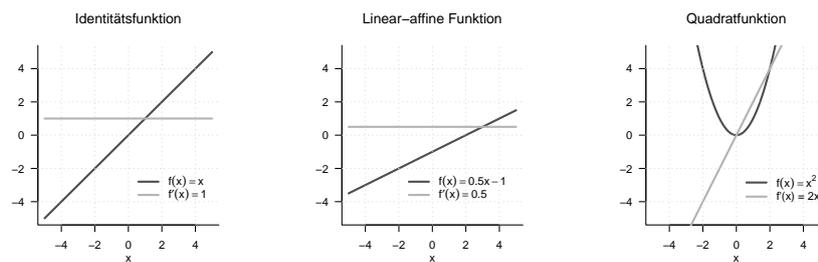


Abbildung 5.1. Ableitungen dreier elementarer Funktionen

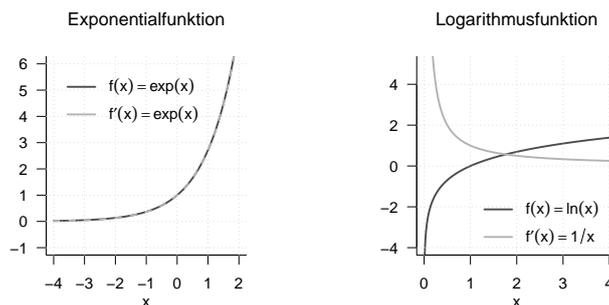


Abbildung 5.2. Ableitungen von Exponentialfunktion und Logarithmusfunktion

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 5.4 (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

$$f(x_0) \leq f(x) \text{ für alle } x \in I \cap U, \quad (5.12)$$

- ein *globales Minimum*, wenn gilt, dass

$$f(x_0) \leq f(x) \text{ für alle } x \in U, \quad (5.13)$$

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

$$f(x_0) \geq f(x) \text{ für alle } x \in I \cap U, \quad (5.14)$$

- ein *lokales Maximum*, wenn gilt, dass

$$f(x_0) \geq f(x) \text{ für alle } x \in U. \quad (5.15)$$

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

$$\operatorname{argmin}_{x \in I \cap U} f(x) \text{ oder } \operatorname{argmax}_{x \in I \cap U} f(x) \quad (5.16)$$

bezeichnet und Extremwerte von Funktionen werden häufig mit

$$\min_{x \in I \cap U} f(x) \text{ oder } \max_{x \in I \cap U} f(x) \quad (5.17)$$

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

$$x_0 \text{ ist Extremstelle von } f \Rightarrow f'(x_0) = 0. \quad (5.18)$$

◦

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

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

$$f'(x_0) = 0 \text{ und } f''(x_0) > 0 \quad (5.19)$$

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

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

$$f'(x_0) = 0 \text{ und } f''(x_0) < 0 \quad (5.20)$$

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

◦

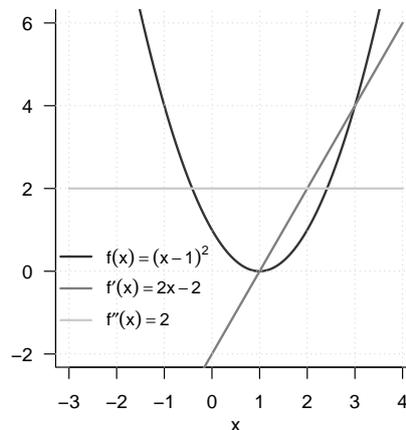


Abbildung 5.3. Analytische Optimierung von $f(x) := (x - 1)^2$

Wir verzichten wiederum auf einen Beweis und verdeutlichen uns die Bedingung an dem in Abbildung 5.3 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 5.4 (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 .

◦

Anstelle eines Beweises betrachten wir beispielhaft die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := (x - 1)^2. \quad (5.21)$$

aus Abbildung 5.3. Die erste Ableitung von f ergibt sich mit der Kettenregel zu

$$f'(x) = \frac{d}{dx} ((x - 1)^2) = 2(x - 1) \cdot \frac{d}{dx} (x - 1) = 2x - 2. \quad (5.22)$$

Die zweite Ableitung von f ergibt sich zu

$$f''(x) = \frac{d}{dx} f'(x) = \frac{d}{dx} (2x - 2) = 2 > 0 \text{ für alle } x \in \mathbb{R}. \quad (5.23)$$

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

$$f'(x^*) = 0 \Leftrightarrow 2x^* - 2 = 0 \Leftrightarrow 2x^* = 2 \Leftrightarrow x^* = 1. \quad (5.24)$$

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

5.3. Differentialrechnung multivariater reellwertiger Funktionen

Wir erinnern zunächst an den Begriff der multivariaten reellwertigen Funktion.

Definition 5.5 (Multivariate reellwertige Funktion). Eine Funktion der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n) \quad (5.25)$$

heißt *multivariate reellwertiger 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

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2 \quad (5.26)$$

Wir visualisieren diese Funktion in Abbildung 5.4. 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.

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

Definition 5.6 (Partielle Differenzierbarkeit und partielle Ableitung). Es sei $D \subseteq \mathbb{R}^n$ eine Menge und

$$f : D \rightarrow \mathbb{R}, x \mapsto f(x) \quad (5.27)$$

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

$$\frac{\partial}{\partial x_i} f(x) := \lim_{h \rightarrow 0} \frac{f(a + he_i) - f(a)}{h} \quad (5.28)$$

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

$$\frac{\partial}{\partial x_i} f : D \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_i} f(x) \quad (5.29)$$

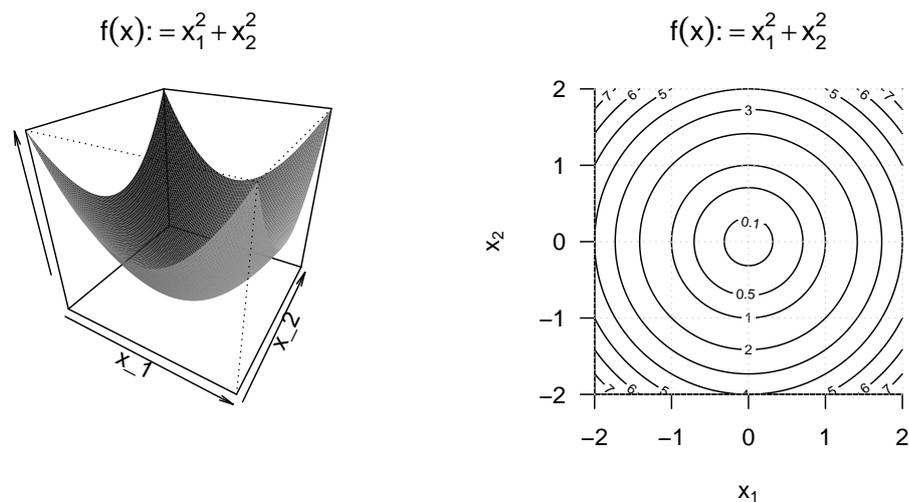


Abbildung 5.4. Visualisierungen einer bivariaten Funktion.

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

$$\frac{f(x + he_i) - f(x)}{h} \tag{5.30}$$

die Änderung $f(x + he_i) - f(x)$ von f pro Strecke h in Richtung e_i . 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

$$\frac{d}{dx_i} \tilde{f}_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n}(x_i) \tag{5.31}$$

der univariaten reellwertigen Funktion

$$\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}, x_i \mapsto \tilde{f}_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n}(x_i) := f(x_1, \dots, x_i, \dots, x_n). \tag{5.32}$$

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

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \quad (5.33)$$

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

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) \quad \text{und} \quad \frac{\partial}{\partial x_2} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x). \quad (5.34)$$

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

$$f_{x_2} : \mathbb{R} \rightarrow \mathbb{R}, x_1 \mapsto f_{x_2}(x_1) := x_1^2 + x_2^2, \quad (5.35)$$

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} ,

$$f'_{x_2}(x) = 2x_1. \quad (5.36)$$

Es ergibt sich also

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) = \frac{\partial}{\partial x_1} (x_1^2 + x_2^2) = f'_{x_2}(x) = 2x_1. \quad (5.37)$$

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

$$\frac{\partial}{\partial x_2} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x) = \frac{\partial}{\partial x_2} (x_1^2 + x_2^2) = f'_{x_1}(x) = 2x_2. \quad (5.38)$$

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

Definition 5.7 (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

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) := \frac{\partial}{\partial x_j} \left(\frac{\partial}{\partial x_i} f \right). \quad (5.39)$$

•

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

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \quad \text{für alle } 1 \leq i, j \leq n. \quad (5.40)$$

◦

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

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \quad (5.41)$$

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

$$\begin{aligned} \frac{\partial^2}{\partial x_1 x_1} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_1} (2x_1) = 2 \\ \frac{\partial^2}{\partial x_1 x_2} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_1} (2x_2) = 0 \\ \frac{\partial^2}{\partial x_2 x_1} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_2} (2x_1) = 0 \\ \frac{\partial^2}{\partial x_2 x_2} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_2} (2x_2) = 2 \end{aligned} \quad (5.42)$$

Offenbar gilt

$$\frac{\partial^2}{\partial x_1 x_2} f(x) = \frac{\partial^2}{\partial x_2 x_1} f(x). \quad (5.43)$$

Beispiel (2)

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

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}. \quad (5.44)$$

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

$$\begin{aligned} \frac{\partial}{\partial x_1} f(x) &= \frac{\partial}{\partial x_1} (x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}) = 2x_1 + x_2, \\ \frac{\partial}{\partial x_2} f(x) &= \frac{\partial}{\partial x_2} (x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}) = x_1 + \sqrt{x_3}, \\ \frac{\partial}{\partial x_3} f(x) &= \frac{\partial}{\partial x_3} (x_1^2 + x_1 x_2 + x_2 \sqrt{x_3}) = \frac{x_2}{2\sqrt{x_3}}. \end{aligned} \quad (5.45)$$

Für die zweiten partiellen Ableitungen hinsichtlich x_1 ergibt sich

$$\begin{aligned} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_1} (2x_1 + x_2) = 2, \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_2} (2x_1 + x_2) = 1, \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_3} (2x_1 + x_2) = 0. \end{aligned} \quad (5.46)$$

Für die zweiten partiellen Ableitungen hinsichtlich x_2 ergibt sich

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_1} (x_1 + \sqrt{x_3}) = 1, \\ \frac{\partial^2}{\partial x_2 \partial x_2} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_2} (x_1 + \sqrt{x_3}) = 0, \\ \frac{\partial^2}{\partial x_3 \partial x_2} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_3} (x_1 + \sqrt{x_3}) = \frac{1}{2\sqrt{x_3}}.\end{aligned}\tag{5.47}$$

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

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_1} \left(\frac{x_2}{2} \sqrt{x_3} \right) = 0, \\ \frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_2} \left(\frac{x_2}{2\sqrt{x_3}} \right) = \frac{1}{2\sqrt{x_3}}, \\ \frac{\partial^2}{\partial x_3 \partial x_3} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_3} \left(x_2 \frac{1}{2} x_3^{-\frac{1}{2}} \right) = -\frac{1}{4} x_2 x_3^{-\frac{3}{2}}.\end{aligned}\tag{5.48}$$

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

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial^2}{\partial x_2 \partial x_1} f(x) = 1, \\ \frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_1} f(x) = 0, \\ \frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_2} f(x) = \frac{1}{2\sqrt{x_3}}.\end{aligned}\tag{5.49}$$

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 5.8 (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

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^n.\tag{5.50}$$

•

Man beachte, dass Gradienten multivariate vektorwertige Funktionen der

$$\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n, x \mapsto \nabla f(x)\tag{5.51}$$

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

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \in \mathbb{R}^2. \quad (5.52)$$

In Abbildung 5.5 visualisieren wir ausgewählte Werte dieses Gradienten für

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

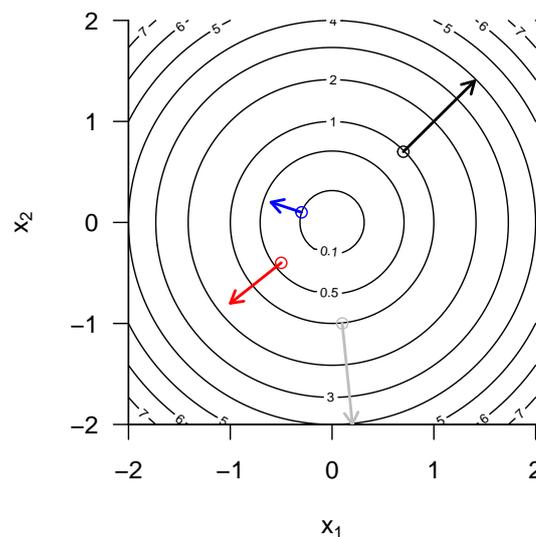


Abbildung 5.5. Exemplarische Gradientenwerte der bivariaten Funktion $f(x) = x_1^2 + x_2^2$.

Beispiel (2)

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

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \frac{\partial}{\partial x_3} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + \sqrt{x_3} \\ \frac{x_2}{2\sqrt{x_3}} \end{pmatrix} \in \mathbb{R}^3. \quad (5.53)$$

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

Definition 5.9 (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

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^{n \times n}. \quad (5.54)$$

•

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

$$\nabla^2 f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}, x \mapsto \nabla^2 f(x) \quad (5.55)$$

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

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x) \text{ für } 1 \leq i, j \leq n \quad (5.56)$$

dass die Hesse-Matrix symmetrisch ist, dass also

$$(\nabla^2 f(x))^T = \nabla^2 f(x) \quad (5.57)$$

gilt.

Beispiel (1)

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

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \in \mathbb{R}^{2 \times 2} \quad (5.58)$$

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

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_3} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \frac{\partial^2}{\partial x_2 \partial x_3} f(x) \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(x) & \frac{\partial^2}{\partial x_3 \partial x_2} f(x) & \frac{\partial^2}{\partial x_3 \partial x_3} f(x) \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & \frac{1}{2\sqrt{3}} \\ 0 & \frac{1}{2\sqrt{3}} & -\frac{1}{4}x_2x_3^{-3/2} \end{pmatrix}. \quad (5.59)$$

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.

5.4. 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

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right). \quad (5.60)$$

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 5.60 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 5.60 an und werten Sie sie in $x = (1, 2)^T$ aus.

6. 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.

6.1. Folgen

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

Definition 6.1 (Reelle Folge). Eine *reelle Folge* ist eine Funktion der Form

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) \quad (6.1)$$

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

$$(x_1, x_2, \dots) \text{ oder } (x_n)_{n=1}^{\infty} \text{ oder } (x_n)_{n \in \mathbb{N}} \text{ oder } (x_n). \quad (6.2)$$

•

Man beachte, dass weil es unendlich viele natürliche Zahlen gibt, eine reelle Folge immer unendlich viele Folenglieder 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

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) := \left(\frac{1}{n}\right)^{\frac{p}{q}} \text{ mit } p, q \in \mathbb{N} \quad (6.3)$$

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

$$\left(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \dots\right). \quad (6.4)$$

(2) Reelle Folgen der Form

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) := q^n \text{ mit } q \in]-1, 1[\quad (6.5)$$

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

$$\begin{aligned} \left(\left(\frac{1}{2} \right)^1, \left(\frac{1}{2} \right)^2, \left(\frac{1}{2} \right)^3, \dots \right) &= \left(\left(\frac{1}{2} \right)^1, \left(\frac{1}{2} \right)^2, \left(\frac{1}{2} \right)^3, \dots \right) \\ &= \left(\frac{1^1}{2^1}, \frac{1^2}{2^2}, \frac{1^3}{2^3}, \dots \right) \\ &= \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots \right) \end{aligned} \quad (6.6)$$

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

Definition 6.2 (Funktionenfolge). Es sei ϕ eine Menge univariater reellwertiger Funktionen mit Definitionsmenge $D \subseteq \mathbb{R}$. Dann ist eine Funktionenfolge eine Funktion der Form

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n). \quad (6.7)$$

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

$$(f_1, f_2, \dots) \text{ oder } (f_n)_{n=1}^{\infty} \text{ oder } (f_n)_{n \in \mathbb{N}} \text{ oder } (f_n). \quad (6.8)$$

•

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

Beispiele für Funktionenfolgen

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

$$\phi := \{f_n \mid f_n : [0, 1] \rightarrow \mathbb{R}, x \mapsto f_n(x) := x^n \text{ für } n \in \mathbb{N}\} \quad (6.9)$$

Dann definiert

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (6.10)$$

eine Funktionenfolge. Für die Funktionswerte der Folgliederglieder von F gilt

$$f_1(x) := x^1, f_2(x) := x^2, f_3(x) := x^3, \dots \quad (6.11)$$

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

$$\phi := \{f_n | f_n : [-a, a] \rightarrow \mathbb{R}, x \mapsto f_n(x) := \sum_{k=0}^n \frac{x^k}{k!} \text{ für } n \in \mathbb{N}\} \quad (6.12)$$

Dann definiert

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (6.13)$$

eine Funktionenfolge. Für die Funktionswerte der Folgenglieder von F gilt

$$f_1(x) := \sum_{k=0}^1 \frac{x^k}{k!}, f_2(x) := \sum_{k=0}^2 \frac{x^k}{k!}, f_3(x) := \sum_{k=0}^3 \frac{x^k}{k!}, \dots \quad (6.14)$$

6.2. 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 6.3 (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

$$|x_n - x| < \epsilon \text{ für alle } n \geq m. \quad (6.15)$$

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

$$\lim_{n \rightarrow \infty} x_n = x \text{ oder } x_n \rightarrow x \text{ für } n \rightarrow \infty \text{ oder } x_n \xrightarrow{n \rightarrow \infty} x. \quad (6.16)$$

•

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

$$f : \mathbb{N} \rightarrow \mathbb{R}, n \mapsto f(n) := n \quad (6.17)$$

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 $p, q \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n}\right)^{\frac{p}{q}} = 0. \quad (6.18)$$

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

$$\lim_{n \rightarrow \infty} q^n = 0. \quad (6.19)$$

Man nennt die harmonischen und geometrischen Folgen entsprechend auch *Nullfolgen*. Für Beweise von Gleichung 6.18 und Gleichung 6.19 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 6.1.

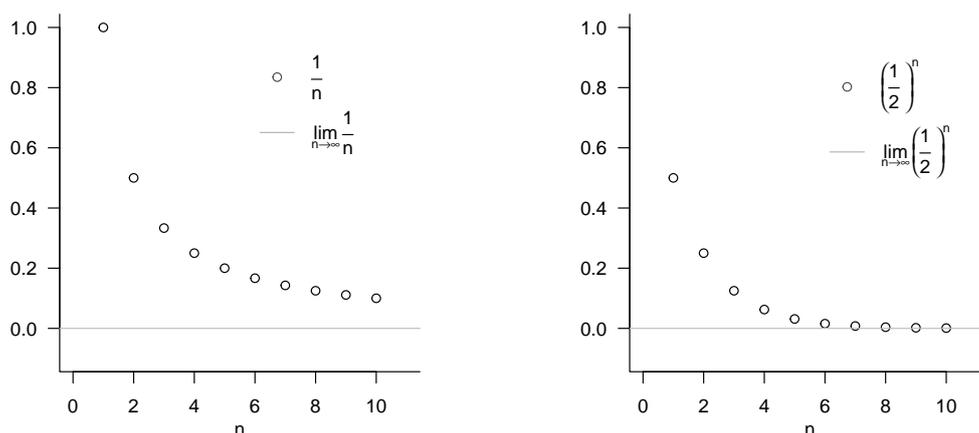


Abbildung 6.1. Beispiele für Grenzwerte reeller Folgen.

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

Definition 6.4 (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

$$f : D \rightarrow \mathbb{R}, x \mapsto f(x) := \lim_{n \rightarrow \infty} f_n(x). \quad (6.20)$$

•

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

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (6.21)$$

mit

$$\phi := \{f_n | f_n : [0, 1] \rightarrow \mathbb{R}, x \mapsto f_n(x) := x^n \text{ für } n \in \mathbb{N}\} \quad (6.22)$$

Dann ist F punktweise konvergent mit Grenzfunktion

$$f : [0, 1] \rightarrow \mathbb{R}, x \mapsto f(x) := \begin{cases} 0, & \text{für } x \in [0, 1[\\ 1, & \text{für } x = 1 \end{cases} \quad (6.23)$$

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

$$F : \mathbb{N} \rightarrow \phi, n \mapsto F(n) \quad (6.24)$$

mit

$$\phi := \{f_n | f_n : [-a, a] \rightarrow \mathbb{R}, x \mapsto f_n(x) := \sum_{k=0}^n \frac{x^k}{k!} \text{ für } n \in \mathbb{N}\} \quad (6.25)$$

Dann ist F punktweise konvergent mit Grenzfunktion

$$f : [-a, a] \rightarrow \mathbb{R}, x \mapsto f(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!} =: \exp(x) \quad (6.26)$$

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

$$\exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (6.27)$$

definiert.

6.3. 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 6.5 (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

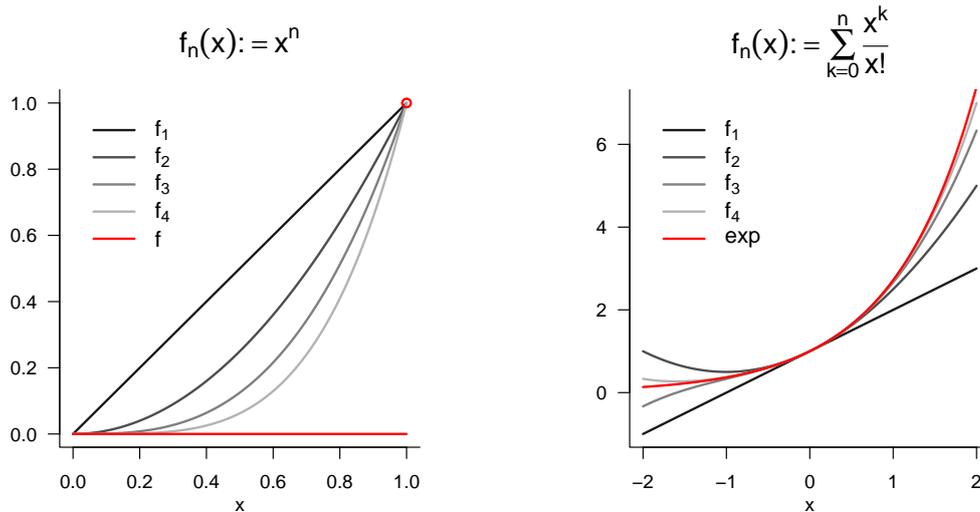


Abbildung 6.2. Beispiele für Grenzwerte von Funktionenfolgen

- (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 6.3 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 $\lim_{x \rightarrow 1} \exp(x) = e$.

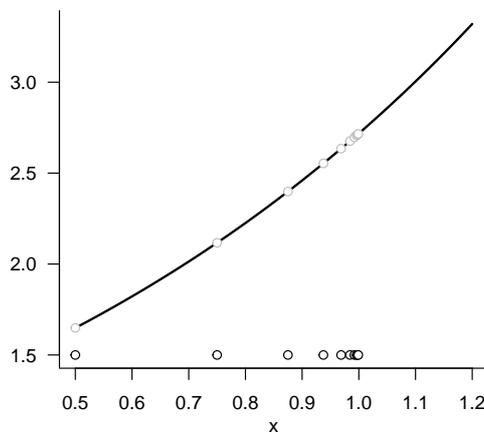


Abbildung 6.3. Beispiele für einen Grenzwert einer Funktion

Wir können nun den Begriff der Stetigkeit einer Funktion definieren.

Definition 6.6 (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

$$\lim_{x \rightarrow a} f(x) = f(a). \quad (6.28)$$

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

$$\lim_{x \rightarrow a} f(x) = f\left(\lim_{x \rightarrow a} x\right) \quad (6.29)$$

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

6.4. Selbstkontrollfragen

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	$f(x) := a$	$F(x) = ax$

Name	Definition	Stammfunktion
Identitätsfunktion	$f(x) := x$	$F(x) = \frac{1}{2}x^2$
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.

□

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.

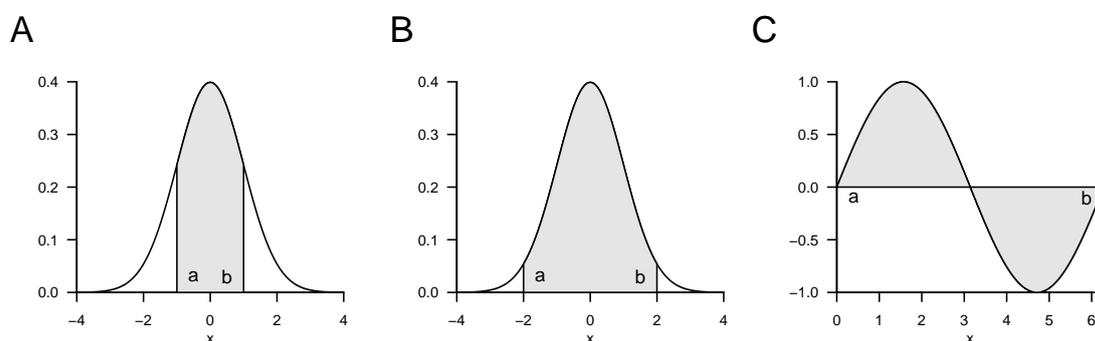


Abbildung 7.1. Beispiele bestimmter Integrale

Um den Begriff des *bestimmten Integrals* im Sinne des *Riemannsches 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

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

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

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

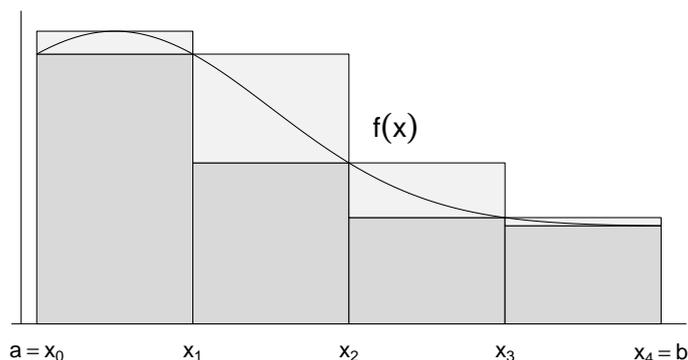


Abbildung 7.2. Riemannsche Summen

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.

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 $\int_a^b f(x) dx$.

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 .

◦

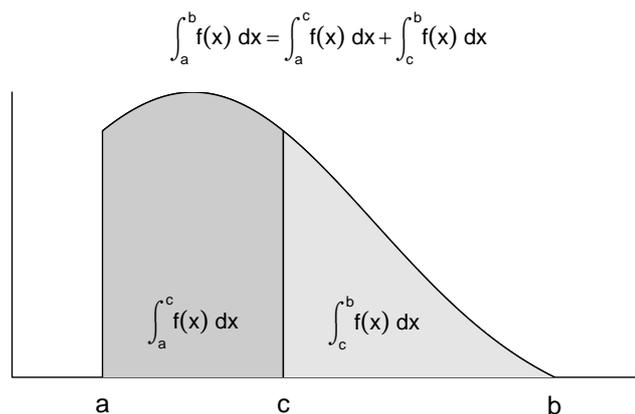


Abbildung 7.3. Additivität bestimmter Integrale

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.

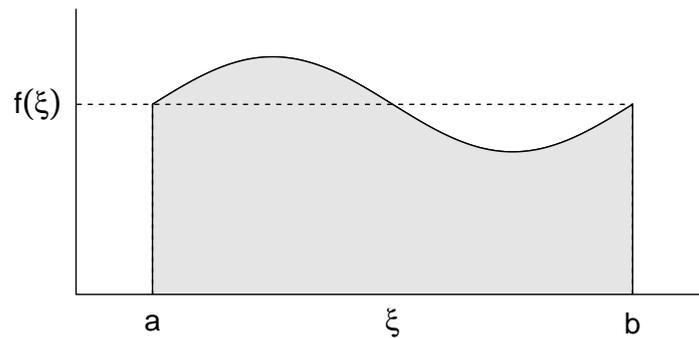


Abbildung 7.4. Zum Mittelwertsatz der Integralrechnung

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

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).

Beispiel (1)

Wir betrachten die Identitätsfunktion

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

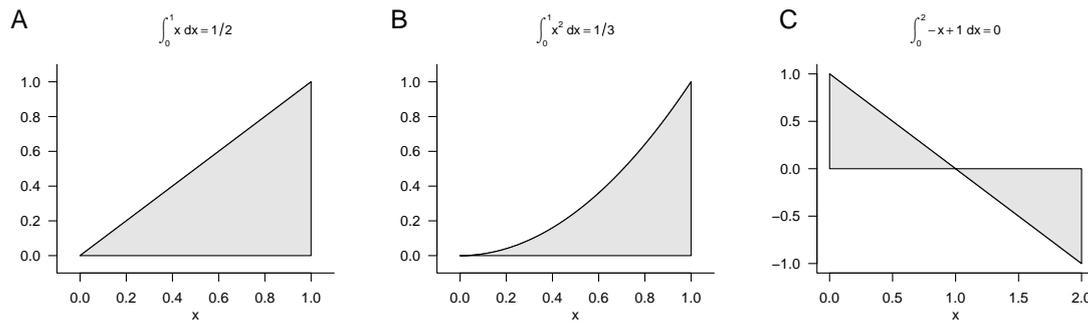


Abbildung 7.5. Beispiele zum Zweiten Hauptsatz der Differential- und Integralrechnung

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 wird 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 lineare 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

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

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

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

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

$$\int_1^\infty \frac{1}{x^2} dx. \tag{7.52}$$

Nach den Festlegungen in der Definition uneigentlicher Integrale gilt

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

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

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

Es ergibt sich also

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

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

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

mehrdimensionale bestimmte Integrale auf Hyperrechtecken. Weiterhin heißen Integrale der Form

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

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

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

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 Riemannsches 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, z.B. einen Selbstauskunftscore, einen Biomarker und eine 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

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

```
3 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 \mathbf{R} implementiert man dieses Beispiel wie folgt

```
1 x = matrix(c(2,3), nrow = 2) # Vektordefinition
2 y = matrix(c(1,3), nrow = 2) # Vektordefinition
3 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 \mathbf{R} implementiert man dieses Beispiel wie folgt

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

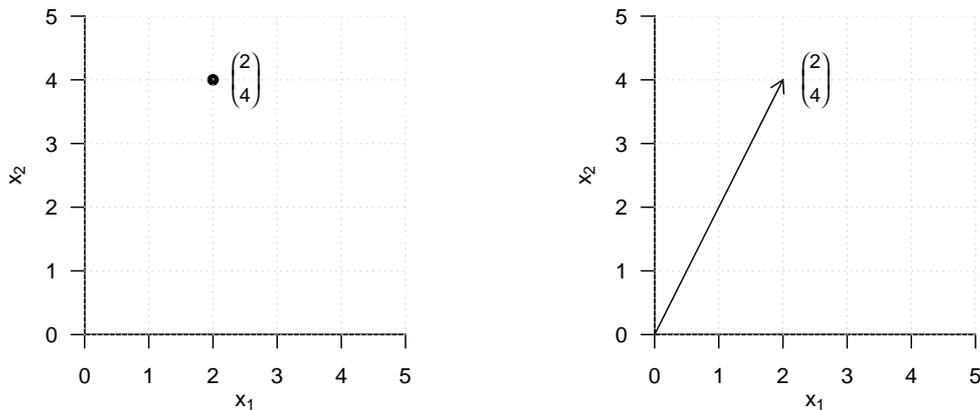


Abbildung 8.1. Visualisierung von Vektoren in \mathbb{R}^2

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

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

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.

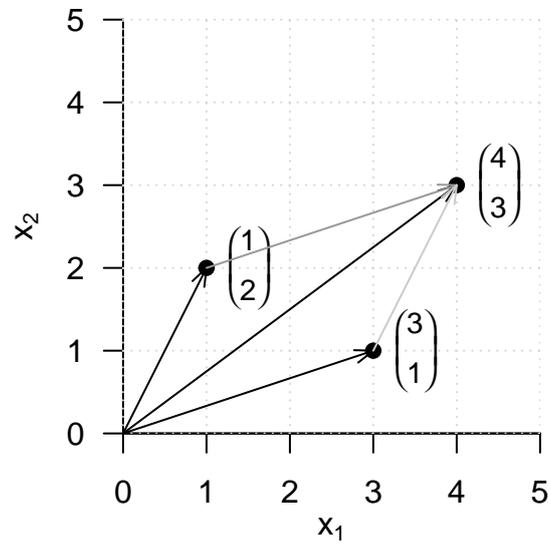
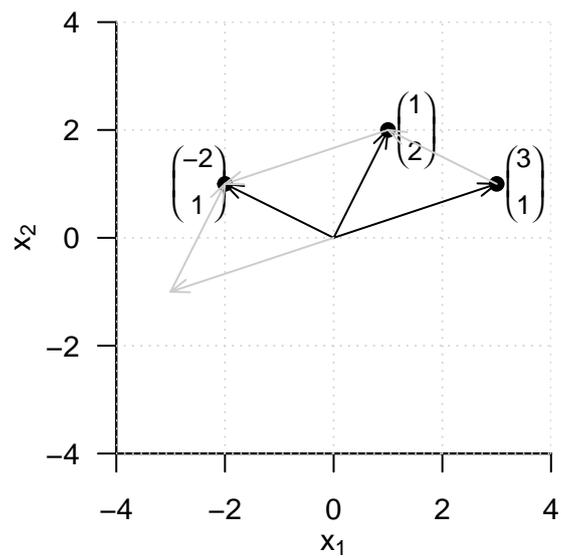
Abbildung 8.4 schließlich visualisiert die Skalarmultiplikation

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

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*

Abbildung 8.2. Vektoraddition in \mathbb{R}^2 Abbildung 8.3. Vektorsubtraktion in \mathbb{R}^2

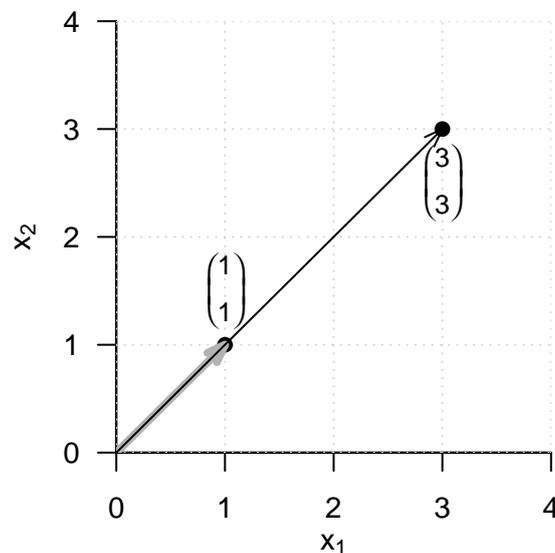


Abbildung 8.4. Skalarmultiplikation in \mathbb{R}^2

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

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

•

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 \mathbf{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 \mathbf{R} gibt es verschiedene Möglichkeiten, ein Skalarprodukt auszuwerten. Wir führen zwei von ihnen für das gegebene Beispiel untenstehend auf.

```

1 # Vektordefinitionen
2 x = matrix(c(1,2,3), nrow = 3)
3 y = matrix(c(2,0,1), nrow = 3)
4
5 # Skalarprodukt mithilfe von R's komponentenweiser Multiplikation und sum() Funktion
6 sum(x*y)

```

```
[1] 5
```

```

1 # Skalarprodukt mithilfe von R's Matrixtransposition und -multiplikation
2 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)

$$\left\| \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\| = \sqrt{\left\langle \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\rangle} = \sqrt{2^2 + 0^2} = \sqrt{4} = 2.00 \quad (8.15)$$

```
1 norm(matrix(c(2,0),nrow = 2), type = "2")           # Vektorlänge = l_2 Norm
```

```
[1] 2
```

Beispiel (2)

$$\left\| \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\| = \sqrt{\left\langle \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\rangle} = \sqrt{2^2 + 2^2} = \sqrt{8} \approx 2.83 \quad (8.16)$$

```
1 norm(matrix(c(2,2),nrow = 2), type = "2")           # Vektorlänge = l_2 Norm
```

```
[1] 2.828427
```

Beispiel (3)

$$\left\| \begin{pmatrix} 2 \\ 4 \end{pmatrix} \right\| = \sqrt{\left\langle \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix} \right\rangle} = \sqrt{2^2 + 4^2} = \sqrt{20} \approx 4.47 \quad (8.17)$$

```
1 norm(matrix(c(2,4),nrow = 2), type = "2")           # Vektorlänge = l_2 Norm
```

```
[1] 4.472136
```

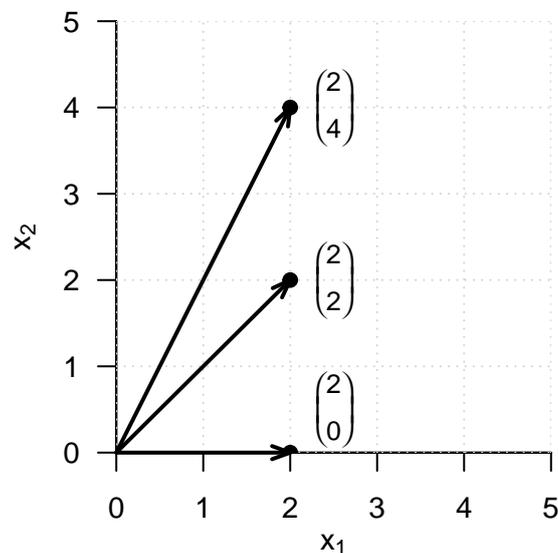
Für den Abstand $d(x, y)$ zweier Vektoren $x, y \in \mathbb{R}^m$ halten wir ohne Beweis fest, dass er zum einen nicht-negativ und symmetrisch ist, also dass

$$d(x, y) \geq 0, d(x, x) = 0 \text{ und } d(x, y) = d(y, x) \quad (8.18)$$

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.

Abbildung 8.5. Vektorlänge in \mathbb{R}^2 **Beispiel (1)**

$$d\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right) = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\| = \left\| \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\| = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} \approx 1.41 \quad (8.20)$$

```
1 norm(matrix(c(1,1),nrow = 2) - matrix(c(2,2),nrow = 2), type = "2")
```

```
[1] 1.414214
```

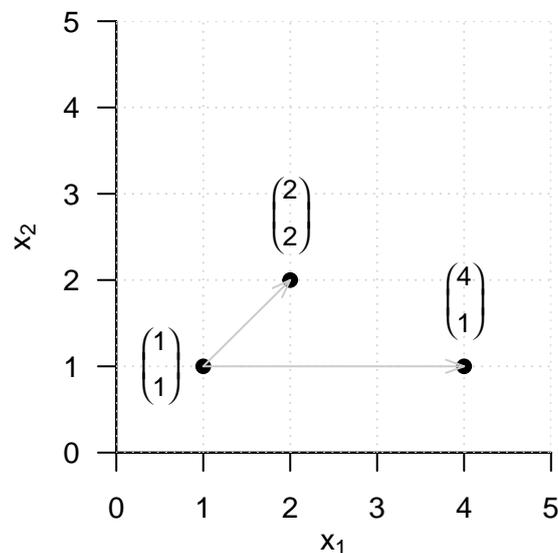
Beispiel (2)

$$d\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \end{pmatrix}\right) = \left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 4 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} -3 \\ 0 \end{pmatrix} \right\| = \sqrt{(-3)^2 + 0^2} = \sqrt{9} = 3 \quad (8.21)$$

```
1 norm(matrix(c(1,1),nrow = 2) - matrix(c(4,1),nrow = 2), type = "2")
```

```
[1] 3
```

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

Abbildung 8.6. Vektorabstände in \mathbb{R}^2

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.

```

1 x = matrix(c(3,0), nrow = 2)           # Vektor 1
2 y = matrix(c(3,3), nrow = 2)         # Vektor 2
3 w = acos(sum(x*y)/(sqrt(sum(x*x))*sqrt(sum(y*y)))) * 180/pi # Winkel in Grad
4 print(w)

```

[1] 45

Beispiel (2)

$$\alpha = \arccos \left(\frac{\left\langle \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\rangle}{\left\| \begin{pmatrix} 3 \\ 0 \end{pmatrix} \right\| \left\| \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\|} \right) = \arccos \left(\frac{3 \cdot 0 + 0 \cdot 3}{\sqrt{3^2 + 0^2} \cdot \sqrt{0^2 + 3^2}} \right) = \arccos \left(\frac{0}{3 \cdot 3} \right) = \frac{\pi}{2} \approx 1.57$$
(8.24)

Die Umrechnung in Grad ergibt dann

$$\frac{\pi}{2} \cdot \frac{180^\circ}{\pi} = 90^\circ$$
(8.25)

Die entsprechende **R** Implementation lautet wie folgt.

