

Mathematik für Ingenieure C4

Wigand Rathmann

Sommersemester 2020

Inhaltsverzeichnis

Einführung	5
1 Wahrscheinlich	8
1.1 Historisches	9
1.2 Gesetz der großen Zahlen	10
1.2.1 Relative Häufigkeiten	12
1.2.2 Wahrscheinlichkeitsrechnung	14
2 Beschreibende Statistik	16
2.1 Lineare Ausgleichsrechnung / Methode der kleinsten Quadrate	25
3 Wahrscheinlichkeitsmodelle	28
3.1 Die Bausteine	28
3.2 Die Ergebnismenge Ω	29
3.3 Zusammengesetzte Ereignisse	29
3.4 Ereignisse, Verknüpfung von Ereignissen	30
3.5 Ereignissystem \mathcal{A}	31
3.6 Darstellung von Ereignissen durch Zufallsvariablen	33
3.7 Relative Häufigkeit und Wahrscheinlichkeit	35
3.8 Weitere Eigenschaften von Wahrscheinlichkeiten	38
3.9 Elementare bedingte Wahrscheinlichkeiten	39
4 Wahrscheinlichkeitsmaße	40
4.1 Diskrete Wahrscheinlichkeitsmaße und Zähldichten	40
4.2 Stetige Wahrscheinlichkeitsmaße und Riemann-Dichten	42
4.3 Verteilungsfunktion	44

5	Mehrstufige Wahrscheinlichkeitsmodelle	47
5.1	Mehrstufige Wahrscheinlichkeitsmodelle	47
5.2	Koppelung stetiger Modelle	48
5.3	Unabhängige Koppelung	48
5.4	Markow-Koppelung	49
5.4.1	Zufälliges Ziehen ohne Zurücklegen	49
6	Zufallsvariablen und Bildmodelle	51
6.1	Zufallsvariablen, Bildmodelle	51
6.2	Bildmodelle und Verteilungen von Zufallsvariablen	52
6.3	Bildmodelle für diskrete Verteilungen	53
6.3.1	Hypergeometrische und Binomial-Verteilung	53
6.3.2	Approximation der Binomial-Verteilung	54
6.3.3	Gelassen Warten	54
6.4	Bildverteilungen stetiger W-Dichten	55
6.5	Randverteilungen und gemeinsame Verteilungen	57
6.6	Stochastische Unabhängigkeit von ZV	58
6.7	Summenverteilung, Faltung	59
7	Kenngrößen	62
7.1	Median und Quantile	62
7.2	Erwartungswert	63
7.3	Erwartungswert: diskrete Modelle	63
7.4	Erwartungswert: stetige und gemischte Modelle	66
7.5	Streuung und Varianz	67
7.6	Kovarianz	68
7.7	Momente	70
7.8	Mehrdimensionale Normalverteilung	71
7.9	Zufällige Summen und bedingte Erwartungswerte	73
7.10	Gesetze der großen Zahlen	74

8 Markow-Ketten	76
8.1 Vorbemerkungen	76
8.2 Grundbegriffe	77
8.3 Gleichgewicht	79
9 Grundgesamtheit und Stichprobe	94
9.1 Rückblick: Beschreibende Statistik	94
9.2 Grundgesamtheit und Stichprobe	95
10 Schätzverfahren	99
10.1 Einleitung	99
10.2 Punktschätzung	99
10.2.1 Begriffe	99
10.2.2 Maximum-Likelihood-Methode	101
10.3 Konfidenzschätzung	102
10.3.1 Begriffe	102
10.3.2 Konfidenzschätzung für den Erwartungswert	103
11 Testen von Hypothesen	107
11.1 Grundbegriffe	107
11.2 Test für $\mathcal{N}(\mu, \sigma^2)$ (EW)	111
11.3 Test für $\mathcal{N}(\mu, \sigma^2)$ (EW, Var unbekannt)	112
11.4 Prüfung der Gleichheit	114
11.5 Nicht vorgestellte Tests	115
11.6 χ^2 -Anpassungstest	115
12 Regressions- und Korrelationsanalyse	117
12.1 Einführung	117
12.2 Regressionsanalyse	117
12.3 Schätzung von β_1, β_2 und σ^2	118
12.4 Prüfen der Parameter β_1, β_2	120
12.5 Korrelationsanalyse	123

A	Anwendungen	A-2
A.1	Kontrastanhebung	A-2
B	Integration im \mathbb{R}^n	B-1
B.1	Parameterintegrale	B-1
B.1.1	Eigentliche Parameterintegrale	B-1
B.1.2	Uneigentliche Parameterintegrale	B-2
B.2	Stetigkeit und Differenzierbarkeit von Funktionen zweier Variablen	B-2
B.2.1	Uneigentliche Parameterintegrale	B-5
B.3	Integration von Funktionen mehrerer Variablen	B-7
B.3.1	Zweidimensionale Integrale	B-8
B.3.2	Substitutionsregel für zweidimensionale Integrale	B-13
B.3.3	Dreidimensionale Integrale	B-14
B.3.4	Substitutionsregel für dreidimensionale Integrale	B-16

Einführung

Skript, Übungsaufgaben

Im StudON

<https://www.studon.fau.de/crs2897780.html>

Beitritt mit Kurspasswort

Kovarianzmatrix

Ausgabe und Abgabe der Übungen

Ausgabe der Übungsblätter: Mo. **vor** der Übung,
1. Blatt am 27.4.

Abgabe der Übungsblätter: bis Do. **bis** 10 Uhr (im Briefkasten oder online, wird bekanntgegeben),
1. Blatt am 28.4.

0.1

Übungsgruppen

<http://univis.uni-erlangen.de/prg?search=lectures&shortname=ingmathc4u&show=plan>

Anmeldung zur Übung

Anmelden unter

Dabei die Gruppen mit Prioritäten anklicken. Die Gruppen werden dann zugewiesen.

Übungsleistung, Bonus

- Für den Erwerb des Scheins werden 50% der Gesamtpunktzahl benötigt. Gesamtpunktzahl steht aber noch nicht fest.
- Bei Erreichen von 75% der Punkte der Gesamtpunktzahl, wird bei Bestehen die Klausurnote um eine Drittelnote verbessert.

0.2

Klausur

- Zugelassene Hilfsmittel:
 - Spickzettel Din A4, handgeschrieben
 - Formelsammlung
- Sonstige Hilfsmittel sind nicht zugelassen (insbesondere elektronische)

0.3

Literatur

Das Kurzsript ist folgenden Büchern und Skripten entnommen:

Hübner, G.: *Stochastik*, Vieweg+Teubner, Wiesbaden, 2009

Beyer, Hackel, Pieper, Tiedge: *Wahrscheinlichkeitsrechnung und mathematische Statistik*, Leipzig, Teubner, 1991

Groß, J.: *Grundlegende Statistik mit R*, Vieweg+Teubner, Wiesbaden, 2010

Graef, F.: *Skript zu Wahrscheinlichkeitsrechnung 1*, FAU AM2, 2003

Weitere Bücher sind

J. Pfanzagl: *Elementare Wahrscheinlichkeitsrechnung*, de Gruyter 1991

A. Leon-Garcia: *Probability and Random Processes for Electrical Engineering*, Addison-Wesley 1994

M. Burkschat, E. Cramer, U. Kramps *Beschreibende Statistik*, 2. Auflage, Springer Spektrum 2012

U. Genschel, C. Becker *Schließende Statistik*, 1. Auflage, Springer 2003

Eine große Auswahl an Stochastik-Büchern steht auch in der Lehrbuchsammlung im TNZB Erlangen.

0.4

Inhalte

- Zufallsexperimente und Wahrscheinlichkeitsräume,
- Laplace-Experimente, Kombinatorik,
- Bedingte Wahrscheinlichkeiten,
- Wahrscheinlichkeiten auf diskreten Ergebnismengen,
- Verteilungen und Zufallsvariable,

- Funktionen von Zufallsvariablen,
- Erwartungswert und Varianz,
- Normalverteilung,
- Grundbegriffe der mathematischen Statistik,
- Grenzwertsätze.

Kapitel 1

Wahrscheinlich

Zentrales Thema der Vorlesung ist die Berechnung der *Wahrscheinlichkeiten von Ereignissen*.

Beispiel 1.1 (Lotto). Wie groß ist die Wahrscheinlichkeit des Ereignisses, dass man beim Lotto „6 aus 49“ eine Fünf mit Zusatzzahl erhält?

Beispiel 1.2 (Fußball). Wie groß war die Wahrscheinlichkeit des Ereignisses, dass beim Bundesligaspiel Fürth gegen Regensburg die Kleeblätter und beim Spiel Ingolstadt gegen Nürnberg der Club gewinnt?

0.6

Antworten:

(i) Lotto: Wenn man nur einen Tipp abgibt, $p = 0.0001107$.

(ii) Bundesliga, bwin-Quoten für Freitag, 13.4.2018, für das Spiel Fürth gegen Regensburg:

	1	0	2
Quoten	2,35	3,20	3,00
W'en	0.40	0.29	0.31

(iii) Bundesliga, bwin-Quoten für Sonntag, 15.4.2018, für das Spiel Ingolstadt gegen Nürnberg:

	1	0	2
Quoten	1,95	3,40	2,75
W'en	0.48	0.27	0.25

Wobei als Gewinn für bwin ca. sieben Prozent des Einsatzes angesetzt wurden. (Stand der Quoten: 11.04.2018, ca 10:30 Uhr)

0.7

Die zentralen Begriffe:

Ereignis:

Eine Situation, von der man nicht sicher vorhersagen kann, ob sie eintreten wird oder nicht.

Wahrscheinlichkeit:

Eine zahlenmäßige Bewertung der Chance, dass das Ereignis eintreten wird.

Frage:

Wie berechnet man Wahrscheinlichkeiten und was hat man davon?

Historischer Rückblick

Dazu werfen wir einen kurzen Blick in die Anfänge der Wahrscheinlichkeitsrechnung.

0.8

1.1 Historisches

Einige Beteiligte

- Chevalier de Méré (1607–1684)
- Blaise Pascal (1623–1662)
- Paul Fermat (1601–1665).

Geschichte

Im Jahr 1654 schrieb de Méré einen Brief an Pascal mit Fragen zu einigen stochastischen Problemen, speziell dem folgenden:

Wenn man beim Würfeln darauf wettet, dass bei 4 Würfeln wenigstens einmal die Augenzahl Sechs erscheint, macht man auf lange Sicht Gewinn ...

Das war unter Zockern damals schon allgemein bekannt und deshalb das Spiel uninteressant geworden. Also suchte de Méré eine Variante mit mindestens den gleichen Chancen, die andere nicht sofort durchschauten.

0.9

Sein Ansatz zur Chancenberechnung von de Meré:

$$\frac{\text{Mögliche Augenzahlen bei einem Wurf}}{\text{Mögliche Würfe}} = \frac{6}{4} = \frac{3}{2}$$

Seine Folgerung:

Zu wetten, dass man bei 24 Würfeln eines Würfelpaars wenigstens einen Sechserpasch bekommt, hat die gleiche Chance, denn

$$\frac{36 \text{ Möglichkeiten beim Werfen von 2 Würfeln}}{24 \text{ Würfe}} = \frac{36}{24} = \frac{3}{2}$$

Seine leidvolle Erfahrung:

Damit macht man auf lange Sicht Verlust!

Seine Schlussfolgerung:

Mathematik bringt nichts fürs tägliche Leben.

0.10

1.2 Gesetz der großen Zahlen

Frage

Was heißt: Auf lange Sicht?

Zufallsexperiment

Das Experiment „Viermaliges Werfen eines Würfels“ kann man beliebig oft wiederholen, aber bei keiner Durchführung lässt sich der Ausgang vorhersagen.

Ein derartiges Experiment nennen wir ein *Zufallsexperiment*.

$E_{6,4}$ bezeichne das Ergebnis, dass bei einer Durchführung dieses Zufallsexperiments wenigstens einmal die Augenzahl Sechs erscheint. Das Zufallsexperiment werde N -mal durchgeführt.

Häufigkeiten

$A_N(E_{6,4})$ sei die (absolute) Häufigkeit des Eintretens von $E_{6,4}$ bei diesen N Durchführungen.

$h_N(E_{6,4}) = \frac{A_N(E_{6,4})}{N}$ nennen wir die *relative Häufigkeit* des Eintretens von $E_{6,4}$ bei diesen N Durchführungen.

0.11

Die Erfahrungen des Herrn de Méré lassen sich dann so formulieren:

Nur im Limes gleich

Bei einer *großen* Anzahl N von Durchführungen dieses Experiments gilt stets

$$h_N(E_{6,4}) = \frac{A_N(E_{6,4})}{N} > \frac{1}{2}$$

und bei der Variante mit den 24 Würfeln ergibt sich

$$h_N(E_{36,24}) = \frac{A_N(E_{36,24})}{N} < \frac{1}{2}.$$

0.12

Allgemeine Erfahrung:

Bei einem Zufallsexperiment wird beobachtet, ob ein Ereignis E eintritt.

Führt man das Zufallsexperiment N -mal durch, wobei N groß ist, so stabilisiert sich die relative Häufigkeit $h_N(E)$ immer bei einer für das Ereignis E charakteristischen Zahl $P(E)$.

$$h_N(E) \rightsquigarrow P(E)$$

Z.B. $h_N(E_{6,4}) \approx 0.52$.

Diese Erscheinung nennt man *das empirische Gesetz der großen Zahlen*.

0.13

Die Säulen, in die eine Kugel fallen kann, seien mit $s = 1, 2, \dots, 10$ nummeriert.

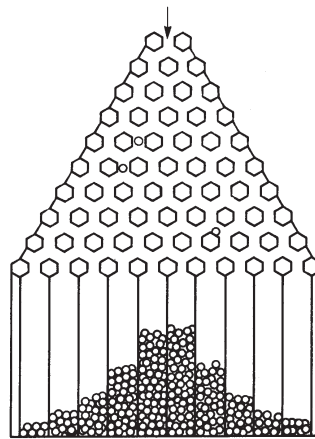
E_s sei das Ereignis „Kugel fällt in Säule s “.

N : Anzahl der Kugeln.

Es gilt

$$h_N(E_s) = \frac{A_N(E_s)}{N} \approx p_s$$

mit stets den gleichen Zahlen p_s .



0.14

Arbeitshypothese:

In jedem Zufallsexperiment steckt ein Naturgesetz in Form einer Funktion P , die jedem dabei beobachtbaren Ereignis E eine Zahl $P(E)$ zuordnet, so dass

$$h_N(E) \rightsquigarrow P(E) \text{ für } N \longrightarrow \infty.$$

Bezeichnungen

$P(E)$ nennen wir die *Wahrscheinlichkeit des Ereignisses E* und diese Art von Konvergenz *das empirische Gesetz der großen Zahlen*.

Frage?

Wie berechnet man $P(E)$?

0.15

1. Ansatz

Approximation durch $h_N(E)$ aus genügend langen Versuchsreihen.

Das ist aber nur selten praktisch durchführbar.

Beispiel 1.3. Fünf mit Zusatzzahl beim Lotto.

2. Ansatz

Wahrscheinlichkeitsrechnung: Berechnung der Wahrscheinlichkeit von *komplexen* Ereignissen unter Kenntnis der Wahrscheinlichkeiten *elementarer* Ereignisse.

0.16

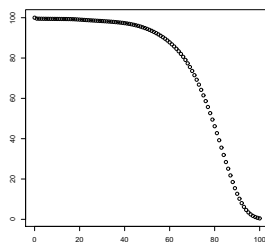
1.2.1 Relative Häufigkeiten

Beispiel 1.4. Sterbetafel

E_n sei das Ereignis, dass ein männlicher Deutscher seinen n -ten Geburtstag erlebt, d.h. mindestens n Jahre alt wird.

Das statistische Bundesamt berechnet die relativen Häufigkeiten $h_N(E_n)$ aus einer großen Anzahl von Personen:

Die Sterbetafel.



Genauer: Es werden die Anzahlen der Überlebenden

$$U_n = 100000 h_N(E_n)$$

aus einer fiktiven Population von 100000 Personen angegeben.

0.17

Beispiel 1.5. Die Sterbetafel ist die Kalkulationsgrundlage für Lebensversicherungen.

Jahresprämie X für eine Auszahlungssumme S auf Endalter 65 für einen 30-jährigen.

Sehr vereinfacht: Ohne Verzinsung, Provisionen und Verwaltungskosten.