```

1 x = matrix(c(3,0), nrow = 2)           # Vektor 1
2 y = matrix(c(0,3), nrow = 2)         # Vektor 2
3 w = acos(sum(x*y)/(sqrt(sum(x*x))*sqrt(sum(y*y)))) * 180/pi # Winkel in Grad
4 print(w)
```

[1] 90

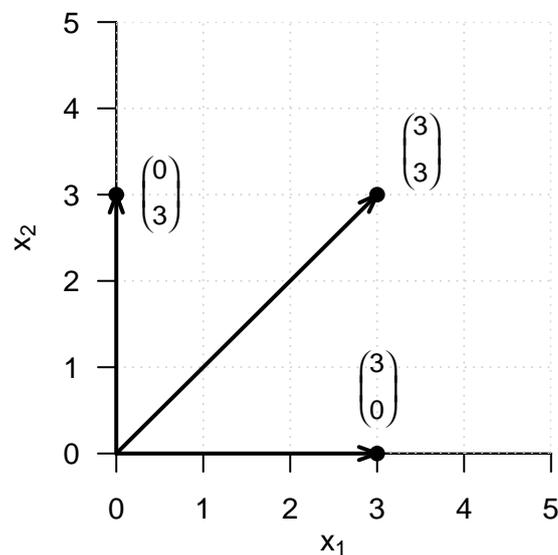


Abbildung 8.7. Winkel in \mathbb{R}^2

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). $((\mathbb{R}^m, +, \cdot), \langle \cdot, \cdot \rangle)$ 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

$$\alpha = \frac{\pi}{2} = 90^\circ. \quad (8.29)$$

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. □

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 $i \neq j$. □

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 unendliche 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

$$v = \sum_{i=1}^m c_i b_i \tag{8.41}$$

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

$$\begin{aligned} v &= a_1 b_1 + \dots + a_m b_m \\ v &= c_1 b_1 + \dots + c_m b_m \end{aligned} \tag{8.42}$$

Subtraktion der unteren von der oberen Gleichung ergibt

$$0 = (a_1 - c_1)b_1 + \dots + (a_m - c_m)b_m \tag{8.43}$$

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. □

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

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases} \tag{8.44}$$

Wir wollen zunächst ein Beispiel für eine Orthonormalbasis betrachten. •

Beispiel (1)

Es ist

$$B_1 := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \tag{8.45}$$

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

$$B := \left\{ e_1, \dots, e_m \mid e_{i_j} = 1 \text{ für } i = j \text{ und } e_{i_j} = 0 \text{ für } i \neq j \right\} \subset \mathbb{R}^m \quad (8.49)$$

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

$$\left\langle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\rangle = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1, \quad (8.52)$$

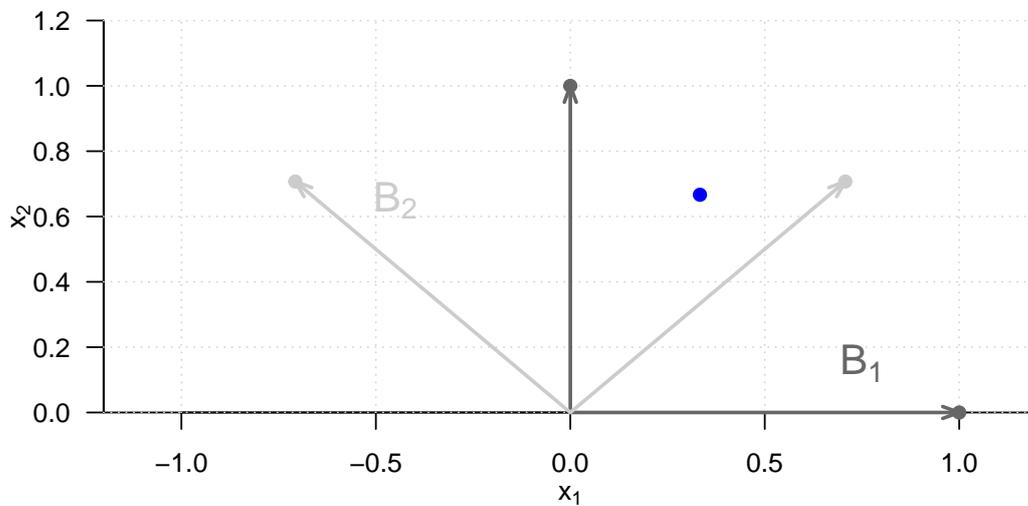
sowie

$$\left\langle \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\rangle = \left(-\frac{1}{\sqrt{2}}\right) \cdot \left(-\frac{1}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1 \quad (8.53)$$

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. Vektorkoordinatentransformation

Im Kapitel 8.4 haben wir festgestellt, dass ein Vektorraum in der Regel viele unterschiedliche Basen hat. Hat man nun mit den Koordinaten eines Vektors bezüglich einer dieser Basen eine Darstellung eines Vektors bezüglich einer dieser Basen vorliegen, so schließt sich die Frage an, wie derselbe Vektor hinsichtlich einer anderen Basis dargestellt werden kann, wie also die Koordinaten desselben Vektors hinsichtlich einer weiteren Basis berechnet werden können. Eine Antwort darauf gibt das *Theorem zur Vektorkoordinatentransformation* (Theorem 8.7) dieses Abschnittes. Bevor wir dieses Theorem anhand der in Kapitel 8.4 diskutierten Basen B_1 und B_2 verdeutlichen, führen wir mit den Begriffen der *Orthogonalprojektion* und den *Vektorkoordinaten bezüglich einer Orthogonalbasis* zwei dazu nötige Konzepte ein. Den Begriff der Orthogonalprojektion definieren wir wie folgt.

Definition 8.15 (Orthogonalprojektion). x und q seien Vektoren im Euklidischen Vektorraum \mathbb{R}^m . Dann ist die *Orthogonalprojektion von x auf q* definiert als der Vektor

$$\tilde{x} = aq \text{ mit } a := \frac{q^T x}{q^T q}, \quad (8.55)$$

wobei der Skalar a *Projektionsfaktor* genannt wird.

•

Sinn von Definition 8.15 ist es, mit $\tilde{x} = aq$ für $a \in \mathbb{R}$ den Punkt in Richtung des Vektors q der dem Punkt x am nächsten ist anzugeben. Wir visualisieren dazu die Orthogonalprojektion eines Vektors x auf einen Vektor q in \mathbb{R}^2 in Abbildung 8.9. Diese

minimierte Distanzeigenschaft im Euklidischen Raum impliziert aber die Orthogonalität von q und $x - \tilde{x}$, woraus die die Formel für a dann direkt folgt, da gilt

$$q^T(x - \tilde{x}) = 0 \Leftrightarrow q^T(x - aq) = 0 \Leftrightarrow q^T x - aq^T q = 0 \Leftrightarrow a = \frac{q^T x}{q^T q}. \quad (8.56)$$

Man beachte, dass wenn q die Länge 1 hat, gilt dass

$$a = \frac{q^T x}{q^T q} = \frac{q^T x}{\|q\|^2} = \frac{q^T x}{1} = q^T x. \quad (8.57)$$

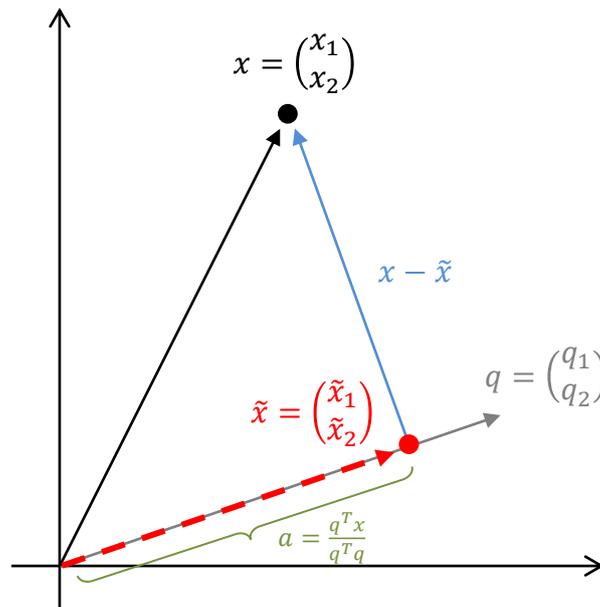


Abbildung 8.9. Orthogonalprojektion eines Vektors x auf einen Vektor q in \mathbb{R}^2 .

Mithilfe der Orthogonalprojektion können die Koordinaten eines Vektors bezüglich einer beliebigen Orthogonalbasis bestimmt werden. Es gilt folgendes Theorem

Theorem 8.6 (Vektorkoordinaten bezüglich einer Orthogonalbasis). *Es sei $x \in \mathbb{R}^m$ und es sei $B := \{q_1, \dots, q_m\}$ eine Orthonormalbasis von \mathbb{R}^m . Dann ergeben sich für $i = 1, \dots, m$ die Koordinaten c_i in der Basisdarstellung von x bezüglich B als die Projektionsfaktoren*

$$c_i = x^T q_i \quad (8.58)$$

in der Orthogonalprojektion von x auf q_i . Äquivalent ist die Basisdarstellung von x bezüglich B gegeben durch

$$x = \sum_{i=1}^m (x^T q_i) q_i. \quad (8.59)$$

◦

Beweis. Für $i = 1, \dots, m$ gilt

$$x = \sum_{j=1}^m c_j q_j \Leftrightarrow q_i^T x = q_i^T \sum_{j=1}^m c_j q_j \Leftrightarrow q_i^T x = \sum_{j=1}^m c_j q_i^T q_j \Leftrightarrow q_i^T x = c_i \Leftrightarrow c_i = x^T q_i. \quad (8.60)$$

□

Man beachte, dass sich hinsichtlich der kanonischen Basis des m -dimensionalen reellen Euklidischen Vektorraums ergibt für jeden Vektor $x \in \mathbb{R}^m$ ergibt, dass

$$c_i = x^T e_i = x_i \text{ für } i = 1, \dots, m. \quad (8.61)$$

Wenn nichts anderes gesagt ist entsprechen die Komponenten eines m -dimensionalen Vektors also in der Regel seinen Koordinaten hinsichtlich der kanonischen Basis des m -dimensionalen reellen Euklidischen Vektorraums. Zur Transformation der Koordinaten eines Vektors zwischen zwei Orthonormalbasen gilt folgendes Theorem, das wir nicht beweisen wollen.

Theorem 8.7 (Vektorkoordinatentransformation). $B_v := \{v_1, \dots, v_m\}$ und $B_w := \{w_1, \dots, w_m\}$ seien zwei Orthonormalbasen eines Vektorraums. $A \in \mathbb{R}^{m \times m}$ sei die Matrix, die durch die spaltenweise Konkatenation der Koordinaten der Vektoren in B_w in der Basisdarstellung bezüglich der Basis B_v ergibt. Dann können die Koordinaten $x_i, i = 1, \dots, m$ eines Vektors x bezüglich der Basis B_v in die Koordinaten $\tilde{x}_1, \dots, \tilde{x}_m$ des Vektors bezüglich der Basis B_w durch

$$\tilde{x} = A^T x \quad (8.62)$$

transformiert werden. Analog können die Koordinaten $\tilde{x}_1, \dots, \tilde{x}_m$ des Vektors hinsichtlich der Basis B_w in die Koordinaten x_1, \dots, x_m des Vektors hinsichtlich B_v durch

$$x = A \tilde{x}. \quad (8.63)$$

transformiert werden.

◦

Zur Berechnung der Koordinaten eines Vektors bezüglich einer alternativen Basis muss im Sinne von Theorem 8.7 also zunächst die Transformationsmatrix A bestimmt werden und dann der zu transformierende Vektor entsprechend multipliziert werden. Man beachte unbedingt, dass ein Vektor hier als fester Punkt in \mathbb{R}^m betrachtet wird, die Komponenten des Vektors dagegen werden als Koordinaten bezüglich einer spezifischen Basis interpretiert. Man vergleiche dazu Abbildung 8.10 zu dem im folgenden erläuterten Beispiel.

Beispiel

Wir nehmen an, dass wir die Koordinaten von $x = \left(\frac{1}{3}, \frac{2}{3}\right)^T \in \mathbb{R}^2$ hinsichtlich der kanonischen Orthonormalbasis $B_v := \{e_1, e_2\}$ in die Koordinaten bezüglich der Basis

$$B_w := \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\} \quad (8.64)$$

transformieren wollen. Die Basisdarstellungen der in Vektoren B_w bezüglich der Basisvektoren in B_v sind

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = a_{11} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ und } \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = a_{12} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_{22} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (8.65)$$

Die Projektionsfaktoren der Orthogonalprojektionen der Vektoren in B_w auf die Vektoren in B_v sind

$$a_{11} = \frac{1}{\sqrt{2}}, a_{21} = \frac{1}{\sqrt{2}}, a_{12} = -\frac{1}{\sqrt{2}}, a_{22} = \frac{1}{\sqrt{2}}. \quad (8.66)$$

Die Transformationsmatrix $A \in \mathbb{R}^{m \times m}$ nach Theorem 8.7 ergibt sich also zu

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}. \quad (8.67)$$

Damit ergibt sich die Vektorkoordinatentransformation von $x \in \mathbb{R}^2$ dann zu

$$\tilde{x} = A^T x = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \approx \begin{pmatrix} 0.70 \\ 0.23 \end{pmatrix}. \quad (8.68)$$

Hinsichtlich der Visualisierung des Beispiels in Abbildung 8.10 beachte man, dass x und \tilde{x} am selben Ort der \mathbb{R}^2 Ebene liegen und sich lediglich die Zahlwerte der Vektorkomponenten durch Orthogonalprojektion auf die entsprechenden Basisvektorachsen unterscheiden.

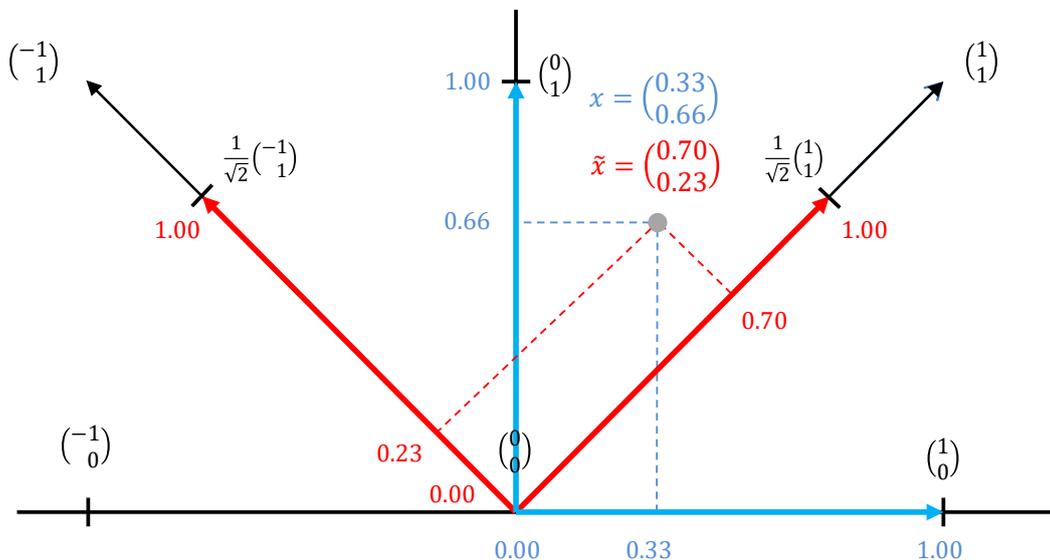


Abbildung 8.10. Vektorkoordinatentransformation eines Vektors x bezüglich zweier Orthonormalbasen des zweidimensionalen kanonischen Euklidischen Vektorraums.

8.6. Selbstkontrollfragen

1. Geben Sie die Definition eines Vektorraums wieder.
2. Geben Sie die Definition des reellen Vektorraums wieder.
3. Es seien

$$x := \begin{pmatrix} 2 \\ 1 \end{pmatrix}, y := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ und } a := 2. \quad (8.69)$$

Berechnen Sie

$$v = a(x + y) \text{ und } w = \frac{1}{a}(y - x) \quad (8.70)$$

4. Geben Sie die Definition des Skalarproduktes auf \mathbb{R}^m wieder.

5. Für

$$x := \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, y := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, z := \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \quad (8.71)$$

berechnen Sie

$$\langle x, y \rangle, \langle x, z \rangle, \langle y, z \rangle \quad (8.72)$$

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 8.71.
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 8.71.
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 8.71.
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.
23. Geben Sie die Definition einer Orthogonalprojektion wieder.
24. Geben Sie das Theorem zu Vektorkoordinaten bezüglich einer Orthogonalbasis wieder.
25. Geben Sie das Vektorkoordinatentransformationstheorem wieder.
26. Erläutern Sie den Nutzen des Vektorkoordinatentransformationstheorems.

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

```
1 # Spaltenweise Definition von A (R default)
2 A = matrix(c(2,1,3,6,0,5), nrow = 2)
3 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.

```
1 # Reihenweise Definition von A (R default)
2 A = matrix(c(2,3,0,
3             1,6,5),
4             nrow = 2,
5             byrow = TRUE)
6 print(A)
```

```
      [,1] [,2] [,3]
[1,]    2    3    0
[2,]    1    6    5
```

```

1 # Zeilenweise Definition von B
2 B = matrix(c(4,1,0,
3             -4,2,0),
4             nrow = 2,
5             byrow = TRUE)
6 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

$$+ : \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (A, B) \mapsto +(A, B) := A + B \tag{9.5}$$

mit

$$\begin{aligned}
 A + B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \\
 &:= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}.
 \end{aligned} \tag{9.6}$$

•

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.

```

1 # Definition
2 A = matrix(c(2, -3, 0,
3             1, 6, 5),
4             nrow = 2,
5             byrow = TRUE)
6 B = matrix(c(4, 1, 0,
7             -4, 2, 0),
8             nrow = 2,
9             byrow = TRUE)
10
11 # Addition
12 C = A + B
13 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,

$$\begin{aligned}
 D = A - B &= \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} - \begin{pmatrix} 4 & 1 & 0 \\ -4 & 2 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} 2-4 & -3-1 & 0-0 \\ 1+4 & 6-2 & 5-0 \end{pmatrix} \\
 &= \begin{pmatrix} -2 & -4 & 0 \\ 5 & 4 & 5 \end{pmatrix}.
 \end{aligned}
 \tag{9.11}$$

In \mathbf{R} führt man diese Rechnung wie folgt aus.

```

1 # Subtraktion
2 D = A - B
3 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

$$\cdot : \mathbb{R} \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}, (c, A) \mapsto \cdot(c, A) := cA
 \tag{9.12}$$

mit

$$cA = c \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} := \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}.
 \tag{9.13}$$

•

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.

```

1 # Definitionen
2 A = matrix(c(3,1,1,
3             5,2,5,
4             2,7,1,
5             3,4,2),
6           nrow = 4,
7           byrow = TRUE)
8 c = -3
9
10 # Skalarmultiplikation
11 B = c*A
12 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

$$-A := (-a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m} \tag{9.16}$$

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

$$\cdot^T : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{m \times n}, A \mapsto \cdot^T(A) := A^T \tag{9.17}$$

mit

$$A^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}^T := \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}. \tag{9.18}$$

•

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
$$A^T = A. \tag{9.19}$$

(2) Es gilt
$$(A^T)^T = A. \tag{9.20}$$

(3) Es gilt
$$(a_{ii})_{1 \leq i \leq \min(n,m)} = (a_{ii})_{1 \leq i \leq \min(n,m)}^T. \tag{9.21}$$

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.

```

1 # Definition
2 A = matrix(c(2,3,0,
3             1,6,5),
4             nrow = 2,
5             byrow = TRUE)
6 print(A)

```

```

      [,1] [,2] [,3]
[1,]    2    3    0
[2,]    1    6    5

```

```

1 # Transposition
2 AT = t(A)
3 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

$$\cdot : \mathbb{R}^{n \times m} \times \mathbb{R}^{m \times k} \rightarrow \mathbb{R}^{n \times k}, (A, B) \mapsto \cdot(A, B) := AB \tag{9.28}$$

mit

$$\begin{aligned} AB &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{pmatrix} \\ &:= \begin{pmatrix} \sum_{i=1}^m a_{1i}b_{i1} & \sum_{i=1}^m a_{1i}b_{i2} & \cdots & \sum_{i=1}^m a_{1i}b_{ik} \\ \sum_{i=1}^m a_{2i}b_{i1} & \sum_{i=1}^m a_{2i}b_{i2} & \cdots & \sum_{i=1}^m a_{2i}b_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ni}b_{i1} & \sum_{i=1}^m a_{ni}b_{i2} & \cdots & \sum_{i=1}^m a_{ni}b_{ik} \end{pmatrix} \\ &= \left(\sum_{i=1}^m a_{ji}b_{il} \right)_{1 \leq j \leq n, 1 \leq l \leq k} \end{aligned} \tag{9.29}$$

•

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

$$(n \times m)(m \times k) = (n \times k). \tag{9.30}$$

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

$$A := \begin{pmatrix} 2 & -3 & 0 \\ 1 & 6 & 5 \end{pmatrix} \text{ und } B := \begin{pmatrix} 4 & 2 \\ -1 & 0 \\ 1 & 3 \end{pmatrix}. \tag{9.31}$$

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

$$(2 \times 3)(3 \times 2) = (2 \times 2) \tag{9.32}$$

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.

```

1 # Definitionen
2 A = matrix(c(2,-3,0,
3             1, 6,5),
4             nrow = 2,
5             byrow = TRUE)
6 B = matrix(c( 4,2,
7             -1,0,
8             1,3),
9             nrow = 3,
10            byrow = TRUE)
11
12 # Matrixmultiplikation
13 C = A %% B
14 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 \mathbf{R} überprüft man diese Rechnung wie folgt.

```

1 # Definitionen
2 A = matrix(c(2,-3,0,
3             1, 6,5),
4             nrow = 2,
5             byrow = TRUE)
6 B = matrix(c( 4,2,
7             -1,0,
8             1,3),
9             nrow = 3,
10            byrow = TRUE)
11
12 # Matrixmultiplikation
13 D = B %*% A
14 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.

```

1 # Beispiel für eine undefinierte Matrixmultiplikation
2 E = t(A) %*% B      # (3 x 2)(3 x 2)

```

Error in t(A) %*% B: nicht passende Argumente

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*

$$A(BC) = (AB)C. \tag{9.41}$$

(2) *Die Kombination von Matrizenmultiplikation und Skalarmultiplikation ist assoziativ,*

$$c(AB) = (cA)B = A(cB). \tag{9.42}$$

◦

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

$$c \left(\sum_{i=1}^m a_{ji} b_{il} \right) = \sum_{i=1}^m (c a_{ji}) b_{il} = \sum_{i=1}^m a_{ji} (c b_{il}). \tag{9.43}$$

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*

$$(A + B)C = AC + BC \tag{9.44}$$

und

$$C^T(A + B) = C^T A + C^T B \tag{9.45}$$

◦

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*

$$AB \neq BA. \tag{9.46}$$

◦

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

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ und } B := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{9.47}$$

Dann gilt

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = BA. \tag{9.48}$$

□

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:

$$Ax = b \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \Leftrightarrow \begin{cases} 1x_1 + 2x_2 = 5 \\ 3x_1 + 4x_2 = 11 \end{cases} \quad (9.53)$$

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 *multiplikativen 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

$$ax = b \Leftrightarrow a^{-1}ax = a^{-1}b \Leftrightarrow 1 \cdot x = a^{-1}b \Leftrightarrow x = \frac{b}{a}. \tag{9.56}$$

Ganz konkret etwa

$$2x = 6 \Leftrightarrow 2^{-1}2x = 2^{-1}6 \Leftrightarrow \frac{1}{2}2x = \frac{1}{2}6 \Leftrightarrow x = 3. \tag{9.57}$$

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

$$A^{-1}A = "1". \tag{9.58}$$

resultiert. Dann hätte man nämlich

$$Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b. \tag{9.59}$$

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

$$I_n := (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{9.60}$$

mit $a_{ij} = 1$ für $i = j$ und $a_{ij} = 0$ für $i \neq j$ heißt *n-dimensionale Einheitsmatrix*. •

In \mathbf{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

$$I_n A = A \text{ und } A I_m = A. \tag{9.61}$$

◦

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 $1 \leq j \leq m$

$$d_{ij} = 0 \cdot a_{1j} + 0 \cdot a_{2j} + \cdots + 0 \cdot a_{i-1,j} + 1 \cdot a_{ij} + \cdots + 0 \cdot a_{i+1,j} + 0 \cdot a_{nj} = a_{ij}. \tag{9.62}$$

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.

```
1 # Definition
2 A = matrix(c(2,1,
3             3,4),
4            nrow = 2,
5            byrow = TRUE)
6
7 # Berechnen von A^{-1}
8 print(solve(A))
```

```
      [,1] [,2]
[1,]  0.8 -0.2
[2,] -0.6  0.4
```

```
1 # Überprüfen der Eigenschaften einer inversen Matrix
2 print(solve(A) %*% A)
```

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

```
1 # Bei der umgekehrten Berechnung ergeben sich kleine Rundungsfehler
2 print(A %*% solve(A))
```

```
      [,1]      [,2]
[1,]    1 -5.551115e-17
[2,]    0  1.000000e+00
```

Nicht-invertierbare Matrizen sind dabei natürlich auch numerisch nicht-invertierbar, wie folgende Fehlermeldung in **R** bezüglich obigen Beispiels einer nicht-invertierbaren Matrix demonstriert.

```
1 B = matrix(c(1,0,
2             0,0),
3            nrow = 2,
4            byrow = 2)
5 solve(B)
```

```
Error in solve.default(B): Lapackroutine dgesv: System ist genau 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

$$|\cdot| : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}, A \mapsto |A|, \tag{9.70}$$

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

$$|A| := a_{11} \quad \text{für } n = 1 \tag{9.71}$$

$$|A| := \sum_{j=1}^n a_{1j}(-1)^{1+j} \det(A_{1j}) \quad \text{für } n > 1 \tag{9.72}$$

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

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \tag{9.73}$$

ergeben sich dabei zum Beispiel folgende Matrizen der Form $A_{ij} \in \mathbb{R}^{3-1 \times 3-1}$:

$$A_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, A_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, A_{21} = \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}, A_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}. \tag{9.74}$$

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

$$|A| = a_{11}a_{22} - a_{12}a_{21}. \tag{9.75}$$

Es sei $A = (a_{ij})_{1 \leq i, j \leq 3} \in \mathbb{R}^{3 \times 3}$. Dann gilt

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \tag{9.76}$$

◦

Beweis. Für $A \in \mathbb{R}^{2 \times 2}$ gilt nach Definition

$$\begin{aligned}
 |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\
 &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| \\
 &= a_{11}|(a_{22})| - a_{12}|(a_{21})| \\
 &= a_{11}a_{22} - a_{12}a_{21}.
 \end{aligned}
 \tag{9.77}$$

Für $A \in \mathbb{R}^{3 \times 3}$ gilt nach Definition und mit der Formel für Determinanten von 2×2 Matrizen

$$\begin{aligned}
 |A| &= \sum_{j=1}^n a_{1j}(-1)^{1+j}|A_{1j}| \\
 &= a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{12}| + a_{13}(-1)^{1+3}|A_{13}| \\
 &= a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| \\
 &= a_{11} \left| \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \right| - a_{12} \left| \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} \right| + a_{13} \left| \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \right| \\
 &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\
 &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\
 &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}.
 \end{aligned}
 \tag{9.78}$$

□

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 Gegendagonalen.”

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.

```

1 # Matrixdefinition und Determinantenberechnung
2 A = matrix(c(2,1,

```

```

3         3,4),
4         nrow = 2,
5         byrow = TRUE)
6 det(A)

```

[1] 5

```

1 # Matrixdefinition und Determinantenberechnung
2 B = matrix(c(1,0,
3             0,0),
4           nrow = 2,
5           byrow = TRUE)
6 det(B)

```

[1] 0

```

1 # Matrixdefinition und Determinantenberechnung
2 C = matrix(c(2,0,0,
3             0,1,0,
4             0,0,3),
5           nrow = 3,
6           byrow = TRUE)
7 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

$$A \text{ ist invertierbar} \Leftrightarrow |A| \neq 0 \text{ und } A \text{ ist nicht invertierbar} \Leftrightarrow |A| = 0. \tag{9.88}$$

◦

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

$$|AB| = |A||B| = |I_n| = 1. \tag{9.89}$$

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

$$A_1 = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, A_3 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \tag{9.90}$$

mit den jeweiligen Determinanten

$$|A_1| = 3 \cdot 2 - 1 \cdot 1 = 5, \quad |A_2| = 2 \cdot 2 - 0 \cdot 0 = 4, \quad |A_3| = 2 \cdot 2 - 2 \cdot 2 = 0. \tag{9.91}$$

Abbildung 9.1 visualisiert die entsprechende Intuition.

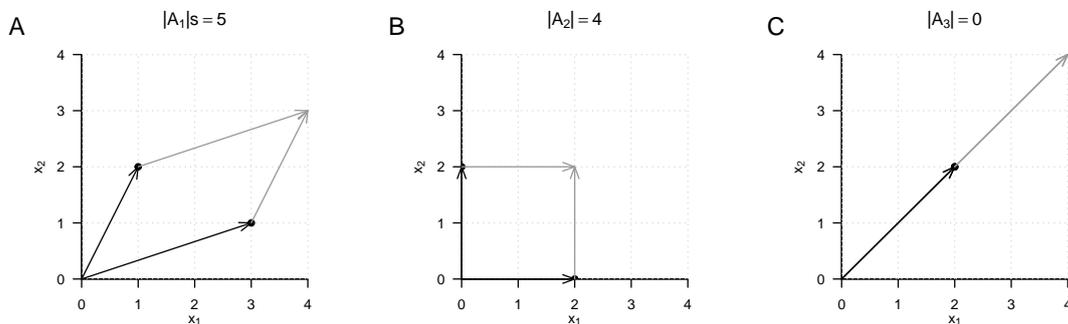


Abbildung 9.1. Determinanten als Parallelotopvolumina.

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

$$I_n := (i_{jk})_{1 \leq j \leq n, 1 \leq k \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ f\u00fcr } j = k \text{ und } i_{jk} = 0 \text{ f\u00fcr } j \neq k. \quad (9.92)$$

Wir bezeichnen die *Einheitsvektoren* $e_i, i = 1, \dots, n$ mit

$$e_i := (e_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{ij} = 1 \text{ f\u00fcr } i = j \text{ und } e_{ij} = 0 \text{ f\u00fcr } i \neq j. \quad (9.93)$$

•

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

$$I_n = (e_1 \ \dots \ e_n) \in \mathbb{R}^{n \times n} \quad (9.94)$$

F\u00fcr $n = 3$ 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 f\u00fcr die Einheitsvektoren bekanntlich f\u00fcr $1 \leq i, j \leq n$

$$e_i^T e_j = 0 \text{ f\u00fcr } i \neq j, e_i^T e_i = 1 \text{ und } e_i^T v = v^T e_i = v_i \text{ f\u00fcr } v \in \mathbb{R}^n. \quad (9.96)$$

9.6.2. Einsmatrizen und Nullmatrizen

Definition 9.11 (Nullmatrizen, Nullvektoren, Einsmatrizen, Einsvektoren). Wir bezeichnen *Nullmatrizen* und *Nullvektoren* mit

$$0_{nm} := (0)_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n. \quad (9.97)$$

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

$$S := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}. \tag{9.105}$$

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

$$S_1 + S_2 = (S_1 + S_2)^T. \tag{9.106}$$

(Inverse.) S sei eine invertierbare symmetrische Matrix und S^{-1} ihre Inverse. Dann ist auch S^{-1} eine symmetrische Matrix, das heißt es gilt

$$(S^{-1})^T = S^{-1}. \tag{9.107}$$

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

$$Q = (q_1 \ \dots \ q_n) \text{ mit } q_i \in \mathbb{R}^n \text{ für } 1 \leq i \leq n, \tag{9.108}$$

dass

$$q_i^T q_j = 0 \text{ für } i \neq j \text{ und } q_i^T q_j = 1 \text{ für } i = j \text{ mit } 1 \leq i, j \leq n. \tag{9.109}$$

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

$$Q^{-1} = Q^T. \tag{9.110}$$

(Transposition) Die Zeilen von Q sind orthonormal, es gilt

$$Q Q^T = I_n \tag{9.111}$$

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-definiter Matrizen zusammen.

Theorem 9.14 (Eigenschaften positiv-definiter 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.

10. 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.

10.1. 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 10.1 (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

$$Av = \lambda v \tag{10.1}$$

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 10.1 (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

$$Av = \lambda v \Leftrightarrow cAv = c\lambda v \Leftrightarrow A(cv) = \lambda(cv). \tag{10.2}$$

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

$$\|v\| = 1. \quad (10.3)$$

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 10.1 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

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (10.4)$$

Dann ist der Vektor der Länge 1

$$v := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (10.5)$$

ein Eigenvektor von A zum Eigenwert $\lambda = 3$, da gilt, dass

$$\begin{aligned} Av &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 3 \\ 3 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= 3 \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \\ &= \lambda v. \end{aligned} \quad (10.6)$$

Inspektion von Abbildung 10.1 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

$$w := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (10.7)$$

dagegen hat zwar die Länge 1, ist aber im Gegensatz zu v kein Eigenvektor von A , da es im Falle von

$$Aw = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad (10.8)$$

keinen Skalar λ geben kann, der mit Null, dem zweiten Eintrag von w , multipliziert einen Wert ungleich Null ergeben kann. Inspektion von Abbildung 10.1 zeigt dementsprechend, dass der aus der Multiplikation von w mit A resultierende Vektor in eine andere Richtung zeigt als w .

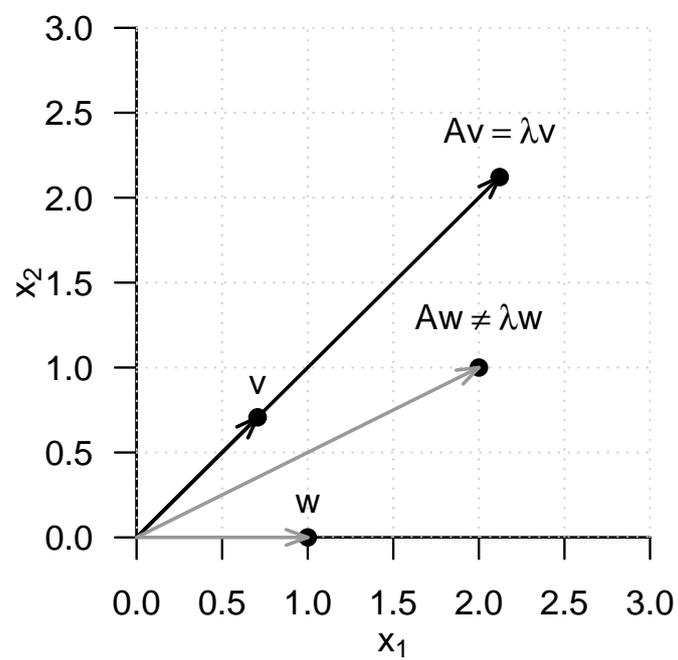


Abbildung 10.1. Eigenvektor einer 2×2 Matrix. Für die Matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (10.9)$$

ist v ein Eigenvektor, w jedoch nicht

Bestimmung von Eigenwerten und Eigenvektoren

Folgendes Theorem besagt, wie die Eigenwerte und Eigenvektoren einer quadratischen Matrix berechnet werden können.

Theorem 10.2 (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

$$\chi_A(\lambda) := |A - \lambda I_m| \quad (10.10)$$

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

$$(A - \lambda_i^* I_m)v_i = 0_m \text{ für } i = 1, 2, \dots \quad (10.11)$$

bestimmt werden.

◦

Beweis. (1) Bestimmen von Eigenwerten

Wir halten zunächst fest, dass mit der Definition von Eigenvektoren und Eigenwerten gilt, dass

$$Av = \lambda v \Leftrightarrow Av - \lambda v = 0_m \Leftrightarrow (A - \lambda I_m)v = 0_m. \quad (10.12)$$

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

$$f : \mathbb{R}^m \rightarrow \mathbb{R}^m, x \mapsto (A - \lambda I_m)x \quad (10.13)$$

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

$$\chi_A(\lambda) = |A - \lambda I_m| = 0 \quad (10.14)$$

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

$$(A - \lambda_i^* I_m)v_i^* = 0_m \quad (10.15)$$

nach v_i^* einen Eigenvektor zum Eigenwert λ_i^* ergibt.

□

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 10.2 hier lediglich anhand eines Beispiels illustrieren.

Beispiel

Dazu sei wiederum

$$A := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad (10.16)$$

Wir wollen zunächst die Eigenwerte von A berechnen. Nach Theorem 10.2 sind dies die Nullstellen des charakteristischen Polynoms von A . Wir berechnen also zunächst das charakteristische Polynom von A durch

$$\chi_A(\lambda) = \left| \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = \left| \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} \right| = (2-\lambda)^2 - 1. \quad (10.17)$$

Mithilfe der pq-Formel zur Lösung quadratischer Gleichungen findet man dann

$$(2 - \lambda_{1/2}^*)^2 - 1 = 0 \Leftrightarrow \lambda_1^* = 3 \text{ oder } \lambda_2^* = 1. \quad (10.18)$$

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

$$(A - \lambda_i I_2)v_i = 0_2. \quad (10.19)$$

Speziell ergibt sich hier, dass für $\lambda_1 = 3$ aus

$$(A - 3I_2)v_1 = 0_2 \Leftrightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} v_{1_1} \\ v_{1_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (10.20)$$

folgt, dass

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (10.21)$$

ein Eigenvektor zum Eigenwert λ_1 ist und dass für $\lambda_2 = 1$ aus

$$(A - 1I_2)v_2 = 0_2 \Leftrightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} v_{2_1} \\ v_{2_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (10.22)$$

folgt, dass

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad (10.23)$$

ein Eigenvektor zum Eigenwert $\lambda_2 = 1$ ist. Weiterhin gelten hier offenbar

$$v_1^T v_2 = 0 \text{ und } \|v_1\| = \|v_2\| = 1. \quad (10.24)$$

Folgender **R** Code demonstriert die Bestimmung der Eigenwerte und Eigenvektoren der hier betrachteten Matrix mithilfe der `eigen()` Funktion.

```

1 # Matrixdefinition
2 A = matrix(c(2,1,
3             1,2),
4             nrow = 2,
5             byrow = TRUE)
6
7 # Eigenanalyse
8 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 diese Theoreme im Kontext der Kanonischen Korrelationsanalyse (Kapitel 15).

Theorem 10.3 (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.

```
1 A = matrix(1:6, nrow = 2, byrow = T) # Matrix A \in \mathbb{R}^{2 x 3}
2 B = matrix(1:6, ncol = 2, byrow = T) # Matrix B \in \mathbb{R}^{3 x 2}
3 EAB = eigen(A %*% B) # Eigenanalyse von AB \in \mathbb{R}^{2 \times 2}
4 EBA = eigen(B %*% A) # Eigenanalyse von BA \in \mathbb{R}^{3 \times 3}
5 w = B %*% EAB$vectors[,1] # Eigenvektor von BA
6 cat("Eigenwerte von AB :", EAB$values[1:2],
7     "\nEigenwerte von BA :", EBA$values[1:2],
8     "\nBAw mit w = Bv :", B %*% A %*% w,
9     "\nlw mit w = Bv :", EBA$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 10.4. *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.

```
1 A = matrix(1:6, nrow = 2, byrow = T) # Matrix A \in \mathbb{R}^{2 x 3}
2 B = matrix(1:8, ncol = 2, byrow = T) # Matrix B \in \mathbb{R}^{4 x 2}
3 a = matrix(1:3, nrow = 3, byrow = T) # Vektor a \in \mathbb{R}^{3 x 1}
4 b = matrix(1:4, nrow = 4, byrow = T) # Vektor b \in \mathbb{R}^{4 x 1}
5 EAabTB = eigen(A %*% a %*% t(b) %*% B) # Eigenanalyse von AabTB \in \mathbb{R}^{4 x 4}
6 cat("Eigenwerte von AabTB :", EAabTB$values,
7     "\nbTBaa :", t(b) %*% B %*% A %*% a,
8     "\nAaa :", A %*% a,
9     "\n(AabTB)Aa :", (A %*% a %*% t(b) %*% B) %*% A %*% a, # Mv
10    "\n(bTBaa)Aa :", as.vector((t(b) %*% B %*% A %*% a)) * (A %*% a)) # = \lambda v