Summe der Einnahmen: $E = X(U_{30} + U_{31} + \dots + U_{64})$

Summe der Ausgaben: $A = S \cdot U_{30}$

Einnahmen gleich Ausgaben $E = A$ ergibt

$$X = \frac{U_{30}}{U_{30} + U_{31} + \dots + U_{64}} S$$

0.18

Beispiel 1.6 (Fußball). Wie groß ist die Wahrscheinlichkeit, dass beim nächsten Heimspiel der Club gewinnt?

Problem

Ein Bundesligaspiel ist kein Zufallsexperiment im vorher definierten Sinn, denn es lässt sich nicht unter gleichen Voraussetzungen wiederholen. Man kann daher keine *objektiven* Wahrscheinlichkeiten aus relativen Häufigkeiten ermitteln.

„Ausweg“

Subjektive Wahrscheinlichkeiten, die ein „Fussballexperte“ den Ereignissen E_1 (Sieg), E_0 (Unentschieden) und E_2 (Niederlage) zuordnet.

Berechnung

Eine Umfrage unter N gleichermaßen ausgewiesenen Experten könnte die relativen Häufigkeiten $h_N(E_i)$ als Schätzwerte für die subjektiven Wahrscheinlichkeiten $\hat{p}_i = P(E_i)$ ergeben.

0.19

Quoten

bwin-Quoten für das Spiel Stuttgart gegen Nürnberg im Jahr 2010:

1	0	2
3.30	3.30	2.10

Quote: Auszahlungsbetrag bei Einsatz von 1 Euro (*in Kontinentaleuropa*)

Frage

Wie hängen die Quoten q_i mit den (subjektiven) Wahrscheinlichkeiten p_i zusammen, die bwin den verschiedenen Ausgängen des Spiels zuordnet?

Quoten

„faire Quote“: $Q_i = \frac{1}{p_i}$, (theoretischer Auszahlungsbetrag)

„faktische Quote“: $q_i = fQ_i$, ($(1 - f)Q_i$ wird als Provision abgezogen, hier $f = 0.924$) Die Quoten von oben ergeben $p_1 = 0.28$, $p_0 = 0.28$, $p_2 = 0.44$.

0.20

Das Wettbüro

Wie muss bwin die Quoten bzw. die Wahrscheinlichkeiten ansetzen, um sichere Einnahmen zu erzielen?

0.21

Überlegungen

Von N Fussballexperten, die je einen Euro setzen, tippen

$$A_N(E_i) = Nh_N(E_i) = N\hat{p}_i$$

auf Ausgang i .

Einnahmen: Von N insgesamt als Einsatz eingezahlten Euro entfallen $N\hat{p}_1$ auf Tipp 1, $N\hat{p}_0$ auf Tipp 0 und $N\hat{p}_2$ auf Tipp 2.

Auszahlung A_i , falls $i = 1, 0, 2$ der tatsächliche Spielausgang ist:

$$A_i = N\hat{p}_i q_i = N \frac{\hat{p}_i}{p_i} f.$$

Falls b_{win} die Quoten so setzt, dass $p_i = \hat{p}_i$ für $i = 1, 0, 2$, ist der gesamte Auszahlungsbetrag unabhängig vom Spielausgang $A_i = fN$ Euro, d.h. $(1 - f)N$ Euro bleiben in der Kasse zurück.

Die Fussballexperten sind die Wetter selbst und die Umfrage die Verteilung der Wetter auf die 3 möglichen Tipps.

0.22

1.2.2 Wahrscheinlichkeitsrechnung

Beispiel 1.7 (Ein Beispiel für die zweite Methode). Berechnung der Zuverlässigkeit von komplexen technischen Systemen, z.B. des mechatronischen Systems eines Flugzeugs.

Verfügbarkeit:

Wahrscheinlichkeit $P(S)$ des Ereignisses S , dass das System über einen bestimmten Zeitraum hinweg nicht ausfällt.

Fragestellung:

Wie muss man das System auslegen, damit $P(S) \geq 1 - \varepsilon$ für ein vorgegebenes kleines ε ?

Die Berechnung von $P(S)$, insbesondere für verschiedene Varianten von S , kann man nicht über relative Häufigkeiten durchführen (z. B. 1000 Flugzeuge mit je 500 Flugstunden).

0.23

Gegeben

Von den Herstellern bekannt sind die Verfügbarkeiten $P(S_k)$ der Einzelkomponenten.

Frage?

Wie berechnet man $P(S)$ aus den $P(S_k)$ bzw. wie kann man die Verfügbarkeit des Gesamtsystems durch Parallelschaltung mehrerer Exemplare einer Komponente über die geforderte Schranke heben?

0.24

Generelles Thema:

Berechnung von Wahrscheinlichkeiten komplexer Ereignisse bei Kenntnis der Wahrscheinlichkeiten von *Elementarereignissen*.

Zutaten

Dazu wird benötigt:

- Ein allgemeines mathematisches Modell eines Zufallsexperiments,
- Rechenregeln für Ereignisse und Wahrscheinlichkeiten und
- theoretische Ansätze für die Wahrscheinlichkeiten von Elementarereignissen in typischen Anwendungssituationen.

0.25

Tipps

- Die mathematischen Hilfsmittel, die für die Wahrscheinlichkeitsrechnung gebraucht werden, sind ziemlich *elementar*.
- Das eigentliche Problem ist der Jargon, der in der Stochastik üblich ist, d.h. wie übersetze ich eine Stochastik-Textaufgabe in eine Mathematik-Aufgabe.

0.26

Kapitel 2

Beschreibende Statistik

Beispiel 2.1. Messung des Gewichts von zehnjährigen Kindern

Folgende Werte wurden gemessen:

34.6	34.5	28.6	37.8	24.8	24.9	27.7	34.0	24.9	30.4
32.0	24.9	35.6	38.1	32.4	33.1	41.5	29.6	37.2	26.6
33.0	36.8	34.1	29.9	29.3	27.6	29.9	35.0	24.1	36.0

Was kann man mit diesen Zahlen anfangen?

- Sortieren
- Gruppieren ($< 30, \geq 30$)
- In Klassen zusammenfassen $[23, 25]$ bis $(41, 43]$.

Definition 2.2. Die aus einer Beobachtung oder Messung aufgezeichneten Daten heißen **Datensatz**. Es wird auch von der **Urliste** gesprochen. Ist der Datensatz aufgrund einer gezielten Teilerhebung entstanden, heißt er **Stichprobe**.

2.27

Definition 2.3. Bezeichnungen

- Der Datensatz wird angegeben als $\mathbf{x} = (x_1, \dots, x_n)$.
- Es werden drei Arten von Daten unterschieden:
 - **Nominal:** qualitative Ausprägung ohne „natürliche“ Ordnung
 - **Ordinal:** qualitative Ausprägung mit „natürliche“ Ordnung
 - **Metrisch/rational:** Ausprägung die durch eine Zahl beschrieben werden kann (besitzt einen Nullpunkt und eine Dimension)
- Die Komprimierung und/oder tabellarische bzw. graphische Darstellung eines Datensatzes nennt man eine **Statistik**.

- Der geordnete Datensatz wird mit $\mathbf{x}_{\square} = (x_{[1]}, \dots, x_{[n]})$ bezeichnet und heißt auch **Ordnungsstatistik**. Die aufsteigend geordnete Auflistung der Beobachtungswerte wird auch **Rangwertreihe** genannt.

$$x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}$$

- Der erste Rangwert $x_{[1]}$ heißt **Minimum**, der letzte Rangwert $x_{[n]}$ heißt **Maximum** des Datensatzes

2.28

Beispiel 2.4. Aus den Messwerten von *Beispiel 2.1* lässt sich folgende Ordnungsstatistik erstellen:

24.1 24.8 24.9 24.9 24.9 26.6 27.6 27.7 28.6 29.3
29.6 29.9 29.9 30.4 32.0 32.4 33.0 33.1 34.0 34.1
34.5 34.6 35.0 35.6 36.0 36.8 37.2 37.8 38.1 41.5

Häufigkeitstabelle

y_i	24	26	28	30	32	34	36	38	40	42
n_i	5	1	3	5	2	6	4	3	0	1

2.29

Definition 2.5 (Rang). $x_{[1]} \leq \dots \leq x_{[n]}$ bezeichne die Rangwertreihe eines Datensatzes mit Werten in \mathbb{R} . Der Rang $R(x_{[j]})$ ist folgendermaßen definiert:

- (i) Tritt ein Beobachtungswert $x_{[j]}$ nur einmal in einem Datensatz auf, so ist die Position j der Rang von $x_{[j]}$

$$R(x_{[j]}) = j.$$

- (ii) Tritt ein Beobachtungswert $x_{[j]}$ mehrfach in einem Datensatz auf, d.h.

$$x_{[r-1]} < x_{[r]} = \dots = x_{[r+s-1]} < x_{[r+s]}$$

so ist der Rang von $x_{[j]}$ definiert als

$$R(x_{[j]}) = \frac{r + (r+1) + \dots + (r+s-1)}{s} = r + \frac{s-1}{2} \quad (2.1)$$

Das mehrfache Auftreten eines Wertes in der Urliste heißt **Bindung**.

2.30

Beispiel 2.6. Der Datensatz aus *Beispiel 2.1* besitzt folgende Rangliste:

Rangliste

24.1	24.8	24.9	26.6	27.6	27.7	28.6	29.3	29.6
1	2	4	6	7	8	9	10	11
29.9	30.4	32.0	32.4	33.0	33.1	34.0	34.1	34.5
12.5	14	15	16	17	18	19	20	21
34.6	35.0	35.6	36.0	36.8	37.2	37.8	38.1	41.5
22	23	24	25	26	27	28	29	30

2.31

Definition 2.7 (kumulierte Häufigkeit). Die Summen von Häufigkeiten einzelner Ausprägungen werden als **kumulierte Häufigkeit** bezeichnet.

Beispiel 2.8. Häufigkeitstabelle für die relativen und kumulierten Häufigkeiten

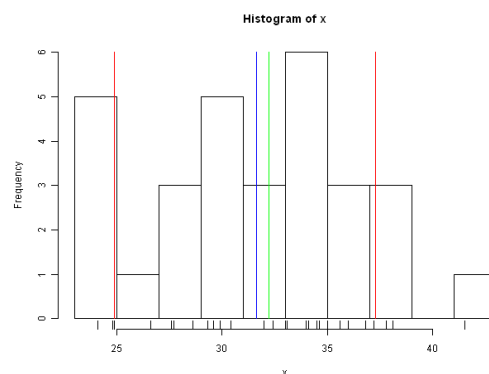
y_i	24	26	28	30	32	34	36	38	40	42
h_i	$\frac{1}{6}$	$\frac{1}{30}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{1}{15}$	$\frac{1}{5}$	$\frac{2}{15}$	$\frac{1}{10}$	0	$\frac{1}{30}$
F_i	$\frac{5}{30}$	$\frac{6}{30}$	$\frac{9}{30}$	$\frac{14}{30}$	$\frac{16}{30}$	$\frac{22}{30}$	$\frac{26}{30}$	$\frac{29}{30}$	$\frac{29}{30}$	$\frac{30}{30}$

Die Auflistung der relativen Häufigkeiten wird auch **Häufigkeitsverteilung** genannt.

2.32

Histogramm

Die Häufigkeitsverteilungen statistischer Daten lassen sich in Form von **Histogrammen** darstellen. Dabei werden die Merkmalsausprägungen auf der Abszisse und die Häufigkeiten auf der Ordinate aufgetragen. Die Häufigkeit jedes Datenwertes wird dann beschrieben durch eine balkenförmige Fläche über dem Messwert:



2.33

Notation

Die Werte der Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ werden zu l Klassen K_i ($i = 1, \dots, l$) zusammengefasst. Dazu werden die Punkte $a = z_0 < z_1 < \dots < z_{l-1} < z_l = b$ aus einem Intervall $[a, b)$ gewählt und es gilt:

$$K_1 = [z_0, z_1] \quad 1 - \text{te Klasse} \quad (2.2)$$

$$K_i = (z_{i-1}, z_i] \quad i - \text{te Klasse für } i = 2, \dots, l \quad (2.3)$$

$$m_i = \frac{z_{i-1} + z_i}{2} \quad i - \text{te Klassenmitte} \quad (2.4)$$

$$h_i = \frac{|\{j | x_j \in K_i\}|}{n} \quad i - \text{te relative Klassenhäufigkeit} \quad (2.5)$$

Werte in einer Klasse K_i werden mit der entsprechenden Klassenmitte m_i identifiziert.

2.34

Spannweite

Der Abstand zwischen dem $x_{[1]}$ und $x_{[n]}$ heißt **Spannweite**

$$v = x_{[n]} - x_{[1]}$$

2.35

Definition 2.9 (Empirische Verteilungsfunktion). In einem aus n Beobachtungen bestehenden Datensatz seien m verschiedenen Merkmalsausprägungen y_1, \dots, y_m aufgetreten. Dazu gehöre die Rangwertreihe $y_{[1]}, \dots, y_{[m]}$ mit den relativen Häufigkeiten h_j ($j = 1, \dots, m$). Die abschnittsweise definierte Funktion

$$F_n = \begin{cases} 0, & x < y_{[1]}, \\ \sum_{j=1}^k h_j & y_{[k]} \leq x < y_{[k+1]}, k \in \{1, \dots, m-1\}, \\ 1, & x \geq y_{[m]} \end{cases} \quad (2.6)$$

heißt **empirische Verteilungsfunktion**.

2.36

Definition 2.10. Es sei ein Datensatz $\mathbf{x} = (x_1, \dots, x_n)$ gegeben. Dann lassen sich die folgenden (ungewichteten) Mittelwerte definieren:

- Das **arithmetische Mittel** \bar{x} des Datensatzes \mathbf{x} (oft auch nur als **Mittelwert**, **Mittel** oder **Durchschnitt** bezeichnet) wird definiert durch

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i.$$

- Der **Median** \tilde{x} von \mathbf{x} ist definiert als

$$\tilde{x} := \begin{cases} x_{[\frac{n+1}{2}]}, & \text{falls } n \text{ ungerade,} \\ \frac{1}{2} \left(x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]} \right), & \text{falls } n \text{ gerade.} \end{cases}$$

- Das **geometrische Mittel** $G(\mathbf{x})$ von \mathbf{x} ist definiert als

$$G(\mathbf{x}) := \sqrt[n]{\prod_{i=1}^n x_i}.$$

- Das **harmonische Mittel** $H(\mathbf{x})$ von \mathbf{x} ist definiert als

$$H(\mathbf{x}) := \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}.$$

2.37

Definition 2.11. Versieht man die Daten x_1, \dots, x_n mit den Gewichten $w_1, \dots, w_n > 0$, so lassen sich die folgenden **gewichteten Mittelwerte** definieren:

- Das **gewichtete (arithmetische) Mittel** $\bar{x}_{(w_1, \dots, w_n)}$ von \mathbf{x} ist definiert als

$$\bar{x}_{(w_1, \dots, w_n)} := \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}.$$

Das geometrische Mittel erhält die ver

- Das **gewichtete geometrische Mittel** $G_{(w_1, \dots, w_n)}(\mathbf{x})$ von \mathbf{x} ist definiert als

$$G_{(w_1, \dots, w_n)}(\mathbf{x}) := \sqrt[w]{\prod_{i=1}^n x_i^{w_i}}, \quad \text{wobei } w := \sum_{i=1}^n w_i.$$

- Das **gewichtete harmonische Mittel** $H_{(w_1, \dots, w_n)}(\mathbf{x})$ von \mathbf{x} ist definiert als

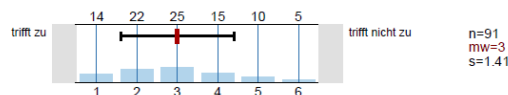
$$H_{(w_1, \dots, w_n)}(\mathbf{x}) := \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n \frac{w_i}{x_i}}.$$

mittelt das Wachstum

Man sieht: Für $w_1 = \dots = w_n = 1$ ergibt sich das jeweilige ungewichtete Mittel.

2.38

Beispiel 2.12. Berechnung des gewichteten Mittelwertes in der Evaluierung der Technischen Fakultät.