Eigenwerte von AabTB : 2620 0
bTBaa : 2620
Aa : 14 32
(AabTB)Aa : 36680 83840
(bTBaa)Aa : 36680 83840
```

10.2. 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 10.5 (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

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \quad (10.25)$$

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

$$ac = bc. \quad (10.26)$$

Um abschließend

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \quad (10.27)$$

zu zeigen, halten wir zunächst fest, dass

$$Sq_i = \lambda_i q_i \Leftrightarrow (Sq_i)^T = (\lambda_i q_i)^T \Leftrightarrow q_i^T S^T = q_i^T \lambda_i \Leftrightarrow q_i^T S = q_i^T \lambda_i \Leftrightarrow q_i^T S q_j = \lambda_i q_i^T q_j \quad (10.28)$$

und

$$Sq_j = \lambda_j q_j \Leftrightarrow q_j^T S = q_j^T \lambda_j \Leftrightarrow q_j^T S q_i = \lambda_j q_j^T q_i \Leftrightarrow (q_j^T S q_i)^T = (\lambda_j q_j^T q_i)^T \Leftrightarrow q_i^T S q_j = \lambda_j q_i^T q_j \quad (10.29)$$

gelten. Sowohl $\lambda_i q_i^T q_j$ als auch $\lambda_j q_i^T q_j$ sind also mit $q_i^T S q_j$ und damit auch miteinander identisch.

□

Offenbar haben wir nur Aussage (2) von Theorem 10.5 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 10.5 angesprochenen orthogonalen Eigenvektoren insbesondere auch orthonormal sind. Mithilfe von Theorem 10.5 können wir nun die Orthonormalzerlegung einer symmetrischen Matrix formulieren und ihre Existenz beweisen.

Theorem 10.6 (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

$$S = Q\Lambda Q^T, \tag{10.30}$$

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

$$Q := (q_1 \quad q_2 \quad \dots \quad q_m) \in \mathbb{R}^{m \times m} \text{ und } \Lambda := \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^{m \times m}, \tag{10.31}$$

folgt dann mit den Definitionen von Eigenwerten und Eigenvektoren zunächst, dass

$$Sq_i = \lambda_i q_i \text{ für } i = 1, \dots, m \Leftrightarrow SQ = Q\Lambda. \tag{10.32}$$

Rechtseitige Multiplikation mit Q^T ergibt dann mit $QQ^T = I_m$, dass

$$SQQ^T = Q\Lambda Q^T \Leftrightarrow SI_m = Q\Lambda Q^T \Leftrightarrow S = Q\Lambda Q^T. \tag{10.33}$$

□

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

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{10.34}$$

mit den oben bestimmten Eigenwerten $\lambda_1 = 3$ und $\lambda_2 = 1$ sowie den zugehörigen orthonormalen Eigenvektoren

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{10.35}$$

seien

$$Q := (v_1 \quad v_2) \text{ und } \Lambda = \text{diag}(\lambda_1, \lambda_2). \tag{10.36}$$

Dann ergibt sich offenbar

$$\begin{aligned} Q\Lambda Q^T &= (v_1 \quad v_2) \text{diag}(\lambda_1, \lambda_2) (v_1 \quad v_2)^T \\ &= \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right) \\ &= \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 3 & -1 \\ 3 & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right) \\ &= \frac{1}{2} \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \\ &= A \end{aligned}$$

und wir haben Theorem 10.6 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 10.2 (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

$$S^{r/s} = Q\Lambda^{r/s}Q^T \text{ mit } \Lambda^{r/s} = \text{diag}(\lambda_i^{r/s}). \quad (10.37)$$

Der Spezialfall $r := 1, s := 2$ wird als *symmetrische Quadratwurzel von S* bezeichnet und hat die Form

$$S^{1/2} = Q\Lambda^{1/2}Q^T \text{ mit } \Lambda^{1/2} = \text{diag}(\lambda_i^{1/2}). \quad (10.38)$$

•

Wir halten fest, dass mit Definition 10.2 offenbar gilt, dass

$$(S^{1/2})^2 = Q\Lambda^{1/2}Q^TQ\Lambda^{1/2}Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = Q\Lambda Q^T = S. \quad (10.39)$$

Weiterhin gilt, dass

$$(S^{-1/2})^2 = Q\Lambda^{-1/2}Q^TQ\Lambda^{-1/2}Q^T = Q\Lambda^{-1/2}\Lambda^{-1/2}Q^T = Q\Lambda^{-1}Q^T = S^{-1}. \quad (10.40)$$

Schließlich gilt, dass

$$\begin{aligned} S^{-1/2}SS^{-1/2} &= Q\Lambda^{-1/2}Q^TQ\Lambda Q^TQ\Lambda^{-1/2}Q^T \\ &= Q\Lambda^{-1/2}\Lambda\Lambda^{-1/2}Q^T \\ &= Q\Lambda\Lambda^{-1}Q^T \\ &= I_m \end{aligned} \quad (10.41)$$

10.3. 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 10.3 (Singularwertzerlegung). $Y \in \mathbb{R}^{m \times n}$ sei eine Matrix. Dann heißt die Zerlegung

$$Y = USV^T, \quad (10.42)$$

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 .

•

Singulärwertzerlegungen werden auf Englisch *singular value decompositions* genannt und entsprechend mit *SVD* abgekürzt. Wir verzichten auf eine Diskussion der Berechnung einer Singulärwertzerlegung und weisen lediglich daraufhin, dass Singulärwertzerlegungen 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 Singulärwertzerlegung und Eigenanalyse und wird an vielen Stellen eingesetzt.

Theorem 10.7 (Singulärwertzerlegung und Eigenanalyse).

$Y \in \mathbb{R}^{m \times n}$ sei eine Matrix und

$$Y = USV^T \quad (10.43)$$

sei ihre Singulärwertzerlegung. Dann gilt:

- Die Spalten von U sind die Eigenvektoren von YY^T ,
- die Spalten von V sind die Eigenvektoren von Y^TY und
- die Singulärwerte sind die Quadratwurzeln der zugehörigen Eigenwerte.

◦

Beweis. Wir halten zunächst fest, dass mit

$$(YY^T)^T = YY^T \text{ und } (Y^TY)^T = Y^TY, \quad (10.44)$$

YY^T und Y^TY symmetrische Matrizen sind und somit Orthonormalzerlegungen haben. Wir halten weiterhin fest, dass mit $V^TV = I_n$, $U^TU = I_m$ und $S^T = S$ gilt, dass

$$YY^T = USV^T(USV^T)^T = USV^TVS^TUT = USSU^T =: U\Lambda U^T \quad (10.45)$$

und

$$Y^TY = (USV^T)^TUSV^T = VS^TUTUS^TV^T =: V\Lambda V^T \quad (10.46)$$

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^TY in Form der Orthonormalzerlegungen

$$YY^T = U\Lambda U^T \text{ und } Y^TY = V\Lambda V^T \quad (10.47)$$

geschrieben, wobei für die Diagonalelemente von Λ gilt, dass sie die quadrierten Werte der Diagonalelemente von S sind.

□

10.4. 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.

10.5. 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.

11. Zufallsvektoren

Zufallsvektoren sind Tupel von Zufallsvariablen. Da jede Zufallsvariable einer Wahrscheinlichkeitsverteilung unterliegt, unterliegt auch ein Zufallsvektor einer Wahrscheinlichkeitsverteilung. Da ein Zufallsvektor aus zwei oder mehr Zufallsvariablen besteht, beschreibt die Verteilung eines Zufallsvektors die *gemeinsame* Verteilung von zwei oder mehr Zufallsvariablen. In diesem Kapitel wollen wir zunächst den Begriff des Zufallsvektors und seiner assoziierten Wahrscheinlichkeitsverteilung, die oft einfach als eine *multivariate Verteilung* bezeichnet wird, einführen. Wir wollen dann diskutieren, wie die Begriffe der WMF, WDF und KVF von Zufallsvariablen auf Zufallsvektoren übertragen werden können. Mit den *marginalen* und *bedingten Verteilungen* führen wir dann nachfolgend Begriffe ein, die in der Betrachtung von Zufallsvariablen nicht auftreten. Schließlich führen wir mit dem Begriff der *unabhängigen Zufallsvariablen* das probabilistische Standardmodell für univariate Datensätze ein, das einen Spezialfall der gemeinsamen Verteilung des durch die Zufallsvariablen konstituierten Zufallsvektors darstellt.

11.1. Definition und multivariate Verteilungen

Die Konstruktion und Definition eines Zufallsvektors ist analog zu der einer Zufallsvariable, mit dem Unterschied, dass es sich bei einer Zufallsvariable um eine skalarwertige, bei einem Zufallsvektor dagegen um eine vektorwertige Abbildung auf dem Ergebnisraum eines Wahrscheinlichkeitsraums handelt.

Definition 11.1 (Zufallsvektor). $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $(\mathcal{X}, \mathcal{S})$ sei ein n -dimensionaler Messraum. Ein n -dimensionaler *Zufallsvektor* ist definiert als eine Abbildung

$$\xi : \Omega \rightarrow \mathcal{X}, \omega \mapsto \xi(\omega) := \begin{pmatrix} \xi_1(\omega) \\ \vdots \\ \xi_n(\omega) \end{pmatrix} \quad (11.1)$$

mit der *Messbarkeitseigenschaft*

$$\{\omega \in \Omega \mid \xi(\omega) \in S\} \in \mathcal{A} \text{ für alle } S \in \mathcal{S}. \quad (11.2)$$

•

Das Standardbeispiel für den Ergebnisraum eines Zufallsvektors ist \mathbb{R}^n , das Standardbeispiel für die auf ihm definierte σ -Algebra ist die n -dimensionale Borelsche σ -Algebra $\mathcal{B}(\mathbb{R}^n)$. Für eine explizite und formale Einführung der n -dimensionalen Borelschen σ -Algebra verweisen wir auf die weiterführende Literatur, beispielsweise Schmidt (2009). Wir begnügen uns hier wieder mit der (weiterhin formal falschen) Intuition der n -dimensionale

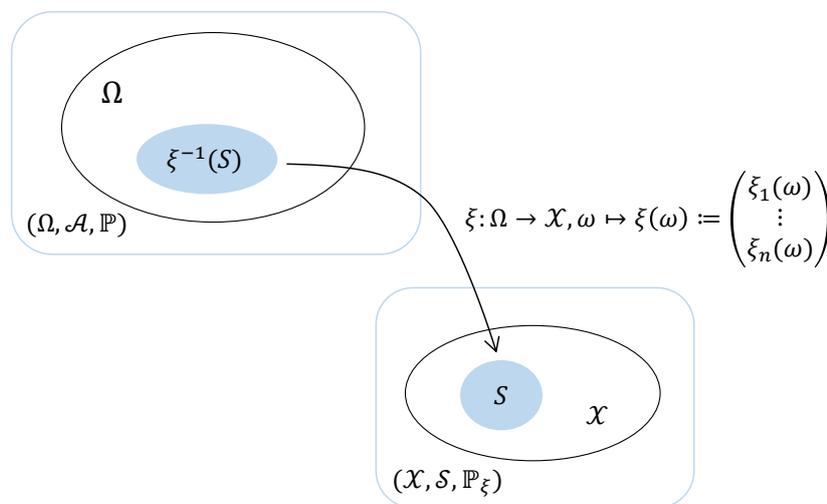
Borelschen σ -Algebra als Menge aller Teilmengen des \mathbb{R}^n . Das Standardbeispiel für einen n -dimensionalen Messraum ist damit $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Wie bei allen vektorwertigen Funktionen nennen wir die den Zufallsvektor konstituierenden Funktionen ξ_i die *Komponentenfunktionen* von ξ . Legen wir den n -dimensionalen Messraum $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ dem Zufallsvektor zugrunde, so haben diese die Form

$$\xi_i : \Omega \rightarrow \mathbb{R}, \omega \mapsto \xi_i(\omega). \tag{11.3}$$

Ohne Beweis halten wir fest, dass der Zufallsvektor ξ messbar ist, wenn für alle $i = 1, \dots, n$ die Funktionen ξ_i messbar sind und umgekehrt. Damit sind die Komponentenfunktionen eines Zufallsvektors letztlich nach Definition Zufallsvariablen. Ein n -dimensionaler Zufallsvektor wird also als eine Konkatenation von n Zufallsvariablen betrachtet, für $n := 1$ entspricht ein Zufallsvektor einer Zufallsvariable.

Zufallsvektoren werden manchmal auch als “multivariate Zufallsvariablen” bezeichnet. Tatsächlich stehen bei der Betrachtung von Zufallsvektoren auch zunächst primär wahrscheinlichkeitstheoretische Aspekte und nicht etwa Aspekte der geometrischen Vektorraumtheorie im Vordergrund. Die Betrachtung von Vektorraumstrukturen ist im Kontext probabilistischer Standardmodelle wie dem Allgemeinen Linearen Modell aber durchaus üblich, so dass wir hier den Begriff des Zufallsvektors bevorzugen (vgl. Christensen (2011)). Trotzdem werden wir Zufallsvektoren, wie in vielen Texten der Wahrscheinlichkeitstheorie und Inferenzstatistik üblich, auch oft in Zeilenform, also etwa als $\xi := (\xi_1, \dots, \xi_n)$, notieren.



$$\mathbb{P}(\xi^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) =: \mathbb{P}_\xi(S)$$

Abbildung 11.1. Konstruktion von Zufallsvektor und multivariater Verteilung.

Das durch die Konstruktion eines Zufallsvektors definierte Bildmaß heißt die *multivariate Verteilung des Zufallsvektors*, wie in folgender Definition ausgeführt (Abbildung 11.1).

Definition 11.2 (Multivariate Verteilung). $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ sei ein n -dimensionaler Messraum und

$$\xi : \Omega \rightarrow \mathcal{X}, \omega \mapsto \xi(\omega) \tag{11.4}$$

sei ein Zufallsvektor. Dann heißt das Wahrscheinlichkeitsmaß \mathbb{P}_ξ , definiert durch

$$\mathbb{P}_\xi : \mathcal{S} \rightarrow [0, 1], S \mapsto \mathbb{P}_\xi(S) := \mathbb{P}(\xi^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) \quad (11.5)$$

die *multivariate Verteilung des Zufallsvektor* ξ .

•

Der Einfachheit halber spricht man oft auch nur von *der Verteilung des Zufallsvektors* ξ oder einer *multivariaten Verteilung*. Die Notationskonventionen für Zufallsvariablen gelten für Zufallsvektoren analog. Zum Beispiel gelten also

$$\begin{aligned} \mathbb{P}_\xi(\xi \in S) &:= \mathbb{P}(\{\xi \in S\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) \\ \mathbb{P}_\xi(\xi = x) &:= \mathbb{P}(\{\xi = x\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) = x\}) \\ \mathbb{P}_\xi(\xi \leq x) &:= \mathbb{P}(\{\xi \leq x\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \leq x\}) \\ \mathbb{P}_\xi(x_1 \leq \xi \leq x_2) &:= \mathbb{P}(\{x_1 \leq \xi \leq x_2\}) = \mathbb{P}(\{\omega \in \Omega \mid x_1 \leq \xi(\omega) \leq x_2\}) \end{aligned} \quad (11.6)$$

wobei die Relationsoperatoren $<, \leq, >, \geq$ hier *komponentenweise* verstanden werden. So heißt beispielsweise $x \leq y$ für $x, y \in \mathbb{R}^n$, dass für alle Komponenten $x_i, y_i, i = 1, \dots, n$ gilt, dass $x_i \leq y_i$. Eben dieser Konvention folgt auch die Definition der *multivariaten kumulativen Verteilungsfunktion* in Generalisierung des Begriffs der kumulativen Verteilungsfunktion einer Zufallsvariable.

Definition 11.3 (Multivariate kumulative Verteilungsfunktionen). ξ sei ein Zufallsvektor mit Ergebnisraum \mathcal{X} . Dann heißt eine Funktion der Form

$$P_\xi : \mathcal{X} \rightarrow [0, 1], x \mapsto P_\xi(x) := \mathbb{P}_\xi(\xi \leq x) \quad (11.7)$$

multivariate kumulative Verteilungsfunktion von ξ .

•

Wie kumulative Verteilungsfunktionen können auch multivariate kumulative Verteilungsfunktionen zur Definition von multivariaten Verteilungen genutzt werden. Häufiger ist allerdings, wie im univariaten Fall, die Definition multivariater Verteilungen durch *multivariate Wahrscheinlichkeitsmassefunktionen* oder *multivariate Wahrscheinlichkeitsdichtefunktionen*. Wir generalisieren deshalb die Definitionen diskreter und kontinuierlicher Zufallsvariablen und ihren assoziierten Wahrscheinlichkeitsmassefunktionen und Wahrscheinlichkeitsdichtefunktionen wie folgt.

Definition 11.4 (Diskreter Zufallsvektor und multivariate Wahrscheinlichkeitsmassefunktion). Ein Zufallsvektor ξ heißt *diskret*, wenn sein Ergebnisraum \mathcal{X} endlich oder abzählbar ist und eine Funktion

$$p_\xi : \mathcal{X} \rightarrow [0, 1], x \mapsto p_\xi(x) \quad (11.8)$$

existiert, für die gilt

- (1) $\sum_{x \in \mathcal{X}} p(x) = 1$ und
- (2) $\mathbb{P}_\xi(\xi = x) = p(x)$ für alle $x \in \mathcal{X}$.

Ein entsprechende Funktion p_ξ heißt *multivariate Wahrscheinlichkeitsmassefunktion (WMF)* von ξ .

•

Der Begriff der multivariaten WMF ist offenbar direkt analog zum Begriff der WMF. Wie univariate WMFen sind multivariate WMFen nicht-negativ und normiert. Der Einfachheit halber spricht man oft einfach von *der WMF eines Zufallsvektors* und verzichtet bei ihrer Bezeichnung, wenn der betreffende Zufallsvektor aus dem Kontext klar ist, auf das ξ Subskript, schreibt also oft einfach p anstatt von p_ξ .

Beispiel

Zur Illustration des Begriffs des diskreten Zufallsvektors und seiner WMF wollen wir ein Beispiel betrachten. Dazu sei $\xi := (\xi_1, \xi_2)$ ein Zufallsvektor, der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annehme, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien. Dann entspricht der Ergebnisraum von ξ der in untenstehender Tabelle spezifizierten Menge an Tupeln (x_1, x_2) :

(x_1, x_2)	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$x_1 = 1$	(1, 1)	(1, 2)	(1, 3)	(1, 4)
$x_1 = 2$	(2, 1)	(2, 2)	(2, 3)	(2, 4)
$x_1 = 3$	(3, 1)	(3, 2)	(3, 3)	(3, 4)

Eine exemplarische bivariate WMF der Form

$$p_\xi : \{1, 2, 3\} \times \{1, 2, 3, 4\} \rightarrow [0, 1], (x_1, x_2) \mapsto p_\xi(x_1, x_2) \quad (11.9)$$

ist dann durch nachfolgende Tabelle definiert:

$p_\xi(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$x_1 = 1$	0.1	0.0	0.2	0.1
$x_1 = 2$	0.1	0.2	0.0	0.0
$x_1 = 3$	0.0	0.1	0.1	0.1

Man beachte, dass die so spezifizierte Funktion p_ξ den Normiertheits- und Nichtnegativitätsansprüchen an eine WMF genügt. Insbesondere gilt hier

$$\sum_{x \in \mathcal{X}} p_\xi(x) = \sum_{x_1=1}^3 \sum_{x_2=1}^4 p_\xi(x_1, x_2) = 1. \quad (11.10)$$

Den Begriff des kontinuierlichen Zufallsvektors und der multivariaten Wahrscheinlichkeitsdichtefunktion definieren wir wie folgt.

Definition 11.5 (Kontinuierlicher Zufallsvektor und multivariate Wahrscheinlichkeitsdichtefunktion). Ein Zufallsvektor ξ heißt *kontinuierlich*, wenn sein Ergebnisraum durch \mathbb{R}^n gegeben ist und eine Funktion

$$p_\xi : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p_\xi(x), \quad (11.11)$$

existiert, für die gilt, dass

- (1) $\int_{\mathbb{R}^n} p_\xi(x) dx = 1$ und
 (2) $\mathbb{P}_\xi(x_1 \leq \xi \leq x_2) = \int_{x_{1_1}}^{x_{2_1}} \cdots \int_{x_{1_n}}^{x_{2_n}} p_\xi(s_1, \dots, s_n) ds_1 \cdots ds_n$.

Eine entsprechende Funktion p_ξ heißt *multivariate Wahrscheinlichkeitsdichtefunktion (WDF)* von ξ .

•

Offenbar ist der der Begriff der multivariaten WDF eines kontinuierlichen Zufallsvektors analog zum Begriff der WDF einer kontinuierlichen Zufallsvariable und wie univariate WDFen sind multivariate WDFen nicht-negativ und normiert. Der Einfachheit halber spricht man auch hier oft einfach von *multivariaten WDFen* und verzichtet auf die den Zufallsvektor identifizieren Subskripte. Wie für kontinuierliche Zufallsvariablen gilt für kontinuierliche Zufallsvektoren

$$\mathbb{P}_\xi(\xi = x) = \mathbb{P}_\xi(x \leq \xi \leq x) = \int_{x_1}^{x_1} \cdots \int_{x_n}^{x_n} p_\xi(s_1, \dots, s_n) ds_1 \cdots ds_n = 0. \quad (11.12)$$

Das Standardbeispiel für eine multivariate WDF ist die *multivariate Normalverteilung*, welcher wir uns ausführlich in Kapitel 11.6 widmen.

11.2. Erwartungswert, Kovarianzmatrix und Korrelationsmatrix

Wie im univariaten Fall ist man manchmal darum bemüht, die Verteilung eines Zufallsvektors mit einigen wenigen Maßzahlen zu charakterisieren. Die multivariaten Analoga der Begriffe des Erwartungswerts und der Varianz von Zufallsvariablen sind der *Erwartungswert* und die *Kovarianzmatrix* eines Zufallsvektors. Letzterer Begriff impliziert, dass man zur Charakterisierung multivariater Verteilungen nicht nur an der Varianz der einzelnen Komponenten untereinander, sondern auch an ihrer paarweisen Kovarianz interessiert sein kann.

Definition 11.6 (Erwartungswert und Kovarianzmatrix von Zufallsvektoren). ξ sei ein n -dimensionaler Zufallsvektor. Dann ist der *Erwartungswert* von ξ definiert als der n -dimensionale reelle Vektor

$$\mathbb{E}(\xi) := \begin{pmatrix} \mathbb{E}(\xi_1) \\ \vdots \\ \mathbb{E}(\xi_n) \end{pmatrix} \quad (11.13)$$

und die *Kovarianzmatrix* von ξ ist definiert als die $n \times n$ Matrix

$$\mathbb{C}(\xi) := \mathbb{E}((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T) \quad (11.14)$$

•

Der Erwartungswert eines Zufallsvektors ξ ist also der Vektor der Erwartungswerte der Komponenten ξ_1, \dots, ξ_n von ξ , ist also direkt im Sinne von Erwartungswerten von Zufallsvariablen definiert. Die Kovarianzmatrix ist in Definition 11.6 formal analog zur Kovarianz zweier Zufallsvariablen definiert. Eine direkte Rückführung des Begriffs der Kovarianzmatrix eines Zufallsvektors auf den Begriff aus dem univariaten Kontext bekannten Begriff der Kovarianz zweier Zufallsvariablen erlaubt folgendes Theorem.

Theorem 11.1 (Kovarianzmatrix eines Zufallsvektors). ξ sei ein n -dimensionaler Zufallsvektor und $\mathbb{C}(\xi)$ sei seine Kovarianzmatrix. Dann gilt

$$\mathbb{C}(\xi) = (\mathbb{C}(\xi_i, \xi_j))_{1 \leq i, j \leq n} = \begin{pmatrix} \mathbb{C}(\xi_1, \xi_1) & \mathbb{C}(\xi_1, \xi_2) & \cdots & \mathbb{C}(\xi_1, \xi_n) \\ \mathbb{C}(\xi_2, \xi_1) & \mathbb{C}(\xi_2, \xi_2) & \cdots & \mathbb{C}(\xi_2, \xi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}(\xi_n, \xi_1) & \mathbb{C}(\xi_n, \xi_2) & \cdots & \mathbb{C}(\xi_n, \xi_n) \end{pmatrix}. \quad (11.15)$$

◦

Beweis. Es gilt

$$\mathbb{C}(\xi) := \mathbb{E}((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T) \quad (11.16)$$

$$= \mathbb{E} \left(\left(\begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} - \begin{pmatrix} \mathbb{E}(\xi_1) \\ \vdots \\ \mathbb{E}(\xi_n) \end{pmatrix} \right) \left(\begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} - \begin{pmatrix} \mathbb{E}(\xi_1) \\ \vdots \\ \mathbb{E}(\xi_n) \end{pmatrix} \right)^T \right) \quad (11.17)$$

$$= \mathbb{E} \left(\begin{pmatrix} \xi_1 - \mathbb{E}(\xi_1) \\ \vdots \\ \xi_n - \mathbb{E}(\xi_n) \end{pmatrix} \begin{pmatrix} \xi_1 - \mathbb{E}(\xi_1) \\ \vdots \\ \xi_n - \mathbb{E}(\xi_n) \end{pmatrix}^T \right) \quad (11.18)$$

$$= \mathbb{E} \left(\begin{pmatrix} \xi_1 - \mathbb{E}(\xi_1) \\ \vdots \\ \xi_n - \mathbb{E}(\xi_n) \end{pmatrix} (\xi_1 - \mathbb{E}(\xi_1) \quad \cdots \quad \xi_n - \mathbb{E}(\xi_n)) \right) \quad (11.19)$$

$$= \mathbb{E} \begin{pmatrix} (\xi_1 - \mathbb{E}(\xi_1))(\xi_1 - \mathbb{E}(\xi_1)) & \cdots & (\xi_1 - \mathbb{E}(\xi_1))(\xi_n - \mathbb{E}(\xi_n)) \\ \vdots & \ddots & \vdots \\ (\xi_n - \mathbb{E}(\xi_n))(\xi_1 - \mathbb{E}(\xi_1)) & \cdots & (\xi_n - \mathbb{E}(\xi_n))(\xi_n - \mathbb{E}(\xi_n)) \end{pmatrix} \quad (11.20)$$

$$= (\mathbb{E}((\xi_i - \mathbb{E}(\xi_i))(\xi_j - \mathbb{E}(\xi_j))))_{1 \leq i, j \leq n} \quad (11.21)$$

$$= (\mathbb{C}(\xi_i, \xi_j))_{1 \leq i, j \leq n}. \quad (11.22)$$

$$(11.23)$$

□

Die Kovarianzmatrix eines Zufallsvektors ξ ist also die Matrix der Kovarianzen der Komponenten von ξ . Damit ist auch die Kovarianzmatrix direkt im Sinne des Begriffs der Kovarianz von Zufallsvektoren gegeben. Da die Kovarianz einer Zufallsvariable mit sich selbst bekanntlich ihre Varianz ist, enthält die Kovarianzmatrix auf ihrer Diagonalen die Varianzen der Komponenten von ξ .

Folgendes Theorem dokumentiert eine Schreibweise für die Kovarianzmatrix eines partitionierten Zufallsvektors im Sinne von Erwartungswerten von Zufallsvektorprodukten an, die zum Beispiel im Rahmen der Kanonischen Korrelationsanalyse hilfreich ist (vgl. Kapitel 15).

Theorem 11.2 (Kovarianzmatrizen von Zufallsvektoren). *Es seien*

$$\zeta = \begin{pmatrix} \xi \\ \nu \end{pmatrix} \text{ mit } \mathbb{E}(\zeta) := 0_m \quad (11.24)$$

ein $m_\xi + m_\nu$ -dimensionaler Zufallsvektor und sein Erwartungswertvektor, respektive. Dann kann die $m \times m$ Kovarianzmatrix von ζ geschrieben werden als

$$\mathbb{C}(\zeta) = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\nu} \\ \Sigma_{\nu\xi} & \Sigma_{\nu\nu} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (11.25)$$

wobei

$$\begin{aligned}
 \Sigma_{\xi\xi} &:= \mathbb{E}(\xi\xi^T) \in \mathbb{R}^{m_\xi \times m_\xi} \\
 \Sigma_{\xi v} &:= \mathbb{E}(\xi v^T) \in \mathbb{R}^{m_\xi \times m_v} \\
 \Sigma_{v\xi} &:= \mathbb{E}(v\xi^T) \in \mathbb{R}^{m_v \times m_\xi} \\
 \Sigma_{vv} &:= \mathbb{E}(vv^T) \in \mathbb{R}^{m_v \times m_v}
 \end{aligned} \tag{11.26}$$

◦

Beweis. Nach Definition der Kovarianzmatrix eines Zufallsvektors gilt

$$\begin{aligned}
 \mathbb{C}(z) &= \mathbb{E}((\zeta - \mathbb{E}(\zeta))(\zeta - \mathbb{E}(\zeta))^T) \\
 &= \mathbb{E}((\zeta - \mathbf{0}_m)(\zeta - \mathbf{0}_m)^T) \\
 &= \mathbb{E}(\zeta\zeta^T) \\
 &= \mathbb{E}\left(\begin{pmatrix} \xi \\ v \end{pmatrix} (\xi^T \quad v^T)\right) \\
 &= \mathbb{E}\left(\begin{pmatrix} \xi\xi^T & \xi v^T \\ v\xi^T & vv^T \end{pmatrix}\right) \\
 &= \begin{pmatrix} \mathbb{E}(\xi\xi^T) & \mathbb{E}(\xi v^T) \\ \mathbb{E}(v\xi^T) & \mathbb{E}(vv^T) \end{pmatrix} \\
 &= \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi v} \\ \Sigma_{v\xi} & \Sigma_{vv} \end{pmatrix}
 \end{aligned} \tag{11.27}$$

□

Schließlich ist man in manchen Anwendungen an einer normalisierten, maßstabsunabhängigen Repräsentation der Kovarianzen eines Zufallsvektors interessiert. Wie im univariaten Fall bietet sich hierfür die Normalisierung der Kovarianz zweier Zufallsvariablen mithilfe ihrer jeweiligen Varianzen im Sinne einer Korrelation an. Diese Überlegung führt auf den Begriff der *Korrelationsmatrix* eines Zufallsvektors.

Definition 11.7 (Korrelationsmatrix). ξ sei ein n -dimensionaler Zufallsvektor. Dann ist die *Korrelationsmatrix* von ξ definiert als die $n \times n$ Matrix

$$\mathbb{R}(\xi) := (\rho_{ij})_{1 \leq i, j \leq n} = \left(\frac{\mathbb{C}(\xi_i, \xi_j)}{\sqrt{\mathbb{V}(\xi_i)}\sqrt{\mathbb{V}(\xi_j)}} \right)_{1 \leq i, j \leq n}. \tag{11.28}$$

•

Da es sich bei den Varianzen der Komponenten von ξ um die Diagonalelemente der Kovarianzmatrix von ξ handelt, ist die Korrelationsmatrix natürlich in der Kovarianzmatrix implizit. Weiterhin gelten, wie immer für Korrelationen, für die Einträge ρ_{ij} , $1 \leq i, j \leq n$ der Korrelationsmatrix, dass

$$\rho_{ij} \in [-1, 1] \text{ für } 1 \leq i, j \in n \text{ und } \rho_{ii} = 1 \text{ für } 1 \leq i \leq n. \tag{11.29}$$

11.3. Marginalverteilungen

Hat man die Verteilung eines Zufallsvektors spezifiziert, so kann man sich fragen, welche Verteilungen daraus für die einzelnen Komponenten des Zufallsvektors, also die Zufallsvariablen, die den Zufallsvektor konstituieren, resultieren. Man nennt diese Verteilungen die *univariaten Marginalverteilungen* des Zufallsvektors. Folgende Definition ist grundlegend.

Definition 11.8 (Univariate Marginalverteilung). $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ sei ein n -dimensionaler Messraum, $\xi : \Omega \rightarrow \mathcal{X}$ sei ein Zufallsvektor, \mathbb{P}_ξ sei die Verteilung von ξ , $\mathcal{X}_i \subset \mathcal{X}$ sei der Ergebnisraum der i ten Komponente ξ_i von ξ , und \mathcal{S}_i sei eine σ -Algebra auf \mathcal{X}_i . Dann heißt die durch

$$\mathbb{P}_{\xi_i} : \mathcal{S}_i \rightarrow [0, 1], S \mapsto \mathbb{P}_\xi(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{i-1} \times S \times \mathcal{X}_{i+1} \times \cdots \times \mathcal{X}_n) \text{ für } S \in \mathcal{S}_i \quad (11.30)$$

definierte Verteilung die *ite univariate Marginalverteilung* von ξ .

•

Konkret kann man sowohl für diskrete als auch für kontinuierliche Zufallsvektoren die WMFen bzw. WDFen ihrer Komponenten direkt aus der entsprechenden multivariaten WMF bzw. WDF bestimmen. Dies ist die Aussage folgenden Theorems.

Theorem 11.3 (Marginale Wahrscheinlichkeitsmassenfunktionen und marginale Wahrscheinlichkeitsdichtefunktionen). (1) $\xi = (\xi_1, \dots, \xi_n)$ sei ein n -dimensionaler diskreter Zufallsvektor mit Wahrscheinlichkeitsmassenfunktion p_ξ und Komponentenergebnisräumen $\mathcal{X}_1, \dots, \mathcal{X}_n$. Dann ergibt sich die Wahrscheinlichkeitsmassenfunktion der i ten Komponente ξ_i von ξ als

$$p_{\xi_i} : \mathcal{X}_i \rightarrow [0, 1], x_i \mapsto p_{\xi_i}(x_i) := \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p_\xi(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n). \quad (11.31)$$

(2) $\xi = (\xi_1, \dots, \xi_n)$ sei ein n -dimensionaler kontinuierlicher Zufallsvektor mit Wahrscheinlichkeitsdichtefunktion p_ξ und Komponentenergebnisraum \mathbb{R} . Dann ergibt sich die Wahrscheinlichkeitsdichtefunktion der i ten Komponente ξ_i von ξ als

$$p_{\xi_i} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_i \mapsto p_{\xi_i}(x_i) := \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_n} p_\xi(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n. \quad (11.32)$$

◦

Die WMFen der univariaten Marginalverteilungen diskreter Zufallsvektoren ergeben sich also durch Summation über alle Werte der zu der jeweils betrachteten Zufallsvariable komplementären Zufallsvariablen und die WDFen der univariaten Marginalverteilungen kontinuierlicher Zufallsvektoren ergeben sich analog durch Integration über alle Werte der zu der jeweils betrachteten Zufallsvariable komplementären Zufallsvariablen. Für einen Beweis von Theorem 11.3 verweisen wir auf die weiterführende Literatur.

Beispiel

In Fortführung des in Kapitel 11.1 betrachteten Beispiels eines zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ ergeben sich für die dort definierte WMF für die marginalen WMFen p_{ξ_1} und p_{ξ_2} anhand von

$$p_{\xi_1}(x_1) = \sum_{x_2=1}^4 p_{\xi}(x_1, x_2) \text{ und } p_{\xi_2}(x_2) = \sum_{x_1=1}^3 p_{\xi}(x_1, x_2) \quad (11.33)$$

zu den an den Rändern der unten spezifizierter Tabelle aufgelisteten WMFen:

$p_{\xi}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{\xi_1}(x_1)$
$x_1 = 1$	0.1	0.0	0.2	0.1	0.4
$x_1 = 2$	0.1	0.2	0.0	0.0	0.3
$x_1 = 3$	0.0	0.1	0.1	0.1	0.3
$p_{\xi_2}(x_2)$	0.2	0.3	0.3	0.2	

Für die Werte von p_{ξ_1} werden die entsprechenden Werte von p_{ξ} also zeilenweise und für die Werte von p_{ξ_2} spaltenweise addiert. Man beachte, dass aus der Normiertheit von p_{ξ} die Normiertheit von p_{ξ_1} und p_{ξ_2} direkt folgt, da sich die Gesamtsumme an Wahrscheinlichkeitsmasse nicht ändert,

$$1 = \sum_{x_1=1}^3 \sum_{x_2=1}^4 p_{\xi}(x_1, x_2) = \sum_{x_1=1}^3 p_{\xi_1}(x_1) = \sum_{x_2=1}^4 p_{\xi_2}(x_2). \quad (11.34)$$

Ein Realisierungsbeispiel mithilfe relativer Häufigkeiten mag den Begriff der marginalen WMF intuitiv verdeutlichen. Nehmen wir an, wir hätten $n = 100$ unabhängige Realisierungen von ξ vorliegen. Um die Wahrscheinlichkeiten $p_{\xi}(x_1, x_2)$ zu schätzen, würden wir die Anzahl der Realisierungen von (x_1, x_2) zählen und durch n teilen. Hätten wir beispielsweise 12 Realisierungen von $(3, 2)$ vorliegen, so würden wir $p_{\xi}(3, 2) \approx 12/100 = 0.12$ schätzen. Die Frage nach der marginalen Wahrscheinlichkeit von $x_2 = 2$ entspräche dann der Frage, wie oft unter den Realisierungen solche zu finden sind, bei denen $x_2 = 2$ ist, irrespektive des Wertes von x_1 . Dies wäre gerade die Anzahl der Realisierungen der Form $(1, 2)$, $(2, 2)$ und $(3, 2)$. Gäbe es von diesen beispielsweise 0, 22 und 12 respektive, so würde man die Wahrscheinlichkeit $p_{\xi_2}(2)$ natürlicherweise durch

$$\frac{0 + 22 + 12}{100} = \frac{0}{100} + \frac{22}{100} + \frac{12}{100} = 0.00 + 0.22 + 0.12 = 0.34 \quad (11.35)$$

schätzen. Anstelle der Wahrscheinlichkeiten $p_{\xi}(1, 2)$, $p_{\xi}(2, 2)$, $p_{\xi}(3, 2)$ addiert man hier also die entsprechenden relativen Häufigkeiten.

Beispiele für marginale Verteilungen von kontinuierlichen Zufallsvektoren behandeln wir am Beispiel der multivariaten Normalverteilung in Kapitel 11.6.

11.4. Bedingte Verteilungen

Hat man die Verteilung eines Zufallsvektors spezifiziert, so kann man sich fragen, welche Verteilung daraus für eine einzelne Komponenten des Zufallsvektors folgt, wenn man den Wert einer anderen Komponente als bekannt annimmt. Dies führt auf den Begriff der *bedingten Verteilung*, welcher sich natürlicherweise aus dem Begriff der bedingten Wahrscheinlichkeit ergibt. Wir erinnern uns zunächst, dass für einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und zwei Ereignisse $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ die bedingte Wahrscheinlichkeit von A gegeben B definiert ist als

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (11.36)$$

Analog wird für zwei Zufallsvariablen ξ_1, ξ_2 mit Ereignisräumen $\mathcal{X}_1, \mathcal{X}_2$ und (messbaren) Mengen $S_1 \in \mathcal{X}_1, S_2 \in \mathcal{X}_2$ die bedingte Verteilung von ξ_1 gegeben ξ_2 mithilfe der Ereignisse

$$A := \{\xi_1 \in S_1\} \text{ und } B := \{\xi_2 \in S_2\} \quad (11.37)$$

definiert. So ergibt sich zum Beispiel die bedingte Wahrscheinlichkeit, dass $\xi_1 \in S_1$ gegeben, dass $\xi_2 \in S_2$ unter der Annahme, dass $\mathbb{P}(\{\xi_2 \in S_2\}) > 0$, zu

$$\mathbb{P}(\{\xi_1 \in S_1\}|\{\xi_2 \in S_2\}) = \frac{\mathbb{P}(\{\xi_1 \in S_1\} \cap \{\xi_2 \in S_2\})}{\mathbb{P}(\{\xi_2 \in S_2\})}. \quad (11.38)$$

Wir betrachten zunächst die Definition der bedingten Verteilungen von diskreten Zufallsvektoren, die lediglich aus zwei Zufallsvariablen bestehen.

Definition 11.9 (Bedingte Wahrscheinlichkeitsmassefunktion und diskrete bedingte Verteilung). $\xi := (\xi_1, \xi_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$, WMF $p_\xi = p_{\xi_1, \xi_2}$ und marginalen WMFen p_{ξ_1} und p_{ξ_2} . Die bedingte WMF von ξ_1 gegeben $\xi_2 = x_2$ ist dann für $p_{\xi_2}(x_2) > 0$ definiert als

$$p_{\xi_1|\xi_2=x_2} : \mathcal{X}_1 \rightarrow [0, 1], x_1 \mapsto p_{\xi_1|\xi_2=x_2}(x_1|x_2) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} \quad (11.39)$$

Analog ist für $p_{\xi_1}(x_1) > 0$ die bedingte WMF von ξ_2 gegeben $\xi_1 = x_1$ definiert als

$$p_{\xi_2|\xi_1=x_1} : \mathcal{X}_2 \rightarrow [0, 1], x_2 \mapsto p_{\xi_2|\xi_1=x_1}(x_2|x_1) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_1}(x_1)} \quad (11.40)$$

Die bedingten Verteilungen mit WMFen $p_{\xi_1|\xi_2=x_2}$ und $p_{\xi_2|\xi_1=x_1}$ heißen dann die *diskreten bedingten Verteilungen von ξ_1 gegeben $\xi_2 = x_2$ und ξ_2 gegeben $\xi_1 = x_1$* , respektive. •

In Analogie zur Definition der bedingten Wahrscheinlichkeit von Ereignissen gilt also

$$p_{\xi_1|\xi_2}(x_1|x_2) = \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} = \frac{\mathbb{P}(\{\xi_1 = x_1\} \cap \{\xi_2 = x_2\})}{\mathbb{P}(\{\xi_2 = x_2\})}. \quad (11.41)$$

Es ist dabei entscheidend zu erkennen, dass bedingte Verteilungen lediglich normalisierte gemeinsame Verteilungen sind.

Beispiel

In Fortführung des in Kapitel 11.1 betrachteten Beispiels eines zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ und seiner in Kapitel 11.3 bestimmten Marginalverteilungen ergeben sich folgende bedingte WMFen für $p_{\xi_2|\xi_1=x_1}$:

$p_{\xi_2 \xi_1}(x_2 x_1)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{\xi_2 \xi_1=1}(x_2 x_1 = 1)$	$\frac{0.1}{0.4} = 0.25$	$\frac{0.0}{0.4} = 0.00$	$\frac{0.2}{0.4} = 0.50$	$\frac{0.1}{0.4} = 0.25$
$p_{\xi_2 \xi_1=2}(x_2 x_1 = 2)$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.2}{0.3} = 0.6\bar{6}$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.0}{0.3} = 0.00$
$p_{\xi_2 \xi_1=3}(x_2 x_1 = 3)$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$

Man beachte, dass zum einen gilt, dass

$$\sum_{x_2=1}^4 p_{\xi_2|\xi_1=x_1}(x_2|x_1) = 1 \text{ für alle } x_1 \in \mathcal{X}_1, \tag{11.42}$$

die bedingten WMFen sind also normiert. Zum anderen beachte man die qualitative Ähnlichkeit der WMFen $p_{\xi_1, \xi_2}(x_1, x_2)$ und $p_{\xi_2|\xi_1}(x_2|x_1)$, die sich einfach daraus ergibt, dass $p_{\xi_1, \xi_2}(x_1, x_2)$ und $p_{\xi_2|\xi_1}(x_2|x_1)$ für alle $x_1 \in \mathcal{X}_1$ bis auf den gemeinsamen Skalierungsfaktor $1/p_{\xi_1}(x_1)$ identisch sind. Im Fall eines kontinuierlichen Zufallsvektors sind die analogen bedingten WDFen definiert wie folgt.

Definition 11.10 (Bedingte Wahrscheinlichkeitsdichtefunktion und kontinuierliche bedingte Verteilung). $\xi := (\xi_1, \xi_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum \mathbb{R}^2 , WDF $p_\xi = p_{\xi_1, \xi_2}$ und marginalen WDFen p_{ξ_1} und p_{ξ_2} . Die bedingte WDF von ξ_1 gegeben $\xi_2 = x_2$ ist dann für $p_{\xi_2}(x_2) > 0$ definiert als

$$p_{\xi_1|\xi_2=x_2} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_1 \mapsto p_{\xi_1|\xi_2=x_2}(x_1|x_2) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} \tag{11.43}$$

Analog ist für $p_{\xi_1}(x_1) > 0$ die bedingte WMF von ξ_2 gegeben $\xi_1 = x_1$ definiert als

$$p_{\xi_2|\xi_1=x_1} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_2 \mapsto p_{\xi_2|\xi_1=x_1}(x_2|x_1) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_1}(x_1)} \tag{11.44}$$

Die Verteilungen mit WDFen $p_{\xi_1|\xi_2=x_2}$ und $p_{\xi_2|\xi_1=x_1}$ heißen dann die *kontinuierlichen bedingten Verteilungen von ξ_1 gegeben $\xi_2 = x_2$ und ξ_2 gegeben $\xi_1 = x_1$* , respektive.

•

Man beachte, dass im kontinuierlichen Fall zwar $\mathbb{P}(\xi = x) = 0$, aber nicht notwendig auch $p_\xi(x) = 0$ gilt. Die bedingten Verteilungen multivariater Normalverteilungen diskutieren wir in Kapitel 11.6.

11.5. Unabhängige Zufallsvariablen

Ähnlich wie die bedingte Wahrscheinlichkeiten von Ereignissen lässt sich auch das Konzept der unabhängigen Ereignisse auf Zufallsvektoren übertragen. Wir definieren zunächst den Begriff der unabhängigen Zufallsvariablen.

Definition 11.11 (Unabhängige Zufallsvariablen). $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $\xi := (\xi_1, \xi_2)$ ein zweidimensionaler Zufallsvektor. Die Zufallsvariablen ξ_1, ξ_2 mit Ergebnisräumen $\mathcal{X}_1, \mathcal{X}_2$ heißen *unabhängig*, wenn für alle $S_1 \subseteq \mathcal{X}_1$ und $S_2 \subseteq \mathcal{X}_2$ gilt, dass

$$\mathbb{P}_\xi(\xi_1 \in S_1, \xi_2 \in S_2) = \mathbb{P}_{\xi_1}(\xi_1 \in S_1)\mathbb{P}_{\xi_2}(\xi_2 \in S_2). \quad (11.45)$$

•

Definition 11.11 besagt, dass die Ereignisse $\{\xi_1 \in S_1\}$ und $\{\xi_2 \in S_2\}$ unabhängig sind. Es gilt also auch, dass

$$\mathbb{P}(\{\xi_1 \in S_1\} | \{\xi_2 \in S_2\}) = \mathbb{P}(\{\xi_1 \in S_1\}) \quad (11.46)$$

und das Wissen um das Eintreten des Ereignisses $\{\xi_2 \in S_2\}$ verändert die Wahrscheinlichkeit des Ereignisses $\{\xi_1 \in S_1\}$ nicht. Das Faktorisierungsprinzip zur Modellierung probabilistischer Unabhängigkeit überträgt sich auf WMFen und WDFen von Zufallsvektoren. Dies ist die Aussagen folgenden Theorems.

Theorem 11.4 (Unabhängigkeit und Faktorisierung).

(1) $\xi := (\xi_1, \xi_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X}_1 \times \mathcal{X}_2$, WMF p_ξ und marginalen WMFen p_{ξ_1}, p_{ξ_2} . Dann gilt

ξ_1 und ξ_2 sind unabhängige Zufallsvariablen \Leftrightarrow

$$p_\xi(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2. \quad (11.47)$$

(2) $\xi := (\xi_1, \xi_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum \mathbb{R}^2 , WDF p_ξ und marginalen WDFen p_{ξ_1}, p_{ξ_2} . Dann gilt

ξ_1 und ξ_2 sind unabhängige Zufallsvariablen \Leftrightarrow

$$p_\xi(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathbb{R}^2. \quad (11.48)$$

◦

Generell ist die Unabhängigkeit zweier Zufallsvariablen also äquivalent zur Faktorisierung ihrer gemeinsamen WMF oder WDF. Für einen Beweis von Theorem 11.4 verweisen wir auf die weiterführende Literatur. Nichtsdestotrotz ist Theorem 11.4 für weite Aspekte der probabilistischen Modellierung grundlegend.

Beispiel

Wir betrachten erneut den zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ aus Kapitel 11.1, dessen gemeinsame und marginale WMFen bekanntlich die untenstehende Form haben

$p_\xi(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{\xi_1}(x_1)$
$x_1 = 1$	0.10	0.00	0.20	0.10	0.40
$x_1 = 2$	0.10	0.20	0.00	0.00	0.30
$x_1 = 3$	0.00	0.10	0.10	0.10	0.30
$p_{\xi_2}(x_2)$	0.20	0.30	0.30	0.20	

Wir fragen zunächst, ob ξ_1 und ξ_2 wohl unabhängig sind. Dies ist nicht der Fall, da hier gilt, dass

$$p_\xi(1, 1) = 0.10 \neq 0.08 = 0.40 \cdot 0.20 = p_{\xi_1}(1)p_{\xi_2}(1). \tag{11.49}$$

Möchten wir basierend auf den Marginalverteilungen von ξ eine gemeinsame Verteilung erzeugen, in der ξ_1 und ξ_2 unabhängig sind, so muss sich jeder Eintrag der gemeinsamen Verteilung $p_\xi(\xi_1, \xi_2)$ aus dem jeweiligen Produkt der Marginalwahrscheinlichkeiten ergeben. Die gemeinsame Verteilung von ξ_1 und ξ_2 unter der Annahme der Unabhängigkeit von ξ_1 und ξ_2 bei gleichen Marginalverteilungen wie im obigen Fall ergibt sich also zu

$p_\xi(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{\xi_1}(x_1)$
$x_1 = 1$	0.08	0.12	0.12	0.08	0.40
$x_1 = 2$	0.06	0.09	0.09	0.06	0.30
$x_1 = 3$	0.06	0.09	0.09	0.06	0.30
$p_{\xi_2}(x_2)$	0.20	0.30	0.30	0.20	

Weiterhin ergeben sich im Falle der Unabhängigkeit von ξ_1 und ξ_2 beispielsweise die bedingten WMFen $p_{\xi_2|\xi_1}$ zu wie folgt:

$p_{\xi_1 \xi_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{\xi_2 \xi_1=1}(x_2 x_1 = 1)$	$\frac{0.08}{0.40} = 0.2$	$\frac{0.12}{0.40} = 0.3$	$\frac{0.12}{0.40} = 0.3$	$\frac{0.08}{0.40} = 0.2$
$p_{\xi_2 \xi_1=2}(x_2 x_1 = 2)$	$\frac{0.06}{0.30} = 0.2$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.06}{0.30} = 0.2$
$p_{\xi_2 \xi_1=3}(x_2 x_1 = 3)$	$\frac{0.06}{0.30} = 0.2$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.06}{0.30} = 0.2$

Im Falle der Unabhängigkeit von ξ_1 und ξ_2 ändert sich die Verteilung von ξ_2 gegeben (oder im Wissen um) den Wert von ξ_1 also nicht und entspricht jeweils der Marginalverteilung von ξ_2 . Dies entspricht natürlich der Intuition der Unabhängigkeit von Ereignissen im Kontext elementarer Wahrscheinlichkeiten.

Wir wollen den Begriff der unabhängigen Zufallsvariablen nun für mehr als zwei Zufallsvariablen definieren.

Definition 11.12 (n unabhängige Zufallsvariablen). $\xi := (\xi_1, \dots, \xi_n)$ sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum $\mathcal{X} = \times_{i=1}^n \mathcal{X}_i$. Die n Zufallsvariablen ξ_1, \dots, ξ_n heißen *unabhängig*, wenn für alle $S_i \in \mathcal{X}_i, i = 1, \dots, n$ gilt, dass

$$\mathbb{P}_\xi(\xi_1 \in S_1, \dots, \xi_n \in S_n) = \prod_{i=1}^n \mathbb{P}_{\xi_i}(\xi_i \in S_i). \tag{11.50}$$

Wenn der Zufallsvektor eine n -dimensionale WMF oder WDF p_ξ mit marginalen WMFen oder WDFen $p_{\xi_i}, i = 1, \dots, n$ besitzt, dann ist die Unabhängigkeit von ξ_1, \dots, ξ_n gleichbedeutend mit der Faktorisierung der gemeinsamen WMF oder WDF, also mit

$$p_\xi(\xi_1, \dots, \xi_n) = \prod_{i=1}^n p_{\xi_i}(x_i). \quad (11.51)$$

•

Es handelt bei Definition 11.12 also um eine direkte Generalisierung des zweidimensionalen Falls.

Sind n Zufallsvariablen nicht nur unabhängig, sondern haben sie auch alle die gleiche Verteilung, so nennt man sie *unabhängig und identisch verteilt (u.i.v.)*:

Definition 11.13 (Unabhängig und identisch verteilte Zufallsvariablen). n Zufallsvariablen ξ_1, \dots, ξ_n heißen *unabhängig und identisch verteilt (u.i.v.)*, wenn

- (1) ξ_1, \dots, ξ_n unabhängige Zufallsvariablen sind, und
- (2) die Marginalverteilungen der ξ_i übereinstimmen, also gilt, dass

$$\mathbb{P}_{\xi_i} = \mathbb{P}_{\xi_j} \text{ für alle } 1 \leq i, j \leq n. \quad (11.52)$$

Wenn die Zufallsvariablen ξ_1, \dots, ξ_n unabhängig und identisch verteilt sind und die i te Marginalverteilung $\mathbb{P}_\xi := \mathbb{P}_{\xi_i}$ ist, so schreibt man auch

$$\xi_1, \dots, \xi_n \sim \mathbb{P}_\xi. \quad (11.53)$$

•

Man sagt kurz, dass “ ξ_1, \dots, ξ_n u.i.v.” sind. Im Englischen spricht man von *independent and identically distributed (i.i.d)* Zufallsvariablen. U.i.v. Zufallsvariablen spielen an vielen Stellen der probabilistischen Modellierung eine wichtige Rolle. So werden, wie wir an späterer Stelle sehen werden, additive Fehlerterme in probabilistischen Modellen meist durch u.i.v. Zufallsvariablen modelliert.

Schließlich halten wir fest, dass n u.i.v. normalverteilte Zufallsvektoren werden als

$$\xi_1, \dots, \xi_n \sim N(\mu, \sigma^2) \quad (11.54)$$

geschrieben werden. In Kapitel 11.6 zeigen wir, wie genau die gemeinsame Verteilung von n u.i.v. normalverteilte Zufallsvektoren beschaffen ist.

11.6. Normalverteilungen

Die multivariate Normalverteilung ist die multivariate Generalisierung der univariaten Normalverteilung. Die Motivation für die verbreiteten Normalverteilungsannahmen in der probabilistischen Modellierung liegt bekanntermaßen im Zentralen Grenzwertsatz: In probabilistischen Modellen repräsentieren probabilistische Terme die Summation sehr vieler Zufallsvorgänge, die durch die deterministischen Bestandteile des jeweiligen Modells, also eine formalisierte wissenschaftliche Theorie, nicht erklärt werden. Nach dem Zentralen Grenzwertsatz ist die Summe dieser nicht erklärten Zufallsvorgänge dann aber gerade normalverteilt.

Über diesen grundlegenden Aspekt hinaus hat die Normalverteilung viele günstige mathematische Eigenschaften, die ihren Einsatz in vielen Bereichen der probabilistischen Modellierung ermöglichen. Anwendungen multivariater Normalverteilungen finden sich damit im Kontext des Allgemeinen Linearen Modells, den Generalisierungen des Allgemeinen Modells zu Hierarchischen Linearen Modellen oder Multivariaten Allgemeinen Linearen Modellen, der Bayesianischen Inferenz bei Normalverteilungsannahmen und nicht zuletzt der Theorie probabilistischer Filter, wie zum Beispiel dem Kalman-Bucy Filter.

Konstruktion

In diesem Abschnitt wollen wir zeigen, wie ein bivariat normalverteilter Zufallsvektor durch Transformation und Konkatenation zweier univariat normalverteilter Zufallsvariablen konstruiert werden kann. Dazu erinnern wir zunächst an den Begriff der normalverteilten Zufallsvariable

Definition 11.14 (Normalverteilte Zufallsvariable). ξ sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (11.55)$$

Dann sagen wir, dass ξ einer *Normalverteilung (oder Gauß-Verteilung)* mit *Erwartungswertparameter* $\mu \in \mathbb{R}$ und *Varianzparameter* $\sigma^2 > 0$ unterliegt und nennen ξ eine *normalverteilte Zufallsvariable*. Wir kürzen dies mit $\xi \sim N(\mu, \sigma^2)$ ab. Die WDF einer normalverteilten Zufallsvariable bezeichnen wir mit

$$N(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (11.56)$$

•

Visuell entspricht der Parameter μ einer normalverteilten Zufallsvariable dem Wert höchster Wahrscheinlichkeitsdichte und der Parameter σ^2 spezifiziert die Breite der WDF (Abbildung 11.2). Weiterhin gelten für den Erwartungswert und die Varianz einer normalverteilten Zufallsvariable bekanntlich

$$\mathbb{E}(\xi) = \mu \text{ und } \mathbb{V}(\xi) = \sigma^2. \quad (11.57)$$

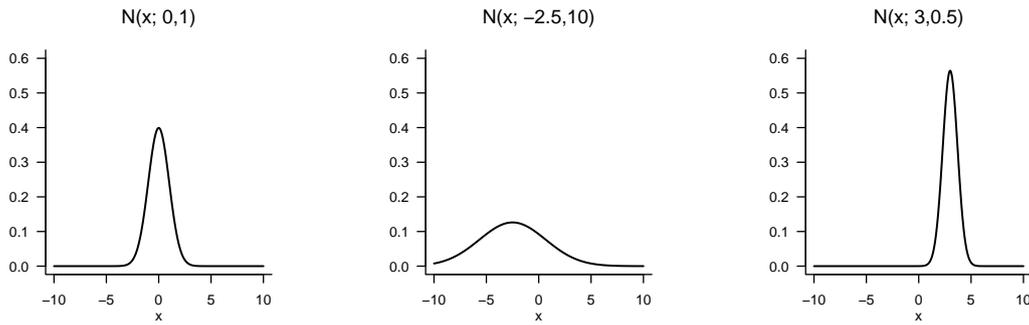


Abbildung 11.2. Wahrscheinlichkeitsdichtefunktionen univariater Normalverteilungen.

Eine normalverteilte Zufallsvariable der Form $\xi \sim N(0, 1)$ schließlich heißt auch *standardnormalverteilt*.

Folgendes Theorem zeigt, wie zwei unabhängige, univariat standardnormalverteilte Zufallsvariablen kombiniert werden können, um einen bivariat verteilten Zufallsvektor zu konstruieren. Die Verteilung eines ebensolchen Zufallsvektors wird dann als *bivariate Normalverteilung* bezeichnet.

Theorem 11.5 (Konstruktion bivariater Normalverteilungen). $\zeta_1 \sim N(0, 1)$ und $\zeta_2 \sim N(0, 1)$ seien zwei unabhängige standardnormalverteilte Zufallsvariablen. Weiterhin seien $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ und $\rho \in]-1, 1[$. Schließlich seien

$$\begin{aligned}\xi_1 &:= \sigma_1 \zeta_1 + \mu_1 \\ \xi_2 &:= \sigma_2 (\rho \zeta_1 + (1 - \rho^2)^{1/2} \zeta_2) + \mu_2.\end{aligned}\tag{11.58}$$

Dann hat die WDF des Zufallsvektors $\xi := (\xi_1, \xi_2)^T$, also der gemeinsamen Verteilung von ξ_1 und ξ_2 , die Form

$$p : \mathbb{R}^2 \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right),\tag{11.59}$$

wobei $n := 2$ und $\mu \in \mathbb{R}^2$ und $\Sigma \in \mathbb{R}^{2 \times 2}$ durch

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ und } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_2 \sigma_1 & \sigma_2^2 \end{pmatrix}\tag{11.60}$$

gegeben sind.

◦

Für einen Beweis des Theorems verweisen wir auf DeGroot & Schervish (2012).

Beispiel

Folgender **R** Code zeichnet das obige Theorem anhand konkreter Beispielwerte für $\mu_1, \mu_2, \sigma_1, \sigma_2$ und ρ nach und gibt die Parameter μ und Σ der resultierenden bivariaten WDF aus.