2.39

Definition 2.13. Es sei $\mathbf{x} = (x_1, \dots, x_n)$ eine Stichprobe und $\mathbf{x}_{[]} = (x_{[1]}, \dots, x_{[n]})$ die zugehörige geordnete Stichprobe. Dann ist für $0 < p < 100$ das **$p\%$ -Quantil** u_p von \mathbf{x} wie folgt definiert:

$$u_{p\%} := \begin{cases} \frac{1}{2} (x_{[n \frac{p}{100}]} + x_{[n \frac{p}{100} + 1]}), & \text{falls } n \frac{p}{100} \text{ ganzzahlig,} \\ x_{[\lceil n \frac{p}{100} \rceil]}, & \text{sonst.} \end{cases}$$

2.40

Bemerkung

- Häufig werden die Paare $(u_{p\%}, u_{(100-p)\%})$ benutzt.
- Der **Median** ist das 50%-Quantil.
- Die 25%- und 75%-Quantile heißen **Quartile** und werden auch **unteres** und **oberes Quartil** genannt.

- Der Abstand zwischen $u_{25\%}$ und $u_{75\%}$ wird als **Interquartilsabstand**

$$IQR := u_{75\%} - u_{25\%} \quad (2.7)$$

bezeichnet.

- Quantile werden auch als Perzentile bezeichnet.

2.41

Beispiel 2.14. Berechnet man für das obige Beispiel die Werte aus Definition 2.10 sowie das untere und das obere Quartil, so ergibt sich:

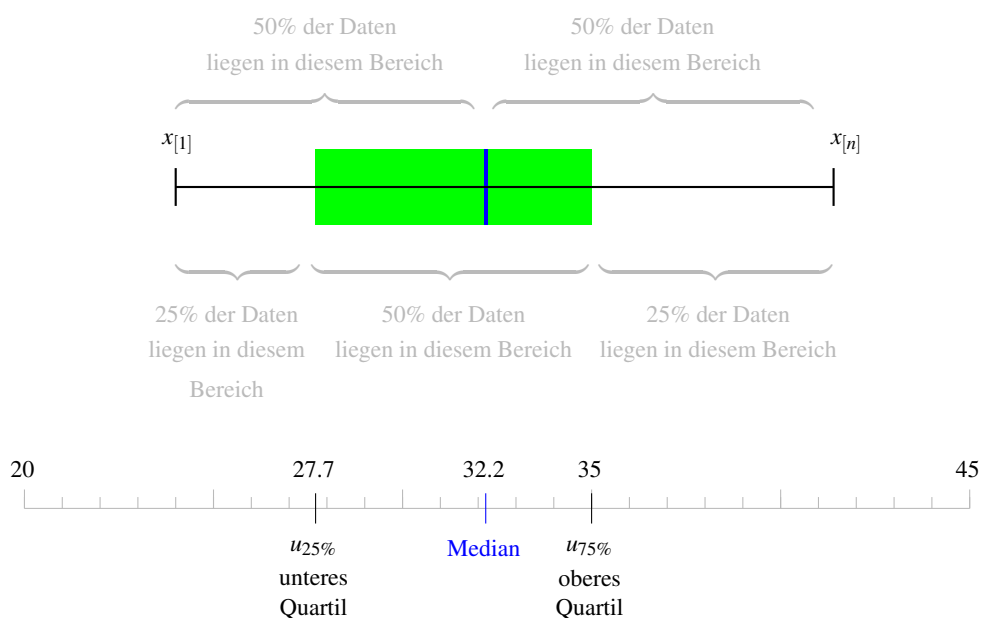
Mittelwert	Median	geometrisches Mittel	harmonisches Mittel	25%-Quantil	75%-Quantil
$\bar{x} = 31,63$	$\tilde{x} = 32,2$	$G(\mathbf{x}) = 31,29$	$H(\mathbf{x}) = 30,95$	$u_{25\%} = 27.7$	$u_{75\%} = 35$

Anhand der errechneten Werte erkennt man die allgemein gültige Ungleichung $H(\mathbf{x}) \leq G(\mathbf{x}) \leq \bar{x}$.

2.42

Boxplot

Die Häufigkeitsverteilung eines Datensatzes lässt sich auch als sogenannter **einfacher Boxplot** darstellen. Betrachten wir noch einmal Beispiel 2.1 (bzw. 2.4), so sieht der zugehörige Boxplot wie folgt aus:



Boxplot zu Beispiel 1.1 (bzw. 1.3)

also median und 25% und 75%-Quantil auzeichnen,box um den Interquar

2.43

Definition 2.15. Die **Streuung** $\sigma_{\mathbf{x}}^n$ eines Datensatzes $\mathbf{x} = (x_1, \dots, x_n)$ ist die quadratische gemittelte Abweichung der x_i von \bar{x} , also

$$\sigma_{\mathbf{x}}^n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Statt Streuung wird auch die Bezeichnung **Standardabweichung** verwendet. Der Ausdruck $\sigma_{\mathbf{x}}^{n^2}$ heißt **Varianz**.

Bemerkung

Die Streuung kann auch durch

$$\sigma_{\mathbf{x}}^n = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}$$

Es müssen nur n , $\sum x_i$ und $\sum x_i^2$ gespeichert werden.

2.44

Beispiel 2.16. In obigen Beispiel ergibt sich $\sum x_i = 948,9$ und $\sum x_i^2 = 30651,01$.

Bemerkung

Bei Teilstichproben interessiert die Schätzung für einen unbekannten Datensatz und es wird die sogenannte **empirische Streuung**

$$s_{\mathbf{x}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

verwendet. Analog zu Definition 2.15 nennt man $s_{\mathbf{x}}^2$ die **empirische Varianz**.

2.45

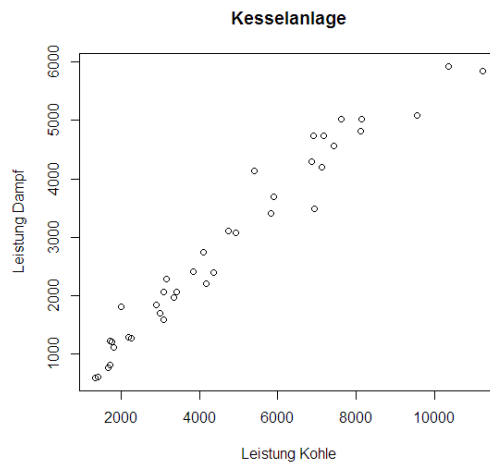
Datenpaare

Oft treten Datenpaare (x_i, y_i) auf. In unserem Beispiel wäre dies (Gewicht, Größe). Es liegt also ein zweidimensionaler Datensatz vor:

$$\mathbf{z} = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)).$$

Beispiel 2.17. Als Beispiel dient eine kohlebefeuerte Kesselanlage, von der in zeitgleichen Meßintervallen die Durchschnittsleistungen (kW) für Kohle und den erzeugten Dampf ermittelt worden sind.¹

¹Der Dozent dankt Herr Dr. Neidhart Kamprath, für die freundliche Unterstützung und diesen Datensatz.



2.46

Quantile-Quantile-Plot

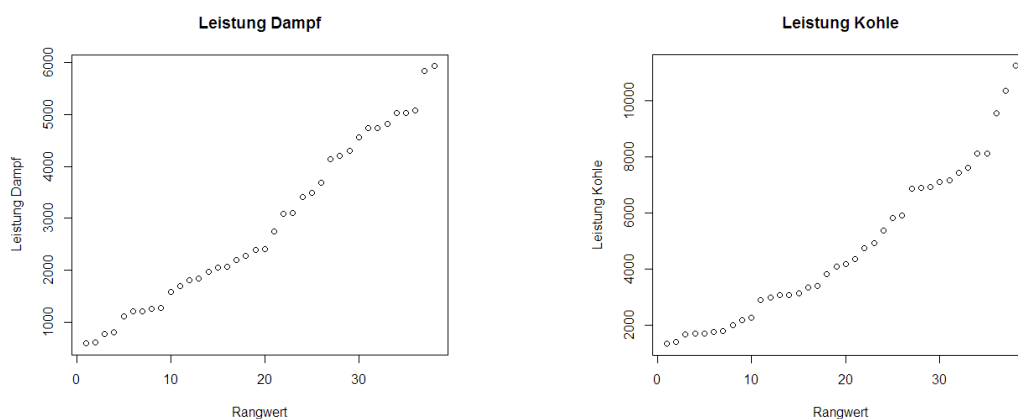
Der **Q-Q-Plot** ist exploratives, grafisches Werkzeug, bei dem die Quantile zweier statistischer Variablen gegeneinander abgetragen werden

Später wird auch der Probability-Probability-Plot betrachtet, bei dem Verteilungsfunktionen zweier statistischer Variablen abgetragen werden.

Vorgehen: Betrachtet werden zwei Stichproben x_{\square} und y_{\square} (am besten gleiche Stichprobengröße)

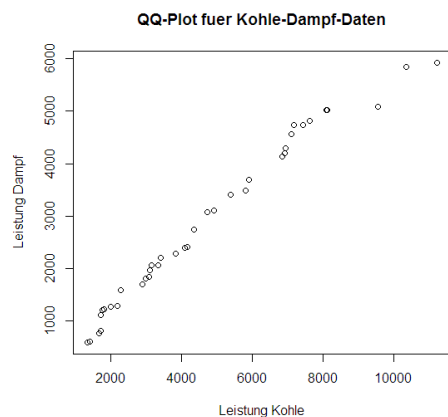
Ziel: Vergleich der Verteilungen zweier statistischer Verteilungen, wobei die Verteilungen nicht bekannt sein müssen.

Fortsetzung Beispiel 2.17



2.47

Fortsetzung Beispiel 2.17



2.48

Zur Beschreibung des Zusammenhangs der beiden obigen Merkmale definieren wir die folgende Kenngröße:

Definition 2.18. Gegeben sind die Datensätze $\mathbf{x} = (x_1, \dots, x_n)$ und $\mathbf{y} = (y_1, \dots, y_n)$. Die **empirische Kovarianz** von \mathbf{x} und \mathbf{y} ist definiert als:

$$s_{xy} := \frac{1}{n-1} \sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y})). \quad (2.8)$$

Dabei ist s_{xx} einfach die **empirische Varianz** s_x^2 .

man zieht eins ab, wegen Bessler-korrektur

Bemerkung

Die Kovarianz lässt sich auch mit der Formel

$$s_{xy} := \frac{1}{n-1} \left(\sum_{i=1}^n (x_i y_i) - n \bar{x} \bar{y} \right).$$

berechnen.

effektiv bedingte Wahrscheinlichkeit: $P(A)P(B) = P(B|A)$ Wenn A,B iid sind. Der Grad

2.49

Definition 2.19. Gegeben sind die Datensätze $\mathbf{x} = (x_1, \dots, x_n)$ und $\mathbf{y} = (y_1, \dots, y_n)$.

- Der **empirische Korrelationskoeffizient** von \mathbf{x} und \mathbf{y} ist definiert als:

$$r_{xy} := \frac{s_{xy}}{s_x s_y}. \quad (2.9)$$

- Zu den Daten sei ein Regressionsmodell r gegeben der Datensatz $\hat{\mathbf{y}}$ auf folgende Weise definiert:

$$\hat{y}_i = r(x_i).$$

Dann ist das **Bestimmtheitsmaß** B definiert durch

$$B := \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

2.50

Satz 2.20. Mit den Bezeichnungen aus der letzten Definition gilt:

$$B = r_{\hat{y}\hat{y}}^2 \quad (2.10)$$

2.51

also kovar zwischen der funktion und dem echten Wert

Satz 2.21. Für einen Datensatz der Form

$$z = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$$

mit $s_x \neq 0$ minimiert die **Regressionsgerade** $y = ax + b$ mit

$$a = \frac{s_{xy}}{s_x^2} \quad \text{und} \quad b = \bar{y} - a\bar{x}$$

die Funktion

$$\min \text{MSE}(\mathbf{x}, \mathbf{y})$$

$$r(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n (y_i - (ax_i + b))^2.$$

2.52

2.1 Lineare Ausgleichsrechnung / Methode der kleinsten Quadrate

Die Fragestellung

Zu gegebenen Datenpaaren (x_j, y_j) , $(j = 1, \dots, n)$ und einem Satz von Funktionen f_i , $(i = 1, \dots, m)$ sollen Parameter gefunden werden, so dass die Gleichungen

$$y_j \stackrel{!}{=} p_1 \cdot f_1(x_j) + p_2 \cdot f_2(x_j) + \dots + p_m \cdot f_m(x_j), \quad j = 1, \dots, n, \quad (2.11)$$

„möglichst gut“ erfüllt sind.

Was dabei möglichst gut meint, muss ebenfalls definiert werden.

Des Residuum

Da die Gleichungen im Allgemeinen nicht erfüllt sind, wird das **Residuum** definiert

$$r_j := y_j - (p_1 \cdot f_1(x_j) + p_2 \cdot f_2(x_j) + \dots + p_m \cdot f_m(x_j)),$$

der Vektor der Residuen $\mathbf{r} \in \mathbb{R}^n$.

also der Fehler zwischen should und does. Der e

Hier sind zwei Dinge zu beobachten:

- Die Residuen hängen von den Parametern p_1, \dots, p_m ab.
- Ein gute Wahl der p_i ist gefunden, wenn $\|\mathbf{r}\|_2$ minimal wird.

2.53

Minimieren des Residuums

Die Bestimmung von p kann also als ein Optimierungsproblem aufgefasst werden:

$$\min_{\mathbf{p} \in \mathbb{R}^m} \frac{1}{2} \|\mathbf{r}\|_2^2 = \frac{1}{2} \sum_{i=1}^n \left(y_i - (p_1 f_1(x_i) + p_2 f_2(x_i) + \dots + p_m f_m(x_i)) \right)^2$$

2.54

Andere Darstellung

Es fällt auf, dass die Werte $f_j(x_i)$ fest sind und mit den p_j linear kombiniert werden. Die Gleichungen

$$\begin{aligned} f_1(x_1)p_1 + \dots + f_m(x_1)p_m &= \hat{y}_1 \\ f_1(x_2)p_1 + \dots + f_m(x_2)p_m &= \hat{y}_2 \\ &\vdots \\ f_1(x_n)p_1 + \dots + f_m(x_n)p_m &= \hat{y}_n \end{aligned}$$

als Matrix-Vektor-Multiplikation der Form

Wobei \mathbf{p} der Vektor aus gewählten und A die gewählten Werte

$$A \cdot \mathbf{p} = \hat{\mathbf{y}}$$

geschrieben werden kann. So erhalten wir die äquivalente Formulierung

$$\min_{\mathbf{p} \in \mathbb{R}^m} \frac{1}{2} \|\mathbf{r}\|_2^2 = \frac{1}{2} \|A\mathbf{p} - \mathbf{y}\|_2^2.$$

2.55

Der Ausdruck $\frac{1}{2} \|A\mathbf{p} - \mathbf{y}\|_2^2$ wird nun im Folgenden betrachtet. Da dieser minimal werden soll, müssen wir den Gradienten dieser Zielfunktion bilden und Null setzen. Fassen wir zunächst die Norm als Skalarprodukt auf:

quadrat kann als skalarprodukt gebaut werden

$$z(\mathbf{p}) := \frac{1}{2} \|A\mathbf{p} - \mathbf{y}\|_2^2 = \frac{1}{2} \langle A\mathbf{p} - \mathbf{y}, A\mathbf{p} - \mathbf{y} \rangle \quad (2.12)$$

$$= \frac{1}{2} (A\mathbf{p} - \mathbf{y})^\top (A\mathbf{p} - \mathbf{y}) \quad (2.13)$$

$$= \frac{1}{2} (\mathbf{p}^\top A^\top A\mathbf{p} - 2\mathbf{p}^\top A^\top \mathbf{y} + \mathbf{y}^\top \mathbf{y}) \quad (2.14)$$

$$A^\top A \mathbf{p} = A^\top \mathbf{y}$$

Damit ergibt sich als notwendige Bedingung 1. Ordnung

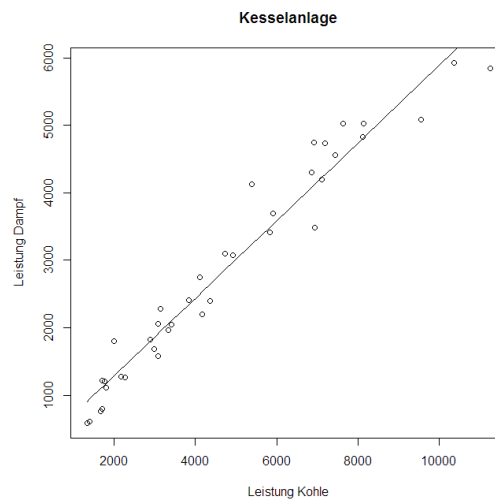
$$\nabla_{\mathbf{p}} z(\mathbf{p}) = A^\top A\mathbf{p} - A^\top \mathbf{y} \stackrel{!}{=} 0.$$

Umstellen: $A^\top(A\mathbf{p} - \mathbf{y}) = 0$ ist null, gdw $(A\mathbf{p} - \mathbf{y}) = 0$

Die Lösung dieser Normalengleichung ist die Lösung des linearen Ausgleichsproblems.

Achtung: Nicht im Kurzschrift

Beispiel 2.22. Wir betrachten nun die Regressionsgerade für das Beispiel Kohle- und Dampfleistung.



$$s_{xy} = 4231700, \quad b = 143,18, \quad a = 0.574$$

2.56

Kapitel 3

Wahrscheinlichkeitsmodelle

3.1 Die Bausteine

Beispiel 3.1. Von einem Terminal soll ein Auftrag an den Zentralrechner abgeschickt werden. Wie lang ist die vorraussichtliche Antwortzeit, ohne Kenntnis über die Auslastung des Servers?

Definition 3.2. Ein **Zufallsexperiment** ist ein (geplanter, gesteuerter oder nur beobachteter) Vorgang, der ein genau abzugrenzendes Ergebnis besitzt, das vom Zufall beeinflusst ist oder beeinflusst sein kann.

Bemerkung

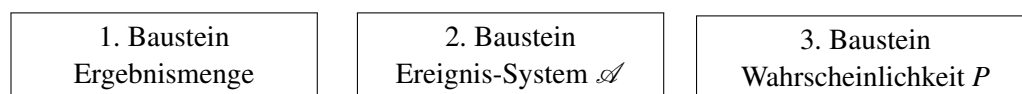
Ein Zufallsexperiment muss reproduzierbar sein, deshalb müssen die Bedingungen genau festgehalten werden.

3.57

Modellierung

- (i) **Aspekt** Die möglichen Ereignisse.
- (ii) **Aspekt** Die möglichen Fragestellungen.
- (iii) **Aspekt** Die zugehörigen Wahrscheinlichkeiten.

Bausteine



3.58

3.2 Die Ergebnismenge Ω

Beispiel 3.3. Antwortzeit

Eine Möglichkeit ist bspw., die Wartezeit zu messen bis eine Antwort eingeht. Die max. Wartezeit könnte 200s betragen. Die Ergebnismenge wäre dann $(0, 200)$.

Definition 3.4. Die **Ergebnismenge** Ω (auch Merkmalsraum, Stichprobenraum, Grundgesamtheit) ist eine nicht-leere Menge mit Elementen $\omega \in \Omega$. Ω gibt die möglichen Ausgänge (**Ergebnisse**) des Zufallsexperiments an.

Beispiel 3.5. Die Ergebnismenge in *Beispiel 3.3* ist $\Omega = (0, 200)$.

Beispiel 3.6. Interessiert einen, ob nach dem Wählen einer Nummer **besetzt** oder **nicht besetzt** ist, so ist $\Omega = \{0, 1\}$.

3.59

Beispiel 3.7. Betrachtet man die Anzahl der Anrufe bei einer bestimmten Telefonnummer zwischen 8 und 9 Uhr, dann könnte $\Omega = \mathbb{N}_0$ oder $\Omega = 0, 1, 2, \dots, 100$ gewählt werden.

Bemerkung

Bei der Wahl von Ω ist stets darauf zu achten, dass alle typischen Fälle erfasst werden.

3.60

3.3 Zusammengesetzte Ereignisse

Beobachtung

Häufig werden mehrere Beobachtungen kombiniert. Z.B. beim Arzt

- Gewicht
- Puls
- Blutdruck
- ...

Wir sprechen dann von zusammengesetzten Merkmalen.

Beispiel 3.8. Drei Funktionen eines Bauteils werden geprüft und als „defekt“ oder „intakt“ bewertet. Ergebnismenge: $\Omega = \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{0, 1\}, i = 1, 2, 3\}$. Die kürzere Alternative wäre: $\Omega = \{0, 1\}^3$.

3.61

Definition 3.9. Das **kartesische Produkt** $\Omega_1 \times \dots \times \Omega_n$ der Mengen $\Omega_1, \dots, \Omega_n$ ist die Menge

$$\Omega = \{(\omega_1, \dots, \omega_n) \text{ mit } \omega_i \in \Omega_i \text{ für alle } i = 1, \dots, n\}.$$

Statt $\Omega_1 \times \dots \times \Omega_n$ wird oft auch $\bigtimes_{i=1}^n \Omega_i$, bei $\Omega_1 = \dots = \Omega_n$ auch Ω_1^n verwendet.

Bemerkung

Das kartesische Produkt findet Anwendung bei der Beschreibung von Systemen, die über die verschiedenen Zeitpunkte $1, \dots, n$ beobachtet werden.

3.62

Beispiel 3.10. Für die Vorhersage von Kursen, kann der Kursverlauf durch

$$\Omega = \prod_{i=1}^{\infty} \mathbb{N}_0$$

modelliert werden.

$\omega \in \Omega$ hat die Form $\omega = (\omega_1, \omega_2, \dots)$ und beschreibt den (ganzzahligen Kurs) am i -ten Tag. Dabei ist jedes ω ein möglicher Kursverlauf.

3.63

3.4 Ereignisse, Verknüpfung von Ereignissen

Beispiel 3.11 (Antwortzeit an einem Rechner-Terminal). Es wird eher nicht gefragt werden, ob die Zeit *genau* 10 Sekunden beträgt, sondern ob die Zeit z.B. *höchstens* 10 Sekunden beträgt.

Definition 3.12. In der Modellebene ist ein **Ereignis** A eine Teilmenge von Ω ($A \subset \Omega$).

Man sagt: „Das Ereignis A tritt ein“, falls ein Ergebnis ω mit $\omega \in A$ beobachtet wird.

Die Menge aller betrachteten Ereignisse heißt **Ereignissystem** \mathcal{A} .

Erinnerung

ω – Ergebnis, Merkmal Ω – Ergebnismenge, Merkmalmenge A – Ereignis \mathcal{A} – Ereignissystem

3.64

Besondere Ereignisse

$A = \emptyset$ \emptyset ist ein unmögliches Ereignis, weil $\omega \in \emptyset$ nie eintritt.

$A = \Omega$ Das Ereignis Ω tritt immer ein.

$A = \{\omega\}$ für $\omega \in \Omega$ $\{\omega\}$ nennt man das **Elementarereignis**.

3.65

Definition 3.13. Die Gesamtheit aller Teilmengen heißt **Potenzmenge von Ω** und wird mit $\mathcal{P}(\Omega)$ bezeichnet.

Bemerkung

Eine n -elementige Menge besitzt genau 2^n Teilmengen.

Beispiel 3.14 (Aktienkurse).

$A =$ „Der Kurs lag gestern über 500“.

$B =$ „Der Kurs ist von gestern auf heute gestiegen“.

$C = A$ und B .

Das Ereignis C tritt ein, wenn A und B eintreten, d.h. $\omega \in A \cap B$.

Ereignis – Mengen

„A oder B oder beide treten ein.“	entspricht	$\omega \in A \cup B$
„A und B treten ein.“	entspricht	$\omega \in A \cap B =: AB$
„A und B treten nie gleichzeitig ein.“	entspricht	$A \cap B = \emptyset$
„A tritt nicht ein.“	entspricht	$\omega \in A^c \Leftrightarrow \omega \notin A$
„A tritt ein, aber B tritt nicht ein.“	entspricht	$\omega \in A \setminus B = A \cap B^c = AB^c$
„Mindestens ein A_i tritt ein.“	entspricht	$\omega \in \bigcup_{i=1}^{\infty} A_i$
„Alle A_i treten ein.“	entspricht	$\omega \in \bigcap_{i=1}^{\infty} A_i$

Vereinbarung

Die Schreibweise $A + B$ bzw. $\sum_{i=1}^{\infty} A_i$ steht für die disjunkte Vereinigung der Mengen A, B bzw. der Mengen A_i . Diese Notation beinhaltet, dass die Mengen A, B disjunkt sind bzw. die A_i alle paarweise disjunkt sind.

De Morgansche Regeln

$$\overline{(\overline{A})} = A$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Daher gilt:

$$\begin{aligned} A, B \in \mathcal{A} &\Rightarrow \overline{A}, \overline{B} \in \mathcal{A} \\ &\Rightarrow \overline{A \cap B} = \overline{A} \cup \overline{B} \in \mathcal{A} \\ &\Rightarrow A \cap B = \overline{(\overline{A \cap B})} \in \mathcal{A} \end{aligned}$$

Definition 3.15. Die **Indikatorfunktion** einer Menge $A \subset \Omega$ ist eine Abbildung $1_A : \Omega \rightarrow \{0, 1\}$ mit

$$1_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A, \\ 0 & \text{falls } \omega \notin A. \end{cases}$$

Die Zuordnung einer Menge A zu ihrer Indikatorfunktion 1_A ist umkehrbar eindeutig.

3.5 Ereignissystem \mathcal{A} **Frage**

Welche Eigenschaften sollte ein System von Ereignissen besitzen?

Definition 3.16. Ein System \mathcal{A} von Teilmengen einer Menge Ω heißt **abgeschlossenes Mengensystem** oder **(Mengen)- σ -Algebra** über Ω , wenn gilt

- (i) $\Omega \in \mathcal{A}$
- (ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- (iii) $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Annahme

Im Folgenden wird immer vorausgesetzt, dass jedes Ereignissystem \mathcal{A} abgeschlossen ist.

Bemerkung

In der Vorlesung werden nur zwei Fälle betrachtet:

- (i) Ω ist endlich oder abzählbar.
- (ii) $\Omega = \mathbb{R}$ oder $\Omega = \mathbb{R}^n$.

3.69

Definition 3.17. Ist \mathcal{E} ein System von Teilmengen von Ω , dann heißt das kleinste abgeschlossene Mengensystem \mathcal{A} (über Ω), das \mathcal{E} enthält $\mathcal{E} \subset \mathcal{A}$, **das von \mathcal{E} (über Ω) erzeugte abgeschlossene Mengensystem – bzw. die von \mathcal{E} erzeugte σ -Algebra** – und wird mit $\mathcal{A}(\mathcal{E})$ bezeichnet. \mathcal{E} heißt der Erzeuger von $\mathcal{A}(\mathcal{E})$.

Folgerung 3.18. Ist Ω abzählbar, dann erzeugt das System \mathcal{E} aller Elementarereignisse $\{\omega\}$ bereits die Potenzmenge $\mathcal{P}(\Omega)$.

3.70

Definition 3.19. Es sei $\mathcal{G}_1 := \{(a, b] : a, b \in \mathbb{R}, a \leq b\}$ die Menge aller halb-offenen Intervalle in \mathbb{R} . Dann heißt die von \mathcal{G}_1 über \mathbb{R} erzeugte σ -Algebra $\mathbb{B} := \mathcal{A}_{\mathbb{R}}(\mathcal{G}_1)$ die **Borel- σ -Algebra über \mathbb{R}** . Die einzelnen Ereignisse aus \mathbb{B} heißen **Borel-Mengen**.

Bemerkung

- (i) Im Fall $a = b$ hat man $(a, a] = \emptyset$. Damit gehören leere Durchschnitte zu \mathcal{G}_1 .
- (ii) Analog zu oben lassen sich die Mengen der Intervalle $[a, b]$, (a, b) und $[a, b)$ definieren. Auch die Mengen dieser Intervalle erzeugen die Borel- σ -Algebra über \mathbb{R} .

3.71

Definition 3.20. Für $\mathbf{a} = (a_1, a_2, \dots, a_n)$ und $\mathbf{b} = (b_1, \dots, b_n)$ in \mathbb{R}^n besteht das **n -dimensionale Intervall $(\mathbf{a}, \mathbf{b}]$** aus allen $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ mit $a_i < x_i \leq b_i$. Das Intervall $(\mathbf{a}, \mathbf{b}]$ heißt **halb-offen**. Die Menge aller halb-offenen Intervalle in \mathbb{R}^n sei $\mathcal{G}_n := \{(\mathbf{a}, \mathbf{b}], \mathbf{a}, \mathbf{b} \in \mathbb{R}^n, \mathbf{a} \leq \mathbf{b}\}$.

Definition 3.21. Die von der Menge \mathcal{G}_n der halb-offenen Intervalle in \mathbb{R}^n erzeugte σ -Algebra $\mathbb{B}^n := \mathcal{A}_{\mathbb{R}^n}(\mathcal{G}_n)$ heißt **Borel- σ -Algebra über \mathbb{R}^n** . Die Elemente von \mathbb{B}^n heißen **Borel-Mengen in \mathbb{R}^n** .

Bemerkung

Die vorherige Bemerkung gilt natürlich auch für den Fall \mathbb{R}^n .

3.72

3.6 Darstellung von Ereignissen durch Zufallsvariablen

Beschreibung von Ereignissen

- A : „Der Kurs lag gestern über 500.“
- A_2 : „Genau zwei der drei Funktionen des Bauteils sind intakt.“

Die verbale Umschreibung von Ereignissen ist auf Dauer unpraktisch.

Kürzere Schreibweise durch Variablen

- X_0 := „Gestriger Aktienkurs“
- Y := „Anzahl der intakten Funktionen“

führt auf

- $X_0 > 500$,
- $Y = 2$.

3.73

Idee

Eine Zufallsvariable ordnet jedem Versuchsausgang einen Wert zu.

Beispiel 3.22 (Funktionen eines Bauteils). Die Zufallsvariable Y bildet den vollständigen Versuchsausgang auf die Teilinformation „Anzahl der Erfolge“ ab:

$$\omega = (1, 0, 1), \quad \omega' = \omega_1 + \omega_2 + \omega_3 = 2$$

$$Y : \Omega \rightarrow \Omega'.$$

3.74

Definition 3.23. Ist X eine Abbildung $\Omega \rightarrow \Omega'$ und $A' \subset \Omega'$, dann wird definiert:

$$\{X \in A'\} := \{\omega \in \Omega : X(\omega) \in A'\}.$$

Eine Teilmenge $A \subset \Omega$ der Form $A := \{X \in A'\}$ heißt **durch X beschreibbar**.

Bemerkung

Analoge Darstellungen wie $\{X \leq b\}$, $\{X = c\}$ sind eingeschlossen.

3.75

Frage

Was passiert, wenn $\mathcal{A} \neq \mathcal{P}(\Omega)$? Ist dann jede Menge $\{X \in A'\}$ ein Ereignis in Ω , wenn $A' \subset \Omega'$?

Definition 3.24. (i) Eine **Zufallsvariable** (kurz ZV) oder **Zufallsgröße** ist eine Abbildung von einer Ergebnismenge Ω mit Ereignissystem \mathcal{A} in eine Bildmenge Ω' mit Ereignissystem \mathcal{A}' .

Für eine Zufallsvariable $X : \Omega \rightarrow \Omega'$ – genauer $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ – wird gefordert :

$$\{X \in A'\} \in \mathcal{A} \text{ für alle } A' \in \mathcal{A}'.$$

(ii) Für eine ZV $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ sei $\mathcal{A}(X)$ das System aller durch X und $A' \in \mathcal{A}'$ beschreibbaren Ereignisse, also die Menge

$$\mathcal{A}(X) := \{ \{X \in A'\} : A' \in \mathcal{A}' \}.$$

3.76

Notiz

$\mathcal{A}(X)$ ist in \mathcal{A} enthalten und abgeschlossen.

3.77

Bemerkung

- (i) ZV werden meist mit X, Y, Z, U, V, W , Ereignisse mit A, B, C, \dots bezeichnet.
- (ii) ZV werden benutzt, um eine in $\omega \in \Omega$ enthaltene Information auszuwählen oder zu komprimieren. (Ggf. durch Koordinaten- oder Parametertransformation).
- (iii) Gelegentlich wird die Identität als ZV benutzt, damit die Bezeichnung einfacher und einheitlicher wird. (Bsp. Antwortzeiten: $Z(\omega) = \omega$, $\{Z \leq 10\}$).
- (iv) Mittels Definition von $\mathcal{A}(X)$ wird die Abgeschlossenheit geprüft:

$$\Omega = \{X \in \Omega'\}, \{X \in A'\}^c = \{X \in A'^c\}, \left\{X \in \bigcup A'_i\right\} = \bigcup \{X \in A'_i\}.$$

3.78

Beispiel

[Laufzeit von Programmen] Um die Laufzeit abzuschätzen, werden zwei nacheinander zu lesende Spuren eines Laufwerks beobachtet. Beschreibt $\omega = (\omega_1, \omega_2)$ die Nummer der beiden Spuren, dann beschreibt die Zufallsvariable $Z(\omega) := |\omega_1 - \omega_2|$ den Abstand der Spuren.