```

1 # Parameterdefinitionen
2 mu_1 = 5.0 # \mu_1
3 mu_2 = 4.0 # \mu_2
4 sig_1 = 1.5 # \sigma_1
5 sig_2 = 1.0 # \sigma_2
6 rho = 0.9 # \rho
7
8 # Realisierungen der standardnormalverteilten ZVen
9 n = 100 # Anzahl Realisierungen
10 zeta_1 = rnorm(n) # \zeta_1 \sim N(0,1)
11 zeta_2 = rnorm(n) # \zeta_2 \sim N(0,1)
12
13 # Evaluation von Realisierungen von \xi_1 und \xi_2
14 xi_1 = sig_1*zeta_1 + mu_1 # Realisierungen von zeta_1
15 xi_2 = sig_2*(rho*zeta_1 + sqrt(1-rho^2)*zeta_2) + mu_2 # Realisierungen von zeta_2
16
17 # Parameter der gemeinsamen Verteilung von \xi_1 und \xi_2
18 mu = matrix(c(mu_1, # \mu \in \mathbb{R}^2
19 mu_2),
20 nrow = 2, byrow = TRUE)
21 Sigma = matrix(c(sig_1^2, rho*sig_1*sig_2, # \Sigma \in \mathbb{R}^{2 \times 2}
22 rho*sig_1*sig_2, sig_2^2),
23 nrow = 2, byrow = TRUE)
24 print(mu)

      [,1]
[1,] 5
[2,] 4

1 print(Sigma)

      [,1] [,2]
[1,] 2.25 1.35
[2,] 1.35 1.00

```

Die durch obigen **R** Code generierten Realisierungen von $\xi = (\xi_1, \xi_2)^T$, sowie die Isokonturen der durch Theorem 11.5 postulierten WDF sind in Abbildung 11.3 dargestellt.

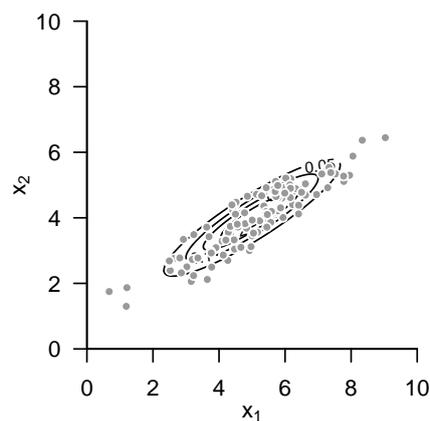


Abbildung 11.3. Konstruktion bivariater Normalverteilungen.

Definition

Wir wollen die multivariate Normalverteilung nun formal einführen und erste Eigenschaften angeben. Wir nutzen dazu folgende Definition.

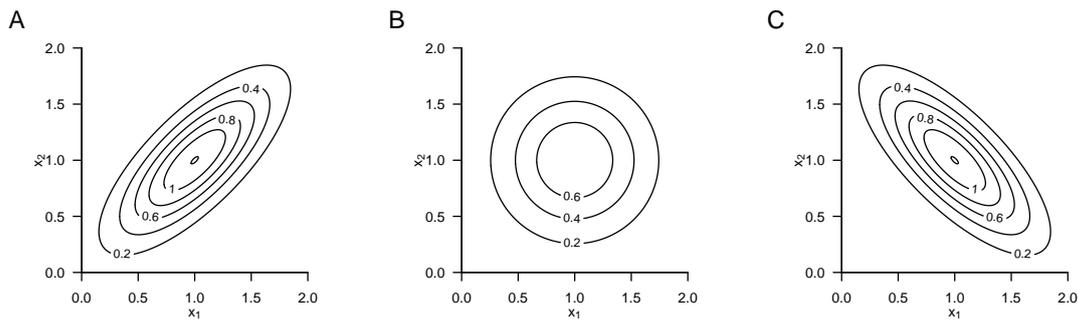


Abbildung 11.4. Wahrscheinlichkeitsdichtefunktionen bivariater Normalverteilungen.

Definition 11.15. ξ sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum \mathbb{R}^n und WDF

$$p : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right). \quad (11.61)$$

Dann sagen wir, dass ξ einer *multivariaten (oder n -dimensionalen) Normalverteilung* mit *Erwartungswertparameter* $\mu \in \mathbb{R}^n$ und positiv-definitem *Kovarianzmatrixparameter* $\Sigma \in \mathbb{R}^{n \times n}$ unterliegt und nennen ξ einen (*multivariat*) *normalverteilten Zufallsvektor*. Wir kürzen dies mit $\xi \sim N(\mu, \Sigma)$ ab. Die WDF eines multivariat normalverteilten Zufallsvektors bezeichnen wir mit

$$N(x; \mu, \Sigma) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right). \quad (11.62)$$

•

Abbildung 11.4 A,B,C zeigen die Isokonturen der WDFen bivariat normalverteilter Zufallsvektoren für $\mu = (1, 1)^T$ und

$$\Sigma_A := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix}, \quad \Sigma_B := \begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.20 \end{pmatrix}, \quad \Sigma_C := \begin{pmatrix} 0.20 & -0.15 \\ -0.15 & 0.20 \end{pmatrix} \quad (11.63)$$

respektive. Abbildung 11.5 A-C zeigen jeweils 200 Realisierungen der entsprechenden Zufallsvektoren.

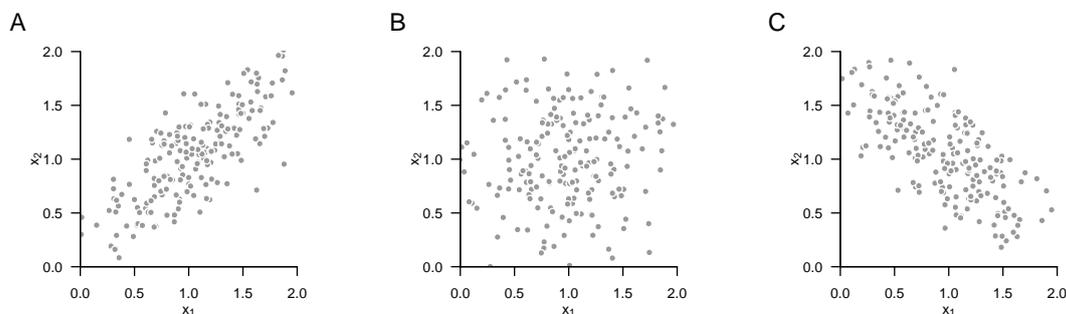


Abbildung 11.5. Realisierungen bivariat normalverteilter Zufallsvektoren.

Ohne Beweis halten wir fest, dass wie im Fall einer univariat normalverteilten Zufallsvariable, der Erwartungswert und die Kovarianzmatrix eines normalverteilten Zufallsvektors durch die entsprechenden Parameter gegeben sind.

Theorem 11.6 (Erwartungswert und Kovarianzmatrix normalverteilter Zufallsvektoren). $\xi \sim N(\mu, \Sigma)$ sei ein multivariat normalverteilter Zufallsvektor mit Erwartungswertparameter $\mu \in \mathbb{R}^n$ und Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{n \times n}$ pd. Dann gelten

$$\mathbb{E}(\xi) = \mu \text{ und } \mathbb{C}(\xi) = \Sigma. \quad (11.64)$$

◦

Wie im Falle der univariat normalverteilten Zufallsvariable entspricht der Parameter $\mu \in \mathbb{R}^n$ dem Wert höchster Wahrscheinlichkeitsdichte der multivariaten Normalverteilung. Analog zum Varianzparameter der univariat normalverteilten Zufallsvariable spezifizieren die Diagonalelemente von $\Sigma \in \mathbb{R}^{n \times n}$ pd die Breite der WDF bezüglich der Zufallsvektorkomponenten ξ_1, \dots, ξ_n . Allgemein spezifiziert im Falle des multivariat normalverteilten Zufallsvektors das i, j te Element von $\Sigma \in \mathbb{R}^{n \times n}$ pd hier nun die Kovarianz der Zufallsvektorkomponenten ξ_i und ξ_j .

Transformationen

In diesem Abschnitt stellen wir einige Resultate zu den Verteilungen transformierter normalverteilter Zufallsvektoren zusammen. Wir verzichten dabei auf Beweise.

Theorem 11.7 (Invertierbare lineare Transformation eines normalverteilten Zufallsvektors). $\xi \sim N(\mu_\xi, \Sigma_\xi)$ sei ein normalverteilter n -dimensionaler Zufallsvektor und es sei $\zeta := A\xi$ mit einer invertierbaren Matrix $A \in \mathbb{R}^{n \times n}$. Dann gilt

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu_\xi \text{ und } \Sigma_\zeta = A\Sigma_\xi A^T. \quad (11.65)$$

◦

Nach Theorem 11.7 ergibt die invertierbare lineare Transformation eines multivariat normalverteilten Zufallsvektors also wiederum einen multivariat normalverteilten Zufallsvektor und die Parameter der Verteilung dieses normalverteilten Zufallsvektors ergeben sich aus den Parametern der Verteilung des ursprünglichen Zufallsvektors und der Transformationsmatrix.

Beispiel

Als Beispiel betrachten wir die invertierbare lineare Transformation eines bivariaten normalverteilten Zufallsvektors ξ . Es seien

$$\mu_\xi := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ und } \Sigma_\xi := \begin{pmatrix} 0.20 & 0.15 \\ 0.15 & 0.20 \end{pmatrix} \quad (11.66)$$

der Erwartungswert- und Kovarianzmatrixparameter von ξ , respektive, und es sei

$$A := \begin{pmatrix} -2 & 1 \\ -1 & 2 \end{pmatrix} \quad (11.67)$$

die Transformationsmatrix. Da $|A| = -3 \neq 0$ ist A invertierbar und es gilt nach Theorem 11.7, dass

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu_\xi = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ und } A\Sigma_\xi A^T = \begin{pmatrix} 0.40 & 0.05 \\ 0.05 & 0.40 \end{pmatrix}. \quad (11.68)$$

Abbildung 11.6 A zeigt Isokonturen der WDF von ξ und Realisierungen $x^{(i)} \in \mathbb{R}^2$ von ξ für $i = 1, \dots, 50$. Abbildung 11.6 B zeigt die transformierten Realisierungen $z^{(i)} = Ax^{(i)} \in \mathbb{R}^2$ von ζ sowie die Isokonturen der WDF von ζ nach Theorem 11.7.

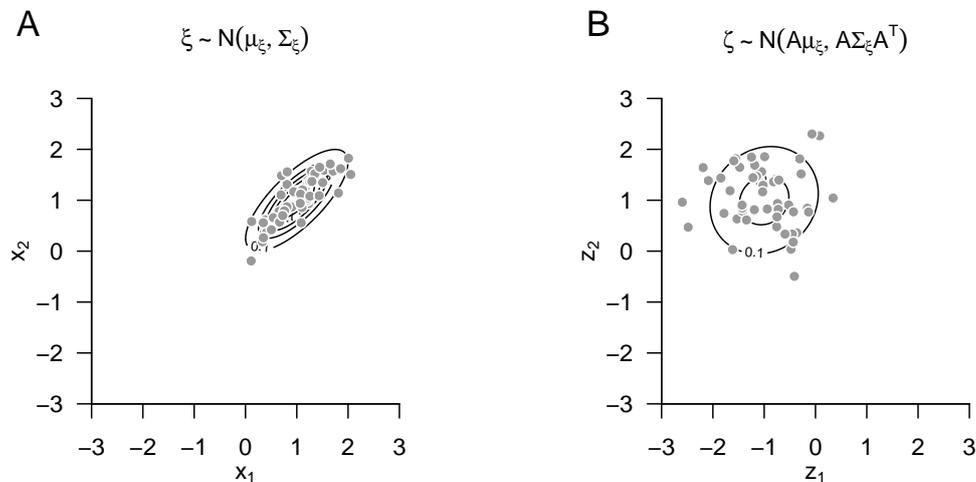


Abbildung 11.6. Invertierbare lineare Transformation eines normalverteilten Zufallsvektors

Die Tatsache, dass ein linear transformierter normalverteilter Zufallsvektor wiederum normalverteilt ist und dass sich die Parameter der Verteilung des transformierten Zufallsvektors aus den Parametern der Verteilung des ursprünglichen Zufallsvektors sowie den Transformationsparametern bestimmen lassen, bleibt auch im Falle einer nicht notwendigerweise invertierbaren linearen Transformation und auch im Falle einer nicht notwendigerweise invertierbaren linear-affinen Transformation wahr. Dies ist die Aussage folgenden zentralen Theorems. Für einen Beweis verweisen wir auf Anderson (2003).

Theorem 11.8 (Linear-affine Transformation eines normalverteilten Zufallsvektors). $\xi \sim N(\mu_\xi, \Sigma_\xi)$ sei ein normalverteilter n -dimensionaler Zufallsvektor und es sei

$$\zeta := Ax + b \text{ mit } A \in \mathbb{R}^{m \times n} \text{ und } b \in \mathbb{R}^m. \quad (11.69)$$

Dann gilt

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \text{ mit } \mu_\zeta = A\mu + b \in \mathbb{R}^m \text{ und } \Sigma_\zeta = A\Sigma A^T \in \mathbb{R}^{m \times m}. \quad (11.70)$$

◦

Sphärizität

Folgendes Theorem ist für die grundlegende Theorie des Allgemeinen Linearen Modells zentral.

Theorem 11.9 (Sphärische multivariate Normalverteilung). Für $i = 1, \dots, n$ seien $N(x_i; \mu_i, \sigma^2)$ die WDFen von n unabhängigen univariaten normalverteilten Zufallsvariablen ξ_1, \dots, ξ_n mit $\mu_1, \dots, \mu_n \in \mathbb{R}$ und $\sigma^2 > 0$. Weiterhin sei $N(x; \mu, \sigma^2 I_n)$ die WDF eines n -variablen normalverteilten Zufallsvektors ξ mit Erwartungswertparameter $\mu := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$. Dann gilt

$$p_\xi(x) = p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i) \quad (11.71)$$

und insbesondere

$$N(x; \mu, \sigma^2 I_n) = \prod_{i=1}^n N(x_i; \mu_i, \sigma^2) \quad (11.72)$$

für alle $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$.

◦

Beweis. Wir zeigen die Identität der multivariaten WDF $N(x; \mu, \sigma^2 I_n)$ mit dem Produkt von n univariaten WDFen $N(x_i; \mu_i, \sigma^2 I_n)$, wobei μ_i der i te Eintrag von $\mu \in \mathbb{R}^n$ ist. Es ergibt sich

$$\begin{aligned} N(x; \mu, \sigma^2 I_n) &= (2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T (\sigma^2 I_n)^{-1} (x - \mu)\right) \\ &= \left(\prod_{i=1}^n 2\pi^{-\frac{1}{2}}\right) (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^T (x - \mu)\right) \\ &= \left(\prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2} (x_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n N(x_i; \mu_i, \sigma^2). \end{aligned} \quad (11.73)$$

□

Einen Kovarianzmatrixparameter der Form $\Sigma = \sigma^2 I_n$ nennt man auch *sphärisch*, da die Isokonturen der WDF eines normalverteilten Zufallsvektors mit einem solchen Kovarianzmatrixparameter *Sphären* bilden (zum Beispiel Kreise bei $n = 2$ und Kugeln bei $n = 3$). Eine multivariate Normalverteilung mit sphärische Kovarianzmatrixparameter nennt man entsprechend eine *sphärische Normalverteilung*. Theorem 11.9 besagt, dass die WDF eines n -dimensionalen normalverteilten Zufallsvektors mit sphärischem Kovarianzparameter der gemeinsamen WDF von n unabhängigen univariat normalverteilten Zufallsvariablen entspricht und umgekehrt. Eine Realisierung eines n -dimensionalen normalverteilten Zufallsvektors entspricht also den Realisierungen von n unabhängigen univariat normalverteilten Zufallsvariablen und umgekehrt. Man beachte, dass die Identität der Verteilungen der $\xi_i, i = 1, \dots, n$ hier nicht vorausgesetzt ist, insbesondere können sich ihre Erwartungswertparameter $\mu_i, i = 1, \dots, n$ explizit unterscheiden.

Marginale und bedingte Verteilungen

Multivariate Normalverteilungen haben die Eigenschaft, dass auch alle anderen assoziierten Verteilungen Normalverteilungen sind und deren Erwartungswert- und Kovarianzmatrixparameter aus den Parametern der jeweils komplementären Verteilung errechnet werden können. Insbesondere gilt zum einen, dass die uni- und multivariaten Marginalverteilungen multivariater Normalverteilungen wiederum Normalverteilungen sind. Zum anderen lassen sich wie alle multivariaten Verteilungen multivariate Normalverteilungen multiplikativ in eine marginale und eine bedingte Verteilung zerlegen. Insbesondere sind nun aber bei multivariaten Normalverteilungen diese Verteilungen wiederum (multivariate) Normalverteilungen, deren Parameter aus den Parametern der gemeinsame Verteilung errechnet werden können und umgekehrt. Wir fassen obige Erkenntnisse formal in den folgenden drei Theoremen zusammen.

Theorem 11.10 (Marginale Normalverteilungen). *Es sei $n := k + l$ und $\xi = (\xi_1, \dots, \xi_n)^T$ sei ein n -dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter*

$$\mu = \begin{pmatrix} \mu_v \\ \mu_\zeta \end{pmatrix} \in \mathbb{R}^n, \quad (11.74)$$

mit $\mu_v \in \mathbb{R}^k$ and $\mu_\zeta \in \mathbb{R}^l$ und Kovarianzmatrixparameter

$$\Sigma = \begin{pmatrix} \Sigma_{vv} & \Sigma_{v\zeta} \\ \Sigma_{\zeta v} & \Sigma_{\zeta\zeta} \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad (11.75)$$

mit $\Sigma_{vv} \in \mathbb{R}^{k \times k}$, $\Sigma_{v\zeta} \in \mathbb{R}^{k \times l}$, $\Sigma_{\zeta v} \in \mathbb{R}^{l \times k}$, und $\Sigma_{\zeta\zeta} \in \mathbb{R}^{l \times l}$. Dann sind $v := (\xi_1, \dots, \xi_k)^T$ und $\zeta := (\xi_{k+1}, \dots, \xi_n)^T$ k - und l -dimensionale normalverteilte Zufallsvektoren, respektive, und es gilt

$$v \sim N(\mu_v, \Sigma_{vv}) \text{ und } \zeta \sim N(\mu_\zeta, \Sigma_{\zeta\zeta}). \quad (11.76)$$

◦

Die Marginalverteilungen einer multivariaten Normalverteilung sind also auch Normalverteilungen und die Parameter der Marginalverteilungen ergeben sich aus den Parametern der gemeinsamen Verteilung. Für Beweise dieses Theorems verweisen wir

auf Mardia et al. (1979) und Anderson (2003). Abbildung 11.7 visualisiert Theorem 11.10 für den Fall $n := 2, k := 1, l := 1$,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2 \text{ und } \Sigma := \begin{pmatrix} 0.10 & 0.08 \\ 0.08 & 0.15 \end{pmatrix} \in \mathbb{R}^{2 \times 2}. \quad (11.77)$$

Abbildung 11.7 A zeigt dabei die WDF des bivariaten Zufallsvektors ξ und Abbildung 11.7 B und C die WDFen der entsprechenden marginalen Zufallsvariablen v und ζ .

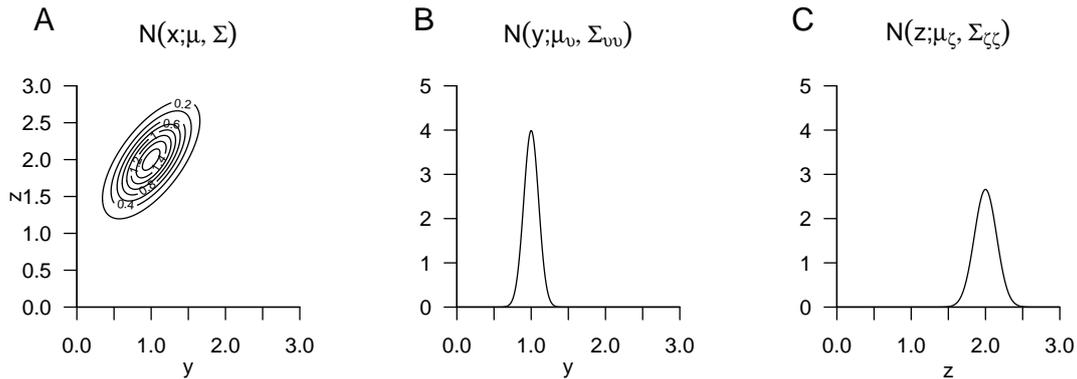


Abbildung 11.7. Marginale Verteilungen eines bivariaten normalverteilten Zufallsvektor.

Mithilfe einer marginalen und einer bedingten multivariaten Normalverteilung lässt sich eine gemeinsame multivariate Normalverteilung konstruieren, deren Parameter sich aus den Parametern der marginalen und bedingten Verteilung ergeben. Dies ist die zentrale Aussage folgenden Theorems.

Theorem 11.11 (Gemeinsame Normalverteilungen). ξ sei ein m -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_\xi : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_\xi(x) := N(x; \mu_\xi, \Sigma_{\xi\xi}) \text{ mit } \mu_\xi \in \mathbb{R}^m, \Sigma_{\xi\xi} \in \mathbb{R}^{m \times m}, \quad (11.78)$$

$A \in \mathbb{R}^{n \times m}$ sei eine Matrix, $b \in \mathbb{R}^n$ sei ein Vektor und v sei ein n -dimensionaler bedingt normalverteilter Zufallsvektor mit bedingter WDF

$$p_{v|\xi}(\cdot|x) : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}, y \mapsto p_{v|\xi}(y|x) := N(y; A\xi + b, \Sigma_{vv}) \text{ mit } \Sigma_{vv} \in \mathbb{R}^{n \times n}. \quad (11.79)$$

Dann ist der $m + n$ -dimensionale Zufallsvektor $(\xi, v)^T$ normalverteilt mit (gemeinsamer) WDF

$$p_{\xi, v} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p_{\xi, v} \left(\begin{pmatrix} x \\ y \end{pmatrix} \right) = N \left(\begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi, v}, \Sigma_{\xi, v} \right), \quad (11.80)$$

mit $\mu_{\xi, v} \in \mathbb{R}^{m+n}$ und $\Sigma_{\xi, v} \in \mathbb{R}^{(m+n) \times (m+n)}$ und insbesondere

$$\mu_{\xi, v} = \begin{pmatrix} \mu_\xi \\ A\mu_\xi + b \end{pmatrix} \text{ und } \Sigma_{\xi, v} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\xi}A^T \\ A\Sigma_{\xi\xi} & \Sigma_{vv} + A\Sigma_{\xi\xi}A^T \end{pmatrix}. \quad (11.81)$$

◦

Insbesondere ergeben sich die Parameter der gemeinsamen Verteilung also als linear-affine Transformation der Parameter der induzierenden Verteilungen. Abbildung 11.8 visualisiert Theorem 11.11 für den Fall $m := 1, n := 1, \mu_\xi := 1, \Sigma_{\xi\xi} := 0.2, A := 1, b := 1$ und $\Sigma_{vv} := 0.1$. Abbildung 11.8 A zeigt dabei die WDF der Zufallsvariable ξ , Abbildung 11.8 B zeigt die WDF der bedingten Verteilung der Zufallsvariable v gegeben ξ und Abbildung 11.7 C schließlich zeigt die WDFen des induzierten bivariaten Zufallsvektors (ξ, v) .

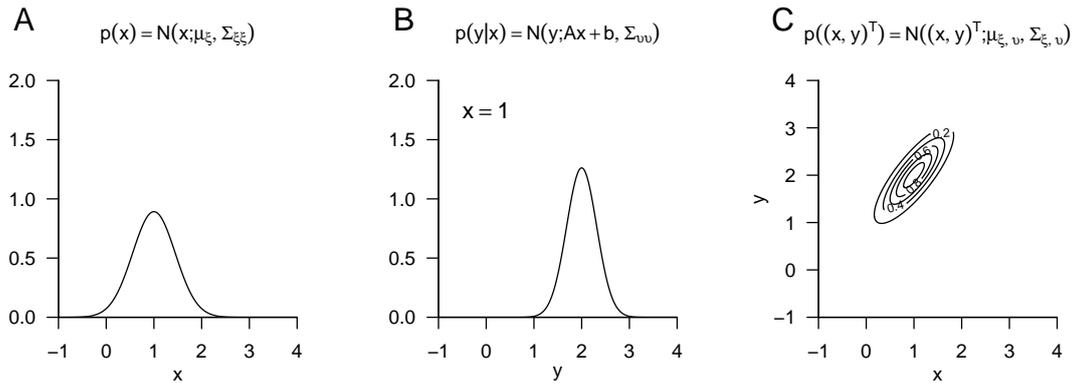


Abbildung 11.8. Gemeinsame Verteilungen einer marginalen und einer auf dieser bedingten normalverteilten Zufallsvariable.

Die Definition einer multivariaten Normalverteilung erlaubt es weiterhin, die bedingten Verteilungen aller Komponenten des entsprechenden Zufallsvektors direkt mithilfe der Parameter der multivariaten Normalverteilung zu bestimmen. Dies ist die zentrale Aussage folgenden Theorems.

Theorem 11.12 (Bedingte Normalverteilungen). (ξ, v) sei ein $m + n$ -dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_{\xi, v} : \mathbb{R}^{m+n} \rightarrow \mathbb{R}_{>0}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto p_{\xi, v} \left(\begin{pmatrix} x \\ y \end{pmatrix} \right) := N \left(\begin{pmatrix} x \\ y \end{pmatrix}; \mu_{\xi, v}, \Sigma_{\xi, v} \right), \quad (11.82)$$

mit

$$\mu_{\xi, v} = \begin{pmatrix} \mu_\xi \\ \mu_v \end{pmatrix}, \Sigma_{\xi, v} = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi v} \\ \Sigma_{v\xi} & \Sigma_{vv} \end{pmatrix}, \quad (11.83)$$

mit $x, \mu_\xi \in \mathbb{R}^m, y, \mu_v \in \mathbb{R}^n$ und $\Sigma_{\xi\xi} \in \mathbb{R}^{m \times m}, \Sigma_{\xi v} \in \mathbb{R}^{m \times n}, \Sigma_{vv} \in \mathbb{R}^{n \times n}$. Dann ist die bedingte Verteilung von ξ gegeben v eine m -dimensionale Normalverteilung mit bedingter WDF

$$p_{\xi|v}(\cdot|y) : \mathbb{R}^m \rightarrow \mathbb{R}_{>0}, x \mapsto p_{\xi|v}(x|y) := N(x; \mu_{\xi|v}, \Sigma_{\xi|v}) \quad (11.84)$$

mit

$$\mu_{\xi|v} = \mu_\xi + \Sigma_{\xi v} \Sigma_{vv}^{-1} (y - \mu_v) \in \mathbb{R}^m \quad (11.85)$$

und

$$\Sigma_{\xi|v} = \Sigma_{\xi\xi} - \Sigma_{\xi v} \Sigma_{vv}^{-1} \Sigma_{v\xi} \in \mathbb{R}^{m \times m}. \quad (11.86)$$

◦

Im Zusammenspiel mit Theorem 11.11 und Theorem 11.10 können die Parameter bedingter und marginaler Normalverteilungen also aus den Parametern der komplementären bedingten und marginalen Normalverteilungen bestimmt werden. Abbildung 11.9 visualisiert Theorem 11.12 für den Fall $m := 2, n := 1$,

$$\mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 0.12 & 0.09 \\ 0.09 & 0.12 \end{pmatrix} \quad (11.87)$$

Dabei zeigt Abbildung 11.9 A die WDF des bivariaten Zufallsvektors $(\xi, v)^T$ und Abbildung 11.9 B und C zeigen die WDF der bedingten Verteilung der Zufallsvariable ξ gegeben $v = 1.5$ und $v = 2.8$, respektive.

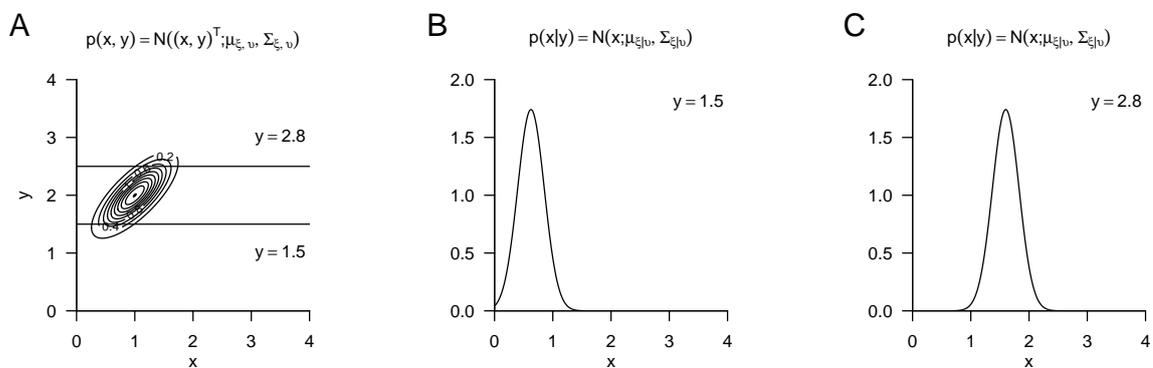


Abbildung 11.9. Bedingte Normalverteilungen

11.7. Literaturhinweise

Die Entwicklung der bivariaten Normalverteilung hat ihre Ursprünge in der statistischen Literatur zur Mitte des 19. Jahrhunderts, insbesondere in den Arbeiten von Francis Galton (1822-1911). Die mathematische Formalisierung der bivariaten Normalverteilung geht dabei wohl insbesondere auf Pearson (1896) zurück (Seal (1967)). Die ursprüngliche Formulierung der multivariaten Normalverteilung wird bei Edgeworth (1892) verortet. Tong (1990) gibt einen umfassenden Überblick zur Theorie und Anwendung der multivariaten Normalverteilung.

11.8. Selbstkontrollfragen

1. Geben Sie Definition eines Zufallsvektors wieder.
2. Geben Sie Definition der multivariaten Verteilung eines Zufallsvektors wieder.
3. Geben Sie Definition einer multivariaten WMF wieder.
4. Geben Sie Definition einer multivariaten WDF wieder.
5. Geben Sie die Definition des Erwartungswerts eines Zufallsvektors wieder.
6. Geben Sie die Definition der Kovarianzmatrix eines Zufallsvektors wieder.
7. Was repräsentieren die Diagonalelemente der Kovarianzmatrix eines Zufallsvektors?
8. Was repräsentieren die Nichtdiagonalelemente der Kovarianzmatrix eines Zufallsvektors?

9. Geben Sie die Definition der Korrelationsmatrix eines Zufallsvektors wieder.
10. Geben Sie die Definition der univariaten Marginalverteilung eines Zufallsvektors wieder.
11. Wie berechnet man die WMF der i ten Komponente eines diskreten Zufallsvektors?
12. Wie berechnet man die WDF der i ten Komponente eines kontinuierlichen Zufallsvektors?
13. Geben Sie Definition der bedingten WMF und der diskreten bedingten Verteilung wieder.
14. Geben Sie Definition der bedingten WDF und der kontinuierlichen bedingten Verteilung wieder.
15. Geben Sie die Definition der WDF eines multivariaten normalverteilten Zufallsvektors wieder.
16. Erläutern Sie die Komponenten der WDF eines multivariaten normalverteilten Zufallsvektors.
17. Geben Sie den Erwartungswert und die Kovarianzmatrix eines normalverteilten Zufallsvektors an.
18. Geben Sie das Theorem zu marginalen Normalverteilungen wieder.
19. Geben Sie das Theorem zu gemeinsamen Normalverteilungen wieder.
20. Geben Sie das Theorem zu bedingten Normalverteilungen wieder.

Teil II.

Frequentistische Inferenz

12. Datendeskription

12.1. Datenanalyseszenarien

Wir wollen hier zunächst die im folgenden zu betrachtenden Datenanalyseverfahren anhand der Dimensionalität ihrer unabhängigen Variablen und abhängigen Variablen klassifizieren. Dazu bezeichnen wir wie üblich eine unabhängige Variable mit x und eine abhängige Variable mit y . Weiterhin sei mit den Subskripten i und j bei x_{ij} und y_{ij} der Wert der j ten univariaten Komponente der jeweiligen Variable, beispielsweise ein Testwert, bei der i ten experimentellen Einheit, beispielsweise einer Proband:in, bezeichnet. Die Gesamtzahl experimenteller Einheiten sei mit n bezeichnet.

Tabelle 12.1 zeigt das Szenario einer univariaten unabhängigen Variable und einer univariaten abhängigen Variable. Typische in diesem Szenario genutzte Inferenzverfahren sind die Bestimmung der *Korrelation* von x_1 und y_1 , die Durchführung einer *einfachen linearen Regression* von y_1 auf x_1 und, wenn x_1 eine kategoriale Faktorvariable ist, *T-Tests* und *Varianzanalysen*.

Tabelle 12.1. Univariete unabhängige Variable x und univariete abhängige Variable y

x_1	y_1
x_{11}	y_{11}
\vdots	\vdots
x_{n1}	y_{n1}

Tabelle 12.2 zeigt das Szenario einer multivariaten unabhängigen Variablen und einer univariaten abhängigen Variablen. Typische in diesem Szenario eingesetzte Inferenzverfahren sind die Bestimmung von *multiplen und partiellen Korrelationen* zwischen x_1, \dots, x_m und y_1 , die Durchführung von *multiplen Regressionsanalysen* und *Kovarianzanalysen* und generell alle datenanalytischen Spezialfälle des *Allgemeinen Linearen Modells*.

Tabelle 12.2. Multivariate unabhängige Variable x und univariete abhängige Variable y

x_1	\dots	x_m	y_1
x_{11}	\dots	x_{1m}	y_{11}
\vdots	\ddots	\vdots	\vdots
x_{n1}	\dots	x_{nm}	y_{nm}

Tabelle 12.3 zeigt das im Kontext von *Einstichproben- T^2 -Tests*, der *einfaktoriellen multivariaten Varianzanalyse* und vielen Szenarien der *prädiktiven Modellierung* relevante Szenario. In diesem Fall ist die unabhängige Variable univariat und kodiert kategorial ein Faktorlevel bzw. eine Gruppenzugehörigkeit, während die abhängige Variable multivariat ist. Insbesondere in prädiktiven Modellierung wird die unabhängige Variable in diesem Kontext auch als *Targetvariable* oder *Label* und die Komponenten der abhängigen Variable als *Features* bezeichnet.

Tabelle 12.3. Univariante unabhängige Variable x und multivariate abhängige Variable y

x_1	y_1	\cdots	y_m
x_{11}	y_{11}	\cdots	y_{1m}
\vdots	\vdots	\vdots	\vdots
x_{n1}	y_{n1}	\cdots	y_{nm}

Tabelle 12.4 schließlich zeigt das Szenario einer multivariaten unabhängigen Variablen und einer multivariaten abhängigen Variablen. Dies ist das datenanalytische Szenario, das wir im Rahmen der *Kanonischen Korrelationsanalyse* genauer betrachten wollen und das generell durch das *Multivariate Allgemeine Lineare Modell* abgebildet und datenanalytisch behandelt werden kann.

Tabelle 12.4. Multivariate unabhängige Variable x und multivariate abhängige Variable y

x_1	\cdots	x_{m_x}	y_1	\cdots	y_{m_y}
x_{11}	\cdots	x_{1m_x}	y_{11}	\cdots	y_{1m_y}
\vdots	\ddots	\vdots	\vdots	\ddots	\vdots
x_{n1}	\cdots	x_{nm_x}	y_{n1}	\cdots	y_{nm_y}

12.2. Deskriptivstatistiken

Wir wollen nun einige Standarddeskriptivstatistiken zur Beschreibung multivariater Datensätze diskutieren. Dazu verallgemeinern wir zunächst die aus der univariaten Deskriptivstatistik bekannten Begriffe des Stichprobenmittels und der Stichprobenvarianz und betrachten dann mit den Mahalanobis-Distanzen multivariate Maße für das Verhältnis von Signal zu Rauschen. Wie immer entwickeln sich diese Begriffe vor dem Hintergrund der Annahme, dass es sich bei beobachteten Daten um Realisierungen entsprechender Zufallsvektoren handelt. Im Gegensatz zu der in Kapitel 12.1 betrachteten und aus dem empirischen Kontext bekannten Organisation von Daten experimenteller Einheiten in Zeilen und ihrer jeweiligen abhängigen Variablenkomponenten in Spalten ist dabei eine Organisation der zu einer experimentellen Einheit gehörenden Variablenkomponenten in Spaltenform zielführender und mit den Schreibweisen des univariaten Falles konsistenter.

Stichprobenmittel, -kovarianzmatrix, und -korrelationsmatrix

Definition 12.1 (Stichprobenmittel, -kovarianzmatrix und -korrelationsmatrix). v_1, \dots, v_n sei eine Menge von m -dimensionalen Zufallsvektoren, genannt *Stichprobe*.

- Das *Stichprobenmittel* der v_1, \dots, v_n ist definiert als der m -dimensionale Vektor

$$\bar{v} := \frac{1}{n} \sum_{i=1}^n v_i. \quad (12.1)$$

- Die *Stichprobenkovarianzmatrix* der v_1, \dots, v_n ist definiert als die $m \times m$ Matrix

$$C := \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})(v_i - \bar{v})^T. \quad (12.2)$$

- Die *Stichprobenkorrelationsmatrix* der v_1, \dots, v_n ist definiert als die $m \times m$ Matrix

$$D := \left(\frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \right)_{1 \leq i, j \leq m}. \quad (12.3)$$

•

Ohne Beweis halten wir fest, dass analog zum univariaten Fall das Stichprobenmittel bei unabhängig und identisch verteilten Zufallsvektoren v_1, \dots, v_n ein unverzerrter Schätzer des Stichprobenvariablenenerwartungswerts $\mathbb{E}(v_i) \in \mathbb{R}^m, i = 1, \dots, n$ ist. Ebenso ist in diesem Fall die Stichprobenkovarianzmatrix ein unverzerrter Schätzer der Stichprobenvariablenkovarianzmatrix $\mathbb{C}(v_i) \in \mathbb{R}^m, i = 1, \dots, n$. Zur konkreten Berechnung von Stichprobenmittel, Stichprobenkovarianzmatrix und Stichprobenkorrelationsmatrix basierend auf einem Datensatz bieten sich die Aussagen des folgenden Theorems an.

Theorem 12.1 (Datenmatrix und Stichprobenstatistiken).

Es sei

$$\Upsilon := \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \quad (12.4)$$

eine $m \times n$ Datenmatrix, die durch die spaltenweise Konkatenation von n m -dimensionalen Zufallsvektoren v_1, \dots, v_n gegeben sei. Dann ergeben sich

- für das Stichprobenmittel

$$\bar{v} = \frac{1}{n} \Upsilon \mathbf{1}_n, \quad (12.5)$$

- für die Stichprobenkovarianzmatrix

$$C = \frac{1}{n-1} \left(\Upsilon \left(I_n - \frac{1}{n} \mathbf{1}_{nn} \right) \Upsilon^T \right), \quad (12.6)$$

- und für Stichprobenkorrelationsmatrix mit

$$D := \text{diag} \left(\sqrt{C_{y_{ii}}}^{-1}, i = 1, \dots, m \right), \quad (12.7)$$

dass

$$R = DCD. \quad (12.8)$$

◦

Beweis. Die Darstellung des Stichprobenmittels ergibt sich aus

$$\begin{aligned} \bar{v} &:= \frac{1}{n} \sum_{i=1}^n v_i \\ &= \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n v_{i1} \\ \vdots \\ \sum_{i=1}^n v_{im} \end{pmatrix} \\ &= \frac{1}{n} \left(\begin{pmatrix} v_{11} & \cdots & v_{n1} \\ \vdots & \ddots & \vdots \\ v_{1m} & \cdots & v_{nm} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right) \\ &= \frac{1}{n} \Upsilon \mathbf{1}_n. \end{aligned} \quad (12.9)$$

Hinsichtlich der Darstellung der Stichprobenkovarianzmatrix halten wir zunächst fest, dass nach Definition gilt, dass

$$\begin{aligned} C &:= \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})(v_i - \bar{v})^T \\ &= \frac{1}{n-1} \sum_{i=1}^n (v_i v_i^T - v_i \bar{v}^T - \bar{v} v_i^T + \bar{v} \bar{v}^T) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n v_i v_i^T - \sum_{i=1}^n v_i \bar{v}^T - \sum_{i=1}^n \bar{v} v_i^T + \sum_{i=1}^n \bar{v} \bar{v}^T \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n v_i v_i^T - n \bar{v} \bar{v}^T - n \bar{v} \bar{v}^T + n \bar{v} \bar{v}^T \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n v_i v_i^T - n \bar{v} \bar{v}^T \right). \end{aligned} \quad (12.10)$$

Mit $\mathbf{1}_n \mathbf{1}_n^T = \mathbf{1}_{nn}$ ergibt sich dann weiterhin

$$\begin{aligned} \Upsilon \left(I_n - \frac{1}{n} \mathbf{1}_{nn} \right) \Upsilon^T &= \left(\Upsilon I_n - \frac{1}{n} \Upsilon \mathbf{1}_{nn} \right) \Upsilon^T \\ &= \Upsilon \Upsilon^T - \frac{1}{n} \Upsilon \mathbf{1}_{nn} \Upsilon^T \\ &= (v_1 \quad \cdots \quad v_n) \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix} - \frac{1}{n} \Upsilon \mathbf{1}_n \mathbf{1}_n^T \Upsilon^T \\ &= \sum_{i=1}^n v_i v_i^T - n \left(\frac{1}{n} \Upsilon \mathbf{1}_n \right) \left(\frac{1}{n} \mathbf{1}_n^T \Upsilon^T \right) \\ &= \sum_{i=1}^n v_i v_i^T - n \bar{v} \bar{v}^T \\ &= \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})(v_i - \bar{v})^T \\ &= C. \end{aligned} \quad (12.11)$$

Hinsichtlich der Korrelationsmatrix ergibt sich nach Definition und für ein beliebiges Indexpaar i, j mit $1 \leq i, j \leq m$ schließlich, dass

$$\begin{aligned} R_{y_{ij}} &= \frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \\ &= \frac{1}{\sqrt{(C)_{ii}}}(C)_{ij} \frac{1}{\sqrt{(C)_{jj}}} \\ &= (DCD)_{ij}. \end{aligned} \tag{12.12}$$

□

Folgender **R** Code wendet die in Theorem 12.1 diskutierten Resultate auf einen Beispieldatensatz mit Datendimensionalität $m = 4$ und Anzahl experimenteller Einheiten $n := 12$ an.