Beispiel

[Aktienkurse]

$X_0 :=$ „Gestriger Aktienkurs“ .

$X_1 :=$ „Heutiger Aktienkurs“ .

$A :=$ „Der Kurs lag gestern über 500“ .

$B :=$ „Der Kurs ist von gestern auf heute gestiegen“ .

$$A = \{X_0 > 500\}$$

$$B = \{X_0 < X_1\}$$

3.79

3.7 Relative Häufigkeit und Wahrscheinlichkeit

3. Baustein

Wie wird ein Ereignis A bewertet?

1. Anhaltspunkt

Aus bekannten Ereignissen wird die relative Häufigkeit bestimmt. Diese wird als Anhaltspunkt für die Wahrscheinlichkeit genommen.

Schreibweise

Wahrscheinlichkeit von A : $P(A)$.

3.80

Empirisches Gesetz der großen Zahlen

Wird ein Zufallsexperiment n -mal mit den Beobachtungswerten x_1, \dots, x_n unter gleichen Bedingungen wiederholt, dann „konvergieren“ die relativen Häufigkeiten

$$h_n(A) := \frac{1}{n} \cdot (\text{Anzahl der } x_i \text{ mit } x_i \in A)$$

für $n \rightarrow \infty$ gegen einen Grenzwert.

Annahme

Jedes Ereignis $A \in \mathcal{A}$ eines genau definierten Zufallsexperiments besitzt eine Wahrscheinlichkeit. Zunächst betrachten wir nur Modelle, bei denen die Wahrscheinlichkeiten gegeben sind.

3.81

Beispiel 3.25. Bei Fußballspielen werden die Seiten mit Hilfe einer vom Schiedsrichter geworfenen Münze ausgelost. Es wird erwartet, dass „Kopf“ und „Zahl“ gleich häufig erscheint.

Beispiel 3.26 (Unfallversicherung). Für Prognosezwecke hat ein Versicherungsunternehmen ein Wahrscheinlichkeitsmodell entwickelt. In diesem Modell sei für jede Person aus einer Gruppe von 10000 Versicherten die Wahrscheinlichkeit für einen Unfall 0,5%. Die relative Häufigkeit ist damit 0,005, die tatsächliche Zahl der Unfälle wird bei 50 liegen.

Additivität

Für zwei disjunkte Ereignisse gilt:

$$h_n(A + B) = h_n(A) + h_n(B) .$$

Es gilt

$$0 \leq h_n(A) \leq 1, \quad h_n(\emptyset) = 0, \quad h_n(\Omega) = 1$$

Wahrscheinlichkeiten

(1)	$P(A) \geq 0$	Nichtnegativität
(1')	$P(A) \leq 1$	
(2)	$P(\Omega) = 1$	Normiertheit
(2')	$P(\emptyset) = 0$	Nulltreue
(3)	$P(A_1 + A_2) = P(A_1) + P(A_2)$	Additivität
(3 _n)	$P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n)$	endliche Additivität
(3')	$P(A_1 + A_2 + \dots) = P(A_1) + P(A_2) + \dots$	σ – Additivität

Definition 3.27. Eine Abbildung $P : \mathcal{A} \rightarrow \mathbb{R}$, wobei \mathcal{A} eine σ -Algebra über Ω ist, heißt **Wahrscheinlichkeitsmaß** (W-Maß) auf \mathcal{A} , wenn folgende Bedingungen erfüllt sind:

- (i) $P(A) \geq 0, \forall A \in \mathcal{A}$, (Nichtnegativität)
- (ii) $P(\Omega) = 1$, (Normiertheit)
- (iii) $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ (σ -Additivität).

Wahrscheinlichkeitsraum

Die drei Bausteine (Ω, \mathcal{A}, P) , nennt man Wahrscheinlichkeitsraum oder Wahrscheinlichkeitsmodell.

Schreibweise

Für $A := \{X \in A'\}$ mit $P(A) = P(\{X \in A'\})$ wird künftig $P(X \in A')$ geschrieben.

Definition 3.28. Ein Zufallsexperiment mit zwei möglichen Ausgängen heißt **Bernoulli-Experiment**. Als Ergebnismenge wird $\Omega = \{0, 1\}$

als W-Modell (Ω, \mathcal{A}, P) mit $\mathcal{A} = \mathcal{P}(\Omega)$

und als W-Maß $P(1) = p$ und $P(0) = 1 - p$ ($0 \leq p \leq 1$) festgelegt.

(Ω, \mathcal{A}, P) heißt **Bernoulli-Modell**, das W-Maß P **Bernoulli-Verteilung mit Parameter p** oder **Bernoulli(p)-Verteilung**, kurz $B(p)$.

Definition 3.29. Ein Zufallsexperiment mit endlich vielen und gleichwertigen Ausgängen heißt **Laplace-Experiment**. Als Ergebnismenge wird $\Omega = \{1, 2, \dots, N\}$ gewählt. Das W-Maß P auf $\mathcal{A} = \mathcal{P}(\Omega)$ ergibt sich aus

$$P(1) = P(2) = \dots = P(N) = \frac{1}{N}.$$

Das W-Maß P heißt auch **Laplace-Verteilung** oder **diskrete Gleichverteilung (über Ω)**, kurz $L(\Omega)$.

Bemerkung

Die Wahrscheinlichkeit $P(A)$ eines beliebigen Ereignisses A berechnet sich aus

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.$$

Einschub: Urnenmodelle

Aus einer Urne mit n durchnummerierten Kugeln werden k Kugeln entnommen. Dabei lassen sich verschiedenen Betrachtungsweisen unterscheiden:

ohne Wiederholungen	$\hat{=}$ ohne Zurücklegen
mit Wiederholungen	$\hat{=}$ mit Zurücklegen
ohne Anordnung	$\hat{=}$ ohne Beachtung der Reihenfolge
mit Anordnung	$\hat{=}$ mit Beachtung der Reihenfolge

Satz 3.30. Die Anzahl der möglichen k -Kombinationen von n Objekten (paarweise verschieden) ergibt sich aus folgender Tabelle:

Kombinationen	ohne Wiederholung ($k \leq n$)	mit Wiederholung
mit Berücksichtigung der Reihenfolge	$V_{n,k} = \frac{n!}{(n-k)!}$	$\bar{V}_{n,k} = n^k$
ohne Berücksichtigung der Reihenfolge	$C_{n,k} = \frac{n!}{k!(n-k)!}$	$\bar{K}_{n,k} = \binom{n+k-1}{k}$

Definition 3.31. Es sei Ω eine Ergebnismenge, \mathcal{A} ein Ereignissystem über Ω und $a \in \Omega$ ein (festes) ausgewähltes Ergebnis. Dann heißt das W-Maß P , definiert durch

$$P(A) = \begin{cases} 1 & \text{für } a \in A \\ 0 & \text{sonst} \end{cases},$$

die **Einpunktverteilung** im Punkt a , kurz $P = \varepsilon_a$.

3.8 Weitere Eigenschaften von Wahrscheinlichkeiten

Rechenregeln

(1)	$P(A) \geq 0$	Nichtnegativität
(1')	$P(A) \leq 1$	
(2)	$P(\Omega) = 1$	Normiertheit
(2')	$P(\emptyset) = 0$	Nulltreue
(3)	$P(A_1 + A_2) = P(A_1) + P(A_2)$	Additivität
(3 _n)	$P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n)$	endliche Additivität
(3')	$P(A_1 + A_2 + \dots) = P(A_1) + P(A_2) + \dots$	σ – Additivität

3.89

Weitere Rechenregeln

(4)	$P(A^c) = 1 - P(A)$	
(5)	$P(A \setminus B) = P(A) - P(AB)$	
(6)	$P(A \cup B) = P(A) + P(B) - P(AB)$	
(7)	$P(A \cup B) \leq P(A) + P(B)$	Subadditivität
(8)	$A \subset B \Rightarrow P(A) \leq P(B)$	Monotonie

Weitere Rechenregeln

(9)	$A_1 \subset A_2 \subset \dots \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$	Stetigkeit von unten
(10)	$A_1 \supset A_2 \supset \dots \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$	Stetigkeit von oben

3.90

Definition 3.32. Ein **Maß auf \mathcal{A} (bzw. über (Ω, \mathcal{A}))** ist eine Abbildung $\mu : \mathcal{A} \rightarrow \mathbb{R} \cup \{\infty\}$ mit

- (1) $\mu(A) \geq 0$,
- (2') $\mu(\emptyset) = 0$,
- (3') $\mu(A_1 + A_2 + \dots) = \mu(A_1) + \mu(A_2) + \dots$

Bemerkung

Ein Maß besitzt die Eigenschaften (3), (3_n), (5)-(10).

3.91

3.9 Elementare bedingte Wahrscheinlichkeiten

Definition 3.33. Es seien A, B Ereignisse in Ω und $P(B) > 0$. Dann heißt

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

die **bedingte Wahrscheinlichkeit** von A unter der Bedingung B . Es gilt

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

Verkettungsregel

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) .$$

Formel der totalen Wahrscheinlichkeit

Es sei $(B_i, i \in I)$ eine abzählbare Zerlegung von Ω und seien $P(B_i)$ und $P(A|B_i)$ für alle $i \in I$ bekannt, dann gilt:

$$P(A) = \sum_{i \in I} P(AB_i) = \sum_{i \in I} P(B_i)P(A|B_i) .$$

3.92

Formel von Bayes

Sei $(B_i, i \in I)$ eine abzählbare Zerlegung von Ω , dann gilt

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i \in I} P(B_i)P(A|B_i)} .$$

Stochastisch Unabhängigkeit

Oft tritt der Fall ein, dass ein Ereignis A **nicht** vom Eintreten eines anderen Ereignisses B abhängt. Dann gilt:

$$P(A|B) = P(A).$$

Dieser Fall heißt **stochastisch unabhängig**. Es gilt

$$P(AB) = P(A)P(B).$$

Definition 3.34. Zwei Ereignisse A und B heißen **stochastisch unabhängig** bzgl. P , wenn gilt:

$$P(AB) = P(A)P(B) .$$

3.93

Definition 3.35. Die Ereignisse A_1, A_2, \dots, A_n in einem W-Raum (Ω, \mathcal{A}, P) heißen **stochastisch unabhängig** (bzgl. P), wenn für alle Teilmengen $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$ von diesen Ereignissen die „Produktformel“ gilt:

$$P(A_{i_1}, A_{i_2}, \dots, A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}) .$$

3.94

Kapitel 4

Wahrscheinlichkeitsmaße

4.1 Diskrete Wahrscheinlichkeitsmaße und Zähldichten

Satz 4.1. Ω sei eine abzählbare Ergebnismenge und $\mathcal{A} = \mathcal{P}(\Omega)$.

(i) Ist P ein W-Maß über (Ω, \mathcal{A}) und $f(\omega) = P(\{\omega\})$ für $\omega \in \Omega$, dann gilt:

$$f(\omega) \geq 0 \quad (\omega \in \Omega), \quad \sum_{\omega \in \Omega} f(\omega) = 1 \quad (4.1)$$

$$P(A) = \sum_{\omega \in A} f(\omega), \quad (A \in \mathcal{A}). \quad (4.2)$$

(ii) Jede Abbildung $f: \Omega \rightarrow \mathbb{R}$ mit den Eigenschaften (4.1) und (4.2) definiert ein W-Maß P auf \mathcal{A} mit der Eigenschaft

$$P(\{\omega\}) = f(\omega) \quad \forall \omega \in \Omega.$$

Die Abbildung f heißt **Zähldichte** oder **Z-Dichte** von P . Die Zuordnung zwischen P und f ist eineindeutig.

3.95

Bemerkung

Bekannt ist die allgemeine binomische Formel

$$(p+q)^n = \sum_{k=1}^n \binom{n}{k} p^k q^{n-k}, \quad p, q \in \mathbb{R}, n = 1, 2, \dots \quad (4.3)$$

Definition 4.2 (Binomial-Verteilung). Gegeben sind $p, q \geq 0$ aus \mathbb{R} mit $p+q=1$ und $\Omega = \{0, 1, \dots, n\}$. Die Dichte

$$f(k) = b(n, p; k) = \binom{n}{k} p^k q^{n-k}$$

über Ω heißt **Binomial-Z-Dichte**.

Das zugehörige W-Maß heißt **Binomial-Verteilung** und wird mit $B(n, p)$ bezeichnet.

Bemerkung

Die geometrische Reihe lautet

$$1 + q + q^2 + \dots = \sum_{i=0}^{\infty} q^i = \frac{1}{1-q}.$$

Sie konvergiert für $|q| < 1$.

Definition 4.3 (Geometrische Verteilung). Gegeben seien $0 < q < 1$ aus \mathbb{R} , $\Omega = \{0, 1, \dots\}$. Die Dichte

$$f(k) = (1-q)q^k$$

über Ω heißt **geometrische Z-Dichte**.

Das zugehörige W-Maß heißt **geometrische Verteilung**.

Definition 4.4 (Poisson-Verteilung). Es sei $\lambda \in (0, \infty)$ und $\Omega = \{0, 1, \dots\}$. Dann heißt die Dichte

$$f(k) = \pi(\lambda; k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

über Ω die **Poisson-Z-Dichte**.

Das zugehörige W-Maß heißt **Poisson-Verteilung** und wird mit $\pi(\lambda)$ bezeichnet.

Bemerkung

Die numerische Berechnung der $B(n, p)$ -Verteilung ist für große Werte von n sehr aufwändig. Sind zudem die Werte von p sehr klein, so bietet sich die sogenannte **Poisson-Approximation** an, um die $B(n, p)$ -Verteilung zu nähern:

Satz 4.5 (Poisson-Approximation der Binomial-Verteilung). Für große n und kleine p_n nähert sich die $B(n, p_n)$ -Verteilung der $\pi(n \cdot p_n)$ -Verteilung an.

D.h.: Für

$$p_n \rightarrow 0 \text{ und } n \cdot p_n \rightarrow \lambda \text{ für } n \rightarrow \infty$$

konvergieren die Werte der Binomial-Z-Dichte

$$b(n, p; k) \text{ für alle } k = 0, 1, 2, \dots \text{ gegen die entsprechenden Werte } \pi(\lambda; k)$$

der Poisson-Z-Dichte,

$$b(n, p; k) \xrightarrow{n \rightarrow \infty} \pi(\lambda; k).$$

Definition 4.6 (Empirische Verteilung). Für einen Datensatz $x = (x_1, \dots, x_n)$ mit Werten in $\Omega \subset \mathbb{R}$ heißt die relative Häufigkeit

$$A \mapsto h_n(A) := \frac{1}{n} \cdot (\text{Anzahl der } x_i \text{ mit } x_i \in A)$$

empirische Verteilung von x .

Sie besitzt die Zähldichte

$$\hat{f}_n^X = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i\}}(x), x \in \Omega.$$

Definition 4.7. Es sei $T \subset \Omega$ eine abzählbare Teilmenge von Ω und $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung mit $f(\omega) \geq 0$ und $\sum_{\omega \in T} f(\omega) = 1$, sowie $f(\omega) = 0$ für $\omega \in T^c$.

Dann heißt f eine **Zähldichte über Ω mit Träger T** .

$$P(A) := \sum_{\omega \in A} f(\omega)$$

erzeugt auf einer beliebigen σ -Algebra \mathcal{A} über Ω ein W-Maß. Dieses wird **diskretes W-Maß** genannt.

4.2 Stetige Wahrscheinlichkeitsmaße und Riemann-Dichten

Definition 4.8. Eine Riemann-integrierbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) \geq 0 \quad (x \in \mathbb{R}) \quad \text{und} \quad \int_{-\infty}^{\infty} f(x) dx = 1 \quad (4.4)$$

heißt **Riemann-Dichte über \mathbb{R} oder auch stetige Dichte**.

Jede R-Dichte definiert eindeutig ein W-Maß p über (\mathbb{R}, \mathbb{B}) mit der Eigenschaft

$$P((a, b]) = P([a, b]) = \int_a^b f(x) dx \quad (4.5)$$

mit $(a \leq b)$ und $P(\{\omega\}) = 0$.

Satz 4.9 (Fortsetzungssatz). *Ist P auf einem geeigneten Erzeuger \mathcal{E} von \mathcal{A} festgelegt und auf \mathcal{E} nicht-negativ, σ -additiv und normiert, dann gibt es eine eindeutige Fortsetzung von P auf \mathcal{A} .*

Borel-Mengen

Für $\Omega = \mathbb{R}$ ist \mathcal{G}_1 ein im Sinne dieses Satzes geeigneter Erzeuger von \mathbb{B} . Die Normiertheit von P wird nachgewiesen durch $P((-n, n]) \rightarrow 1$ für $n \rightarrow \infty$.

Definition 4.10 (Rechteck-Verteilung). Ist

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in (a, b), \\ 0 & \text{sonst,} \end{cases}$$

dann ist f eine Riemann-Dichte über $\Omega = \mathbb{R}$. Das zugehörige W-Maß heißt **stetige Gleichverteilung** oder **Rechteck (a, b) -Verteilung**, kurz $\mathcal{R}(a, b)$.

Definition 4.11 (Exponential-Verteilung $\text{Exp}(\alpha)$). Ist $\alpha > 0$, so ist

$$f(x) = \alpha e^{-\alpha x} 1_{(0, \infty)}(x) = \begin{cases} \alpha e^{-\alpha x} & x > 0, \\ 0 & x \leq 0. \end{cases} \quad (4.6)$$

eine Riemann-Dichte. Das zugehörige W-Maß nennt man **Exponential-Verteilung** und wird mit $\text{Exp}(\alpha)$ bezeichnet.

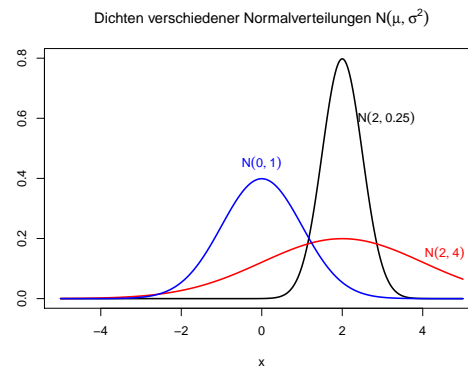
Definition 4.12 (Normalverteilung $\mathcal{N}(a, \sigma^2)$). Für jeden Wert $a \in \mathbb{R}$ und $\sigma > 0$ ist

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \quad (4.7)$$

eine Riemann-Dichte. Das zugehörige W-Maß heißt **Normalverteilung** mit „Mittelwert“ a und der „Streuung“ σ ; kurz $\mathcal{N}(a, \sigma^2)$.

Standardnormalverteilung $\mathcal{N}(0, 1)$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}. \quad (4.8)$$



Definition 4.13 (Gamma (α, ν) - Verteilung). Ein W-Maß $\Gamma_{\alpha, \nu}$ mit $\alpha, \nu > 0$ und der Riemann-Dichte

$$\gamma(x) := \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (4.9)$$

heißt **Gamma** (α, ν) -Verteilung.

Eigenschaften der Gamma-Funktion

Die **Gamma-Funktion** ist definiert als

$$\Gamma(\nu) := \int_0^\infty x^{\nu-1} e^{-x} dx, \quad \nu > 0 \quad (4.10)$$

und besitzt die folgenden Eigenschaften:

- (1) $\Gamma(1) = 1$,
- (2) $\Gamma(\nu + 1) = \nu \cdot \Gamma(\nu)$,

(3) $v \mapsto \log \Gamma(v)$ ist eine konvexe Funktion, d.h. Γ ist logarithmisch konvex.

Ein nützlicher Wert im Zusammenhang mit Eigenschaft (2) ist:

$$(4) \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Falls v eine natürliche Zahl ist, so lässt sich die Gamma-Funktion berechnen durch:

$$(5) \quad \Gamma(v) = (v-1)!, \quad v \in \mathbb{N}.$$

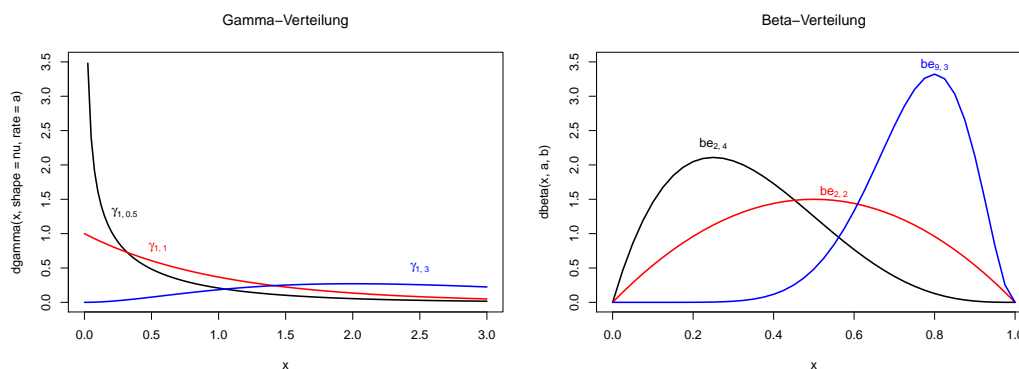
3.107

Definition 4.14 (Beta(μ, v)-Verteilung). Ein W-Maß $Be(\mu, v)$ mit $\mu, v > 0$ und der Riemann-Dichte

$$be_{\mu,v}(x) := \begin{cases} \frac{\Gamma(\mu+v)}{\Gamma(\mu)\Gamma(v)} x^{\mu-1} (1-x)^{v-1} & 0 < x < 1 \\ 0 & x \leq 0, \end{cases} \quad (4.11)$$

heißt **Beta(μ, v)-Verteilung**.

3.108



Definition 4.15. Eine n -dimensionale Riemann-integrierbare Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $f(x) \geq 0$ für $x \in \mathbb{R}^n$ und $\int_{\mathbb{R}^n} f(x) dx = 1$ heißt **Riemann-Dichte in \mathbb{R}^n** .

Anmerkung

Auch hier ergibt sich eindeutig ein W-Maß P über $(\mathbb{R}^n, \mathbb{B}^n)$, das für geeignete Ereignisse A als Riemann-Gebietsintegral ausgewertet werden kann.

3.109

4.3 Verteilungsfunktion

Definition 4.16 (Verteilungsfunktion). Ist P ein beliebiges W-Maß über (\mathbb{R}, \mathbb{B}) , dann heißt die Abbildung $F : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$F(x) := P((-\infty, x]), \quad x \in \mathbb{R} \quad (4.12)$$

die **Verteilungsfunktion von P** . Es gilt

$$P((a, b]) = F(b) - F(a), \quad a, b \in \mathbb{R}, \quad a \leq b. \quad (4.13)$$

Berechnung von Verteilungsfunktionen

Für ein W-Maß über \mathbb{R} mit der Riemann-Dichte f gilt mit dieser Definition

$$F(x) = \int_{-\infty}^x f(t) dt \quad \text{und} \quad P((a, b]) = \int_a^b f(t) dt = F(b) - F(a). \quad (4.14)$$

3.110**Verteilungsfunktion der Rechtecksverteilung $\mathcal{R}(a, b)$**

$$F(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a} & a < x \leq b, \\ 1 & x > b. \end{cases}$$

Verteilungsfunktion von $\text{Exp}(\alpha)$

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\alpha x} & x > 0. \end{cases}$$

3.111**Verteilungsfunktion Φ von $\mathcal{N}(0, 1)$**

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}.$$

Verteilungsfunktion Φ von $\mathcal{N}(0, 1)$ falls $T \sim \mathcal{N}(a, \sigma^2)$

$$F_{a, \sigma^2} = \int_{-\infty}^x \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}} = \Phi\left(\frac{x-a}{\sigma}\right), \quad x \in \mathbb{R}.$$

mit der Substitution $u = \frac{t-a}{\sigma}$ und $du = \frac{1}{\sigma} dt$.

3.112**Empirische Verteilungsfunktion**

Zur empirischen Verteilung eines Datensatzes $\mathbf{x} = (x_1, \dots, x_n)$ gehört die **empirische Verteilungsfunktion**

$$\hat{F}_n^{\mathbf{x}}(x) := \frac{1}{n} \sum_{i=1}^n 1_{[x_i, \infty)}(x), \quad x \in \mathbb{R}. \quad (4.15)$$

3.113

Definition 4.17 (Gemischte Verteilungen). Hat das W-Maß P über \mathbb{R} die Darstellung

$$P(A) = \alpha_d P_d(A) + \alpha_s P_s(A), \quad A \in \mathbb{B}, \quad (4.16)$$

mit einer diskreten Verteilung P_d , einer stetigen Verteilung P_s und Gewichten $\alpha_d \in [0, 1]$ und $\alpha_s = 1 - \alpha_d$, dann heißt P eine **gemischte Verteilung**.

3.114

Folgerung 4.18. Ist F die VF eines W-Maßes P über (\mathbb{R}, \mathbb{B}) , dann gilt:

- (i) F ist isoton, d. h. monoton nicht fallend.

(ii) F ist „normiert“, d. h. F besitzt die Grenzwerte

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0,$$
$$F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1.$$

(iii) F ist rechtsseitig stetig.

(iv) F besitzt linksseitige Grenzwerte

$$F(x-) = \lim_{h \rightarrow 0^+} F(x-h) = P((-\infty, x)).$$

(v) Für Einpunktmengen $\{x\}$ gilt:

$$P(\{x\}) = F(x) - F(x-).$$

3.115

Bemerkung

Jede Abbildung $F : \mathbb{R} \rightarrow \mathbb{R}$ mit den Eigenschaften (i),(ii),(iii) ist eine Verteilungsfunktion. Die Zuordnung von Verteilungsfunktionen über \mathbb{R} und W-Maßen über (\mathbb{R}, \mathbb{B}) die durch (4.12) und (4.13) hergestellt wird, ist eineindeutig.

3.116

Kapitel 5

Mehrstufige Wahrscheinlichkeitsmodelle

5.1 Mehrstufige Wahrscheinlichkeitsmodelle

Mehrstufiger Versuch

Wird bei n -Merkmalen $\omega_1, \dots, \omega_n$ mit $\omega_i \in \Omega_i$ die Wahrscheinlichkeit der Merkmale stufenweise in Abhängigkeit von den vorangehenden Ergebnissen bewertet,

$$\begin{array}{ll} \omega_1 \mapsto f_1(\omega_1) & \text{eine Z-Dichte} \\ \omega_2 \mapsto f_2^1(\omega_1; \omega_2) & \text{eine von } \omega_1 \text{ abhängige Z-Dichte} \\ \omega_3 \mapsto f_3^2(\omega_1, \omega_2; \omega_3) & \text{eine von } (\omega_1, \omega_2) \text{ abhängige Z-Dichte} \\ \vdots & \end{array}$$

dann wird der Gesamtversuch bewertet durch die Z-Dichte

$$(\omega_1, \dots, \omega_n) \mapsto f(\omega_1, \dots, \omega_n) := f_1(\omega_1) f_2^1(\omega_1, \omega_2) \cdots f_n^{n-1}(\omega_1, \dots, \omega_{n-1}; \omega_n). \quad (5.1)$$

Dass f eine Z-Dichte ist, folgt aus

$$f_i^{i-1}(\omega_1, \dots, \omega_{i-1}; \omega_i) \geq 0 \quad \text{und} \quad \sum_{\omega_i \in \Omega_i} f_i^{i-1}(\omega_1, \dots, \omega_{i-1}; \omega_i) = 1. \quad (5.2)$$

5.117

Definition 5.1. (i) Die Z-Dichten $f_i^{i-1}(\omega_1, \dots, \omega_{i-1}; \omega_i)$ heißen **Übergangszähldichten** von $\Omega_1 \times \cdots \times \Omega_{i-1}$ nach Ω_i . Die jeweils vorausgehenden Beobachtungen $(\omega_1, \dots, \omega_{i-1})$ heißen **Vorgeschichte zur Stufe i** .

(ii) Die durch (5.1) definierte Gesamtdichte f wird als **Koppelung** von $f_1, f_2^1, \dots, f_n^{n-1}$ bezeichnet. Schreibweise:

$$f = f_1 \otimes f_2^1 \otimes \cdots \otimes f_n^{n-1}. \quad (5.3)$$

Bemerkung

Zu jeder **ÜZ-Dichte** f_i^{i-1} gehört ein von $(\omega_1, \dots, \omega_{i-1})$ abhängiges W-Maß P_i^{i-1} . Es heißt **Übergangs-W-Maß**.

Das W-Maß zur Gesamt-Z-Dichte wird mit

$$P = P_1 \otimes P_2^1 \otimes \dots \otimes P_n^{n-1}$$

bezeichnet.

5.118

5.2 Koppelung stetiger Modelle

Übergang zu stetigen Modellen

Anstatt der Z-Dichten werden R-Dichten $f_i^{i-1}(x_1, \dots, x_{i-1}; x_i)$ betrachtet. Die Gesamtdichte $f = f_1 \otimes f_2^1 \otimes \dots \otimes f_n^{n-1}$ wird definiert durch

$$f(x_1, \dots, x_n) := f_1(x_1) \cdots f_n^{n-1}(x_1, \dots, x_{n-1}; x_n). \quad (5.4)$$

Es muss sichergestellt werden, dass f auch Riemann-integrierbar ist.

5.119

5.3 Unabhängige Koppelung

Definition 5.2 (Unabhängige Koppelung, Produktdichte). Hängen bei einem mehrstufigen Versuch die ÜZ-Dichten oder ÜR-Dichten f_2^1, \dots, f_n^{n-1} nicht von den jeweiligen Vorgeschieden ab, so spricht man von **unabhängiger Koppelung**. Die Übergangsdichten sind dann einfache Z- bzw. R-Dichten f_2, \dots, f_n . Die Dichte f des Gesamtversuchs ist dann das Produkt der Einzeldichten, also

$$f(\omega_1, \dots, \omega_n) = f_1(\omega_1) f_2(\omega_2) \cdots f_n(\omega_n). \quad (5.5)$$

f wird als **Produktdichte** bezeichnet.

5.120

Folgerung 5.3 (Produktformel für $A = A_1 \times \dots \times A_n$). In einem n -stufigen unabhängig gekoppelten W-Modell mit den „einstufigen“ W-Maßen P_1, \dots, P_n gilt für ein „Produktereignis“ der Form $A = A_1 \times \dots \times A_n$ Folgendes:

$$P(A) = P(A_1 \times \dots \times A_n) = P_1(A_1) \cdots P_n(A_n). \quad (5.6)$$

Folgerung 5.4 (n -faches Laplace-Experiment). Werden mehrere Laplace-Versuche ohne gegenseitige Beeinflussung durchgeführt, dann ist der Gesamtversuch auch ein Laplace-Experiment. Die Z-Dichten der Einzelversuche sind $f_i(\omega_i) = \frac{1}{|\Omega_i|}$, ($\omega_i \in \Omega_i$). Durch unabhängige Koppelung erhält man

$$f(\omega_1, \dots, \omega_n) = \frac{1}{|\Omega_1|} \cdots \frac{1}{|\Omega_n|} = \frac{1}{|\Omega|} \quad \text{mit} \quad \Omega = \Omega_1 \times \dots \times \Omega_n. \quad (5.7)$$

5.121

Definition 5.5 (*n*-faches Bernoulli-Experiment). Die *n*-fache unabhängige Wiederholung eines Bernoulli(*p*)-Experiments ($0 \leq p \leq 1$, $n = 1, 2, 3, \dots$) heißt ***n*-faches Bernoulli(*p*)-Experiment**. Die Ergebnismenge ist $\Omega = \{0, 1\}^n$. Für die Z-Dichte gilt

$$f(\omega_1, \dots, \omega_n) = p^k (1-p)^{n-k} \quad \text{mit } k = \sum_{i=1}^n \omega_i. \quad (5.8)$$

Das zugehörige W-Maß wird mit $B_n(p)$ bezeichnet.

5.122

Definition 5.6 (Standard-Normalverteilung in \mathbb{R}^n). Die *n*-fache unabhängige Koppelung der Standard-Normalverteilungen $\mathcal{N}(0, 1)$ mit den R-Dichten $f_i(x_i) = \phi_i(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}$ heißt ***n*-dimensionale Standard-Normalverteilung** und hat als R-Dichte:

$$f(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)}. \quad (5.9)$$

5.123

5.4 Markow-Koppelung

Definition 5.7 (Markow-Koppelung). Hängen bei einem mehrstufigen Versuch die ÜZ- oder ÜR-Dichten nicht von der vollen Vorgeschichte ab, sondern nur vom letzten beobachteten Wert, so wird von einer **Markow-Koppelung** gesprochen. Die Folge der Beobachtungen heißt dann **Markow-Prozess**, im diskreten Fall auch **Markow-Kette**.