```

1 # Simulation eines Datensatzes
2 library(MASS) # multivariate Normalverteilung
3 m = 4 # Datenvektordimension
4 n = 12 # Anzahl Datenvektoren
5 mu = matrix(c(1,2,3,4)) # Erwartungswertparameter
6 Sigma = diag(4) # Kovarianzmatrixparameter
7 Y = t(mvrnorm(n, mu, Sigma)) # Datensatzrealisierung
8 print(Y)

      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 0.380757 3.206102 0.7449730 -0.4244947 0.8556004 1.207538 3.3079784
[2,] 1.681932 1.070638 0.5125397 0.9248077 3.0000288 1.378733 0.6155732
[3,] 1.085641 4.176583 1.3350276 2.5364696 1.8840799 2.249181 5.0871665
[4,] 3.318340 3.675730 4.0601604 3.4111055 4.5314962 2.481606 4.3065579
      [,8] [,9] [,10] [,11] [,12]
[1,] 1.105802 1.456999 0.9228471 0.6659992 0.965274
[2,] 3.869291 2.425100 1.7613529 3.0584830 2.886423
[3,] 3.017396 1.713699 1.3593945 3.4501871 2.981440
[4,] 2.463550 3.699024 3.4717201 3.3479052 3.943103

1 # Auswertung von Deskriptivstatistiken
2 n = ncol(Y) # Anzahl Datenvektorealisierungen
3 I_n = diag(n) # Einheitsmatrix I_n
4 J_n = matrix(rep(1,n^2), nrow = n) # 1_{nn}
5 y_bar = (1/n)*Y %*% J_n[,1] # Stichprobenmittel
6 C = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y)) # Stichprobenkovarianzmatrix
7 D = diag(1/sqrt(diag(C))) # Kov-Korr-Transformationsmatrix
8 R = D %*% C %*% D # Stichprobenkorrelationsmatrix
9
10 # Ausgabe
11 print(y_bar)

      [,1]
[1,] 1.199615
[2,] 1.932075
[3,] 2.573022
[4,] 3.559191

1 print(C)

      [,1] [,2] [,3] [,4]
[1,] 1.1450788 -0.29291622 0.91837375 0.16929650
[2,] -0.2929162 1.20170927 -0.09630563 -0.16923746
[3,] 0.9183737 -0.09630563 1.50573951 0.08718263
[4,] 0.1692965 -0.16923746 0.08718263 0.40266165

1 print(R)

      [,1] [,2] [,3] [,4]
[1,] 1.0000000 -0.24970431 0.69940198 0.2493218
[2,] -0.2497043 1.00000000 -0.07159407 -0.2432913
[3,] 0.6994020 -0.07159407 1.00000000 0.1119657
[4,] 0.2493218 -0.24329134 0.11196568 1.0000000

```

Mahalanobis-Distanzen

Abschließend wollen wir mit den sogenannten *Mahalanobis-Distanzen* multivariate Generalisierungen von aus der univariaten Anwendung bekannten Signal-zu-Rauschen-Maßen betrachten. Wir definieren den Begriff der Mahalanobis-Distanz wie folgt.

Definition 12.2 (Mahalanobis-Distanzen). ξ_1 sei ein Zufallsvektor, eine Realisation eines Zufallsvektors, ein multivariater Erwartungswert oder ein multivariates Stichprobenmittel, ξ_2 sei ein Zufallsvektor, eine Realisation eines Zufallsvektors, ein multivariater Erwartungswert oder ein multivariates Stichprobenmittel und Ξ sei eine Kovarianzmatrix oder eine Stichprobenkovarianzmatrix. Dann heißt

$$D = (\xi_1 - \xi_2)^T \Xi^{-1} (\xi_1 - \xi_2) \quad (12.13)$$

Mahalanobis-Distanz von ξ_1 und ξ_2 hinsichtlich Ξ .

•

Eine Mahalanobis-Distanz ist damit eine durch eine Kovarianzmatrix normalisierte quadrierte Euklidische Distanz (vgl. Kapitel 8.2). Ähnliche Maße für das Verhältnis eines Abstandes und einer Variabilität sind bekanntlich die *z-Transformation* $z = (y - \mu)/\sigma$ für $y \in \mathbb{R}$ und die Parameter $\mu, \sigma^2 > 0$ einer univariaten Normalverteilung sowie *Cohen's d* $d = (\bar{v}_1 - \bar{v}_2)/s_{12}$ für zwei Stichprobenmittel \bar{v}_1 und \bar{v}_2 und ihre korrespondierende gepoolte Stichprobenstandardabweichung s_{12} . Im Unterschied zur Mahalanobis-Distanz sind diese Maße allerdings nicht quadriert und damit Vorzeichen behaftet. In Analogie zur *z-Transformation* oder zu Cohen's *d* wird allerdings auch bei Mahalanobis-Distanzen ein Abstand in Einheiten von Variabilität gemessen. Bei Cohen's *d* bedeutet ja ein Wert von $d = 1$ gerade, dass der Abstand von \bar{v}_1 und \bar{v}_2 eine gepoolte Standardabweichung beträgt. Ebenso verhält es sich mit den Mahalanobis-Distanzen.

Anhand von Abbildung 12.1 und Abbildung 12.2 wollen wir den Einfluss der Varianz und der Kovarianz von Komponenten der ξ_1 und ξ_2 auf ihre Mahalanobis-Distanz noch etwas genauer betrachten. Die Titel der Unterabbildungen von Abbildung 12.1 zeigen die Mahalanobis-Distanzen der Vektoren $\xi_1 := (-1, -1)^T$ und $\xi_2 := (1, 1)^T$ bei Kovarianzmatrizen von

$$\Sigma_1 := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \Sigma_2 := \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \text{ und } \Sigma_3 := \begin{pmatrix} 1.5 & 0.0 \\ 0.0 & 1.5 \end{pmatrix}, \quad (12.14)$$

die mithilfe von Normalverteilungsisokonturen dargestellt sind. Für Σ_1 entspricht die Mahalanobis-Distanz dabei der quadrierten Euklidischen Distanz von ξ_1 und ξ_2 . An der Darstellung zu Σ_2 erkennt man, dass im Fall sphärischer Kovarianzmatrizen eine geringere Komponentenvarianz von ξ_1 und ξ_2 zu einer größeren Mahalanobis-Distanz führt. Umgekehrt erkennt man an der Darstellung zu Σ_3 , dass im Fall sphärischer Kovarianzmatrizen eine höhere Komponentenvarianz von ξ_1 und ξ_2 in einer kleineren Mahalanobis-Distanz resultiert. Intuitiv nähert die Komponentenvarianz die Komponenten also an.

Die Titel der Unterabbildungen von Abbildung 12.2 zeigen die Mahalanobis-Distanzen derselben Vektoren bei Kovarianzmatrizen von

$$\Sigma_1 := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}, \Sigma_2 := \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix} \text{ und } \Sigma_3 := \begin{pmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{pmatrix}, \quad (12.15)$$

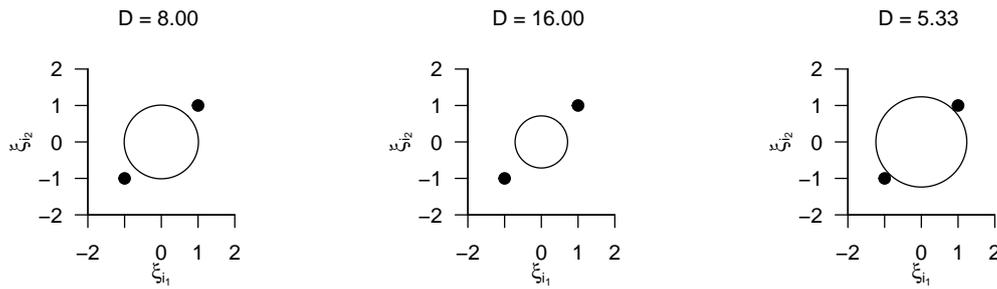


Abbildung 12.1. Mahalanobis-Distanzen als Funktion von Komponentenvarianzen

Für Σ_1 entspricht dabei wiederum die die Mahalanobis-Distanz der quadrierten Euklidischen Distanz von ξ_1 und ξ_2 . An der Darstellung zu Σ_2 erkennt man, dass eine stark positive Kovarianz der Komponenten von ξ_1 und ξ_2 in einer kleineren Mahalanobis-Distanz resultiert. Umgekehrt erkennt man an der Darstellung zu Σ_3 , dass eine stark negative Kovarianz der Komponenten von ξ_1 und ξ_2 zu einer größeren Mahalanobis-Distanz führt. Intuitiv nähert also auch die Kovarianz von Komponenten ξ_1 und ξ_2 an. Alternativ kann man die Höhe einer Mahalanobis-Distanz dabei auch als ein Maß für die Unwahrscheinlichkeit der Realisierung zweier Werte eines Zufallsvektors bei einer gegebenen Kovarianzmatrix verstehen.

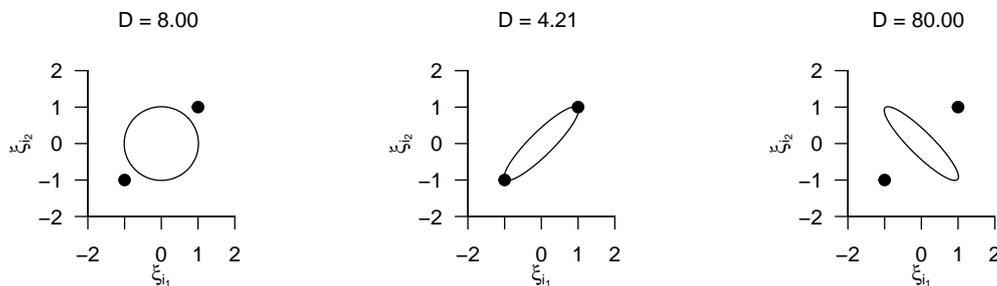


Abbildung 12.2. Mahalanobis-Distanzen als Funktion von Komponentenkovarianzen

12.3. Literaturhinweise

Die Resultate zur Matrixdarstellung von Stichprobenmittel, Stichprobenkovarianzmatrix und Stichprobenkorrelationsmatrix folgen Rencher & Christensen (2012). Der Begriff der Mahalanobis-Distanz geht zurück auf Mahalanobis (1936).

12.4. Selbstkontrollfragen

1. Erläutern Sie vier prinzipielle Datenanalyseszenarien anhand der Dimensionalität ihrer unabhängigen und abhängigen Variablen. Nennen Sie Beispiele für die in dem jeweiligen Szenario häufig eingesetzten Datenanalyseverfahren.

2. Geben Sie die Definition des Stichprobenmittels wieder.
3. Geben Sie die Definition der Stichprobenkovarianzmatrix wieder.
4. Geben Sie die Definition der Stichprobenkorrelationsmatrix wieder.
5. Geben Sie das Theorem zu Datenmatrix und Stichprobenstatistiken wieder.
6. Geben Sie die Definition einer Mahalanobis-Distanz wieder.
7. Erläutern Sie die intuitive Bedeutung einer Mahalanobis-Distanz.

13. Einstichproben-T²-Tests

13.1. Anwendungsszenario

Wie im univariaten Fall ist das Anwendungsszenario eines Einstichproben-T²-Tests dadurch gekennzeichnet, dass n Datenpunkte einer Stichprobe (Gruppe) randomisierter experimenteller Einheiten betrachtet werden. In Generalisierung des univariaten Falls sind die n Datenpunkte allerdings multivariat, jeder Datenpunkt besteht also aus zwei oder mehr Zahlen und kann als Vektor in \mathbb{R}^m mit $m > 1$ betrachtet werden. In Analogie zum univariaten Fall wird von den n Datenpunkten angenommen, dass sie Realisierungen von n unabhängigen und identisch multivariat normalverteilten Zufallsvektoren sind. Hinsichtlich der identischen multivariaten Normalverteilung $N(\mu, \Sigma)$ dieser Zufallsvektoren wird angenommen, dass sowohl der wahre Erwartungswertparameter μ als auch der wahre Kovarianzmatrixparameter Σ unbekannt sind. Schließlich wird vorausgesetzt, dass ein Interesse an einem inferentiellen Vergleich des wahren, aber unbekanntem, Erwartungswertparameters μ mit einem vorgegebenen Wert μ_0 , beispielsweise $\mu_0 := 0_m$, besteht. Wie im univariaten Fall ergeben sich auch in diesem Anwendungsszenario eine Reihe möglicher Hypothesenszenarien mit jeweils unterschiedlichen Testgütefunktionen und damit Herangehensweisen an Testumfangkontrolle und Stichprobengrößenoptimierung. Wir wollen im Rahmen dieser Einführung nur das Szenario einer einfachen Nullhypothese und einer zusammengesetzten Alternativhypothese,

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R}^m \setminus \{\mu_0\} \quad (13.1)$$

genauer untersuchen.

Anwendungsbeispiel

Für ein konkretes Anwendungsbeispiel betrachten wir die Analyse simulierter Prä-Post-Interventions-Differenzwerte von BDI Scores (dBDI) und Glukokortikoidplasmaleveln (dGLU), die, wie in Tabelle 14.1 dargestellt, an einer Gruppe von $n = 20$ Patient:innen erhoben worden sein könnten. Positive Werte von dBDI und dGLU entsprächen dabei einer Reduktion der Depressionssymptomatik, negative Werte zeigen eine Verschlechterung des Depressionszustandes an.

Bei der Anwendung eines Einstichproben-T²-Tests auf die Daten dieses simulierten Datensatzes nehmen wir an, dass die zweidimensionalen Datenvektoren (dBDI, dGLU) Realisierungen von $n = 20$ unabhängig normalverteilten zweidimensionalen Zufallsvektoren $v_i \sim N(\mu, \Sigma)$ sind. Wir nehmen weiterhin an, dass wir daran interessiert sind, unsere Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekanntem, Erwartungswertparameters $\mu \in \mathbb{R}^2$ mit einem Vergleichswert $\mu_0 \in \mathbb{R}^2$, etwa einem Therapieerfolgsnormwert, zu quantifizieren.

Tabelle 13.1. Prä-Post-Interventions-Differenzwerte von BDI Scores und Glukokortikoidplasmaleveln von $n = 20$ Patient:innen

dBDI	dGLU
35	6.1
25	4.0
20	1.7
29	2.6
29	1.9
17	0.9
33	2.0
28	4.1
26	3.9
31	3.8
14	2.1
18	2.0
19	5.0
28	2.6
20	2.1
35	4.4
28	4.0
32	3.9
32	1.0
25	1.9

Unabhängig von diesem inferenzstatistischen Vorgehen betrachten wir zunächst zu diesem Datensatz einige Deskriptivstatistiken wie durch folgenden **R** Code ausgewertet und in Abbildung 13.1 dargestellt. Verglichen mit einem Therapienormwert von $\mu_0 := (30, 3.5)^T$ fallen die Komponenten des Stichprobenmittels mit $\bar{v} = (26.3, 3.0)^T$ etwas geringer aus, allerdings bei einer nicht zu vernachlässigen Datenvariabilität, die sich in einer Mahalanobisdistanz von $D = 0.4$ des Stichprobenmittels vom Therapienormwert in bezug auf die Stichprobenkovarianzmatrix des Datensatzes widerspiegelt.

```

1 # Datendesektion
2 D = read.csv("_data/202-Einstichproben-T2-Tests.csv") # Daten einlesen
3 Y = rbind(D$y_1i, D$y_2i) # Datenmatrix
4 mu_0 = matrix(c(30,3.5), nrow = 2) # Normwert
5 n = ncol(Y) # Anzahl Datenpunkte
6 j_n = matrix(rep(1,n), nrow = n) # 1_n
7 I_n = diag(n) # I_n
8 J_n = matrix(rep(1,n^2), nrow = n) # 1_{nn}
9 Y_bar = (1/n)*Y %*% j_n # Stichprobenmittel
10 C = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y)) # Stichprobenkovarianzmatrix
11 D = t(Y_bar - mu_0) %*% solve(C) %*% (Y_bar - mu_0) # Mahalanobis Distanz
12 cat("Y_bar =", Y_bar, "\nD = ", D) # Ausgabe

```

```

Y_bar = 26.25615 2.991039
D = 0.3773184

```

13.2. Modellformulierung und Modellevaluation

Wir definieren zunächst das Einstichproben-T²-Test-Modell wie folgt.

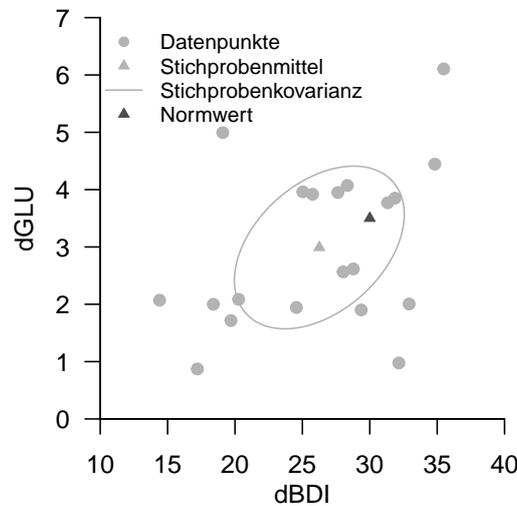


Abbildung 13.1. Deskriptivstatistiken der dBDI, dGLU Daten des Beispieldatensatzes. Jeder Punkt visualisiert die Daten einer Patient:in, die Stichprobenkovarianz ist durch die 0.4 Isokontur einer zweidimensionalen Normalverteilung mit Erwartungswertparameter und Kovarianzmatrixparameter entsprechend dem Stichprobenmittel und der Stichprobenkovarianz dargestellt

Definition 13.1 (Einstichproben- T^2 -Test-Modell). Für $i = 1, \dots, n$ seien v_i m -dimensionale Zufallsvektoren, die die n Datenpunkte eines Einstichproben- T^2 -Test Szenarios modellieren. Dann hat das *Einstichproben- T^2 -Test-Modell* die strukturelle Form

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0_m, \Sigma) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R}^m, \Sigma \in \mathbb{R}^{m \times m} \text{ pd} \quad (13.2)$$

und die Datenverteilungsform

$$v_i \sim N(\mu, \Sigma) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ mit } \mu \in \mathbb{R}^m, \Sigma \in \mathbb{R}^{m \times m} \text{ pd.} \quad (13.3)$$

•

Die Äquivalenz von struktureller Form und Datenverteilungsform des Einstichproben- T^2 -Test-Modells folgt dabei direkt mit Theorem 11.8 durch Transformation der Zufallsvektoren ε_i anhand von Gleichung 13.2.

13.3. Modellevaluation

Teststatistik und Test

Wir definieren als nächstes eine Teststatistik für das Einstichproben- T^2 -Test Szenario.

Definition 13.2 (Einstichproben- T^2 -Teststatistik). Gegeben seien das Einstichproben- T^2 -Test-Modell und ein Nullhypothesenparameter $\mu_0 \in \mathbb{R}^m$. Dann ist die Einstichproben- T^2 -Teststatistik definiert als

$$T^2 := n(\bar{v} - \mu_0)^T C^{-1}(\bar{v} - \mu_0), \tag{13.4}$$

wobei \bar{v} und C das Stichprobenmittel und die Stichprobenkovarianzmatrix der v_1, \dots, v_n bezeichnen.

•

Die Einstichproben- T^2 -Teststatistik ist offenbar die mit dem Stichprobenumfang n skalierte Mahalanobis-Distanz von \bar{v} und μ_0 hinsichtlich C (vgl. Kapitel 12.2). Damit gilt entsprechend, dass bei konstanter Stichprobenkovarianzmatrix die Einstichproben- T^2 -Test Teststatistik T^2 größere Werte für eine größere Euklidische Distanz von \bar{v} und μ_0 annimmt und bei konstanter Euklidischer Distanz von \bar{v} und μ_0 der Wert der Teststatistik T^2 von der Höhe der Datenvariabilität abhängt. Hinsichtlich der Verteilung der Einstichproben- T^2 -Teststatistik halten wir zunächst folgendes Theorem fest, das wir nicht beweisen wollen.

Theorem 13.1 (Verteilung der skalierten Einstichproben- T^2 -Teststatistik). *Es seien $v_1, \dots, v_n \sim N(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^m$ und $\Sigma \in \mathbb{R}^{m \times m}$ pd,*

$$\nu := \frac{n - m}{(n - 1)m} \tag{13.5}$$

und für $\mu \in \mathbb{R}^m$ sei die Einstichproben- T^2 -Teststatistik definiert als

$$T^2 := n(\bar{v} - \mu_0)^T C^{-1}(\bar{v} - \mu_0). \tag{13.6}$$

Dann gilt

$$\nu T^2 \sim f(\delta, m, n - m), \tag{13.7}$$

wobei $f(\delta, m, n - m)$ die nichtzentrale f -Verteilung mit Nichtzentralitätsparameter

$$\delta := n(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0) \tag{13.8}$$

sowie mit Freiheitsgradparametern m und $n - m$ bezeichnet.

◦

Für einen Beweis von Theorem 13.1 verweisen wir auf Hotelling (1931) und Anderson (2003). Wir erinnern in diesem Zusammenhang an die Begriffe der f -Zufallsvariable und der nichtzentralen f -Zufallsvariable, für die wir exemplarische WDFen in Abbildung 13.2 und Abbildung 13.3 darstellen.

Definition 13.3 (f -Zufallsvariable). ξ sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p_\xi : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p_\xi(x) := \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1}{2} - 1}}{(\nu_1 x + \nu_2)^{\frac{\nu_1 + \nu_2}{2}}}, \tag{13.9}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass ξ einer f -Verteilung mit Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen ξ eine f -Zufallsvariable mit

Freiheitsgradparametern ν_1 und ν_2 . Wir kürzen dies mit $\xi \sim f(\nu_1, \nu_2)$ ab. Die WDF einer f -Zufallsvariable bezeichnen wir mit $f(x; \nu_1, \nu_2)$, die KVF einer f -Zufallsvariable bezeichnen wir mit $F(x; \nu_1, \nu_2)$, und die inverse KVF einer f -Zufallsvariable bezeichnen wir mit $F^{-1}(x; \nu_1, \nu_2)$.

•

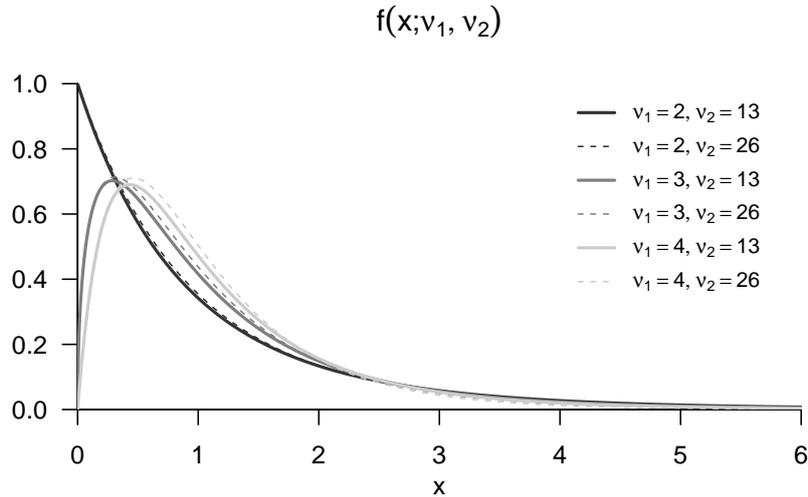


Abbildung 13.2. Exemplarische WDFen einer f -Zufallsvariable

Definition 13.4 (Nichtzentrale f -Zufallsvariable). ξ sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p_\xi : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p_\xi(x) := \sum_{k=0}^{\infty} \frac{e^{-\delta/2} (\delta/2)^k}{\frac{\Gamma(\nu_2/2)\Gamma(\nu_1/2+k)}{\Gamma(\nu_2/2+\nu_1/2+k)} k!} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2+k} \left(\frac{\nu_2}{\nu_2 + \nu_1 x}\right)^{(\nu_1+\nu_2)/2+k} x^{\nu_1/2-1+k} \quad (13.10)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass ξ einer nichtzentralen f -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 unterliegt und nennen ξ eine nichtzentrale f -Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparametern ν_1 und ν_2 . Wir kürzen dies mit $\xi \sim f(\delta, \nu_1, \nu_2)$ ab. Die WDF einer f -Zufallsvariable bezeichnen wir mit $f(x; \delta, \nu_1, \nu_2)$, die KVF einer nichtzentralen f -Zufallsvariable bezeichnen wir mit $F(x; \delta, \nu_1, \nu_2)$, und die inverse KVF einer nichtzentralen f -Zufallsvariable bezeichnen wir mit $F^{-1}(x; \delta, \nu_1, \nu_2)$.

•

Im univariaten Fall sind bekanntlich die F -Statistiken der Varianzanalyse bei Zutreffen der Nullhypothese f -verteilt und bei Zutreffen der Alternativhypothese nichtzentral- f -verteilt. Für den Fall $\mu = \mu_0$, dass also der wahre, aber unbekannte Erwartungswertparameter mit dem Nullhypothesenparameter identisch ist, gilt nach Gleichung 13.8 $\delta = 0$ und $f(\delta, m, n - m)$ entspricht der f -Verteilung $f(m, n - m)$.

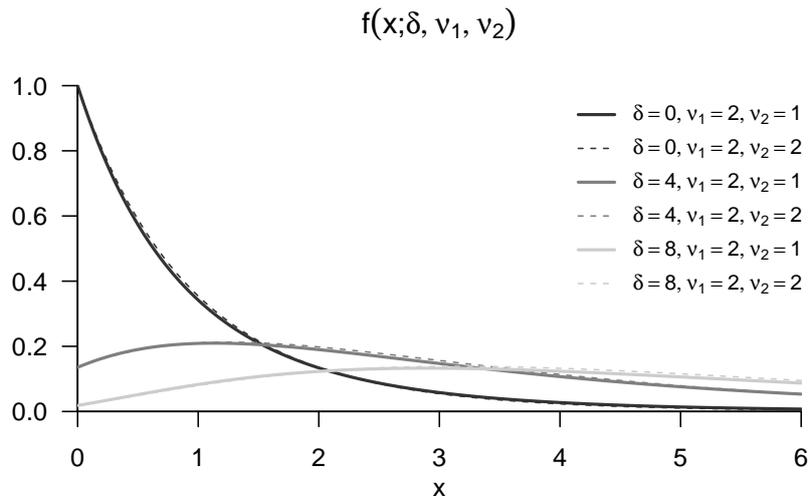


Abbildung 13.3. Exemplarische WDFen einer nichtzentralen f -Zufallsvariable

Wir halten weiterhin fest, dass aus Theorem 13.1 im univariaten Fall $m := 1$ aus Gleichung 13.5 folgt, dass

$$\nu = \frac{n - 1}{(n - 1) \cdot 1} = 1 \tag{13.11}$$

und mit der Stichprobenvarianz S^2 einer univariaten Stichprobe entsprechend folgt, dass

$$T^2 = n \frac{(\bar{v} - \mu_0)^2}{S^2} = \left(\sqrt{n} \frac{\bar{v} - \mu_0}{S} \right)^2. \tag{13.12}$$

Dies ist offenbar das Quadrat der bekannten univariaten Einstichproben-T-Test Teststatistik. Damit ist nach Theorem 13.1 das Quadrat der univariaten Einstichproben-T-Test Teststatistik nach $f(\delta, 1, n - 1)$ verteilt. Intuitiv und verkürzt ausgedrückt ist also eine quadrierte t -Zufallsvariable eine f -Zufallsvariable.

Aus Theorem 13.1 folgen nun für Einstichproben- T^2 -Teststatistik unmittelbar folgende Formen für ihre WDF und KVF.

Theorem 13.2 (WDF und KDF der Einstichproben- T^2 -Teststatistik). *Im Einstichproben- T^2 -Test Szenario sei*

$$\nu := \frac{n - m}{(n - 1)m} \tag{13.13}$$

Dann ist eine WDF der Einstichproben- T^2 -Teststatistik gegeben durch

$$p_{T^2} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, t^2 \mapsto p_{T^2}(t^2) := \nu f(\nu t^2; \delta, m, n - m) \tag{13.14}$$

und eine KVF der Einstichproben- T^2 -Teststatistik ist gegeben durch

$$P_{T^2} : \mathbb{R}_{\geq 0} \rightarrow [0, 1], t^2 \mapsto P_{T^2}(t^2) := F(\nu t^2; \delta, m, n - m) \tag{13.15}$$

◦

Beweis. Wir halten zunächst fest, dass das Theorem zur univariaten WDF Transformation bei linear-affinen Abbildungen besagt, dass für eine Zufallsvariable ξ mit WDF p_ξ und der Definition $v = f(\xi)$ mit $f(\xi) := a\xi + b$ für $a \neq 0$ eine WDF von v definiert ist durch $p_v(y) := (1/|a|)p_\xi((y-b)/a)$. Im vorliegenden Fall ist $\xi = \nu T^2$ mit WDF $f(\delta, m, n - m)$ und $v := T^2 = \frac{1}{\nu}\nu T^2$, also $a = 1/\nu$ und $b = 0$. Mit $\nu > 0$ ergibt sich (13.14) also aus

$$p_{T^2}(t^2) = \frac{1}{a} p_{\nu T^2} \left(\frac{t^2}{a} \right) = \nu f(\nu t^2; m, n - m). \tag{13.16}$$

(13.15) folgt dann damit, dass WDFen bei kontinuierlichen Zufallsvariablen die Ableitungen der entsprechenden KVfen sind. Mit der Kettenregel der Differentiation ergibt sich

$$\begin{aligned} \frac{d}{dt^2} P_{T^2}(t^2) &= \frac{d}{dt^2} (F(\nu t^2; m, n - m, \delta)) \\ &= \frac{d}{dt^2} F(\nu t^2; m, n - m, \delta) \frac{d}{dt^2} (\nu t^2) \\ &= \nu f(\nu t^2; m, n - m, \delta) \\ &= p_{T^2}(t^2). \end{aligned} \tag{13.17}$$

□

Wir halten fest, dass die skalierte Einstichproben- T^2 -Test Teststatistik νT^2 nach $f(\delta, m, n - m)$ nichtzentral f -verteilt ist, die WDF der Einstichproben- T^2 -Test Teststatistik T^2 selbst dagegen durch $\nu f(\nu t^2; \delta, m, n - m)$ gegeben ist. Wir simulieren diese Verteilung mithilfe folgenden **R** Codes und visualisieren diese Simulation in Abbildung 13.4.