5.124

5.4.1 Zufälliges Ziehen ohne Zurücklegen

Gleichverteilung mehrfacher Ziehungen

Beim mehrfachen zufälligen Ziehen ohne Zurücklegen aus unterschiedlichen Objekten hat jedes mögliche Ergebnis die gleiche Wahrscheinlichkeit.

Anzahl möglicher Permutationen

von *n* aus *N* Objekten *mit* Berücksichtigung der Reihenfolge *ohne* Wiederholung

$$|\Omega_{\neq}| = N(N-1)(N-2) \cdots (N-n+1) =: (N)_n \quad (5.10)$$

$(N)_n$ wird als ***n*-faches absteigendes Produkt** bezeichnet. Es ist für $n \in \mathbb{N}_0$ und $N \in \mathbb{R}$ definiert. Es gilt $(N)_0 = 1$ und $(n)_n = n!$.

Anzahl möglicher Kombinationen

von *n* aus *N* Objekten *ohne* Berücksichtigung der Reihenfolge *ohne* Wiederholung

$$|\Omega'| = \frac{(N)_n}{n!} =: \binom{N}{n}. \quad (5.11)$$

Diese Definition ist auch für $N \in \mathbb{R}$, insbesondere $N < n$ gültig.

5.125

Verallgemeinerung

$$f(\omega_1, \dots, \omega_n) = \frac{(K)_k (N-K)_{n-k}}{(N)_n}, \quad \text{mit } k = \sum_{i=1}^n \omega_i, \quad (5.12)$$

$$P(B_k) = \binom{n}{k} \frac{(K)_k (N-K)_{n-k}}{(N)_n} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}. \quad (5.13)$$

5.126

Kapitel 6

Zufallsvariablen und Bildmodelle

6.1 Zufallsvariablen, Bildmodelle

Zur Erinnerung (vgl. Definition 3.24)

Eine **Zufallsvariable** ist eine Abbildung $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ mit der Eigenschaft:

$$\{X \in A'\} \in \mathcal{A} \text{ für alle } A' \in \mathcal{A}'.$$

$\{X \in A'\} = \{\omega \in \Omega | X(\omega) \in A'\}$ ist ein durch X beschreibbares Ereignis.

6.127

Definition 6.1. Für jede Abbildung X heißt $A := \{X \in A'\}$ das **Urbild** von A' . Schreibweise: $X^{-1}(A')$.

Die Zuordnung $A' \mapsto X^{-1}(A') = \{X \in A'\}$ von $\mathcal{P}(\Omega')$ \rightarrow $\mathcal{P}(\Omega)$ heißt **Urbildfunktion** X^{-1} .

Bemerkung

X^{-1} ist eine Mengenabbildung und im Gegensatz zur Umkehrfunktion immer definiert.

6.128

Definition 6.2. • Gilt für das Paar (Ω, \mathcal{A}) $\Omega \neq \emptyset$ und \mathcal{A} ist eine σ -Algebra über Ω gilt, dann heißt (Ω, \mathcal{A}) **Messraum**.

- Sind (Ω, \mathcal{A}) und (Ω', \mathcal{A}') Messräume und gilt $X : \Omega \rightarrow \Omega'$ mit

$$X^{-1}(A') = \{X \in A'\} \quad \forall A' \in \mathcal{A}'$$

dann heißt X **messbar** (auch \mathcal{A} - \mathcal{A}' -messbar).

Idee

- W-Maße von (Ω, \mathcal{A}) auf (Ω', \mathcal{A}') übertragen
- es genügt sich bei $A' \in \mathcal{A}'$ auf Erzeuger in Bildraum zu beschränken
- in VL ist die Messbarkeitseigenschaft immer erfüllt.

Beispiel 6.3. Langskript: [messbare Abbildungen] Folgende Abbildungen sind immer messbar:

- (i) alle Abbildungen $\Omega \rightarrow \Omega'$ mit $\mathcal{A} = \mathcal{P}(\Omega)$
- (ii) Indikatorfunktion: 1_A mit $A \in \mathcal{A}$.
- (iii) $f: \mathbb{R}^k \rightarrow \mathbb{R}^n$ mit $k, n \geq 1$, falls $\mathcal{A} = \mathbb{B}_k$ und $\mathcal{A}' = \mathbb{B}_n$
- (iv) alle Vielfachen, Summen, Produkte, Quotienten (wohldefiniert), Extrema von ZV
- (v) alle Suprema, Infima, Grenzwerte von Folgen von ZV
- (vi) alle messbaren Funktionen von ZV

6.2 Bildmodelle und Verteilungen von Zufallsvariablen

Beispiel 6.4. Langskript: [Qualitätskontrolle] Neuer Aspekt Erzeugen von W-Modellen

- N Objekte, K sind markiert, n Ziehungen W-Modell für Fertigungsprotokoll $\Omega = \{0, 1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$, Z-Dichte

$$f(\omega_1, \dots, \omega_n) = \frac{\binom{K}{\sum \omega_i} \binom{N-K}{n - \sum \omega_i}}{\binom{N}{n}}$$

- ZV Z_n beschreibt die Anzahl der gezogenen markierten Stücke
- $Z_n: \Omega \rightarrow \Omega'$ mit $(\omega_1, \dots, \omega_n) \mapsto \sum \omega_i$, $\Omega' = \{0, 1, \dots, n\}$

Welchen Nutzen hat nun Z_n ?

- Z_n verdichtet die Information, Ziehungsreihenfolge wird nicht mehr berücksichtigt.
- $B_k = \{\text{Es werden } k \text{ Stücke gezogen}\}: B_k = \{Z_n = k\}$
- Z-Dichte

$$f(k) = P'(\{k\}) = P(\{Z_n = k\}) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Allgemeiner:

$$P'(A') = P(Z_n \in A') := P(\underbrace{\{Z_n \in A'\}}_{Z_n^{-1}(A')}) = \sum_{\omega \in \Omega: Z_n(\omega) \in A'} P(\omega)$$

Definition 6.5. Ist (Ω, \mathcal{A}, P) ein W-Raum, Ω' eine (nicht leere) Menge, \mathcal{A}' ein Ereignissystem über Ω' und $X : \Omega \rightarrow \Omega'$ eine Zufallsvariable, dann ist die Zuordnung

$$A' \mapsto P^X(A') := P(X^{-1}(A')) = P(X \in A') \quad (6.1)$$

mit $A' \in \mathcal{A}'$ ein W-Maß über (Ω', \mathcal{A}') .

P^X heißt **Bildmaß von P unter X** oder **Verteilung von X** (bzgl. P). $(\Omega', \mathcal{A}', P^X)$ ist das **Bildmodell** von (Ω, \mathcal{A}, P) unter X .

Folgerung 6.6. (a) Ist $X : \Omega \rightarrow \Omega'$ eine ZV und Ω' oder $T' = X(\Omega)$ abzählbar, dann hat P^X die Z-Dichte f^X mit

$$f^X(\omega') = P(X = \omega'), \quad \omega' \in \Omega'. \quad (6.2)$$

(b) Ist X eine reellwertige Zufallsvariable, dann hat P^X die Verteilungsfunktion F^X mit

$$F^X(t) = P(X \leq t), \quad t \in \mathbb{R}. \quad (6.3)$$

6.3 Bildmodelle für diskrete Verteilungen

6.3.1 Hypergeometrische und Binomial-Verteilung

Definition 6.7 (Hypergeometrische Verteilung $H(N, K, n)$). Das **hypergeometrische Modell** besteht aus $\Omega' = \{0, 1, \dots, n\}$, $\mathcal{A}' = \mathcal{P}(\Omega')$ und P^{Z_n} , angegeben durch die Z-Dichte

$$h(N, K, n; k) := f^{Z_n}(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad 0 \leq k \leq n. \quad (6.4)$$

Das W-Maß P^{Z_n} heißt **hypergeometrische Verteilung** und wird mit $H(N, K, n)$ bezeichnet.

Definition 6.8. Das **Binomial-Modell** mit den Parametern $n \in \mathbb{N}$ und $0 \leq p \leq 1$ besteht aus $\Omega' = \{0, 1, \dots, n\}$, $\mathcal{A}' = \mathcal{P}(\Omega')$ und P^{S_n} , angegeben durch die Z-Dichte

$$b(n, p; k) := f^{S_n}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n. \quad (6.5)$$

Das W-Maß P^{S_n} heißt **Binomial(n, p)-Verteilung**, kurz $B(n, p)$.

- Hypergeometrische Verteilung Anwendung: N Elemente, $K < N$ markierte Elemente, n Ziehungen, k gezogenen sind markiert Dichtefunktion: hypergeometrische Verteilung > Ziehen ohne Zurücklegen mit Zurücklegen Bernoulli(p), Anzahl der Erfolge Binomial-verteilt
- Approximation der Binormalverteilung Poisson-Approximation: großes N und kleines p Normalverteilung: Hintergrund
- Negative Binomialverteilung
- geometrische Verteilung

6.134

6.3.2 Approximation der Binomial-Verteilung

Poisson-Approximation der Binomial-Verteilung

Im Satz 4.5 haben wir bereits gesehen, dass die Binomial-Verteilung durch die Poisson-Verteilung approximiert werden kann. Es gilt aber nur für $n \rightarrow \infty$ und p sehr klein. Dabei wird dann $\lambda = np$ gesetzt.

6.135

Die Normal-Approximation der Binomial-Verteilung

Satz 6.9 (Zentraler Grenzwertsatz). *Die Summe vieler kleiner und voneinander unabhängiger zufälliger Ereignisse verhält sich näherungsweise – und für wachsende Anzahl der Summanden mit zunehmender Genauigkeit – wie eine Normalverteilung.*

Satz 6.10. *Ist F^{S_n} die Verteilungsfunktion der Binomial(n, p)-Verteilung, und Φ die Verteilungsfunktion der Standard-Normalverteilung, dann gilt*

$$F^{S_n}(x) \approx \Phi\left(\frac{x-a}{\sigma}\right), \quad x \in \mathbb{R}, \quad (6.6)$$

wobei $a = np$ und $\sigma = \sqrt{np(1-p)}$ der approximierenden Normalverteilung ist.

6.136

6.3.3 Gelassen Warten

Definition 6.11 (Geometrische Verteilungen $\text{Geo}^{+/0}(p)$). Für $0 < p < 1$ und $q := 1 - p$ definierten wir die **geometrische Verteilung** $\text{Geo}^+(p)$ durch die Z-Dichte

$$\text{geo}^+(p; k) := pq^{k-1}, \quad k = 1, 2, 3, \dots \quad (6.7)$$

und die **geometrische Verteilung** $\text{Geo}^0(p)$ durch die Z-Dichte

$$\text{geo}^0(p; k) := pq^k, \quad k = 0, 1, 2, \dots \quad (6.8)$$

Beide Verteilungen beschreiben die Wartezeit bis zum ersten Erfolg bei einer unendlichen Folge von Bernoulli(p)-Versuchen. Bei $\text{Geo}^+(p)$ wird der erfolgreiche Versuch mitgezählt, bei $\text{Geo}^0(p)$ nicht.

6.137

Folgerung 6.12 (Verteilungsfunktion der geometrischen Verteilung). Die $\text{Geo}^+(p)$ -Verteilung besitzt die Verteilungsfunktion

$$F^{W_1}(x) = P(W_1 \leq x) = 1 - (1 - p)^{\lfloor x \rfloor}, x \geq 0, \quad (6.9)$$

die $\text{Geo}^0(p)$ -Verteilung besitzt entsprechend die Verteilungsfunktion

$$F^{W_1-1}(x) = P(W_1 - 1 \leq x) = 1 - (1 - p)^{\lfloor x+1 \rfloor}, x \geq 0, \quad (6.10)$$

6.138

Definition 6.13 (Negative Binomialverteilung). Die **negative Binomialverteilung** $\text{Nb}^+(r, p)$ die die Anzahl W_r der Versuche bis zum r -ten Erfolg beschreibt, besitzt die Z-Dichte

$$f^{W_r}(k) = \text{nb}^+(r, p; k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, k = r, r+1, \dots \quad (6.11)$$

Werden nur die Misserfolge gezählt, dann ergibt sich $\text{Nb}^0(r, p)$ mit der Z-Dichte

$$f^{W_r-1}(k) = \text{nb}^0(r, p; k) = \binom{k+r-1}{r-1} p^r (1-p)^k, k = 0, 1, 2, \dots \quad (6.12)$$

6.139

6.4 Bildverteilungen stetiger W-Dichten

Satz 6.14. Es sei P^X eine Verteilung über (\mathbb{R}, \mathbb{B}) und die Zufallsvariable $Y = a + bX$ eine lineare Funktion von X mit $a, b \in \mathbb{R}$, $b \neq 0$, hier $b > 0$.

(i) Besitzt P^X die VF F^X , dann besitzt P^Y die Verteilungsfunktion

$$F^Y(y) = F^X\left(\frac{y-a}{b}\right), \quad y \in \mathbb{R}. \quad (6.13)$$

(ii) Besitzt P^X die R-Dichte f^X , dann besitzt P^Y die R-Dichte

$$f^Y(y) = \frac{1}{b} f^X\left(\frac{y-a}{b}\right), \quad y \in \mathbb{R}. \quad (6.14)$$

- (iii) Ist P^X die Standardnormalverteilung $\mathcal{N}(0, 1)$ mit VF Φ und R-Dichte ϕ , dann hat $Y = a + bX$ die VF

$$F^Y(y) = \Phi\left(\frac{y-a}{b}\right)$$

und die R-Dichte

$$f^Y(y) = \frac{1}{b} \phi\left(\frac{y-a}{b}\right).$$

Y entspricht der Normalverteilung $\mathcal{N}(a, b^2)$.

6.140

Folgerung 6.15. (i) Ist X eine Zufallsvariable mit Werten in \mathbb{R} und der VF F^X , dann besitzt $Y = X^2$ die Verteilungsfunktion

$$F^Y(y) = F^{X^2}(y) = (F^X(\sqrt{y}) - F^X(-\sqrt{y})) 1_{[0, \infty)}(y), \quad y \in \mathbb{R}, \quad (6.15)$$

(siehe auch Folgerung 4.18).

- (ii) Besitzt X eine R-Dichte f^X , dann hat $Y = X^2$ die R-Dichte

$$f^Y(y) = \frac{1}{2\sqrt{y}} (f^X(-\sqrt{y}) + f^X(\sqrt{y})) 1_{[0, \infty)}(y), \quad y \in \mathbb{R}. \quad (6.16)$$

6.141

Satz 6.16. Ist P^X die Standard-Normalverteilung $\mathcal{N}(0, 1)$, dann besitzt die Verteilung P^{X^2} die VF

$$F^{X^2}(y) = [2\Phi(\sqrt{y}) - 1] 1_{[0, \infty)}(y), \quad y \in \mathbb{R}. \quad (6.17)$$

und die R-Dichte

$$f^{X^2}(y) = \frac{1}{\sqrt{y}} \phi(\sqrt{y}) 1_{[0, \infty)}(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}} 1_{[0, \infty)}(y), \quad y \in \mathbb{R}. \quad (6.18)$$

Die Verteilung P^{X^2} heißt **Chi(1)-Quadrat-Verteilung**, kurz χ_1^2 , und ist eine spezielle Gamma-Verteilung $\Gamma_{\frac{1}{2}, \frac{1}{2}}$.

6.142

Transformationen von ZV ($Y = g(X)$)

Besitzt die ZV X eine stetige Verteilung über \mathbb{R}^2 mit R-Dichte f^X und ist $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ eine Abbildung. Dann gilt für die VF F^Y der ZV $Y = g(X)$

$$F^Y(y) = P(Y \leq y) = \int_{B_y} f^X(x_1, x_2) dx_1 dx_2, \quad y \in \mathbb{R}, \quad (6.19)$$

mit $B_y := \{(x_1, x_2) \in \mathbb{R}^2 : g(x_1, x_2) \leq y\}$.

Dies lässt sich auf mehr als zwei Dimensionen übertragen.