```

1 # Modellparameter
2 m = 2 # Dimensionalität der Zufallsvektoren/Daten
3 n = 15 # Anzahl der Datenpunkte
4 mu_0 = matrix(c(1,1), nrow = 2) # Nullhypothesenparameter
5 mu = matrix(c(2,2), nrow = 2) # wahrer, aber unbekannter, Erwartungswertparameter
6 Sigma = matrix(c(0.5,0.3, 0.3,0.5), nrow = 2, byrow = TRUE) # wahrer, aber unbekannter, Kovarianzmatrixparameter
7
8 # Simulation
9 library(MASS) # R Paket für multivariate Normalverteilungen
10 nsim = 1e4 # Anzahl Simulationen/Datensatzrealisierungen
11 Yb = matrix(rep(NA,n*nsim), nrow = 2) # Stichprobenmittelarray
12 T2 = rep(NA,nsim) # Einstichproben-T^2-Teststatistik Array
13 j_n = matrix(rep(1,n), nrow = n) # I_n
14 I_n = diag(n) # I_n
15 J_n = matrix(rep(1,n^2), nrow = n) # I_{(nn)}
16 for(s in 1:nsim){
17 Y = t(mvnrnorm(n,mu,Sigma)) # \upsilon_{sim} i \sim N(\mu, \Sigma), i = 1, \dots, n
18 Y_bar = (1/n)*(Y %>% j_n) # Stichprobenmittel
19 C = (1/(n-1))*(Y %>% (I_n - (1/n)*J_n) %>% t(Y)) # Stichprobenkovarianzmatrix
20 T2[s] = n*t(Y_bar - mu_0) %>% solve(C) %>% (Y_bar - mu_0) # Einstichproben-T^2-Teststatistik
21 Yb[,s] = Y_bar # Stichprobenmittel für Visualisierung
22 }

```

Den Einstichproben- T^2 -Test definieren wir schließlich als einen kritischen Wert-basierten Test wie folgt.

Definition 13.5 (Einstichproben- T^2 -Test). Gegeben seien das Einstichproben- T^2 -Modell und die Einstichproben- T^2 -Teststatistik. Dann ist für einen kritischen Wert $k \geq 0$ der Einstichproben- T^2 -Test definiert als der kritische Wert-basierte Test

$$\phi(v) := 1_{\{T^2 > k\}} := \begin{cases} 1 & T^2 > k \\ 0 & T^2 \leq k \end{cases}. \tag{13.19}$$

•

In Definition 13.5 repräsentiert wie üblich $\phi(v) = 1$ den Vorgang des Ablehnens von H_0 und $\phi(v) = 0$ den Vorgang des Nichtablehnens von H_0 .

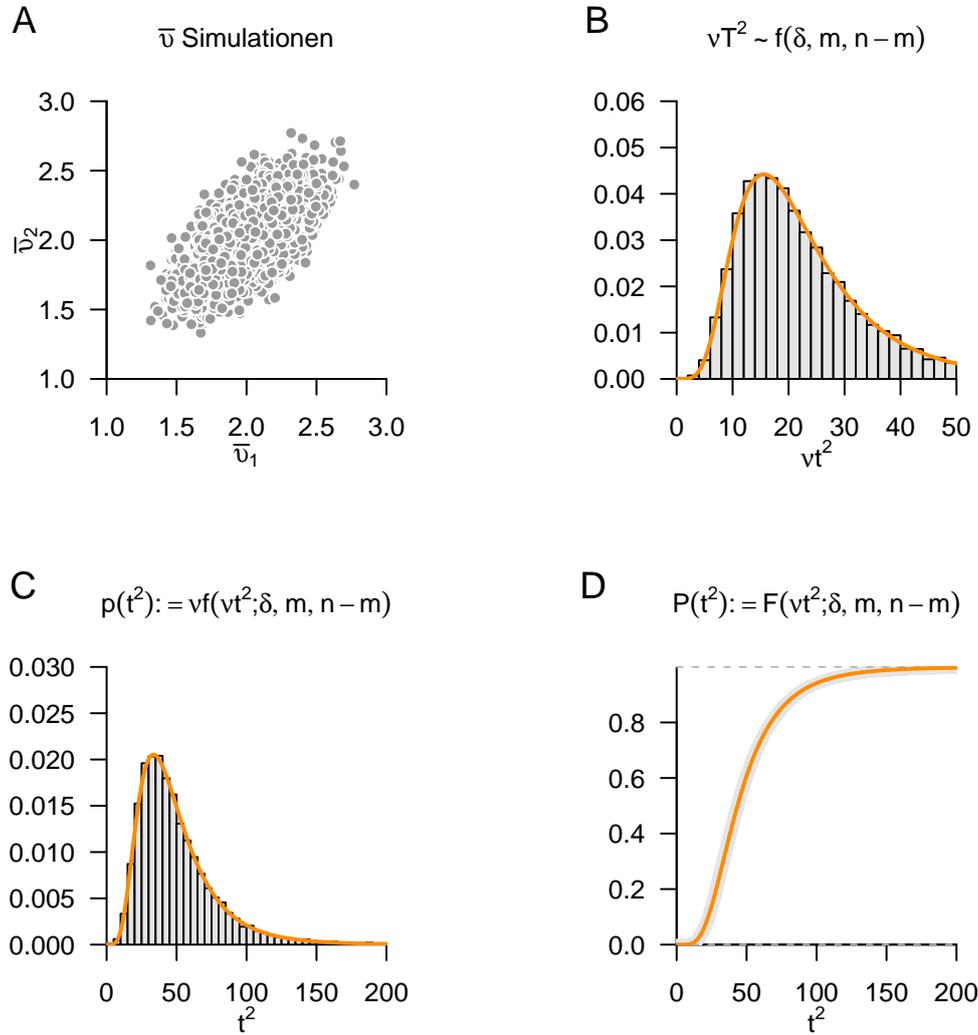


Abbildung 13.4. Verteilung der Einstichproben- T^2 -Teststatistik. **A** Stichprobenmittel von 10.000 Realisierungen eines Einstichproben- T^2 -Test-Modells mit $m := 2$ und $n := 15$ und wahren, aber unbekanntem, Parametern

$$\mu := (2, 2)^T \text{ und } \Sigma = \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}. \tag{13.18}$$

B Histogramm der entsprechenden Realisierungen der skalierten Einstichproben- T^2 -Teststatistik für $\mu_0 := (1, 1)^T$ (grau) und analytische Form dieser Verteilung (orange). **C** Histogramm der entsprechenden Realisierungen der Einstichproben- T^2 -Teststatistik (grau) und ihre analytische Form (orange). **D** Empirische KVF der entsprechenden Realisierungen der Einstichproben- T^2 -Teststatistik (grau) und ihre analytische Form (orange).

Analyse der Testgütefunktion

Um für den in Definition 13.5 definierten Test Prozeduren zur Testumfangkontrolle (Typ I Fehlerbegrenzung) und zur Stichprobengrößenoptimierung (Typ II Fehlerbegrenzung) zu entwickeln, betrachten wir zunächst seine Testgütefunktion. Es gilt folgendes Theorem.

Theorem 13.3 (Testgütefunktion des Einstichproben- T^2 -Tests). ϕ sei der Einstichproben- T^2 -Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R}^m \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - F(\nu k; \delta_\mu, m, n - m) \tag{13.20}$$

wobei $F(\cdot; \delta_\mu, m, n - m)$ die KVF der nichtzentralen f -Verteilung mit Freiheitsgradparametern m und $n - m$ sowie mit Nichtzentralitätsparameter

$$\delta_\mu := n(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) \tag{13.21}$$

bezeichnet.

◦

Beweis. Die Testgütefunktion des betrachteten Tests ist definiert als

$$q_\phi : \mathbb{R}^m \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \tag{13.22}$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt, gleich sind, benötigen wir also zunächst die Verteilung der Teststatistik. Wir haben oben aber bereits gesehen, dass

$$\frac{n - m}{m(n - 1)} T^2 \sim f(m, n - m, \delta_\mu) \text{ mit } \delta_\mu := n(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) \tag{13.23}$$

gilt. Der Ablehnungsbereich des betrachteten Tests ist $A :=]k, \infty[$. Also ergibt sich

$$\begin{aligned} q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\ &= \mathbb{P}_\mu(T^2 \in]k, \infty[) \\ &= \mathbb{P}_\mu(T^2 > k) \\ &= 1 - \mathbb{P}_\mu(T^2 \leq k) \\ &= 1 - F(\nu k; \delta_\mu, m, m - n) \end{aligned} \tag{13.24}$$

□

Wir wollen diese Testgütefunktion beispielhaft für zwei Szenarien mit $m := 2$ und $n := 15$ in Abhängigkeit des kritischen Wertes k betrachten. Abbildung 13.5 und Abbildung 13.6 visualisieren q_ϕ in diesen Szenarien für einen Nullhypothesenparameter $\mu_0 := (1, 1)^T$ und die wahren, aber unbekanntem, Kovarianzmatrixparameter

$$\Sigma_1 := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \text{ und } \Sigma_2 := \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix}, \tag{13.25}$$

respektive. In beiden Fällen und unabhängig von k resultiert eine größere Distanz des wahren, aber unbekanntem, Erwartungswertparameters μ vom Nullhypothesenparameter μ_0 in einer höheren Wahrscheinlichkeit dafür, dass der Test ϕ den Wert 1 annimmt, also die Nullhypothese abgelehnt wird. Die Zunahme dieser Wahrscheinlichkeit ist im ersten Szenario isotropisch, im zweiten dagegen aufgrund der Form des wahren, aber unbekanntem, Kovarianzparameters nicht. Bei einem kleinen kritischen Wert k werden hohe

Wahrscheinlichkeiten für eine Ablehnung der Nullhypothese schon bei geringen Distanzen zwischen μ und μ_0 erreicht, bei einem größeren kritischen Wert k dagegen erst für größere Distanzen. Untenstehender **R** Code demonstriert das Vorgehen zur Evaluation dieser Testgütefunktionen

```

1 # Modellparameter
2 m = 2 # m
3 n = 15 # n
4 nu = (n-m)/((n-1)*m) # \nu
5 Sigma = diag(m) # \Sigma = I_2
6 iSigma = solve(Sigma) # \Sigma^{-1}
7
8 # Testparameter
9 mu_0 = matrix(c(1,1), nrow = 2) # \mu_0
10 k_all = c(2,4,6) # k <-> \phi
11 n_k = length(k_all) # Anzahl k Werte/Tests
12
13 # q_\phi(\mu) Evaluation
14 mu_min = 0 # \mu_i Minimum
15 mu_max = 2 # \mu_i Maximum
16 mu_res = 1e3 # \mu_i Auflösung
17 mu_i = seq(mu_min, mu_max, len = mu_res) # \mu_i
18 q_phi = array(dim = c(mu_res, mu_res, length(k_all))) # q_\phi Array
19 for(k in 1:n_k){
20   for(i in 1:mu_res){
21     for(j in 1:mu_res){
22       mu = matrix(c(mu_i[i], mu_i[j]), nrow = 2) # \mu
23       delta_mu = n*(mu - mu_0) %*% iSigma %*% (mu - mu_0) # \delta_\mu
24       q_phi[i,j,k] = 1 - pf(nu*k_all[k], m, n-m, delta_mu)} } }

```

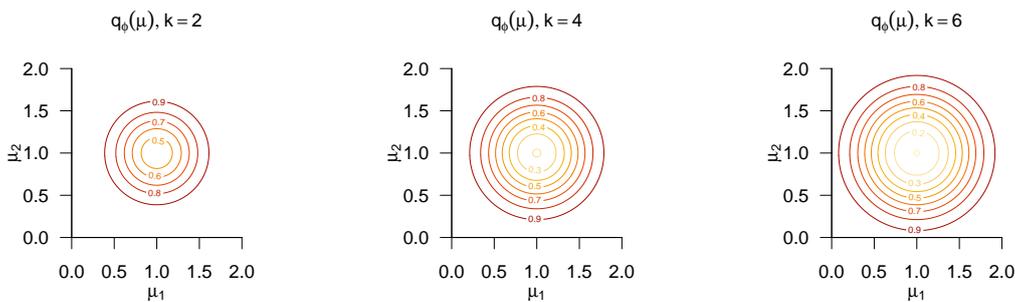


Abbildung 13.5. Einstichproben-T²-Test Testgütefunktionen für kritische Werte $k = 2$, $k = 4$ und $k = 6$ im Szenario

$$m := 2, n := 15, \mu_0 := (1, 1)^T, \Sigma_1 := \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix} \tag{13.26}$$

Testumfangkontrolle

Bekanntlich erlaubt die Testumfangkontrolle die Begrenzung der größtmöglichen Wahrscheinlichkeit für einen Typ I Fehler. Im aktuellen Testszenario haben wir folgendes Theorem.

Theorem 13.4 (Testumfangkontrolle des Einstichproben-T²- Tests). ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m), \tag{13.28}$$

wobei $\nu := (n - m)/((n - 1)m)$ und $F^{-1}(\cdot; m, n - m)$ die inverse KVF der f -Verteilung mit Freiheitsgradparametern m und $n - m$ ist.

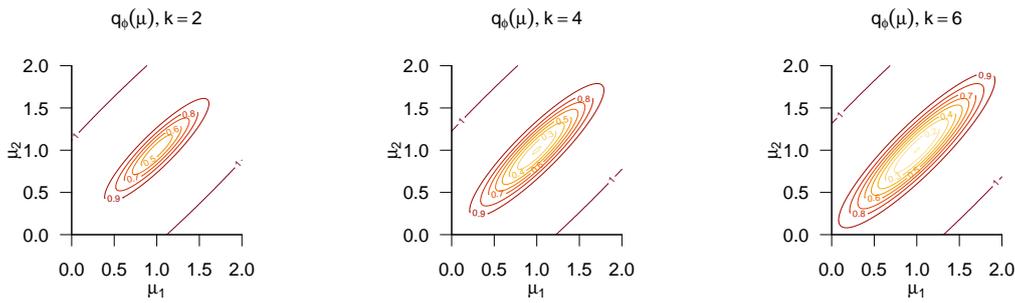


Abbildung 13.6. Einstichproben-T²-Test Testgütefunktionen für kritische Werte $k = 2$, $k = 4$ und $k = 6$ im Szenario

$$m := 2, n := 15, \mu_0 := (1, 1)^T, \Sigma_1 := \begin{pmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{pmatrix} \tag{13.27}$$

◦

Beweis. Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$ gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wird zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$q_\phi(\mu_0) = 1 - F(\nu k; \delta, m, n - m) = 1 - F(\nu k; 0, m, n - m) = 1 - F(\nu k; m, n - m), \tag{13.29}$$

wobei $F(\nu k; \delta, m, n - m)$ und $F(\nu k; m, n - m)$ die KVF der nichtzentralen f -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparametern m und $n - m$ sowie der f -Verteilung mit Freiheitsgradparametern m und $n - m$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned} q_\phi(\mu_0) &= 1 - F(\nu k_{\alpha_0}; m, n - m) \\ &= 1 - F(\nu \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m); m, n - m) \\ &= 1 - F(F^{-1}(1 - \alpha_0; m, n - m); m, n - m) \\ &= 1 - (1 - \alpha_0) = \alpha_0. \end{aligned} \tag{13.30}$$

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ gilt und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

□

Wir visualisieren die Wahl von

$$k_{\alpha_0} = \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m) \tag{13.31}$$

für den Fall $m = 2, n = 15$ und ein Signifikanzlevel von $\alpha_0 := 0.05$ in [Abbildung 13.7](#). Untenstehender **R** Code simuliert die Testumfangkontrolle für ein Einstichproben-T²-Test Szenario mit

$$m := 2, n := 15, \mu := \mu_0 := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ und } \Sigma := \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}. \tag{13.32}$$

Der auf Grundlage von 10^4 Datensatzrealisationen geschätzte Testumfang stimmt gut mit dem Signifikanzlevel überein.

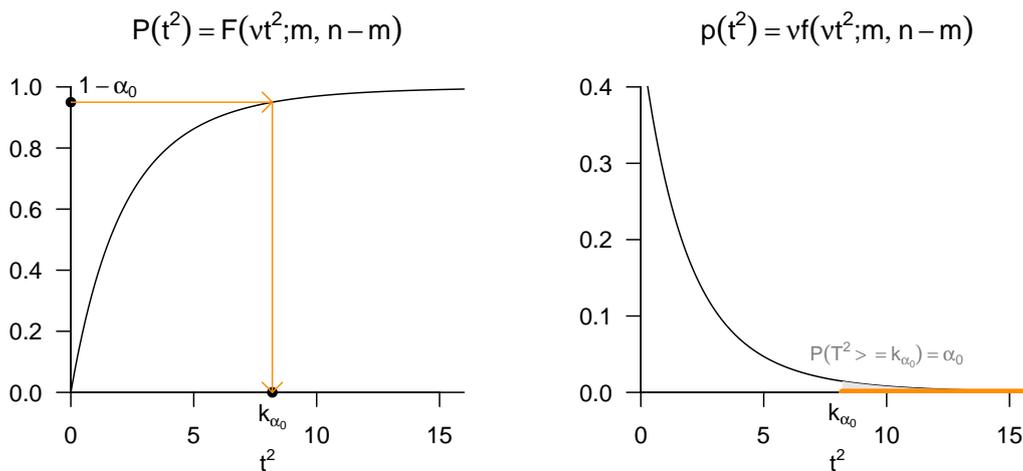


Abbildung 13.7. Testumfangkontrolle durch Selektion eines α_0 -abhängigen kritischen Wertes für den Einstichproben- T^2 -Test anhand von KVF und WDF der Einstichproben- T^2 -Teststatistik.

```

1 # Modellparameter
2 m = 2 # Dimensionalität der Zufallsvektoren/Daten
3 n = 15 # Anzahl der Datenpunkte
4 nu = (n-m)/(m*(n-1)) # Parameter
5 mu_0 = matrix(c(1,1), nrow = 2) # Nullhypotheseparameter
6 mu = mu_0 # w.a.u. Erwartungswertparameter bei Zutreffen von H0
7 Sigma = matrix(c(0.5,0.3, 0.3,0.5), nrow = 2, byrow = TRUE) # wahrer, aber unbekannter, Kovarianzmatrixparameter
8
9 # Testparameter
10 alpha_0 = 0.05 # Signifikanzlevel
11 k_alpha_0 = (1/nu)*qf(1-alpha_0, m,n-m) # kritischer Wert
12
13 # Simulation der Testumfangkontrolle
14 library(MASS) # R Paket für multivariate Normalverteilungen
15 nsim = 1e4 # Testentscheidungsarray
16 phi = rep(NaN,nsim) # Testentscheidungsarray
17 j_n = matrix(rep(1,n), nrow = n) # I_n
18 I_n = diag(n) # I_n
19 J_n = matrix(rep(1,n^2), nrow = n) # I_{nn}
20 for(s in 1:nsim){
21 Y = t(mvnrnorm(n,mu,Sigma)) # Simulationssituationen
22 Y_bar = (1/n)*(Y %>% j_n) # Y_i \sim N(mu,Sigma), i = 1,...,n
23 C = (1/(n-1))*(Y %>% (I_n-(1/n)*J_n) %>% t(Y)) # Stichprobenkovarianzmatrix
24 T2 = n*t(Y_bar - mu_0) %>% solve(C) %>% (Y_bar - mu_0) # Einstichproben-T^2-Test Statistik
25 if(T2 > k_alpha_0){ # Test 1_{T^2 >= k_alpha_0}
26 phi[s] = 1 # Ablehnen von H_0
27 } else {
28 phi[s] = 0} # Nicht Ablehnen von H_0
29 cat("\nKritischer Wert = ", k_alpha_0,
30 "\nGeschätzter Testumfang alpha = ", mean(phi)) # Ausgabe

```

Kritischer Wert = 8.196602
 Geschätzter Testumfang alpha = 0.0509

In der Praxis entsprechen obige Ergebnisse dann folgendem Vorgehen bei der Durchführung eines Einstichproben- T^2 -Tests. Man unterstellt, dass ein vorliegender Datensatz von m -dimensionalen Datenvektoren eine Realisation von n u.i.v. m -dimensionalen Zufallsvektoren $v_1, \dots, v_n \sim N(\mu, \Sigma)$ mit unbekanntem Parametern $\mu \in \mathbb{R}^m$ und $\Sigma \in \mathbb{R}^{m \times m}$ pd ist und möchte entscheiden, ob für ein $\mu_0 \in \mathbb{R}^m$ eher die Nullhypothese $H_0 : \mu = \mu_0$ oder die Alternativhypothese $H_1 : \mu \neq \mu_0$ zutrifft. Zu diesem Zweck wählt man zunächst ein Signifikanzlevel α_0 und bestimmt dann den zugehörigen kritischen Wert k_{α_0} . Beispielsweise gilt für $m = 2$ und $n = 15$ bei Wahl von $\alpha_0 := 0.05$, dass $k_{0.05} = \nu^{-1}F^{-1}(1 - 0.05; 2, 13) \approx 8.2$ ist. Anhand von m, n, μ_0 , dem Stichprobenmittel

\bar{v} und der Stichprobenkovarianzmatrix C berechnet man dann die Realisierung der Einstichproben- T^2 -Teststatistik

$$T^2 := n(\bar{v} - \mu_0)^T C^{-1}(\bar{v} - \mu_0). \tag{13.33}$$

Wenn das berechnete T^2 größer als k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht. Die oben entwickelte Theorie zur Testumfangkontrolle des Einstichproben- T^2 -Test garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

p-Wert

Wir erinnern daran, dass per Definition der p-Wert das kleinste Signifikanzlevel α_0 ist, bei welchem man die Nullhypothese basierend auf einem vorliegenden Wert der Teststatistik ablehnen würde. Wir haben folgendes Theorem

Theorem 13.5. *Für den p-Wert des in Definition 13.5 definierten Test gilt*

$$p\text{-Wert} = \mathbb{P}(T^2 \geq t^2) = 1 - F(\nu t^2; m, n - m). \tag{13.34}$$

◦

Beweis. Bei einem beobachteten Wert t^2 der Einstichproben- T^2 -Teststatistik T^2 würde H_0 für jedes α_0 mit $t^2 \geq \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt

$$\alpha_0 \geq \mathbb{P}(T^2 \geq t^2). \tag{13.35}$$

Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq \mathbb{P}(T^2 \geq t^2)$ ist dann $\alpha_0 = \mathbb{P}(T^2 \geq t^2)$, also folgt

$$p\text{-Wert} = \mathbb{P}(T^2 \geq t^2) = 1 - F(\nu t^2; m, n - m). \tag{13.36}$$

Es bleibt zu zeigen, dass gilt

$$\begin{aligned} t^2 &\geq \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m) \\ \Leftrightarrow \nu t^2 &\geq F^{-1}(1 - \alpha_0; m, n - m) \\ \Leftrightarrow \alpha_0 &\geq \mathbb{P}(T^2 \geq t^2). \end{aligned} \tag{13.37}$$

Dies aber folgt aus

$$\begin{aligned} t^2 &\geq \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m) \\ \nu t^2 &\geq F^{-1}(1 - \alpha_0; m, n - m) \\ F(\nu t^2; m, n - m) &\geq F(F^{-1}(1 - \alpha_0; m, n - m); m, n - m) \\ F(\nu t^2; m, n - m) &\geq 1 - \alpha_0 \\ \mathbb{P}(T^2 \leq t^2) &\geq 1 - \alpha_0 \\ \alpha_0 &\geq 1 - \mathbb{P}(T^2 \leq t^2). \end{aligned} \tag{13.38}$$

□

Zum Beispiel ergeben sich bei $m = 2$ und $n = 15$ der p-Wert für $t^2 = 7.00$ zu 0.071 und bei $m = 4$ und $n = 15$ der p-Wert für $t^2 = 7.00$ zu 0.304. Die gleiche Anzahl an Datenpunkten resultiert bei höherer Datendimensionalität also in einem höheren p-Wert. Weiterhin ergeben sich bei $m = 2$ und $n = 15$ der p-Wert für $t^2 = 9.00$ zu 0.040 und bei $m = 2$ und $n = 99$ der p-Wert für $t^2 = 7.00$ zu 0.035. Geringere Verhältnisse von geschätzter Nullhypothesenabweichung und geschätzter Daten(ko)varianz können also hinsichtlich des p-Wertes durch eine höhere Anzahl an Datenpunkten ausgeglichen werden.

Analyse der Powerfunktion

Bekanntlich ist man manchmal an der Optimierung der Stichprobengröße vor der Durchführung einer Studie interessiert. Zu diesem Zweck betrachtet man die Testgütefunktion

$$q_\phi : \mathbb{R}^m \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - F(\nu k; \delta_\mu, m, n - m) \quad (13.39)$$

bei kontrolliertem Testumfang, also für

$$k_{\alpha_0} := \nu^{-1} F^{-1}(1 - \alpha_0; m, n - m) \quad (13.40)$$

mit festem α_0 als Funktion des Nichtzentralitätsparameters, also der wahren, aber unbekanntem, Effektstärke und des Stichprobenumfangs. Insbesondere hängt hier k_{α_0} auch von n ab. Es ergibt sich dabei die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\delta_\mu, n) \mapsto \pi(\delta_\mu, n) := 1 - F(\nu k_{\alpha_0}; \delta_\mu, m, n - m). \quad (13.41)$$

Bei festgelegtem α_0 hängt diese sogenannten *Powerfunktion des Einstichproben- T^2 -Tests* also vom wahren, aber unbekanntem, Nichtzentralitätsparameter δ_μ , der Datendimensionalität m und von der Stichprobengröße n ab. Wir evaluieren diese Abhängigkeiten mithilfe untenstehenden **R** Codes und visualisieren sie exemplarisch in Abbildung 13.8.

```

1 # Szenariospezifikationen
2 a_0_all = c(0.05, 0.01) # \alpha_0 R. utilsaum
3 d_mu_min = 0 # \delta_\mu Minimum
4 d_mu_max = 20 # \delta_\mu Maximum
5 d_mu_res = 30 # \delta_\mu Auflösung
6 d_mu_all = seq(d_mu_min, d_mu_max, len = d_mu_res) # \delta_\mu d Raum
7 n_min = 5 # n Minimum
8 n_max = 20 # n Maximum
9 n_res = 30 # n Auflösung
10 n_all = seq(n_min, n_max, len = n_res) # n Raum
11 m_all = c(2, 4) # m Raum
12
13 # Evaluation der Powerfunktion
14 pi = array(dim = c(d_mu_res, n_res, 2, 2)) # Powerfunktionsarray
15 for (a in 1:length(a_0_all)){
16   for (l in 1:length(m_all)){
17     for (i in 1:length(d_mu_all)){
18       for (j in 1:length(n_all)){
19         m = m_all[l] # Datendimensionalität
20         n = n_all[j] # Stichprobenumfang
21         d_mu = d_mu_all[i] # wahrer, aber unbekannter, Parameter
22         nu = (n-m)/(m*(n-1)) # Parameter
23         alpha_0 = a_0_all[a] # Signifikanzlevel
24         k_alpha_0 = (1/nu)*qf(1-alpha_0, m, n-m) # kritischer Wert
25         pi[l, j, l, a] = 1 - pf(nu*k_alpha_0, m, n-m, d_mu)} } } } # Powerfunktionswert

```

Generell lässt sich aus der Perspektive der Anwendung festhalten, dass π als Funktion von n monoton steig. Ein größerer Stichprobenumfang resultiert damit also im Allgemeinen in einer kleineren Wahrscheinlichkeit für einen Typ II Fehler. Dabei bleiben allerdings mögliche weitere Kosten für die Erhöhung des Stichprobenumfangs unberücksichtigt. Weiterhin hängen die Werte der Powerfunktion π offensichtlich vom wahren, aber unbekanntem, Nichtzentralitätsparameterwert

$$\delta_\mu = n(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0) \quad (13.42)$$

ab. Würde man diesen Wert schon mit großer Präzision kennen, so gäbe es keinen Grund eine Studie und ihren Stichprobenumfang zu planen. Es wird deshalb zur Stichprobengrößenoptimierung im Vorfeld einer Studie im Allgemeinen folgendes Vorgehen favorisiert:

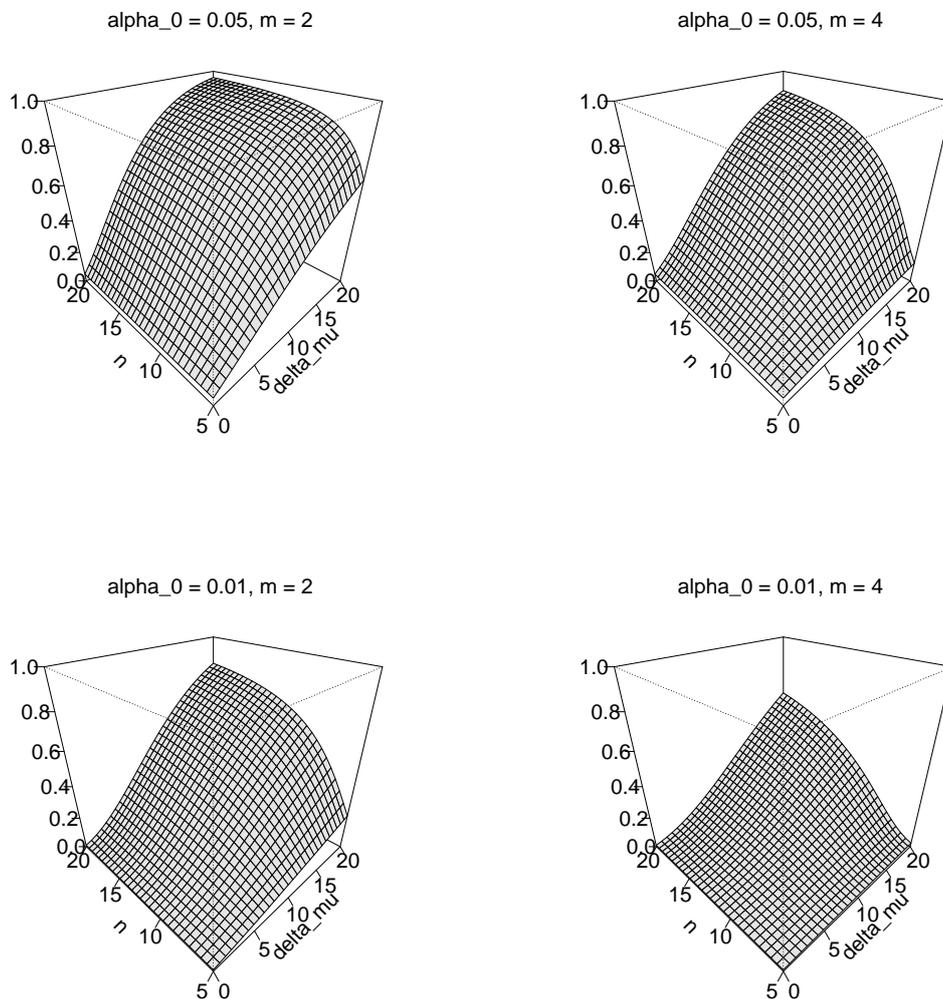


Abbildung 13.8. Powerfunktionen des Einstichproben- T^2 -Tests. Die Abbildungen zeigen die Wahrscheinlichkeit dafür, dass der Einstichproben- T^2 -Test den Wert 1 annimmt, dass also die Nullhypothese abgelehnt wird, als Funktion des wahren, aber unbekanntem, Nichtzentralitätsparameters δ_μ sowie des Stichprobenumfangs n . Dabei bilden die Abbildungen der ersten Zeile die Powerfunktionen des Einstichproben- T^2 -Tests für das Signifikanzlevel $\alpha_0 = 0.05$ und die Abbildungen der zweiten Zeile die Powerfunktionen des Einstichproben- T^2 -Tests für das Signifikanzlevel $\alpha_0 = 0.01$ ab. Ein niedrigeres Signifikanzlevel resultiert wie üblich in einer geringeren Wahrscheinlichkeit für die Ablehnung der Nullhypothese über den gesamten Bereich von δ_μ und n . Die erste Spalte der Abbildung zeigt die beiden Signifikanzlevelszenerarien für eine Datendimensionalität von $m := 2$, die zweite Spalte für eine Datendimensionalität von $m := 4$. Eine Erhöhung der Datendimensionalität führt, bei Konstanthalten aller weiteren Parameter, zu einer Reduktion der Wahrscheinlichkeit für das Ablehnen der Nullhypothese. Äquivalent sind bei höherer Datendimensionalität höhere Stichprobenumfänge nötig um bei vergleichbarem Nichtzentralitätsparameter die Nullhypothese gleichwahrscheinlich abzulehnen.

- (1) Man legt zunächst das Signifikanzlevel α_0 zur Kontrolle der Wahrscheinlichkeit eines Typ 1 Fehlers fest und evaluiert die entsprechende Powerfunktion.
- (2) Man wählt einen Mindestparameterwert δ_μ^* , den man mit einer Wahrscheinlichkeit von

$$\pi(\delta_\mu, n) = \beta \quad (13.43)$$

detektieren möchte, bei dem man also die Nullhypothese ablehnen möchte. Der Wert von δ_μ^* ergibt sich dabei aus problemspezifischen Überlegungen, wie zum Beispiel der Frage nach einem klinisch bedeutsamen Wert. Ein konventioneller Wert für die gewünschte Detektionswahrscheinlichkeit ist $\beta := 0.8$.

- (3) Basierend auf der evaluierten Powerfunktion liest man die für

$$\pi(\delta_\mu = \delta_\mu^*, n) = \beta \quad (13.44)$$

minimal nötige Stichprobengröße n ab. Größere Stichprobengrößen führen aufgrund der Monotonie von π als Funktion von n sicher zu einer gleichen oder höheren Wahrscheinlichkeit für das Ablehnen der Nullhypothese.

Für eine Datendimensionalität von $m := 2$ und Mindestparameterwert von $\delta_\mu^* = 12$ evaluiert untenstehender **R** Code wie in Abbildung 13.9 dargestellt die minimale Stichprobengröße um mit einer Wahrscheinlichkeit von $\beta = 0.8$ die Nullhypothese abzulehnen.

```

1 # Szenariospezifikation
2 n_min = 5 # n Minimum
3 n_max = 20 # n Maximum
4 n_res = 1e2 # n Auflösung
5 n = seq(n_min, n_max, len = n_res) # n Raum
6 alpha_0 = 0.05 # Signifikanzlevel
7
8 # Poweranalyse
9 m = 2 # Datendimensionalität
10 d_mu_fix = 12 # fester Nichtzentralitätsparameter
11 nu = (n-m)/(m*(n-1)) # Parameter
12 k_alpha_0 = (1/nu)*qf(1-alpha_0, m, n-m, d_mu_fix) # kritischer Wert
13 pi_n = 1 - pf(nu*k_alpha_0, m, n-m, d_mu_fix) # Powerfunktionswert
14 beta = 0.8 # gewünschter Powerfunktionswert
15 i = 1 # Indexinitialisierung
16 n_min = NaN # minimales n Initialisierung
17 while(pi_n[i] < beta){ # Solange \pi(\delta_\mu, n) < \beta
18   n_min = n[i] # Aufnahme des minimal nötigen ns
19   i = i + 1 # und Erhöhung des Indexes
20 }
21 cat("Minimal nötiges n =", ceiling(n_min)) # Ausgabe

```

Minimal nötiges n = 17

13.4. Anwendungsbeispiel

Wir betrachten das eingangs diskutierte Anwendungsbeispiel eines simulierten zweidimensionalen Datensatzes dreier Studiengruppen. Wir wollen abschließen für diesen Datensatz die Nullhypothese H_0 für ein Abweichen des wahren, aber unbekanntes, Erwartungswertparameters der Daten von μ_0 ist. Wir betrachten also weiterhin die einfache Nullhypothese $H_0 : \mu = \mu_0$ und die zusammengesetzte Alternativhypothese $H_1 : \mu \neq \mu_0$. Folgender **R** Code implementiert das praktische Vorgehen für ein Signifikanzlevel von $\alpha_0 := 0.05$.

```

1 # Datenbereitstellung
2 D = read.csv("../data/202-Einstichproben-T2-Tests.csv") # Datensatzeinlesen
3 Y = rbind(D$y_1i, D$y_2i) # Datenmatrix

```

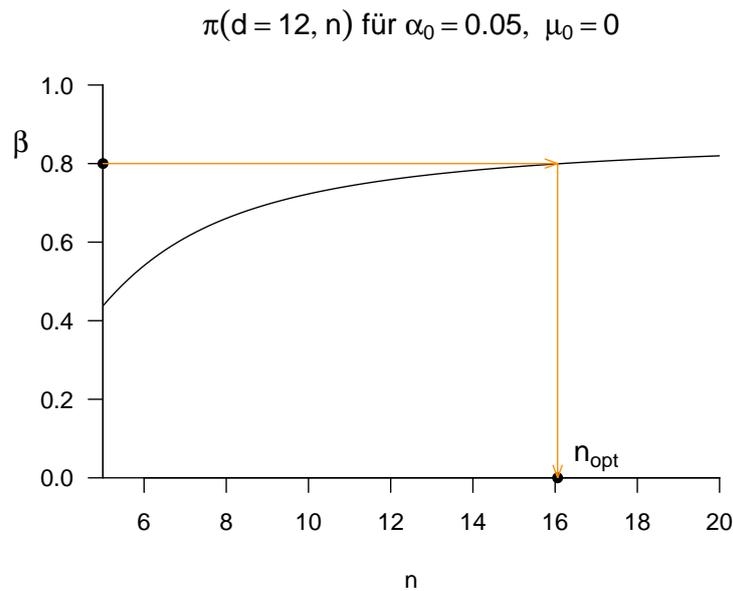


Abbildung 13.9. Bestimmung eines minimalen Stichprobenumfangs zur Detektion eines Nichtzentralitätsmindestparameterwert von $\delta_\mu = 12$. Die Abbildung zeigt die entsprechende Schnittfunktion der in Abbildung 13.8 dargestellten Funktion.

```

4
5 # Testparameter
6 m = nrow(Y)           # Dimensionalität der Zufallsvektoren/Daten
7 n = ncol(Y)           # Anzahl der Datenpunkte
8 nu = (n-m)/(m*(n-1))  # Parameter
9 mu_0 = matrix(c(30,3.5), nrow = 2) # H0 Hypothesenparameter ("Normwert")
10 alpha_0 = 0.05       # Signifikanzlevel
11 k_alpha_0 = (1/nu)*qf(1-alpha_0,m,n-m) # kritischer Wert
12
13 # Testevaluation
14 j_n = matrix(rep(1,n), nrow = n) # I_n
15 I_n = diag(n)                # I_n
16 J_n = matrix(rep(1,n^2), nrow = n) # I_{nn}
17 Y_bar = (1/n)*(Y %*% j_n)     # Stichprobenmittel
18 C = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y)) # Stichprobenkovarianzmatrix
19 T2 = n*(Y_bar - mu_0) %*% solve(C) %*% (Y_bar - mu_0) # T^2 Statistik
20 if(T2 > k_alpha_0){          # Test 1_{T^2 >= k_alpha_0}
21   phi = 1                    # Ablehnen von H_0
22 } else {
23   phi = 0                    # Nicht Ablehnen von H_0
24 }
25 p = 1 - pf(nu*T2,m,n-m)     # p-Wert
26
27 # Ausgabe
28 cat("Y_bar = ", Y_bar,
29     "\nC = ", C,
30     "\nT^2 = ", T2,
31     "\nalpha_0 = ", alpha_0,
32     "\nk = ", k_alpha_0,
33     "\nphi = ", phi,
34     "\np = ", p)

```

```

Y_bar = 26.25615 2.991039
C = 38.8981 3.549813 3.549813 1.972143
T^2 = 7.546368
alpha_0 = 0.05
k = 7.504065
phi = 1
p = 0.04928746

```

Im vorliegenden Fall nimmt die Einstichproben- T^2 -Teststatistik einen größeren Wert als der kritische Wert an, es gilt damit $\phi(v) = 1$ und man lehnt die Nullhypothese ab. Der korrespondierende p-Wert ist durch 0.049 gegeben.

13.5. Literaturhinweise

Die Theorie des Einstichproben- T^2 -Tests geht zurück auf Hotelling (1931).

13.6. Selbstkontrollfragen

1. Beschreiben Sie das Anwendungsszenario für einen Einstichproben- T^2 -Test.
2. Geben Sie die Definition des Einstichproben- T^2 -Test Modells wieder
3. Geben Sie die Definition der Einstichproben- T^2 -Teststatistik wieder.
4. Erläutern Sie, wann die Einstichproben- T^2 -Teststatistik hohe Werte annimmt.
5. Geben Sie das Theorem zu WDF und KDF der Einstichproben- T^2 -Teststatistik wieder.
6. Geben Sie das Theorem zur Testumfangkontrolle eines Einstichproben- T^2 -Tests wieder.
7. Erläutern Sie das praktische Vorgehen bei der Durchführung eines Einstichproben- T^2 -Tests.
8. Geben Sie das Theorem zum p-Wert eines Einstichproben- T^2 -Test an und erläutern Sie die Komponenten des entsprechenden Ausdrucks.

14. Einfaktorielle Varianzanalyse

14.1. Anwendungsszenario

Das Anwendungsszenario einer einfaktoriellen multivariaten Varianzanalyse ist durch das Vorliegen von multivariaten Datenpunkten von zwei oder mehr Gruppen randomisierter experimenteller Einheiten gekennzeichnet, die sich hinsichtlich der Level eines experimentellen Faktors unterscheiden. Ist die Anzahl an Datenpunkten in jeder Gruppe gleich, so spricht von einem balancierten einfaktoriellen multivariaten Varianzanalysedesign. Von den Datenpunkten der i ten Gruppe bzw. des i ten Faktorlevels wird angenommen, dass sie Realisierungen von jeweils n_i unabhängigen und identisch multivariat normalverteilten Zufallsvektoren sind, deren wahre, aber unbekannte, Erwartungswertparameter sich potentiell über die Gruppen hinweg unterscheiden und deren wahrer, aber unbekannter, Kovarianzmatrixparameter über die Gruppen hinweg identisch ist. In diesen Grundannahmen handelt es beim Anwendungsszenario der einfaktoriellen multivariaten Varianzanalyse also um die Generalisierung des Einstichproben- T^2 -Test Szenarios zu zwei oder mehr Gruppen experimenteller Einheiten. Grundlegend wird vorausgesetzt, dass ein Interesse am einem inferentiellen Vergleich, der wahren, aber unbekanntem, faktorlevelspezifischen Erwartungswertparameter besteht.

Anwendungsbeispiel

Als konkretes Anwendungsbeispiel betrachten wir die Analyse von Prä-Post-Interventions-BDI-Score und Prä-Post-Interventions-Glukokortikoidplasmalevel Differenzwerten von drei Gruppen von jeweils 15 Patient:innen, die unterschiedliche Psychotherapiesettings (Face-to-face und Online) bzw. eine Wartelistenkontrollbedingung durchlaufen haben. Wir stellen dazu in Tabelle 14.1 einen simulierten Beispieldatensatz dar. Die erste Spalte von Tabelle 14.1 (COND) listet das spezifische Therapiesetting (F2F: Face-to-face, ONL: online, WLC: waitlist control) der Patient:innen auf. Die zweite Spalte (dBDI) listet die entsprechenden BDI-Score-Differenzwerte und die dritte Spalte (dGLU) die entsprechenden Glukokortikoidplasmalevel-Differenzwerte auf. In beiden Fällen zeigen positive Werte eine Abnahme der Depressionssymptomatik, negative Werte dagegen einer Zunahme der Depressionssymptomatik an. Abbildung 14.1 visualisiert diesen Datensatz sowie die gruppenspezifischen Stichprobenmittel und Stichprobenkovarianzen als Normalverteilungsiskonturen.

Tabelle 14.1. Prä-Post-Interventions-BDI-II-Score und -Glukokortikoidplasmalevel Differenzenwerte von drei Studiengruppen (F2F: Face-to-face, ONL: online, WLC: waitlist control) jeweils 15 Patient:innen

COND	dBDI	dGLU
F2F	11	4.3
F2F	10	3.9
F2F	12	3.5
F2F	7	2.6
F2F	10	3.3
F2F	12	3.5
F2F	9	3.1
F2F	9	3.6
F2F	9	5.6
F2F	11	3.6
F2F	7	3.4
F2F	9	4.0
F2F	11	5.6
F2F	14	5.3
F2F	8	3.2
ONL	6	3.1
ONL	8	2.7
ONL	7	2.1
ONL	8	3.1
ONL	11	2.8
ONL	9	2.8
ONL	9	3.7
ONL	8	2.7
ONL	6	3.6
ONL	7	1.4
ONL	9	1.3
ONL	8	3.4
ONL	7	3.7
ONL	7	2.3
ONL	9	3.4
WLC	-2	0.7
WLC	2	1.4
WLC	1	1.0
WLC	2	0.9
WLC	3	1.6
WLC	2	1.6
WLC	5	1.1
WLC	-1	-0.2
WLC	2	2.5
WLC	-2	0.4
WLC	1	0.7
WLC	3	-0.1
WLC	1	1.3
WLC	4	0.6
WLC	4	0.7

Folgender **R** Code demonstriert die Auswertung gruppenspezifischer Deskriptivstatistiken für diesen Datensatz.

```

1 # Studiengruppenspezifische Deskriptivstatistiken
2 D = read.csv("../data/203-Einfaktorielle-Varianzanalyse.csv") # Dateneinlesen
3 m = 2 # Datendimension von Interesse
4 p = 3 # Anzahl Gruppen
5 k = 15 # Anzahl Datenpunkte pro Gruppe
6 Y = array(dim = c(m,k,p)) # Datenarrayinitialisierung
7 Y[, ,1] = rbind(D$dBDI[D$COND == "F2F"], # F2F dBDI Werte
8               D$dGLU[D$COND == "F2F"]) # F2F dGLU Werte
9 Y[, ,2] = rbind(D$dBDI[D$COND == "ONL"], # ONL dBDI Werte
10              D$dGLU[D$COND == "ONL"]) # ONL dGLU Werte
11 Y[, ,3] = rbind(D$dBDI[D$COND == "WLC"], # WLC dBDI Werte
12              D$dGLU[D$COND == "WLC"]) # WLC dGLU Werte
13 y_bar_i = array(dim = c(m,p)) # Stichprobenmittelarray
14 C_i = array(dim = c(m,m,p)) # Stichprobenkovarianzmatrizenarray
15 j_k = matrix(rep(1,k), nrow = k) # 1_{\cdot}
16 I_k = diag(k) # Einheitsmatrix I_{\cdot}
17 J_k = matrix(rep(1,k^2), nrow = k) # 1_{\cdot\cdot}
18 for (i in 1:p){ # Gruppeniterationen
19   y_bar_i[,i] = (1/k)*(Y[, ,i] %*% j_k) # Stichprobenmittel \bar{\mu}_i
20   C_i[, ,i] = (1/(k-1))*(Y[, ,i] %*% (I_k - (1/k)*J_k) %*% t(Y[, ,i])) # Stichprobenkovarianzmatrix C_i

```

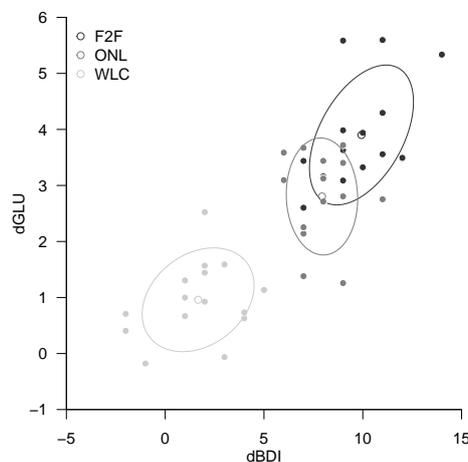


Abbildung 14.1. Deskriptivstatistiken der dBDI, dGLU Daten des Beispieldatensatzes. Jeder Punkt visualisiert die Daten einer Patient:in. Die Stichprobenkovarianz ist durch die 0.7 Isokontur einer zweidimensionalen Normalverteilung mit Erwartungswertparameter und Kovarianzmatrixparameter entsprechend dem Stichprobenmittel und der Stichprobenkovarianz der jeweiligen Gruppe dargestellt.

14.2. Modellformulierung und Modellschätzung

Wir definieren das Modell einfaktoriellen multivariaten Varianzanalyse wie folgt.

Definition 14.1 (Modell der einfaktoriellen multivariaten Varianzanalyse). Für $i = 1, \dots, p$ und $j = 1, \dots, n_i$ seien v_{ij} m -dimensionale Zufallsvektoren, die die $n := \sum_{i=1}^p n_i$ m -dimensionalen Datenpunkte eines einfaktoriellen multivariaten Varianzanalyseszenarios modellieren. Dann hat das Modell der einfaktoriellen multivariaten Varianzanalyse die strukturelle Form

$$v_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0_m, \Sigma) \text{ u.i.v. mit } \mu_i \in \mathbb{R}^m \text{ und } \Sigma \in \mathbb{R}^{m \times m} \text{ pd} \quad (14.1)$$

und die Datenverteilungsform

$$v_{ij} \sim N(\mu_i, \Sigma) \text{ u.v. mit } \mu_i \in \mathbb{R}^m \text{ und } \Sigma \in \mathbb{R}^{m \times m} \text{ pd.} \quad (14.2)$$

In Definition 14.1 bezeichnet n_i die Anzahl der Zufallsvektoren v_{ij} der i ten von p Gruppen experimenteller Einheiten. Im Falle eines balancierten Designs gilt offenbar $n_1 = \dots = n_p$. In diesem Fall setzen wir der Einfachheit halber $k := n_i$ für $i = 1, \dots, p$. In diesem Fall gilt für die Gesamtanzahl an Zufallsvektoren dann $n = pk$. Die Äquivalenz von struktureller Form und Datenverteilungsform der einfaktoriellen multivariaten Varianzanalyse ergibt sich mit Theorem 11.8 durch Transformation der ε_{ij} unter Multiplikation mit der Einheitsmatrix und unter Addition der jeweiligen gruppenspezifischen Erwartungswertparameter. Die wahren, aber unbekanntenen, Parameter $\mu_i, i = 1, \dots, p$ und Σ des einfaktoriellen multivariaten Varianzanalysemodells können anhand der in folgendem Theorem definierten Schätzer geschätzt werden.

Theorem 14.1 (Parameterschätzer der einfaktoriellen multivariaten Varianzanalyse). Gegeben sei das Modell der einfaktoriellen multivariaten Varianzanalyse. Dann ist für $i = 1, \dots, p$

$$\hat{\mu}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \quad (14.3)$$

ein unverzerrte Schätzer des gruppenspezifischen Erwartungswertparameters μ_i und

$$\hat{\Sigma} := \frac{1}{n-p} \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \hat{\mu}_i)(v_{ij} - \hat{\mu}_i)^T \quad (14.4)$$

ein unverzerrter Schätzer des Kovarianzmatrixparameters Σ .

In Theorem 14.1 ist $\hat{\mu}_i$ offenbar das Stichprobenmittel der Zufallsvektoren der i ten Gruppe. $\hat{\Sigma}$ ist die gruppenunspezifische Stichprobenkovarianzmatrix aller Zufallsvektoren und entspricht der mit $1/(n-p)$ skalierten *Within-Group Sum-of-Squares Matrix*, die wir in Theorem 14.2 einführen werden. Folgender **R** demonstriert die Evaluation der Parameterschätzer mithilfe einer **R** Funktion.

```

1 estimate = function(Y){
2
3   # Diese Funktion evaluiert die Parameterschätzer einer einfaktoriellen
4   # multivariaten Varianzanalyse basierend auf einen m x k x p Datensatz Y.
5   #
6   # Input
7   #   Y           : m x k x p Datenarray
8   #
9   # Output
10  #   $mu_hat     : m x p \mu_i Parameterschätzer
11  #   $$Sigma_hat : m x m \Sigma Parameterschätzer
12  # -----
13  # Dimensionsparameter
14  d   = dim(Y)           # Datensatzdimensionen
15  m   = d[1]            # Datendimension
16  k   = d[2]            # Anzahl Datenpunkte pro Gruppe
17  p   = d[3]            # Anzahl Gruppen
18
19  # Erwartungswertparameterschätzer
20  mu_hat_i = matrix(apply(Y,3,rowMeans), nrow = m)
21
22  # Kovarianzmatrixparameterschätzer
23  Sigma_hat = matrix(rep(0,m*m), nrow = m)
24  for(i in 1:p){
25    for(j in 1:k){
26      Sigma_hat = Sigma_hat + (1/(k*p-p))*(Y[,j,i] - mu_hat_i[,i]) %*% t(Y[,j,i] - mu_hat_i[,i])
27    }
28  }
29
30  # Outputspezifikation

```

```
31 return(list(mu_hat_i = mu_hat_i, Sigma_hat = Sigma_hat))
```

Anstelle eines Beweises validieren wir die Aussage von Theorem 14.1 beispielhaft mithilfe folgender **R** Simulation, in der wir die Erwartungswerte der Schätzer durch ihre Stichprobenmittelwerte über Datensatzrealisierungen hinweg approximieren.

```
1 # Modellparameter
2 library(MASS) # multivariate Normalverteilungen
3 p = 3 # Anzahl Gruppen
4 k = 15 # Anzahl Datenpunkte pro Gruppe
5 m = 2 # Datendimension
6 mu_i = matrix(c(1,2,2,1,3,2.5), ncol = p) # Erwartungswertparameter
7 Sigma = matrix(c(1,.5,.5,1), ncol = m) # Kovarianzmatrixparameter
8
9 # Simulationsparameter und Arrays
10 nsm = 1e2 # Anzahl Simulation
11 mu_hat_is = array(dim = c(m,p,nsm)) # \hat{\mu}_i Array
12 Sigma_hats = array(dim = c(m,m,nsm)) # \hat{\Sigma} Array
13
14 # Simulationen
15 for(s in 1:nsm){
16
17 # Datengeneration
18 Y = array(dim = c(m,k,p)) # Datenarray
19 for(i in 1:p){
20 Y[, , i] = t(mvrnorm(k, mu_i[,i], Sigma)) # Datengeneration
21 }
22 S = estimate(Y) # Parameterschätzung
23 mu_hat_is[, , s] = S$mu_hat_i # \hat{\mu}_i
24 Sigma_hats[, , s] = S$Sigma_hat # \hat{\Sigma}
25 }
26
27 # Schätzererwartungswertschätzung
28 E_hat_mu_i_hat = apply(mu_hat_is, c(1,2), mean)
29 E_hat_Sigma_hat = apply(Sigma_hats, c(1,2), mean)
```

Wie in Abbildung 14.2 ergibt sich hier auch schon bei einem recht geringen Simulationsaufwand von 100 Datensatzrealisierungen eine gute Korrespondenz zwischen wahren, aber unbekanntem, Parameterwerten μ_i , $i = 1, \dots, p$ und den approximierten Erwartungswertparametern $\mathbb{E}(\hat{\mu}_i)$ für $i = 1, \dots, p$ und $\mathbb{E}(\hat{\Sigma})$.

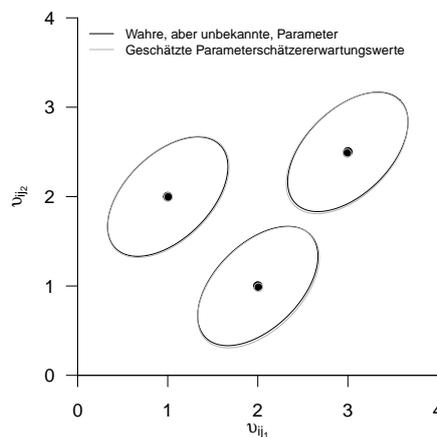


Abbildung 14.2. Simulationsbasierte Validierung des Theorems zu den Parameterschätzern der einfaktoriellen multivariaten Varianzanalyse am Beispiel von $m := 2$, $p := 3$, $k := 15$ und 100 Realisierungen der entsprechenden multivariat normalverteilten Zufallsvektoren

Die Anwendung der Parameterschätzung auf die Daten des Beispieldatensatzes in Tabelle 14.1 ergibt folgende Resultate.

```

1 S = estimate(Y) # Parameterschätzung
2 print(S$mu_hat_i) # Ausgabe

```

```

      [,1] [,2] [,3]
[1,] 9.933333 7.933333 1.666667
[2,] 3.900525 2.803179 0.9594878

```

```

1 print(S$Sigma_hat) # Ausgabe

```

```

      [,1] [,2]
[1,] 3.3142857 0.4022937
[2,] 0.4022937 0.6381336

```

14.3. Modellevaluation

Primäres Ziel einer einfaktoriellen multifaktoriellen Varianzanalyse ist meist das Testen der Nullhypothese

$$H_0 : \mu_1 = \dots = \mu_p. \quad (14.5)$$

Diese Nullhypothese besagt, dass keine Unterschiede zwischen den wahren, aber unbekanntem, Erwartungswertparametern der Stichprobengruppen bestehen. Die Alternativhypothese lautet somit

$$H_1 : \mu_{i_l} \neq \mu_{j_l} \text{ für mindestens ein Paar } i, j \text{ mit } i \neq j, 1 \leq i, j \leq p \text{ und mindestens ein } l \text{ mit } 1 \leq l \leq m. \quad (14.6)$$

Die Alternativhypothese besagt also, dass sich mindestens zwei wahre, aber unbekanntem, Erwartungswertparameter in mindestens einer ihrer Komponenten unterscheiden. Wie aus dem Kontext der univariaten einfaktorielle Varianzanalyse bekannt impliziert das Ablehnen der Nullhypothese auch hier keine Aussage über die genaue Form des inferierten Erwartungswertparameterunterschiedes.

Im Rahmen einfaktoriellen multivariaten Varianzanalyse können Tests der Nullhypothese mit verschiedenen Teststatistiken konstruiert werden. Diesen Teststatistiken ist gemein, dass sie auf eine Generalisierung der aus dem univariaten Fall bekannten Quadratsummenzerlegung der einfaktoriellen Varianzanalyse zurückgehen. Wir führen im nächsten Abschnitt zunächst diese sogenannte *Kreuzproduktsummenmatrizenzerlegung* der einfaktoriellen multivariaten Varianzanalyse ein. Nachfolgend betrachten wir dann die Modellevaluation mithilfe der *Wilks'- Λ -Statistik*.

Kreuzproduktsummenmatrizenzerlegung

Folgendes Theorem generalisiert die Quadratsummenzerlegung der einfaktoriellen Varianzanalyse auf das multivariate Anwendungsszenario.

Theorem 14.2 (Kreuzproduktsummenmatrizenzerlegung).

Gegeben sei das Modell der einfaktoriellen multivariaten Varianzanalyse. Weiterhin seien

$$\bar{v} := \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} v_{ij} \text{ und } \bar{v}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \quad (14.7)$$

das Gesamtstichprobenmittel und das i te Gruppenstichprobenmittel, respektive. Schließlich seien

$$T := \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v})(v_{ij} - \bar{v})^T \quad \text{die Totale Sum-of-Squares Matrix}$$

$$B := \sum_{i=1}^p n_i (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T \quad \text{die Between-Group Sum-of-Squares Matrix}$$

$$W := \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T \quad \text{die Within-Group Sum-of-Squares Matrix.}$$

Dann gilt

$$T = B + W. \quad (14.8)$$

◦

Beweis. Es gilt

$$\begin{aligned} T &= \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v})(v_{ij} - \bar{v})^T \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i + \bar{v}_i - \bar{v})(v_{ij} - \bar{v}_i + \bar{v}_i - \bar{v})^T \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} ((v_{ij} - \bar{v}_i) + (\bar{v}_i - \bar{v}))((v_{ij} - \bar{v}_i) + (\bar{v}_i - \bar{v}))^T \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} ((v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T + 2(v_{ij} - \bar{v}_i)(\bar{v}_i - \bar{v})^T + (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T + \sum_{j=1}^{n_i} 2(v_{ij} - \bar{v}_i)(\bar{v}_i - \bar{v})^T + \sum_{j=1}^{n_i} (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T + 2 \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i) \right) (\bar{v}_i - \bar{v})^T + n_i (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T + 2 \left(\sum_{j=1}^{n_i} \left(v_{ij} - \frac{1}{n_i} \sum_{j=1}^{n_i} v_{ij} \right) \right) (\bar{v}_i - \bar{v})^T + n_i (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T + 2 \left(\sum_{j=1}^{n_i} v_{ij} - \sum_{j=1}^{n_i} v_{ij} \right) (\bar{v}_i - \bar{v})^T + n_i (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T \right) \\ &= \sum_{i=1}^p \left(\sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T + n_i (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T \right) \\ &= \sum_{i=1}^p n_i (\bar{v}_i - \bar{v})(\bar{v}_i - \bar{v})^T + \sum_{i=1}^p \sum_{j=1}^{n_i} (v_{ij} - \bar{v}_i)(v_{ij} - \bar{v}_i)^T \\ &= B + W. \end{aligned} \quad (14.9)$$

□

Die intuitive Interpretation der Totalen, Between-Group, und Within-Group Sum-of-Squares Matrizen ist analog zu den aus dem univariaten Szenario bekannten Begriffen: Die Matrix T repräsentiert die totale Variabilität der Datenvektoren um das Gesamtstichprobenmittel, die Matrix B repräsentiert die Variabilität der Gruppenstichprobenmittel um das Gesamtstichprobenmittel und die Matrix W repräsentiert die Variabilität der Datenvektoren um ihre jeweiligen Gruppenstichprobenmittel.

Wie im univariaten Fall wird auch hier also die Gesamtdatenvariabilität additiv in zwei unabhängige Beiträge zerlegt. Dabei kann die Matrix W auch als Maß für die Residualvariabilität verstanden werden, weil sie die verbleibende Variabilität nach Schätzung der Gruppenerwartungswertparameter quantifiziert. Offenbar gilt für den Schätzer $\hat{\Sigma}$ des gemeinsamen Stichprobenkovarianzmatrixparameters aus Theorem 14.1

$$W = (n - p)\hat{\Sigma}. \quad (14.10)$$

Folgender **R** demonstriert die Evaluation der in Theorem 14.2 definierten Matrizen mithilfe einer **R** Funktion.

```

1  sos = function(Y){
2
3  # Diese Funktion evaluiert die Kreuzproduktsummenmatrizen T,B,W einer
4  # einfaktoriellen Varianzanalyse basierend auf einen m x k x p Datensatz Y.
5  #
6  # Input
7  #   Y      : m x k x p Datenarray
8  #
9  # Output
10 #   $y_bar   : m x 1 Gesamtmittelwert
11 #   $y_bar_i : m x p Gruppenmittelwerte
12 #   $T       : m x m Total Sum of Squares Matrix
13 #   $B       : m x m Between-Group Sum-of-Squares Matrix
14 #   $W       : m x m Within Group Sum of Squares Matrix
15 # -----
16 d   = dim(Y)                # Datensatzdimensionen
17 m   = d[1]                  # Datendimension
18 k   = d[2]                  # Anzahl Datenpunkte pro Gruppe
19 p   = d[3]                  # Anzahl Gruppen
20
21 # Mittelwerte
22 y_bar_i = matrix(apply(Y,3,rowMeans), nrow = m) # Gruppenstichprobenmittel
23 y_bar   = matrix(rowMeans(y_bar_i), nrow = m)  # Gesamtstichprobenmittel
24
25 # Totale Sum-of-Squares Matrix
26 T = matrix(rep(0,m*m), nrow = m)
27 for(i in 1:p){
28   for(j in 1:k){
29     T = T + (Y[,j,i] - y_bar) %*% t(Y[,j,i] - y_bar)}
30
31 # Between Sum of Squares Matrix
32 B = matrix(rep(0,m*m), nrow = m)
33 for(i in 1:p){
34   B = B + k*(y_bar_i[,i] - y_bar) %*% t(y_bar_i[,i] - y_bar)}
35
36 # Within Sum of Squares Matrix
37 W = matrix(rep(0,m*m), nrow = m)
38 for(i in 1:p){
39   for(j in 1:k){
40     W = W + (Y[,j,i] - y_bar_i[,i]) %*% t(Y[,j,i] - y_bar_i[,i])}
41
42 # Outputspezifikation
43 return(list(y_bar_i = y_bar_i, y_bar = y_bar, T = T, B = B, W = W))}

```

Modellevaluation mit der Wilks'- Λ -Statistik

Basierend auf der Kreuzproduktsummenmatrizenzerlegung in Theorem 14.2 wurde eine Reihe von Teststatistiken für die einfaktorielle multivariate Varianzanalyse vorgeschlagen (Wilks (1932), Pillai (1955), Roy (1953), Hotelling (1951)). Wir betrachten hier exemplarisch lediglich die *Wilks'- Λ -Statistik* nach Wilks (1932). Im Gegensatz zur F -Teststatistik der univariaten einfaktoriellen Varianzanalyse sind die Frequentistischen Verteilungen von der Wilks'- Λ -Statistik bei Zutreffen der Nullhypothese nur für bestimmte Anwendungsszenarien, insbesondere bei kleinen Werte der Datendimension m und der Gruppenanzahl p , analytisch exakt zu bestimmen. In diesen Szenarien ist man auf f -Verteilungen mit m - und p -abhängigen Freiheitsgradparametern geführt. Für Anwendungsszenarien mit größeren Werten von m und/oder p existieren lediglich Approximationen der Frequentistischen Verteilungen der Wilks'- Λ -Statistik, die nur asymptotisch für unendlich große Stichprobenumfänge $n \rightarrow \infty$ exakt sind.

Auch diese Approximationen sind wiederum durch f -Verteilungen mit m - und p -abhängigen Freiheitsgradparametern gegeben. In der Anwendung unterscheiden sich Testentscheidungen basierend auf exakten oder approximativen Verteilungen der verschiedenen Teststatistiken meist nicht. Zur Absicherung dieser Aussage mögen im konkreten Fall von Datendimension und Gruppengröße Simulationen helfen, mögliche Unterschiede zwischen der Wilks'- Λ -Statistik und anderen Teststatistiken sowie der Approximation ihrer Verteilungen abzuschätzen.

Wir definieren zunächst die Wilks'- Λ -Statistik.

Definition 14.2 (Wilks'- Λ -Statistik). Gegeben sei das Modell der einfaktoriellen multivariaten Varianzanalyse sowie die Between-Group Sum-of-Squares Matrix B und die Within-Group Sum-of-Squares Matrix W . Dann ist *Wilks'- Λ -Statistik* definiert als

$$\Lambda := \frac{|W|}{|T|} = \frac{|W|}{|B + W|}. \quad (14.11)$$

•

Intuitiv misst Λ das Verhältnis von Residualdatenvariabilität, repräsentiert durch die Determinante von W , und Gesamtdatenvariabilität, repräsentiert durch die Determinante von $T = B + W$. Λ ist damit also analog zum Effektstärkenmaß η^2 der univariaten einfaktoriellen Varianzanalyse definiert. Im Falle der Gleichheit der gruppenspezifischen Stichprobenmittel gilt für Λ insbesondere

$$\bar{v}_1 = \dots = \bar{v}_p = \bar{v} \Rightarrow B = 0_{mm} \Rightarrow \Lambda = \frac{|W|}{|T|} = \frac{|W|}{|B + W|} = \frac{|W|}{|0_{mm} + W|} = \frac{|W|}{|W|} = 1. \quad (14.12)$$

Für ansteigende Unterschiede zwischen den gruppenspezifischen Stichprobenmittel \bar{v}_i nimmt $|B+W|$ gegenüber $|W|$ zu und Λ somit ab. Ohne Beweis wollen wir dabei festhalten, dass $0 \leq \Lambda \leq 1$. Im Gegensatz zu den meisten bekannten Teststatistiken sprechen hier also *kleine Werte der Teststatistik Λ für eine Abweichung von der Nullhypothese*.

Äquivalent kann Λ auch in Form der Eigenwerte der Matrix $W^{-1}B$ angegeben werden. Dabei ist die Matrix $W^{-1}B$ das multivariate Analogon zum dem aus der univariaten einfaktoriellen Varianzanalyse bekanntem Verhältnis von Between-Group Sum-of-Squares und Within-Group Sum-of-Squares, welches die Grundlage für F -Teststatistik im univariaten Fall bildet. Es gilt folgendes Theorem.

Theorem 14.3 (Eigenwertform von der Wilks'- Λ -Statistik). *Es seien das Modell der einfaktoriellen multivariaten Varianzanalyse, die Between-Group Sum-of-Squares Matrix B , die Within-Group Sum-of-Squares Matrix W und die Wilks'- Λ -Statistik definiert wie oben. Weiterhin seien $\lambda_1, \dots, \lambda_s$ die Eigenwerte von $W^{-1}B$. Dann gilt*

$$\Lambda = \prod_{i=1}^s \frac{1}{1 + \lambda_i}. \quad (14.13)$$

◦

Wir verzichten auf einen Beweis.

Für spezielle, durch die Datendimension m und die Gruppenanzahl p charakterisierte Anwendungsszenarien stellt Wilks (1932) analytische Formen für die Verteilungen von Transformationen von Λ unter der Nullhypothese bereit. Dabei entsprechen diese analytischen Formen den Verteilungen von f -Zufallsvariablen mit bestimmten, durch die Datendimension und Gruppenanzahl des Anwendungsszenario vorgegebenen Freiheitsgradparametern. Wir fassen die Anwendungsszenarien, die betreffende Transformation von Λ und die analytischen Formen ihrer Frequentistischen Verteilungen bei Zutreffen der Nullhypothese in folgendem Theorem zusammen.

Theorem 14.4 (Spezielle H_0 Verteilungen von Transformationen der Wilks'- Λ -Statistik).

Es seien das Modell der einfaktoriellen Varianzanalyse und die Wilks'- Λ -Statistik definiert wie oben und es gelte außerdem die Nullhypothese

$$H_0 : \mu_1 = \dots = \mu_p. \tag{14.14}$$

Dann sind für die in den ersten beiden Tabellenspalten aufgeführten Spezialfällen die in der dritten Tabellenspalte aufgeführten Statistiken f -Zufallsvariablen und zwar mit den in der vierten Tabellenspalte aufgeführten Freiheitsgradparametern.

Datendimension m	Gruppenanzahl p	Statistik	f -Verteilungsparameter
Beliebig	2	$\frac{1-\Lambda}{\Lambda} \frac{n-p-m+1}{m}$	$m, n-p-m+1$
Beliebig	3	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{n-p-m+1}{m}$	$2m, 2(n-p-m+1)$
1	Beliebig	$\frac{1-\Lambda}{\Lambda} \frac{n-p}{p-1}$	$p-1, n-p$
2	Beliebig	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{n-p-1}{p-1}$	$2(p-1), 2(n-p-1)$

◦

In Abbildung 14.3 visualisieren wir exemplarische simulationsbasierte Validierungen der in Theorem 14.4 angegebenen Verteilungen.

Rao (1951) hat für allgemeine Anwendungsszenarien die in folgendem Theorem angegebenen Approximationen von Verteilungen von Transformationen der Wilks'- Λ -Statistik vorgeschlagen. Man beachte, dass sich die in diesem Theorem definierte Teststatistik τ wie in Abbildung 14.4 gezeigt zu Λ reziprok verhält. Geringe Werte von Λ als Evidenz gegen die Nullhypothese entsprechen also hohen Werten von τ .

Theorem 14.5 (Approximative H_0 Verteilungen von Transformationen der Wilks'- Λ -Statistik).

Es seien das Modell der einfaktoriellen Varianzanalyse und die Wilks'- Λ -Statistik definiert wie oben und es gelte außerdem die Nullhypothese

$$H_0 : \mu_1 = \dots = \mu_p. \tag{14.15}$$

Dann ist die Statistik

$$\tau := \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \frac{\nu_2}{\nu_1} \tag{14.16}$$

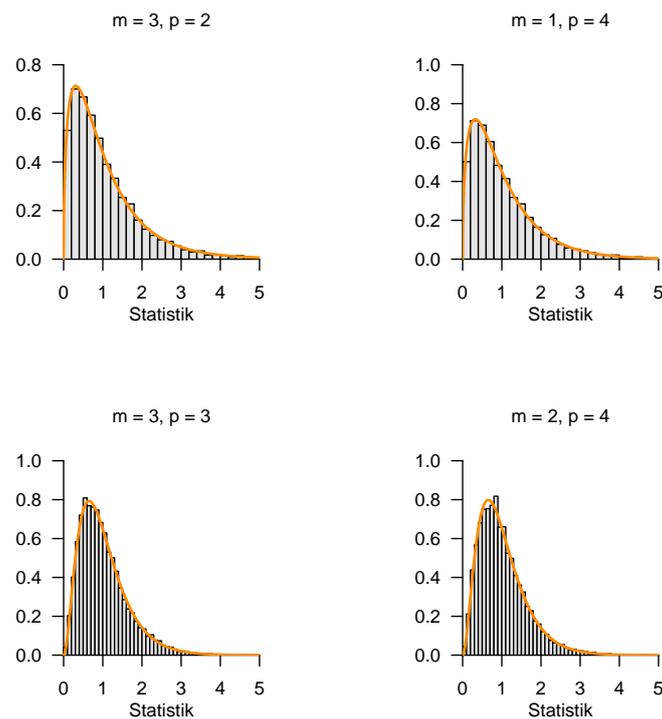


Abbildung 14.3. Simulation spezieller Verteilungen der Wilks'- Λ -Statistik Transformationen bei Zutreffen der Nullhypothese.

mit

$$\nu_1 := m(p - 1) \text{ und } \nu_2 := wt - \frac{1}{2}(m(p - 1) - 2) \tag{14.17}$$

sowie

$$w := n - 1 - \frac{1}{2}(m + p) \text{ und } t := \sqrt{\frac{m^2(p - 1)^2 - 4}{m^2 + (p - 1)^2 - 5}} \tag{14.18}$$

approximativ f -verteilt mit Freiheitsgradparametern ν_1 und ν_2 .

◦

In [Abbildung 14.5](#) visualisieren wir exemplarische simulationsbasierte Validierungen der in [Theorem 14.5](#) angegebenen Verteilungen.

Mithilfe der von Rao (1951) bestimmten approximativen Verteilungen der transformierten Wilks'- Λ -Statistik τ bei Zutreffen der Nullhypothese können wir nun einen Hypothesentest für die einfaktorielle multivariate Varianzanalyse angeben.

Theorem 14.6 (Wilks'- Λ -basierter Test, Testumfangkontrolle, p-Wert). *Gegeben seien Modell der einfaktoriellen multivariaten Varianzanalyse und die transformierte Wilks'- Λ -Statistik τ mit Verteilungsparametern ν_1, ν_2 wie oben definiert. Weiterhin sei für $\Upsilon = (v_1, \dots, v_n)$ der kritische Wert-basierte Test*

$$\phi(\Upsilon) := 1_{\{\tau > k\}} := \begin{cases} 1 & \tau > k \\ 0 & \tau \leq k \end{cases} \tag{14.19}$$

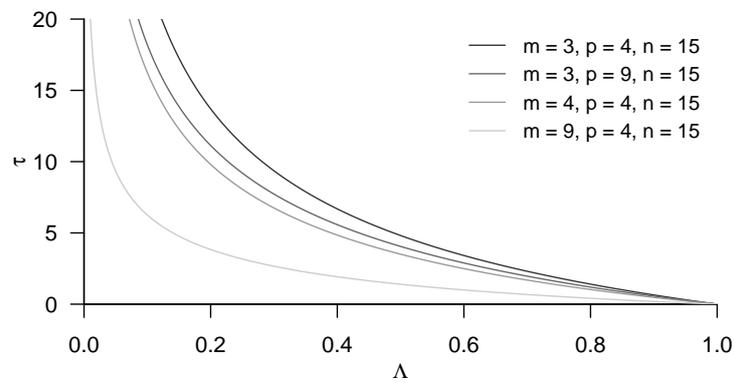


Abbildung 14.4. Beziehung der Wilks'- Λ -Statistik und der von Rao (1951) betrachteten τ Statistik in exemplarischen Anwendungsszenarien.

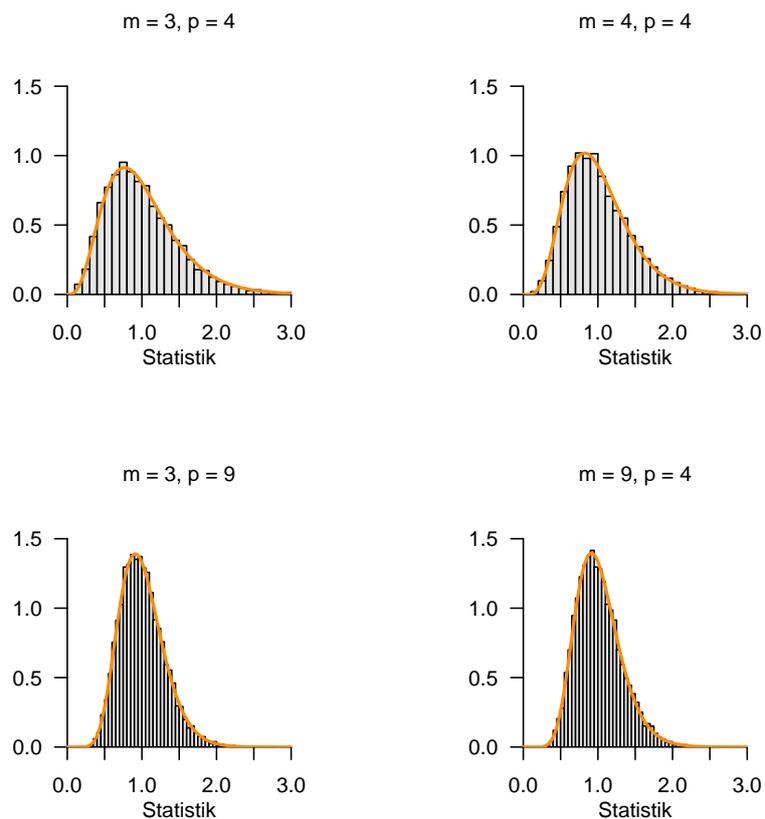


Abbildung 14.5. Simulation approximativer Verteilungen der Wilks'- Λ -Statistik Transformationen bei Zutreffen der Nullhypothese.

definiert. Dann ist ϕ genau dann ein Level- α_0 -Test mit Testumfang α , wenn

$$k := k_{\alpha_0} := F^{-1}(1 - \alpha_0; \nu_1, \nu_2) \tag{14.20}$$

ist und der p -Wert einer realisierten τ -Teststatistik $\tilde{\tau}$ ergibt sich zu

$$p\text{-Wert} = \mathbb{P}(\tau \geq \tilde{\tau}) = 1 - F(\tilde{\tau}; \nu_1, \nu_2) \tag{14.21}$$

◦

Wir verzichten auf einen Beweis von Theorem 14.6, welcher analog zu den entsprechenden Beweisen zum Beispiel beim Einstichproben-T²-Test geführt werden kann. Die in Theorem 14.6 implizite Wahl eines kritischen Wertes zur Testumfangkontrolle im Anwendungsszenario $m := 3, p := 4$ und $n_i = 15$ für $i = 1, \dots, p$ und damit $\nu_1 = 9$ und $\nu_2 = 132$ bei einem Signifikanzlevel von $\alpha_0 = 0.05$ visualisieren wir in Abbildung 14.6.

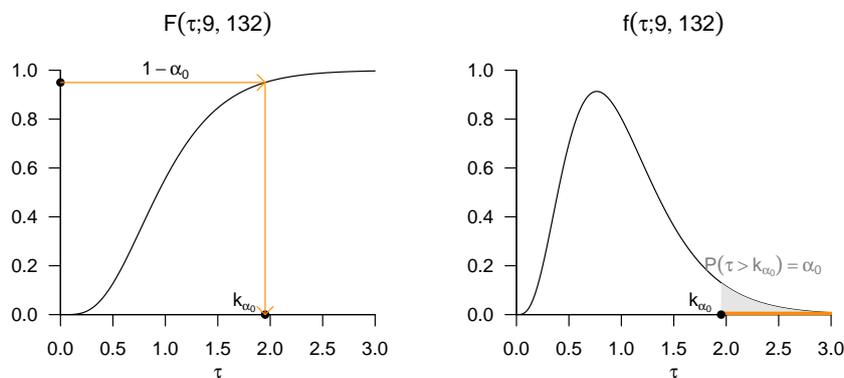


Abbildung 14.6. Testumfangkontrolle durch Selektion eines α_0 -abhängigen kritischen Wertes für den Wilks'Λ-Statistik-basierten Test anhand von KVF und WDF der transformierten Wilks'Λ-Statistik τ .

Praktisches Vorgehen

In der Praxis entsprechen obige Ergebnisse dann folgendem Vorgehen bei der Durchführung einer einfaktoriellen multivariaten Varianzanalyse: Man unterstellt, dass ein vorliegender Datensatz von $i = 1, \dots, p$ Gruppen von m -dimensionalen Datenvektoren für jeweils $j = 1, \dots, n_i$ eine Realisation von Zufallsvektoren $v_{ij} \sim N(\mu_i, \Sigma)$ mit unbekanntem, gruppenspezifischen Erwartungswertparametern $\mu_i \in \mathbb{R}^m$ und gruppenunspecifischem Kovarianzmatrixparameter $\Sigma \in \mathbb{R}^{m \times m}$ pd ist. Man möchte entscheiden, ob eher die Nullhypothese $H_0 : \mu_1 = \dots = \mu_p$ identischer wahrer, aber unbekannter, Erwartungswertparameter zutrifft oder eher nicht. Zu diesem Zweck wählt man zunächst ein Signifikanzlevel α_0 und bestimmt dann den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ im Szenario von dreidimensionalen Datenvektoren ($m = 3$), vier Gruppen ($p = 4$) und $n_i = 15$ experimentellen Einheiten pro Gruppe und damit einer Gesamtzahl von $n = 60$ Datenpunkten, dass $\nu_1 = 9$ und $\nu_2 = 132$ sind und sich der in Theorem 14.6 definierte kritische Wert zu $k_{\alpha_0} = F^{-1}(1 - 0.05; 9, 132) \approx 1.95$ ergibt. Basierend auf den realisierten Datensatz berechnet man dann zunächst die

Wilk's- Λ -Statistik und den resultierenden, m, p, n -abhängigen realisierten Wert der transformierten Wilk's- Λ -Statistik τ . Wenn der berechnete Wert von τ größer als k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht. Die oben entwickelte Theorie zur Testumfangkontrolle bei der einfaktoriellen multivariaten Varianzanalyse auf Grundlage von Wilks'- Λ -Statistik garantiert dann, dass man im Mittel in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Folgender **R** Code demonstriert die Anwendung des in Theorem 14.6 definierten Hypothesentests auf bei Zutreffen der Nullhypothese unter dem Modell der einfaktoriellen multivariaten Varianzanalyse generierte Daten und validiert seine Kontrolle des Testumfangs für die in Abbildung 14.5 skizzierten Anwendungsszenarien.

```

1 # Szenarioparameter
2 library(MASS) # R Paket für multivariate Normalverteilungen
3 nsm = 1e4 # Datensimulationsanzahl
4 M = c(3,3,4,9) # Datendimension
5 P = c(4,9,4,4) # Gruppenanzahl
6 k = 15 # Datenpunkte pro Gruppe
7 N = k*P # Gesamtanzahl Datenpunkte
8 alpha_0 = 0.05 # \alpha_0
9 nsc = length(M) # Szenarienzahl
10 TAU = matrix(rep(NaN,nsm*nsc), ncol = nsc) # Statistik Array
11 NU = matrix(rep(NaN,2*nsc), ncol = nsc) # Parameter Array
12 KA = rep(NaN, nsc) # Kritische Werte
13 PHI = matrix(rep(0,nsm*nsc), ncol = nsc) # Testarray
14
15 # Simulationen
16 for(sc in 1:nsc){ # Szenarioiterationen
17
18 # Modellparameter
19 m = M[sc] # Datendimension
20 p = P[sc] # Gruppenanzahl
21 n = N[sc] # Gesamtanzahl Datenpunkte
22 mu_i = matrix(rep(0,m), nrow = m) # Identische Gruppenerwartungswertparameter bei H_0
23 Sigma = diag(m) # Identische Gruppenkovarianzmatrixparameter
24
25 # Varianzanalyse Parameter
26 w = n-1-(1/2)*(m*p) # w
27 t = sqrt((m^2*(p-1)^2-4)/(m^2+(p-1)^2-5)) # t
28 nu_1 = m*(p-1) # \nu_1
29 nu_2 = w*t-(1/2)*(m*(p-1)-2) # \nu_2
30 KA[sc] = qf(1-alpha_0,nu_1,nu_2) # kritischer Wert
31
32 # Datensimulationen
33 for(sm in 1:nsm){
34 Y = array(dim = c(m,k,p)) # Datenarrayinitialisierung
35 for(i in 1:p){ # Gruppeniterationen
36 Y[,i] = t(mvrnorm(k,mu_i,Sigma)) # Datensimulation
37 S = sos(Y) # Stichprobenmittel und Sum of Squares Matrizen
38 L = det(SSW)/det(SSW + SSB) # Wilks' Lambda
39 tau = ((1-L^(1/t))/L^(1/t))*(nu_2/nu_1) # Statistik
40 PHI[sm,sc] = tau > KA[sc]} # Test
41 cat("Kritische Werte : ", KA,
42 "\nGeschätzte Testumfänge: ", apply(PHI, 2, mean))

```

```

Kritische Werte : 1.951729 1.547641 1.821661 1.565152
Geschätzte Testumfänge: 0.0472 0.0496 0.0522 0.0488

```

14.4. Anwendungsbeispiel

Wir betrachten das eingangs diskutierte Anwendungsbeispiel eines simulierten zweidimensionalen Datensatzes dreier Therapiegruppen. Wir wollen nun mithilfe einer einfaktoriellen Varianzanalyse für diesen Datensatz die Nullhypothese identischer Gruppenerwartungswertparameter überprüfen. Folgender **R** Code implementiert das praktische Vorgehen für ein Signifikanzlevel von $\alpha_0 := 0.05$.

```

1 # Einlesen und Präprozessierung des Datensatzes
2 D = read.csv("../data/203-Einfaktorielle-Varianzanalyse.csv") # Dateneinlesen
3 m = 2 # Datendimension von Interesse
4 p = 3 # Anzahl Gruppen
5 k = 15 # Anzahl Datenpunkte pro Gruppe
6 n = p*k # Gesamtanzahl Datenpunkte

```

```

7 Y = array(dim = c(m,k,p)) # Datenarrayinitialisierung
8 Y[, ,1] = rbind(D$dBDI[D$COND == "F2F"], # F2F dBDI Werte
9 D$dGLU[D$COND == "F2F"]) # F2F dGLU Werte
10 Y[, ,2] = rbind(D$dBDI[D$COND == "ONL"], # ONL dBDI Werte
11 D$dGLU[D$COND == "ONL"]) # ONL dGLU Werte
12 Y[, ,3] = rbind(D$dBDI[D$COND == "WLC"], # WLC dBDI Werte
13 D$dGLU[D$COND == "WLC"]) # WLC dGLU Werte
14
15 # Einfaktorielle Varianzanalyse
16 alpha_0 = 0.05 # Signifikanzlevel
17 S = sos(Y) # Sum of Squares Matrizen
18 L = det(S$W)/det(S$W + S$B) # Wilks' Lambda
19 w = n-1-(1/2)*(m+p) # w
20 t = sqrt((m^2*(p-1)^2-4)/(m^2+(p-1)^2-5)) # t
21 nu_1 = m*(p-1) # \nu_1
22 nu_2 = w*t-(1/2)*(m*(p-1)-2) # \nu_2
23 tau = ((1-L^(1/t))/L^(1/t))*(nu_2/nu_1) # Teststatistik
24 k_alpha_0 = qf(1-alpha_0, nu_1, nu_2) # kritischer Wert
25 phi = as.numeric(tau > k_alpha_0) # Nullhypothesentest
26 P = 1-pf(tau, nu_1, nu_2) # Überschreitungswahrscheinlichkeit
27
28 # Ausgabe
29 cat("Wilks' Lambda ", L,
30 "\ntau ", tau,
31 "\nnu_1 ", nu_1,
32 "\nnu_2 ", nu_2,
33 "\nphi ", phi,
34 "\nP(tau > tau_tilde)", P)

```

```

Wilks' Lambda      0.1569049
tau                 31.25301
nu_1                4
nu_2                82
phi                 1
P(tau > tau_tilde) 8.881784e-16

```

Im vorliegenden Fall wird die Nullhypothese identischer Gruppenerwartungswertparameter also verworfen. Schließlich validieren wir obige Analyse im Sinne eines Black-Box-Verfahrens mithilfe der **R** `lm()` und `Manova()` Funktionen.

```

1 library(car)
2 D = read.csv("../data/203-Einfaktorielle-Varianzanalyse.csv") # Dateneinlesen
3 model = lm(Cbind(D$dBDI,D$dGLU) ~ D$COND, D) # Modellspezifikation
4 Manova(model, test.statistic = "Wilks") # Einfaktorielle Varianzanalyse

```

```

Type II MANOVA Tests: Wilks test statistic
      Df test stat approx F num Df den Df Pr(>F)
D$COND 2 0.1569 31.253 4 82 8.346e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

14.5. Literaturhinweise

Die Theorie der einfaktoriellen multivariaten Varianzanalyse geht zurück auf Wilks (1932). Anderson (2003) gibt eine Einführung in die Approximationstheorie für multivariate Modelle, das Wissen um die exakten Verteilungen der Teststatistiken um 1970 wird von Rao (1972) zusammengefasst.

14.6. Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario einer einfaktoriellen multivariaten Varianzanalyse.
2. Geben Sie die Definition des Modells der einfaktoriellen multivariaten Varianzanalyse wieder.
3. Geben Sie das Theorem zu den Parameterschätzern der einfaktoriellen multivariaten Varianzanalyse wieder.
4. Erläutern Sie die Null- und Alternativhypothesen einer einfaktoriellen multivariaten Varianzanalyse.
5. Geben Sie das Theorem zur Kreuzproduktschätzmatrixzerlegung wieder.

6. Was messen die Totale, Between-Group und die Within-Group Sum-of-Squares Matrizen, respektive?
7. Geben Sie die Definition der Wilks'- Λ -Statistik wieder.
8. Erläutern Sie Gemeinsamkeiten und Unterschiede zwischen speziellen und approximativen H_0 Verteilungen von Wilks- Λ -Transformationen bei der einfaktoriellen multivariaten Varianzanalyse.
9. Geben Sie das Theorem zum Wilks- Λ -Statistik-basierten Test im Rahmen der einfaktoriellen multivariaten Varianzanalyse wieder.
10. Erläutern Sie das praktische Vorgehen zur Durchführung eines Wilks- Λ -Statistik-basierten Test im Rahmen der einfaktoriellen multivariaten Varianzanalyse.

15. Kanonische Korrelationsanalyse

15.1. Anwendungsszenario und Grundzüge

Ausgangspunkt einer kanonischen Korrelationsanalyse ist die exploratorische Frage nach der Stärke des Zusammenhangs einer multivariaten unabhängigen Variablen x und einer multivariaten abhängigen Variable y . Im Kontext der kanonischen Korrelationsanalyse werden diese Variablen auch oft als *Prädiktoren* und *Kriterien*, respektive, bezeichnet. Dabei werden vorliegende Daten von sowohl der Prädiktoren als auch der Kriterien als Realisierungen von unabhängig und identisch verteilten m_x - bzw. m_y -dimensionalen Zufallsvektoren betrachtet. Um die Frage nach der Stärke des Zusammenhangs von Prädiktoren und Kriterien zu beantworten, betrachtet die kanonische Korrelationsanalyse Linearkombinationen der Komponenten der Prädiktoren und Kriterien. Wir bezeichnen die Linearkombinationen von x und y mit Parametervektoren $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ im Folgenden mit

$$\xi := a^T x \text{ und } v := b^T y \quad (15.1)$$

ξ und v sind dann als Linearkombinationen von Zufallsvariablen selbst Zufallsvariablen, deren Korrelation $\rho(\xi, v)$ berechnet werden kann. Die kanonische Korrelationsanalyse bestimmt dann die Parametervektoren a und b so, dass die Korrelation von ξ und v so groß wie möglich wird. Ist eine solche Parametervektorkombination und ihre entsprechende Korrelation, die dann als kanonische Korrelation bezeichnet gefunden, so kann ξ als bester Prädiktor und v als “am besten prädizierbares Kriterium” interpretiert werden.

Für Skalare α und β sind die Korrelationen $\rho(\xi, v)$ und $\rho(\alpha\xi, \beta v)$ allerdings, wie im Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen gezeigt, identisch. Die kanonische Korrelationsanalyse sucht deshalb speziell Parametervektoren a und b , für die $\rho(\xi, v)$ einerseits maximal ist und für die ξ und v gleichzeitig jeweils eine standardisierte Varianz von 1 haben. Da aufgrund des Theorem zu Kovarianz und Korrelation bei linear-affinen Transformationen die Varianzen zu verschiedenen skalaren Vielfachen von ξ und v verschieden sind, legt diese die Parametervektorwerte a und b , für die $\rho(\xi, v)$ maximal ist, eindeutig fest. Zur Bestimmung einer kanonischen Korrelation und der Parametervektoren von a und b ist man also auf ein Optimierungsproblem mit Nebenbedingungen geführt.

In unserer Darstellung kanonischen Korrelationsanalyse folgen wir Mardia et al. (1979). Dabei werden die Prädiktor- und Kriterienzufallsvektoren x und y in einem Zufallsvektor

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \quad (15.2)$$

zusammengefasst, für den wir durchgängig annehmen wollen, dass $\mathbb{E}(z) = 0_m$ mit $m = m_x + m_y$ gilt. Im Anwendungskontext entspricht dies der Annahme, dass vor Durchführung der kanonischen Korrelationsanalyse die Stichprobenmittel des Stichprobenmittels von den beobachteten Daten vor Durchführung der kanonischen Korrelationsanalyse subtrahiert

werden. Da, wie wir sehen werden, die Schätzung der kanonischen Korrelation allerdings lediglich auf den Stichprobenkovarianzmatrizen beruht ist dieser Schritt verzichtbar.

Der mathematische Fokus der Entwicklung der kanonischen Korrelationsanalyse nach Mardia et al. (1979) liegt damit auf der Kovarianzmatrix $\mathbb{C}(z)$. Speziell ergeben sich die Kovarianzen von Linearkombinationen von x und y aus Matrixprodukten von $\mathbb{C}(z)$ und es können einige Matrixtheoreme, die in Kapitel 11 und Kapitel 10 dokumentiert sind angewendet werden. Generell wird in der Entwicklung nach Mardia et al. (1979) ein Optimierungsansatz mithilfe von Lagrangefunktionen, wie in den Originalarbeiten von Hotelling (1935) und Hotelling (1936) gewählt, zugunsten der Eigenanalyse von Matrixprodukten supprimiert. Für die Entwicklung mit einem Lagrangeansatz verweisen wir auf zum Beispiel Anderson (2003).

Im Folgenden diskutieren wir in Kapitel 15.2 zunächst zwei Theoreme, die direkt durch die kanonische Korrelationsanalyse motiviert sind und die, zusammen mit den oben erwähnten Theoremen in Kapitel 11 und Kapitel 10 das analytische und probabilistische Fundament der kanonischen Korrelationsanalyse bilden. In Kapitel 15.3 führen wir mit den *kanonischen Korrelationen*, den *kanonischen Variaten* und den *kanonischen Koeffizientenvektoren* dann die zentralen Begriffe der kanonischen Korrelationsanalyse ein. In Kapitel 15.4 diskutieren wir dann, wie diese Größen auf Basis der Stichprobenkovarianzmatrix eines Datensatzes von Prädiktoren- und Kriterienrealisierungen geschätzt werden können. In Kapitel 15.5 schließlich wenden wir die kanonische Korrelationsanalyse auf das unten skizzierte Anwendungsbeispiel an.

Anwendungsbeispiel

Als konkretes Anwendungsbeispiel für eine kanonische Korrelationsanalyse betrachten wir einen simulierten Datensatz zur Effektivität einer Psychotherapie bei Depressionen. Dabei seien pro Patient:in vier Variablen verfügbar: Zum einen als Maße für die Therapiequalität die Dauer der Psychotherapie (DUR) und die klinische Erfahrung der behandelnden Psychotherapeut:in (EXP); zum anderen als Maße für die Reduktion der Depressionssymptomatik sowohl BDI Score und Glukokortikoidplasmalevel Differenzwerte zwischen Beginn und Ende der Therapie (dBDI und dGLU), wobei positive Werte eine Verringerung der Depressionssymptomatik anzeigen sollen. Wir stellen uns vor, dass man im Sinne einer exploratorischen Analyse daran interessiert ist, inwieweit die Variablen DUR (x_1) und EXP (x_2) als Prädiktoren (unabhängige Variablen) mit den Variablen dBDI (y_1) und dGLU (y_2) als Kriterien (abhängige Variablen) zusammenhängen. Im Sinne obiger Bezeichner gilt hier also $m_x = m_y = 2$. Tabelle 15.1 zeigt dazu einen Beispieldatensatz für $n = 20$ Patient:innen.

Tabelle 15.1. Beispielhafte Prädiktoren- und Kriterienrealisierungen im Kontext der Psychotherapie bei Depressionen

DUR	EXP	dBDI	dGLU
27.9	7.8	35.5	6.1
15.3	9.3	25.0	4.0
17.4	2.1	19.7	1.7
21.5	6.5	28.8	2.6
28.2	1.3	29.4	1.9
14.0	2.7	17.2	0.9
28.0	3.9	32.9	2.0
28.9	0.1	28.3	4.1
23.2	3.8	25.8	3.9
22.6	8.7	31.3	3.8
11.2	3.4	14.4	2.1
14.1	4.8	18.4	2.0
13.5	6.0	19.1	5.0
23.7	4.9	28.0	2.6
17.7	1.9	20.3	2.1
25.4	8.3	34.8	4.4
20.0	6.7	27.6	4.0
24.4	7.9	31.9	3.9
29.8	1.1	32.2	1.0
17.6	7.2	24.6	1.9

15.2. Mathematischer Hintergrund

Folgendes Theorem bildet das analytische Fundament der kanonischen Korrelationsanalyse.

Theorem 15.1 (Maximierung quadratischer Formen mit Nebenbedingungen). $A \in \mathbb{R}^{m \times m}, B \in \mathbb{R}^{m \times m}$ *pd* seien symmetrische Matrizen und λ_1 sei der größte Eigenwert von $B^{-1}A$ mit assoziiertem Eigenvektor $v_1 \in \mathbb{R}^m$. Dann ist λ_1 eine Lösung des Optimierungsproblems

$$\max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \tag{15.3}$$

◦

Beweis. $B^{1/2}$ sei die symmetrische Quadratwurzel von B und es sei

$$y := B^{1/2}x \Leftrightarrow x = B^{-1/2}y \tag{15.4}$$

Dann kann mit der symmetrischen Matrix

$$K := B^{-1/2}AB^{-1/2} \in \mathbb{R}^{m \times m} \tag{15.5}$$

das Optimierungsproblem (15.3) geschrieben werden als

$$\max_y y^T K y \text{ unter der Nebenbedingung } y^T y = 1. \tag{15.6}$$

Dies gilt, weil

$$\max_x x^T A x \Leftrightarrow \max_y (B^{-1/2}y)^T A (B^{-1/2}y) \Leftrightarrow \max_y y^T B^{-1/2}AB^{-1/2}y \Leftrightarrow \max_y y^T K y \tag{15.7}$$

und

$$x^T Bx = 1 \Leftrightarrow y^T B^{-1/2} B B^{-1/2} y = 1 \Leftrightarrow y^T y = 1. \quad (15.8)$$

Weil K eine symmetrische Matrix ist, existiert die Orthonormalzerlegung (vgl. (2) Eigenanalyse)

$$K = Q \Lambda Q^T, \quad (15.9)$$

wobei die Spalten der orthogonalen Matrix Q die Eigenvektoren von K und die Diagonalelemente von Λ die zugehörigen Eigenwerte von K sind. Mit der orthogonalen Matrix Q aus obiger Orthonormalzerlegung sei nun

$$z := Q^T y \Leftrightarrow y := Qz. \quad (15.10)$$

Dann kann das Optimierungsproblem (15.6) geschrieben werden als

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \text{ unter der Nebenbedingung } z^T z = 1, \quad (15.11)$$

weil

$$\max_y y^T K y \Leftrightarrow \max_z (Qz)^T K (Qz) \Leftrightarrow \max_z z^T Q^T Q \Lambda Q^T Q z \Leftrightarrow \max_z z^T \Lambda z \Leftrightarrow \max_z \sum_{i=1}^m \lambda_i z_i^2 \quad (15.12)$$

und

$$y^T y = 1 \Leftrightarrow (Qz)^T Qz = 1 \Leftrightarrow z^T Q^T Q z = 1 \Leftrightarrow z^T z = 1. \quad (15.13)$$

Die Eigenwerte von K seien nun absteigend sortiert, also $\lambda_1 \geq \dots \geq \lambda_m$. Dann gilt für das Optimierungsproblem (15.11), dass

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \lambda_1, \quad (15.14)$$

weil

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 \leq \max_z \sum_{i=1}^m \lambda_1 z_i^2 = \lambda_1 \max_z \sum_{i=1}^m z_i^2 = \lambda_1 \quad (15.15)$$

wobei sich die letzte Gleichung aus der Nebenbedingung $z^T z = 1$ ergibt. Schließlich gilt

$$\max_z \sum_{i=1}^m \lambda_i z_i^2 = \lambda_1, \quad (15.16)$$

für $z := e_1 = (1, 0, \dots, 0)^T$. Zusammenfassend heißt das, dass $z = e_1$ eine Lösung des Optimierungsproblem (15.11) ist und das λ_1 das entsprechende Maximum ist. Damit ergibt sich aber sofort, dass dann

$$y = Qz = Qe_1 = q_1 \text{ und } x = B^{-1/2} q_1 \quad (15.17)$$

Lösungen der äquivalenten Optimierungsprobleme (15.6) und (15.3), respektive, sind. Nach Konstruktion ist q_1 ein Eigenvektor von $B^{-1/2} A B^{-1/2}$ und nach obigem Theorem zu Eigenwerten und Eigenvektoren von Matrixprodukten damit auch ein Eigenvektor von

$$B^{-1/2} B^{-1/2} A = B^{-1} A \quad (15.18)$$

und die zugehörigen Eigenwerte sind gleich. Damit aber folgt, dass der größte Eigenwert von $B^{-1} A$ und sein assoziierter Eigenvektor eine Lösung von

$$\max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \quad (15.19)$$

ist.

□

Nach Wortlaut des Theorems gilt also für die Funktion

$$f : \mathbb{R}^m \rightarrow \mathbb{R}, x \mapsto f(x) := x^T A x, \quad (15.20)$$

dass

$$v_1 = \arg \max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1 \quad (15.21)$$

und dass

$$\lambda_1 = \max_x x^T A x \text{ unter der Nebenbedingung } x^T B x = 1. \quad (15.22)$$

Das folgende Theorem setzt die für die kanonischen Korrelationsanalyse zentralen Größen der Varianzen von Linearkombinationen von Zufallsvektoren und der Korrelation von Linearkombinationen von Zufallsvektoren in Bezug zur Kovarianzmatrix der gemeinsamen Verteilung der Zufallsvektoren und bildet das probabilistische Fundament der kanonischen Korrelationsanalyse.

Theorem 15.2 (Linearkombinationen von Zufallsvektorpartitionen).

Es sei

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) = 0_m \text{ und } \mathbb{C}(z) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad (15.23)$$

ein m -dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswertvektor und seine Kovarianzmatrix, respektive. Weiterhin seien für $a \in \mathbb{R}^{m_x}$ und $b \in \mathbb{R}^{m_y}$ die Zufallsvariablen

$$\xi := a^T x \text{ und } v := b^T y \quad (15.24)$$

als Linearkombinationen der Komponenten von x und y definiert. Dann gelten

- (1) $\mathbb{V}(\xi) = a^T \Sigma_{xx} a$
- (2) $\mathbb{V}(v) = b^T \Sigma_{yy} b$
- (2) $\rho(\xi, v) = a^T \Sigma_{xy} b$, wenn $\mathbb{V}(\xi) = 1$ und $\mathbb{V}(v) = 1$.

◦

Beweis. Wir betrachten zunächst die Varianz von ξ . Mit dem Varianzverschiebungssatz gilt

$$\begin{aligned} \mathbb{V}(\xi) &= \mathbb{E}(\xi\xi) - \mathbb{E}(\xi)\mathbb{E}(\xi) \\ &= \mathbb{E}((a^T x)(a^T x)) - \mathbb{E}(a^T x)\mathbb{E}(a^T x) \\ &= \mathbb{E}((a^T x)(a^T x)^T) - \mathbb{E}(a^T x)\mathbb{E}(a^T x) \\ &= \mathbb{E}(a^T x x^T a) - \mathbb{E}(a^T x)\mathbb{E}(a^T x) \\ &= a^T \mathbb{E}(x x^T) a - a^T \mathbb{E}(x) a^T \mathbb{E}(x) \\ &= a^T \mathbb{E}(x x^T) a - a^T 0_{m_x} a^T 0_{m_x} \\ &= a^T \Sigma_{xx} a. \end{aligned} \quad (15.25)$$

Der Beweis zur Varianz von v folgt dann analog. Mit der Definition der Korrelation von Zufallsvariablen und mit $\mathbb{V}(\xi) = \mathbb{V}(v) = 1$ und dem Kovarianzverschiebungssatz gilt

$$\begin{aligned} \rho(\xi, v) &= \frac{\mathbb{C}(\xi, v)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(v)}} \\ &= \frac{\mathbb{C}(\xi, v)}{\sqrt{1}\sqrt{1}} \\ &= \mathbb{C}(\xi, v) \\ &= \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v) \\ &= \mathbb{E}((a^T x)(b^T y)) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\ &= \mathbb{E}((a^T x)(b^T y)^T) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\ &= \mathbb{E}(a^T x y^T b) - \mathbb{E}(a^T x)\mathbb{E}(b^T y) \\ &= a^T \mathbb{E}(x y^T) b - a^T \mathbb{E}(x) b^T \mathbb{E}(y) \\ &= a^T \mathbb{E}(x y^T) b - a^T 0_{m_x} b^T 0_{m_y} \\ &= a^T \Sigma_{xy} b. \end{aligned} \quad (15.26)$$

□

Die Varianz der Zufallsvariable $a^T x$ ergibt sich also als die “quadrierte” Linearkombination von Σ_{xx} und die Varianz der Zufallsvariable $b^T y$ ergibt sich als die “quadrierte” Linearkombination von Σ_{yy} . Die Korrelation der Zufallsvariablen $a^T x$ und $b^T y$ schließlich ergibt sich “quadrierte” Linearkombination von Σ_{xy} .

15.3. Modellformulierung

Mit den beiden in Kapitel 15.2 diskutierten Theoremen ist es nun möglich, die Begriffe der *kanonischen Koeffizientenvektoren*, der *kanonischen Variate*, und schließlich der *kanonischen Korrelationen* zu definieren.

Definition 15.1 (Kanonische Koeffizientenvektoren, Variate, Korrelationen).

Es seien

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \text{ und } \mathbb{C}(z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (15.27)$$

ein m -dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswert und seine Kovarianzmatrix, respektive. Weiterhin sei

$$K := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \quad (15.28)$$

mit der Singulärwertzerlegung

$$K = A \Lambda B^T, \quad (15.29)$$

wobei

$$A := (\alpha_1 \ \dots \ \alpha_k) \in \mathbb{R}^{m_x \times m_y} \text{ und } B := (\beta_1 \ \dots \ \beta_k) \in \mathbb{R}^{m_y \times m_y} \quad (15.30)$$

die orthogonale Matrix der Eigenvektoren von KK^T und die orthogonale Matrix der Eigenvektoren von $K^T K$, respektive, bezeichnen und

$$\Lambda := \text{diag}(\lambda_1^{1/2}, \dots, \lambda_k^{1/2}) \in \mathbb{R}^{m_y \times m_y}, \quad (15.31)$$

die Diagonalmatrix der Quadratwurzeln der zugehörigen absteigend geordneten Eigenwerte bezeichnet. Schließlich seien für $i = 1, \dots, k$

$$a_i := \Sigma_{xx}^{-1/2} \alpha_i \in \mathbb{R}^{m_x} \text{ und } b_i := \Sigma_{yy}^{-1/2} \beta_i \in \mathbb{R}^{m_y}. \quad (15.32)$$

Dann heißen für $i = 1, \dots, k$

- (1) $a_i \in \mathbb{R}^{m_x}$ und $b_i \in \mathbb{R}^{m_y}$ die *iten kanonischen Koeffizientenvektoren*,
- (2) die Zufallsvektoren $\xi_i := a_i^T x$ und $v_i := b_i^T y$ die *iten kanonischen Variaten* und
- (3) $\rho_i := \lambda_i^{1/2}$ die *ite kanonische Korrelation*.

•

Ihre Bedeutsamkeit erlangen diese Begriffe durch ihre Eigenschaften, die wir im folgenden Theorem zusammenfassen.

Theorem 15.3 (Eigenschaften kanonischer Korrelationen und Variaten). *Es seien*

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \text{ mit } \mathbb{E}(z) := 0_m \text{ und } \mathbb{C}(z) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (15.33)$$

ein m -dimensionaler partitionierter Zufallsvektor sowie sein Erwartungswert und seine Kovarianzmatrix, respektive. Weiterhin seien für $i = 1, \dots, k$ die kanonischen Koeffizientenvektoren a_i, b_i , die kanonischen Variaten ξ, v_i und die kanonischen Korrelationen ρ_i definiert wie oben. Dann gilt, dass für $1 \leq r \leq k$ das Maximum des r ten restringierten Optimierungsproblems

$$\phi_r = \max_{a,b} a^T \Sigma_{xy} b \quad (15.34)$$

unter den Nebenbedingungen

$$a^T \Sigma_{xx} a = 1, \quad b^T \Sigma_{yy} b = 1, \quad a_i^T \Sigma_{xx} a = 0 \text{ für } i = 1, \dots, r-1 \quad (15.35)$$

(1) den Wert $\phi_r = \rho_r$ hat und (2) bei $a = a_r$ und $b = b_r$ angenommen wird.

◦

Beweis. Wir betrachten das restringierte Optimierungsproblem

$$\phi_r^2 = \max_{a,b} (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } a^T \Sigma_{xx} a = 1, b^T \Sigma_{yy} b = 1, a_i^T \Sigma_{xx} a = 0, i = 1, \dots, r-1 \quad (15.36)$$

Wir folgen Mardia et al. (1979), S. 284 und gehen schrittweise vor, d.h. wir lösen das restringierte Optimierungsproblem

$$\phi_r^2 = \max_a \left(\max_b (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \right) \text{ u.d.N. } a^T \Sigma_{xx} a = 1, a_i^T \Sigma_{xx} a = 0, i = 1, \dots, r-1 \quad (15.37)$$

von innen nach außen.

Schritt (1)

Wir wählen wir zunächst ein festes $a \in \mathbb{R}^m$ und betrachten das restringierte Optimierungsproblem

$$\max_b (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \quad (15.38)$$

Dieses Optimierungsproblem kann geschrieben werden als

$$\max_b b^T \Sigma_{yx} a a^T \Sigma_{xy} b \text{ u.d.N. } b^T \Sigma_{yy} b = 1, \quad (15.39)$$

weil gilt, dass

$$(a^T \Sigma_{xy} b)^2 = (a^T \Sigma_{xy} b) (a^T \Sigma_{xy} b) = (a^T \Sigma_{xy} b)^T a^T \Sigma_{xy} b = b^T \Sigma_{yx} a a^T \Sigma_{xy} b. \quad (15.40)$$

Das Optimierungsproblem (15.39) kann nun mithilfe des Theorems zur Maximierung quadratischer Formen mit Nebenbedingungen gelöst werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yx} a a^T \Sigma_{xy} \text{ und } B := \Sigma_{yy}. \quad (15.41)$$

Dann hat (15.39) die Form

$$\max_b b^T A b \text{ unter der Nebenbedingung } b^T B b = 1, \quad (15.42)$$

Das Maximum von (15.42) entspricht nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen dem größten Eigenwert von

$$B^{-1} A = \Sigma_{yy}^{-1} \Sigma_{yx} a a^T \Sigma_{xy} \quad (15.43)$$

Der größte Eigenwert von $\Sigma_{yy}^{-1}\Sigma_{yx}aa^T\Sigma_{xy}$ wiederum kann mithilfe des Theorems zum Eigenwert und Eigenvektor eines Matrixvektorprodukts bestimmt werden. Im Sinne dieses Theorems setzen wir dazu

$$A := \Sigma_{yy}^{-1}\Sigma_{yx}, \quad b := a, \quad B := \Sigma_{xy} \quad (15.44)$$

und erhalten für den betreffenden Eigenwert

$$\lambda_a = b^T B A a = a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a. \quad (15.45)$$

als Lösung (Maximum) des restringierten Optimierungsproblems

$$\max_b (a^T \Sigma_{xy} b)^2 \text{ u.d.N. } b^T \Sigma_{yy} b = 1 \quad (15.46)$$

Schritt (2)

Basierend auf Schritt (1) verbleibt die Lösung des restringierten Optimierungsproblem

$$\phi_r^2 = \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, \quad a_i^T \Sigma_{xx} a = 0, \quad i = 1, \dots, r-1 \quad (15.47)$$

Dazu halten wir zunächst fest, dass (15.47) mit den Definitionen von α_i und K in der Definition der Kanonischen Koeffizientenvektoren, Variaten, und Korrelationen geschrieben werden kann als

$$\phi_r^2 = \max_\alpha \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0, \quad i = 1, \dots, r-1, \quad (15.48)$$

denn

$$\begin{aligned} \phi_r^2 &= \max_a a^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } a^T \Sigma_{xx} a = 1, \quad a_i^T \Sigma_{xx} a = 0 \Leftrightarrow \\ \phi_r^2 &= \max_\alpha \alpha^T \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} a \text{ u.d.N. } \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 1, \quad \alpha_i^T \Sigma_{xx}^{-1/2} \Sigma_{xx} \Sigma_{xx}^{-1/2} \alpha = 0 \\ \phi_r^2 &= \max_\alpha \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0 \\ \phi_r^2 &= \max_\alpha \alpha^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2} \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0 \\ \phi_r^2 &= \max_\alpha \alpha^T K K^T \alpha \text{ u.d.N. } \alpha^T \alpha = 1, \quad \alpha_i^T \alpha = 0 \end{aligned} \quad (15.49)$$

Dabei sind nach der betreffenden Definition die α_i die Eigenvektoren von $K K^T$ mit den $i = 1, \dots, r-1$ größten Eigenwerten. Nach dem Theorem zur Maximierung quadratischer Formen mit Nebenbedingungen ist die Lösung von (15.48) der größte Eigenwert von $K K^T$ mit seinem assoziierten Eigenvektor. Die Nebenbedingung $\alpha_i^T \alpha = 0$ schränkt diese Wahl auf den r -ten größten Eigenwert und seinen assoziierten Eigenvektor α_r ein. Mit der Definition von Eigenwerten und Eigenvektoren gilt also

$$\phi_r^2 = \alpha_r^T K K^T \alpha_r = \alpha_r^T \lambda_r \alpha_r = \lambda_r \alpha_r^T \alpha_r = \lambda_r. \quad (15.50)$$

Wir haben also gezeigt, dass das restringierte Optimierungsproblem des Theorems den Maximumwert $\phi_r = \lambda_r^{1/2}$ hat. Es bleibt zu zeigen, dass dieser Maximumwert für a_r und b_r angenommen wird.

Schritt (3)

Einsetzen von a_r und b_r in $a^T \Sigma_{xy} b$ ergibt mit

$$K = A \Lambda B^T \Leftrightarrow K B = A \Lambda B^T B \Leftrightarrow K B = A \Lambda \Leftrightarrow K \beta_r = \alpha_r \lambda_r^{1/2} \quad (15.51)$$

dass

$$a_r^T \Sigma_{xy} b_r = \alpha_r^T \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \beta_r = \alpha_r^T K \beta_r = \alpha_r^T \alpha_r \lambda_r^{1/2} = \rho_r \quad (15.52)$$

Also nimmt $a^T \Sigma_{xy} b$ bei a_r und b_r seinen restringierten Maximalwert λ_r an.

□

ϕ_1 ist also die größtmögliche Korrelation von

$$\xi = a^T x \text{ und } \nu = b^T y \quad (15.53)$$

unter den Nebenbedingungen

$$\mathbb{V}(\xi) = 1 \text{ und } \mathbb{V}(v) = 1 \quad (15.54)$$

und erfüllt damit die Forderungen an *die* kanonische Korrelatione. ϕ_r mit $r > 1$ ist die größtmögliche Korrelation von

$$\xi = a^T x \text{ und } v = b^T y \quad (15.55)$$

unter den Nebenbedingungen

$$\mathbb{V}(\xi) = 1, \mathbb{V}(v) = 1 \text{ und } \mathbb{C}(\xi_i, \xi) = 0 \text{ für die kanonischen Variate } \xi_i \text{ mit } i = 1, \dots, r - 1. \quad (15.56)$$

Simulationsbeispiel

Wir betrachten das Beispiel (vgl. Uurtio et al. (2018))

$$p(x) = N(x; 0_4, I_4) \text{ und } p(y|x) = N(y; Lx, G) \quad (15.57)$$

mit

$$L := \begin{pmatrix} 0.0 & 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 \end{pmatrix} \text{ und } G := \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \quad (15.58)$$

Hier gilt offenbar $m_x = 4, m_y = 3, m = 7$ und

$$\begin{aligned} y_1 &= x_3 + \varepsilon_1 \\ y_2 &= x_1 + \varepsilon_2 \\ y_3 &= -x_4 + \varepsilon_3 \end{aligned} \quad (15.59)$$

mit

$$x_1 \sim N(0, 1), x_3 \sim N(0, 1), x_4 \sim N(0, 1) \quad (15.60)$$

und

$$\varepsilon_1 \sim N(0, 0.2), \varepsilon_2 \sim N(0, 0.4), \varepsilon_3 \sim N(0, 0.3) \quad (15.61)$$

Mit dem Theorem zu gemeinsamen Normalverteilungen (vgl. Kapitel 11.6) ergibt sich, dass

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N(0_7, \Sigma) \quad (15.62)$$

mit

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \quad (15.63)$$

wobei

$$\Sigma_{xx} = I_4, \quad \Sigma_{xy} = L^T, \quad \Sigma_{yx} = L \text{ und } \Sigma_{yy} = G + LL^T. \quad (15.64)$$

Explizit ergibt sich also

$$\Sigma = \begin{pmatrix} I_4 & L^T \\ L & G + LL^T \end{pmatrix} = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 1.2 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.4 & 0.0 \\ 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & 1.3 \end{pmatrix} \quad (15.65)$$

Folgender **R** Code definiert zunächst den Kovarianzmatrixparameter der gemeinsamen Verteilung von x und y .

```

1 # R Pakete für Matrizenrechnung
2
3 library(expm)
4
5 # Modellparameter
6 L = matrix(c(0,0,1, 0,
7             1,0,0, 0,
8             0,0,0,-1),
9           nrow = 3,
10          byrow = T)
11 G = diag(c(0.2,0.4,0.3))
12
13 # Kovarianzmatrixpartition
14 Sigma_xx = diag(4)
15 Sigma_xy = t(L)
16 Sigma_yx = L
17 Sigma_yy = G + L %>% t(L)
18 Sigma = rbind(cbind(Sigma_xx, Sigma_xy), cbind(Sigma_yx, Sigma_yy))
19 print(Sigma)

```

```

      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]  1.0  0.0  0.0  0.0  0.0  1.0  0.0
[2,]  0.0  1.0  0.0  0.0  0.0  0.0  0.0
[3,]  0.0  0.0  1.0  0.0  1.0  0.0  0.0
[4,]  0.0  0.0  0.0  1.0  0.0  0.0 -1.0
[5,]  0.0  0.0  1.0  0.0  1.2  0.0  0.0
[6,]  1.0  0.0  0.0  0.0  0.0  1.4  0.0
[7,]  0.0  0.0  0.0 -1.0  0.0  0.0  1.3

```

Anhand von Definition 15.1 bestimmt folgender **R** Code dann basierend auf obigem Kovarianzmatrixparameter die kanonischen Korrelationen und kanonischen Koeffizientenvektoren.

```

1 # Evaluation der iten kanonischen Koeffizientenvektoren und Korrelationen
2 K = sqrtm(solve(Sigma_xx)) %>% Sigma_xy %>% sqrtm(solve(Sigma_yy)) # K
3 ALB = svd(K) # K = A\Lambda V
4 A = ALB$u # A
5 Lambda = ALB$d # Lambda
6 B = ALB$v # B
7 rho = Lambda # \rho_i = \lambda_i^{1/2}
8 a = sqrtm(solve(Sigma_xx)) %>% A # a_i = \Sigma_{xx}^{-1/2} \alpha_i
9 b = sqrtm(solve(Sigma_yy)) %>% B # b_i = \Sigma_{yy}^{-1/2} \beta_i

```

```

rho_1 = 0.9128709 , a_1^T = ( 0 0 -1 0 ) , b_1^T = ( -0.9128709 0 0 )
rho_2 = 0.877058 , a_2^T = ( 0 0 0 1 ) , b_2^T = ( 0 0 -0.877058 )
rho_3 = 0.8451543 , a_3^T = ( -1 0 0 0 ) , b_3^T = ( 0 -0.8451543 0 )

```

15.4. Modellschätzung

Zur Schätzung einer kanonischen Korrelationsanalyse wird die Kovarianzmatrix $\mathbb{C}(z)$ des gemeinsamen Zufallsvektors z von Prädiktoren und Kriterien durch ihr Stichprobenäquivalent C ersetzt. Dies ist die Aussage folgender Definition.

Definition 15.2 (Schätzer der kanonischen Korrelationsanalyse). Für $i = 1, \dots, n$ seien

$$z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \text{ mit } \mathbb{E}(z_i) := 0_m \text{ und } \mathbb{C}(z_i) := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (15.66)$$

unabhängig und identisch verteilte m -dimensionale partitionierte Zufallsvektoren sowie ihr Erwartungswert und ihre Kovarianzmatrix, respektive, und

$$C := \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (15.67)$$

sei ihre Stichprobenkovarianzmatrix. Dann sind für $i = 1, \dots, k := \min\{m_x, m_y\}$

$$\hat{a}_i := C_{xx}^{-1/2} \hat{\alpha}_i \in \mathbb{R}^{m_x}, \quad \hat{b}_i := C_{yy}^{-1/2} \hat{\beta}_i \in \mathbb{R}^{m_y} \text{ und } \hat{\rho}_i := \hat{\lambda}_i^{1/2} \quad (15.68)$$

Schätzer der i ten kanonischen Koeffizientenvektoren und kanonischen Korrelationen, respektive. Dabei sind mit

$$\hat{K} := C_{xx}^{-1/2} C_{xy} C_{yy}^{-1/2} \in \mathbb{R}^{m_x \times m_y} \quad (15.69)$$

$\hat{\alpha}_i$ und $\hat{\lambda}_i$ der i te Eigenvektor und sein zugehöriger Eigenwert von $\hat{K} \hat{K}^T$ und $\hat{\beta}_i$ der entsprechende Eigenvektor von $\hat{K}^T \hat{K}$.

•

Wir verzichten auf eine Diskussion der Güte dieser Schätzung.

Simulationsbeispiel

Mithilfe folgenden **R** Codes verdeutlichen wir uns die Schätzung kanonischer Korrelationen und Koeffizientenvektoren in dem oben betrachteten Beispiel. Dazu generieren wir Realisierungen der Prädiktoren und Kriterien bei Stichprobenumfängen zwischen $n = 100$ und $n = 1000$. Abbildung 15.1 visualisiert die wahren, aber unbekannt, kanonischen Korrelationen ρ_1, ρ_2, ρ_3 und ihre für jede Simulation basierend auf der Stichprobenkovarianzmatrix des realisierten Datensatzes geschätzte Äquivalente $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$. Die Variabilität der Schätzung nimmt mit zunehmenden Stichprobenumfang ab. Abbildung 15.2 visualisiert die Absolutwerte des wahren, aber unbekannt, ersten kanonischen Koeffizientenvektors a_1 sowie ihre entsprechenden Schätzungen als \hat{a}_1 . Auch hier nimmt die Variabilität der Schätzung mit zunehmenden Stichprobenumfang ab. Allerdings ist zu beachten, dass es sich hierbei um die Absolutwerte des kanonischen Koeffizientenvektors handelt und das Vorzeichen abhängig von der Schätzung auch im Widerspruch zum wahren, aber unbekannt, Wert des kanonischen Koeffizientenvektors stehen kann.

```

1 # R Pakete
2 library(MASS)
3 library(expm)
4
5 # Modellparameter
6 m_x = 4
7 m_y = 3
8 k = min(m_x, m_y)
9 L = matrix(c(0,0,1,0,1,0,0,0,0,0,0,-1), nrow = 3, byrow = 3)
10 G = diag(c(0.2,0.4,0.3))
11 Sigma_xx = diag(4)

```

```

12 Sigma_xy = t(L)
13 Sigma_yx = L
14 Sigma_yy = G + L %*% t(L)
15 Sigma = rbind(cbind(Sigma_xx, Sigma_xy), cbind(Sigma_yx, Sigma_yy))
16 K = sqrtm(solve(Sigma_xx)) %*% Sigma_xy %*% sqrtm(solve(Sigma_yy))
17 ALB = svd(K)
18 A = ALB$u
19 Lambda = ALB$d
20 B = ALB$v
21 rho = Lambda
22 a = sqrtm(solve(Sigma_xx)) %*% A
23 b = sqrtm(solve(Sigma_yy)) %*% B

1 # Simulationen
2 n = 1e1:1e3
3 rho_hat = matrix(rep(NaN, length(n)*k), nrow = k)
4 a_1_hat = matrix(rep(NaN, length(n)*m_x), nrow = m_x)
5 for(i in 1:length(n)){
6
7   # Datengeneration
8   Y = t(mvnrnorm(n[i], rep(0, m_x+m_y), Sigma))
9   I_n = diag(n[i])
10  J_n = matrix(rep(1, n[i]^2), nrow = n[i])
11
12  # Stichprobenkovarianzmatrixpartition
13  C = (1/(n[i]-1))*Y %*% (I_n - (1/n[i])*J_n) %*% t(Y)
14  C_xx = C[1:m_x, 1:m_x]
15  C_xy = C[1:m_x, (m_x+1):(m_x+m_y)]
16  C_yx = C[(m_x+1):(m_x+m_y), 1:m_x]
17  C_yy = C[(m_x+1):(m_x+m_y), (m_x+1):(m_x+m_y)]
18
19  # Kanonische Korrelationsanalyse
20  K_hat = sqrtm(solve(C_xx)) %*% C_xy %*% sqrtm(solve(C_yy))
21  ALB_hat = svd(K_hat)
22  A_hat = ALB_hat$u
23  Lambda_hat = ALB_hat$d
24  B_hat = ALB_hat$v
25  a_hat = sqrtm(solve(C_xx)) %*% A_hat
26  b_hat = sqrtm(solve(C_yy)) %*% B_hat
27  rho_hat[,i] = as.matrix(Lambda_hat)
28  a_1_hat[,i] = a_hat[,1]
29 }

```

Schätzung kanonischer Korrelationen

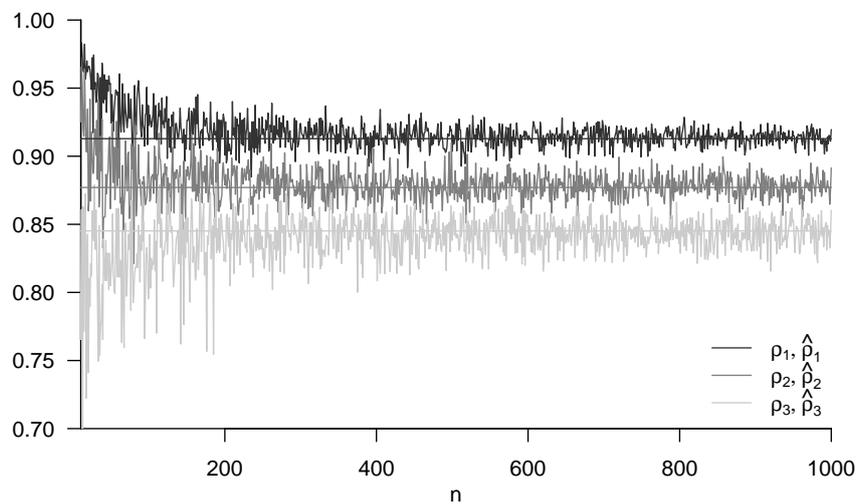


Abbildung 15.1. Schätzung kanonischer Korrelationen im Simulationsbeispiel

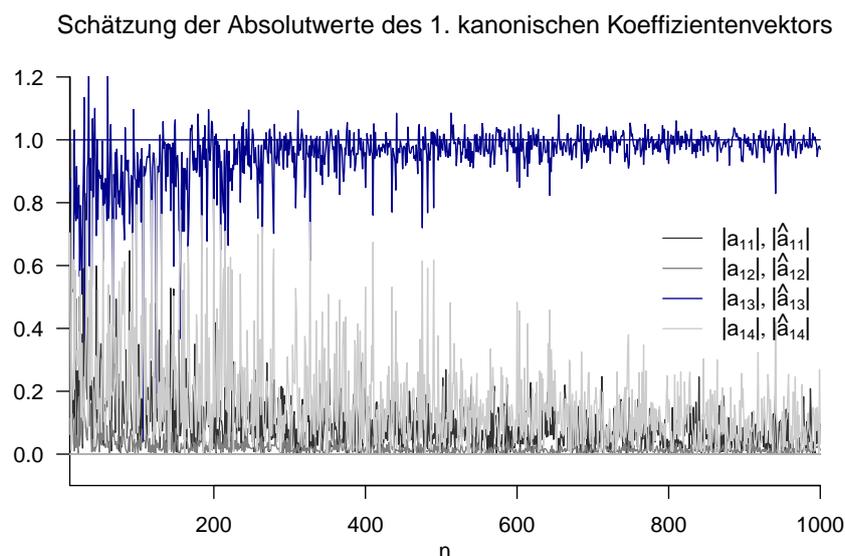


Abbildung 15.2. Schätzung der Absolutwerte des ersten kanonischen Koeffizientenvektors im Simulationsbeispiel

15.5. Anwendungsbeispiel

Zuletzt wollen wir die konkrete Berechnung einer kanonischen Korrelationsanalyse im Kontext des in Kapitel 15.1 diskutierten Anwendungsbeispiels demonstrieren. Folgender **R** Code implementiert die Berechnung der kanonischen Korrelationen und kanonischen Koeffizientenvektoren für diesen Datensatz.

```

1  # R Paket
2  library(expm)
3
4  # Datenpräprozessierung
5  fname = "./_data/204-kanonische-korrelationsanalyse.csv"
6  D = read.table(fname, sep = ",", header = TRUE)
7  x = as.matrix(cbind(D$DUR, D$EXP))
8  y = as.matrix(cbind(D$dBDI, D$dGLU))
9  n = nrow(x)
10 m_x = ncol(x)
11 m_y = ncol(y)
12 Y = t(cbind(x,y))
13
14 # Stichprobenkovarianzmatrixpartition
15 I_n = diag(n)
16 J_n = matrix(rep(1,n^2), nrow = n)
17 C = (1/(n-1))*(Y %>% (I_n-(1/n)*J_n) %>% t(Y))
18 C_xx = C[1:m_x,1:m_x]
19 C_xy = C[1:m_x,(m_x+1):(m_x+m_y)]
20 C_yx = C[(m_x+1):(m_x+m_y),1:m_x]
21 C_yy = C[(m_x+1):(m_x+m_y),(m_x+1):(m_x+m_y)]
22
23 # Kanonische Korrelationsanalyse
24 K_hat = sqrtm(solve(C_xx)) %>% C_xy %>% sqrtm(solve(C_yy))
25 ALB_hat = svd(K_hat)
26 A_hat = ALB_hat$u
27 Lambda_hat = ALB_hat$d
28 B_hat = ALB_hat$v
29 a_hat = sqrtm(solve(C_xx)) %>% A_hat
30 b_hat = sqrtm(solve(C_yy)) %>% B_hat
31 rho_hat = as.matrix(Lambda_hat)

```

```

rho_hat_1 : 0.9950575
a_hat_1   : -0.1623409 -0.173979
b_hat_1   : -0.1554175 -0.05025419
rho_hat_2 : 0.5010358
a_hat_2   : -0.06026274 0.3118808
b_hat_2   : -0.08128072 0.7773036

```

Neben der Implementation mithilfe einer Singulärwertzerlegung bietet **R** auch eine direkte Bestimmung mithilfe der `cancor()` Funktion an. Folgender **R** Code demonstriert das entsprechende Vorgehen.

```
1 # Datenpräprozessierung
2 fname = "../data/204-kanonische-korrelationsanalyse.csv"
3 D = read.table(fname, sep = ",", header = TRUE)
4 x = as.matrix(cbind(D$DUR, D$EXP))
5 y = as.matrix(cbind(D$dBDI, D$dGLU))
6 cca = cancor(x,y)
```

```
rho_hat_1 : 0.9950575
rho_hat_2 : 0.5010358
```

Man findet also, dass die geschätzte maximale Korrelation einer Linearkombinationen der Prädiktorvariablen `DUR` und `EXP` mit einer Linearkombination der Kriterien `dBDI` und `dGLU` mit $\hat{\rho}_1 = 0.99$ sehr hoch ist. Man kann daraus schließen, dass in diesem Fall die Prädiktorvariablen gemeinschaftlich hoch mit den Kriterien assoziiert sind. Es ergeben sich hier insbesondere die Linearkombinationen

$$\xi = 0.16 \text{ DUR} + 0.17 \text{ EXP} \text{ und } v = 0.15 \text{ dBDI} + 0.05 \text{ dGLU} \quad (15.70)$$

als bester Prädiktor und als am besten prädizierbares Kriterium, respektive. Die Dauer der Psychotherapie und die Erfahrung der behandelnden Psychotherapeut:in scheinen, bei der aktuellen Datenskalierung zur bestmöglichen Prädiktion der Therapiegröße also in etwa gleichbedeutend, bei dem bestprädizierbarem Kriterium der Therapieeffizienz trägt bei der aktuellen Datenskalierung die BDI Score Reduktion etwas mehr bei als die Glukokortikoidplasmalevel Reduktion bei.

15.6. Literaturhinweise

Die kanonische Korrelationsanalyse geht zurück auf Hotelling (1935) und Hotelling (1936).

15.7. Selbstkontrollfragen

Referenzen

- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed). Wiley-Interscience.
- 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>
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. <https://doi.org/10.1080/01621459.2017.1285773>
- 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>
- Christensen, R. (2011). *Plane Answers to Complex Questions*. Springer New York. <https://doi.org/10.1007/978-1-4419-9816-3>
- DeGroot, M. H., & Schervish, M. J. (2012). *Probability and Statistics* (4th ed). Addison-Wesley.
- Edgeworth, F. Y. (1892). The Law of Error and Correlated Averages. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 34(210), 429–438. <https://doi.org/10.1080/14786449208620355>
- 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>
- Friston, K., Da Costa, L., Sakthivadivel, D. A. R., Heins, C., Pavliotis, G. A., Ramstead, M., & Parr, T. (2023). Path Integrals, Particular Kinds, and Strange Things. *Physics of Life Reviews*, 47, 35–62. <https://doi.org/10.1016/j.plrev.2023.08.016>
- Golub, G. H., & Van Loan, C. F. (2013). *Matrix Computations* (Fourth edition). The Johns Hopkins University Press.
- 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
- Hotelling, H. (1931). The Generalization of Student's Ratio. *The Annals of Mathematical Statistics*, 2(3), 360–378. <https://doi.org/10.1214/aoms/1177732979>
- Hotelling, H. (1935). The Most Predictable Criterion. *Journal of Educational Psychology*, 26(2), 139–142. <https://doi.org/10.1037/h0058165>
- Hotelling, H. (1936). Relations Between Two Sets of Variates. *Biometrika*, 28(3/4), 321. <https://doi.org/10.2307/2333955>
- Hotelling, H. (1951). *A Generalized T Test and Measure of Multivariate Dispersion*.
- Mahalanobis, S. A. (1936). *Reprint of: Mahalanobis, P.C. (1936) "On the Generalised*

- Distance in Statistics.* 80(S1), 1–7. <https://doi.org/10.1007/s13171-019-00164-5>
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
- Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica*. Royal Society.
- Ostwald, D., Kirilina, E., Starke, L., & Blankenburg, F. (2014). A Tutorial on Variational Bayes for Latent Linear Stochastic Time-Series Models. *Journal of Mathematical Psychology*, 60, 1–19. <https://doi.org/10.1016/j.jmp.2014.04.003>
- Pearson, K. (1896). Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 18, 253–318. <https://www.jstor.org/stable/90707>
- Pillai, K. C. S. (1955). Some New Test Criteria in Multivariate Analysis. *The Annals of Mathematical Statistics*, 26(1), 117–121. <https://doi.org/10.1214/aoms/1177728599>
- Rao, C. R. (1951). An Asymptotic Expansion of the Distribution of Wilk's Criterion. *Bulletin of the International Statistical Institute*, 33(2), 177–180.
- Rao, C. R. (1972). Recent Trends of Research Work in Multivariate Analysis. *Biometrics*, 28(1), 3. <https://doi.org/10.2307/2528958>
- Rencher, A. C., & Christensen, W. F. (2012). *Methods of Multivariate Analysis* (Third Edition). Wiley.
- 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>
- Roy, S. N. (1953). On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis. *The Annals of Mathematical Statistics*, 24(2), 220–238. <https://doi.org/10.1214/aoms/1177729029>
- Schmidt, K. D. (2009). *Maß und Wahrscheinlichkeit*. Springer.
- Seal, H. L. (1967). Studies in the History of Probability and Statistics. XV: The Historical Development of the Gauss Linear Model. *Biometrika*, 54(1/2), 1. <https://doi.org/10.2307/2333849>
- Searle, S. (1982). *Matrix Algebra Useful for Statistics*. Wiley-Interscience.
- Strang, G. (2009). *Introduction to Linear Algebra*. Cambridge University Press.
- Tong, Y. L. (1990). *Multivariate Normal Distribution*. Springer.
- Unger, L. (2000). *Grundkurs Mathematik*.
- Uurtio, V., Monteiro, J. M., Kandola, J., Shawe-Taylor, J., Fernandez-Reyes, D., & Rousu, J. (2018). A Tutorial on Canonical Correlation Methods. *ACM Computing Surveys*, 50(6), 1–33. <https://doi.org/10.1145/3136624>
- Wilks, S. S. (1932). Certain Generalizations in the Analysis of Variance. *Biometrika*, 24(3-4), 471–494. <https://doi.org/10.1093/biomet/24.3-4.471>