6.143

6.5 Randverteilungen und gemeinsame Verteilungen

Beispiel 6.17. Verteilung für die Auftragsarten 1,2,3 und die Auftragsdauer $d = 5, 10, 15, 20$:

t	Z-Dichte $f(t, d)$			
	$d = 5$	$d = 10$	$d = 15$	$d = 20$
1	0,06	0,12	0,02	0,00
2	0,10	0,20	0,15	0,05
3	0,03	0,09	0,12	0,06

6.144

Beispiel 6.18. Verteilung für die Auftragsarten 1,2,3 und die Auftragsdauer $d = 5, 10, 15, 20$:

t	Z-Dichte $f(t, d)$				$\Sigma = f^T(t)$
	$d = 5$	$d = 10$	$d = 15$	$d = 20$	
1	0,06	0,12	0,02	0,00	0,20
2	0,10	0,20	0,15	0,05	0,50
3	0,03	0,09	0,12	0,06	0,30
$\Sigma = f^D(d)$	0,19	0,41	0,29	0,11	$\Sigma = 1$

Definition 6.19. Ist $\Omega = \Omega_1 \times \cdots \times \Omega_n$, dann heißt die ZV $X_i : \Omega \rightarrow \Omega_i$, $\omega \mapsto \omega_i$ die **i-te Projektion** oder auch die i-te Koordinatenvariable. Die Verteilung P^{X_i} von X_i heißt die **i-te Randverteilung**.

6.145

Folgerung 6.20. (i) Die i-te Randverteilung P^{X_i} ergibt sich aus

$$P^{X_i}(A_i) = P(X \in A_i) = P(\Omega_1 \times \cdots \times A_i \times \cdots \times \Omega_n)$$

für $A_i \in \mathcal{A}_i$.

(ii) Ist Ω abzählbar und f eine Z-Dichte von P , dann besitzt P^{X_i} die Z-Dichte f^{X_i} , welche gegeben ist durch

$$f^{X_i}(\omega_i) = \sum_{\omega_1 \in \Omega_1} \cdots \sum_{\omega_{i-1} \in \Omega_{i-1}} \sum_{\omega_{i+1} \in \Omega_{i+1}} \cdots \sum_{\omega_n \in \Omega_n} f(\omega_1, \dots, \omega_n).$$

Sie wird auch **i-te Randdichte** genannt.

(iii) Ist $\Omega_i = \mathbb{R}$, $\mathcal{A}_i = \mathbb{B}$ und $\mathcal{A} = \mathbb{B}^n$ und besitzt P eine R-Dichte f , dann hat auch P^{X_i} eine R-Dichte f^{X_i} mit

$$f^{X_i}(x_i) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

6.146

Folgerung 6.21. In unabhängig gekoppelten W-Modellen gilt für P^{X_i}

$$P^{X_i}(A_i) = P(\Omega_1 \times \cdots \times A_i \times \cdots \times \Omega_n) = P_1(\Omega_1) \cdots P_i(A_i) \cdots P_n(\Omega_n) = P_i(A_i).$$

Definition 6.22. Seien die $Y_i : \Omega \rightarrow \Omega_i$ ZV, dann ist $Y = (Y_1, \dots, Y_n)$ eine ZV mit $Y : \Omega \rightarrow \Omega_1 \times \cdots \times \Omega_n$. Die Verteilung $P^Y = P^{(Y_1, \dots, Y_n)}$ von Y heißt dann **gemeinsame Verteilung** von Y_1, \dots, Y_n .

Folgerung 6.23. Die i -te Randverteilung einer gemeinsamen Verteilung $P^{(Y_1, \dots, Y_n)}$ ist die Verteilung P^{Y_i} , die Verteilung von Y_i .

6.147

Folgerung 6.24. Bei der n -fachen Ziehung ohne Zurücklegen aus N Objekten, wovon K markiert sind, sei (X_1, \dots, X_n) das vollständige Ergebnis. $(X_i(\omega) \in \{0, 1\})$. Dann gilt, dass jede Permutation X_{i_1}, \dots, X_{i_n} von X_1, \dots, X_n dieselbe gemeinsame Verteilung hat und dass deshalb die Randverteilungen $P^{(X_i)}$ bzw. $P^{(X_{i_1}, \dots, X_{i_k})}$ ($k < n$) für alle (i_1, \dots, i_k) mit $(i_l \neq i_m)$ jeweils übereinstimmen. ZV mit dieser Eigenschaft heißen austauschbar.

6.148

6.6 Stochastische Unabhängigkeit von ZV

Beispiel 6.25. Die Produktdichte für das Beispiel 6.8

t	$f(t, d) = f_1(t) f_2^1(t; d)$			
	$d = 5$	$d = 10$	$d = 15$	$d = 20$
1	0,2 · 0,3	0,2 · 0,6	0,2 · 0,1	0,2 · 0,0
2	0,5 · 0,2	0,5 · 0,4	0,5 · 0,3	0,5 · 0,1
3	0,3 · 0,1	0,3 · 0,3	0,3 · 0,4	0,3 · 0,2

6.149

Satz 6.26. Jede gemeinsame Verteilung $P^{(Y_1, \dots, Y_n)}$ mit R-Dichte $f^{(Y_1, \dots, Y_n)}$ lässt sich als Kopplungsmodell mit der R-Dichte

$$f^{(Y_1, \dots, Y_n)}(y_1, \dots, y_n) = f_1(y_1) f_2^1(y_1; y_2) \cdots f_n^{n-1}(y_1, \dots, y_{n-1}; y_n)$$

darstellen. Dazu werden die Randdichten

$$f^{(Y_1, \dots, Y_{n-1})}, f^{(Y_1, \dots, Y_{n-2})}, \dots, f^{(Y_1, Y_2)}, f^{(Y_1)} = f_1$$

durch Integration über y_n, y_{n-1}, \dots, y_2 berechnet und es ergibt sich daraus

$$f_i^{i-1}(y_1, \dots, y_n) = \frac{f^{(Y_1, \dots, Y_i)}(y_1, \dots, y_i)}{f^{(Y_1, \dots, Y_{i-1})}(y_1, \dots, y_{i-1})}. \quad (6.20)$$

6.150

Definition 6.27. Die Übergangsdichten

$$f_i^{i-1}(y_1, \dots, y_{i-1}; y_i) = P(Y_i = y_i | Y_1 = y_1, \dots, Y_{i-1} = y_{i-1})$$

heißen **bedingte Dichten**. Die zugehörigen Übergangs-W-Maße heißen **bedingte Verteilungen**. Es wird auch $f^{Y_i | (Y_1, \dots, Y_{i-1})}$ bzw. $P^{Y_i | (Y_1, \dots, Y_{i-1})}$ geschrieben.

6.151

Definition 6.28. Die Zufallsvariablen Y_1, \dots, Y_n mit $Y_i : \Omega \rightarrow \Omega_i$ heißen **stochastisch unabhängig** wenn für die gemeinsame Verteilung die Produktformel

$$P^{Y_1, \dots, Y_n}(A_1 \times \dots \times A_n) = P^{Y_1}(A_1) \cdots P^{Y_n}(A_n) \quad (6.21)$$

für beliebige Ereignisse $A_i \in \Omega_i$ gilt.

Folgerung 6.29. Besitzen die Zufallsvariablen Y_1, \dots, Y_n mit $Y_i : \Omega \rightarrow \Omega_i$ R-Dichten, dann ist die stochastische Unabhängigkeit äquivalent dazu, dass die gemeinsame Verteilung eine Produktdichte besitzt.

Weiß man, dass die n Zufallsvariablen Y_1, \dots, Y_n stochastisch unabhängig sind, so lassen sich oftmals direkt Entscheidungen über die stochastische Unabhängigkeit von mit Y_1, \dots, Y_n zusammenhängenden Zufallsvariablen treffen:

Satz 6.30. Sind die Zufallsvariablen Y_1, \dots, Y_n stochastisch unabhängig, dann sind auch stochastisch unabhängig:

- (i) Umstellungen von Y_1, \dots, Y_n , z.B. Y_3, Y_1, Y_5, Y_4, Y_2 ,
- (ii) Teilmengen von Y_1, \dots, Y_n , z.B. Y_1, Y_3, Y_4 ,
- (iii) disjunkte Gruppen von stoch. unabh. ZV, z.B. $Z_1 = (Y_1, Y_3)$, $Z_2 = (Y_4, Y_5)$,
- (iv) messbare Funktionen von stoch.unabh. ZV, z.B. $g(Z_1) = Y_1 + 2Y_3^2$, $h(Z_2) = Y_4 \cdot e^{Y_5}$.

Außerdem gilt:

- (i) Jede konstante ZV ist von anderen ZV stoch.unabh.
- (ii) Sind Y_1, \dots, Y_{n-1} stoch.unabh. und sind $(Y_1, \dots, Y_{n-1}), Y_n$ stoch.unabh., dann sind auch Y_1, \dots, Y_n stochastisch unabhängig.

Folgerung 6.31. Die Ereignisse A_1, \dots, A_n sind stoch.unabh. genau dann, wenn die ZV $1_{A_1}, \dots, 1_{A_n}$ stoch.unabh. sind.

6.7 Summenverteilung, Faltung

Satz 6.32. (a) X und Y seien zwei ZV über (Ω, \mathcal{A}, P) mit Werten in \mathbb{Z} und gemeinsamer Z-Dichte $f^{(X,Y)}(x,y)$, dann besitzt $X+Y$ die Z-Dichte

$$f^{X+Y}(z) = \sum_{x \in \mathbb{Z}} f^{(X,Y)}(x, z-x), \quad z \in \mathbb{Z}. \quad (6.22)$$

(b) X und Y seien zwei ZV über (Ω, \mathcal{A}, P) mit Werten in \mathbb{R} und gemeinsamer R-Dichte $f^{(X,Y)}(x,y)$, dann besitzt $X+Y$ die R-Dichte

$$f^{X+Y}(z) = \int_{\mathbb{R}} f^{(X,Y)}(x, z-x) dx, \quad z \in \mathbb{R}. \quad (6.23)$$

Definition 6.33. Die Summe von stochastisch unabhängigen Zufallsvariablen X und Y heißt **Faltung** der Einzelverteilungen und ist definiert als:

$$P^X * P^Y := P^{X+Y} \text{ und } f^X * f^Y := f^{X+Y}. \quad (6.24)$$

Folgerung 6.34. Für nicht-negative und stochastisch unabhängige Zufallsvariablen X und Y mit Dichten f^X und f^Y wird die Faltung nach folgender Formel berechnet:

(a) für Z-Dichten

$$f^{X+Y}(z) = (f^X * f^Y)(z) = \sum_{x=0}^z f^X(x) f^Y(z-x), \quad z \in \mathbb{N}_0, \quad (6.25)$$

(b) für R-Dichten

$$f^{X+Y}(z) = (f^X * f^Y)(z) = \int_0^z f^X(x) f^Y(z-x) dx, \quad z \geq 0. \quad (6.26)$$

6.155**Faltung von Binomialverteilungen**

Die Binomialverteilung entspricht der Summenverteilung von n stoch. unabh. Bernoulli(p)-Zufallsvariablen X_1, X_2, \dots, X_n . $B(n, p)$ ist die n -fache Faltung von $B(p)$ -Verteilungen, d.h.:

$$B(n, p) = \underbrace{B(p) * \dots * B(p)}_{n \text{ Faktoren}}. \quad (6.27)$$

Damit gilt

$$B(n+m, p) = B(n, p) * B(m, p). \quad (6.28)$$

Faltung von Poisson-Verteilungen

Wie aus Satz 4.5 bekannt ist, lässt sich die Poisson-Verteilung durch $B(n, p)$ mit $n \cdot p_n \rightarrow \lambda$ approximieren. Die Faltungsformel lautet

$$\pi(\lambda_1) * \pi(\lambda_2) = \pi(\lambda_1 + \lambda_2). \quad (6.29)$$

6.156**Faltung von geometrischen Verteilungen**

Betrachten wir eine Folge von stoch.unabh. Bernoulli(p)-Versuchen, so sind die jeweiligen Zwischenwartezeiten auf den nächsten Erfolg (Anzahl der Bernoulli(p)-Versuche, die notwendig sind, um den nächsten Erfolg zu erzielen) auch wieder stoch. unabh. und $\text{Geo}^+(p)$ -verteilt. Die Verteilung für die Wartezeit auf den r -ten Erfolg ergibt sich durch r -fache Faltung der $\text{Geo}^+(p)$ -Verteilung. Diese liefert die negative Binomialverteilung

$$\text{Nb}^+(r, p) = \underbrace{\text{Geo}^+(p) * \text{Geo}^+(p) * \dots * \text{Geo}^+(p)}_{r \text{ Faktoren}} \quad (6.30)$$

und somit auch

$$\text{Nb}^+(r_1 + r_2, p) = \text{Nb}^+(r_1, p) * \text{Nb}^+(r_2, p). \quad (6.31)$$

6.157**Faltung der Normalverteilung**

Die Faltung zweier beliebiger Normalverteilungen ergibt wieder eine Normalverteilung

$$\mathcal{N}(a, \sigma^2) * \mathcal{N}(b, \tau^2) = \mathcal{N}(a+b, \sigma^2 + \tau^2). \quad (6.32)$$

Es addieren sich die Mittelwerte a, b und die Quadrate der Streuungen σ, τ (siehe (7.27)). Dies ist der Grund für die Notation $\mathcal{N}(a, \sigma^2)$.

6.158

Faltung von Gamma-Verteilungen

Die Faltung von zwei Gamma-Verteilungen mit gleichem Parameter α ergibt

$$\Gamma_{\alpha, \mu} * \Gamma_{\alpha, \nu} = \Gamma_{\alpha, \mu + \nu}. \quad (6.33)$$

Bemerkung

Es gelten die folgenden Spezialfälle:

$$\text{Exp}(\alpha) = \Gamma_{\alpha, 1} \quad (6.34)$$

$$\chi_1^2 = \Gamma_{\frac{1}{2}, \frac{1}{2}} \quad (6.35)$$

Folgerung 6.35. Es gilt mit der obigen Bemerkung:

$$\Gamma_{\alpha, n} = \underbrace{\text{Exp}(\alpha) * \text{Exp}(\alpha) * \cdots * \text{Exp}(\alpha)}_{n \text{ Faktoren}} \quad (6.36)$$

$$\chi_n^2 := \Gamma_{\frac{1}{2}, \frac{n}{2}} = \underbrace{\chi_1^2 * \chi_1^2 * \cdots * \chi_1^2}_{n \text{ Faktoren}} \quad (6.37)$$

6.159

Kapitel 7

Kenngrößen

7.1 Median und Quantile

Zur Erinnerung

Eine **Zufallsvariable** ist eine Abbildung $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ mit der Eigenschaft:

$$\{X \in A'\} \in \mathcal{A} \text{ für alle } A' \in \mathcal{A}'.$$

$\{X \in A'\} = \{\omega \in \Omega | X(\omega) \in A'\}$ ist ein durch X beschreibbares Ereignis.

Weitere Annahmen

Zur ZV X gehören die Z-Dichte oder R-Dichte f^X und die Verteilungsfunktion F^X .

7.160

Definition 7.1 (Median). Ein **Median** von X (oder P^X) ist jeder Wert $m \in \mathbb{R}$, an dem die Verteilungsfunktion F^X den Wert $\frac{1}{2}$ erreicht oder überschreitet, d.h. für den gilt

$$F^X(m-) \leq \frac{1}{2} \leq F^X(m). \quad (7.1)$$

Bemerkung

(i) Links und rechts von m liegen jeweils höchstens die Hälfte der Wahrscheinlichkeit:
 $P(X < m) = F^X(m-) \leq \frac{1}{2}.$

(ii) Der Median ist nicht eindeutig, ggf. wird der Mittelpunkt eines Intervalls verwendet.

(iii) Der Median hat zunehmende Bedeutung, fehlende Eindeutigkeit ist aber negativ.

7.161

Definition 7.2 (Quantil). Ein Wert $u_\alpha \in \mathbb{R}, 0 < \alpha < 1$, heißt **α -Quantil** von P^X , wenn für die VF F^X gilt

$$F^X(u_\alpha-) \leq \alpha \leq F^X(u_\alpha).$$

Es wird auch vom $p\%$ -Quantil gesprochen. Üblich ist die Schreibweise

$$u_\alpha \text{ oder } u_{p\%}.$$

Spezielle Quantile

- (i) Das 50%-Quantil $u_{50\%}$ entspricht dem Median.
- (ii) Die 25%- und 75%-Quantile heißen auch **Quartile**.

7.162**Anwendung der Quantile**

- (i) Angabe der „Breite“ einer Verteilung, z.B. $u_{5\%}$ und $u_{95\%}$.
- (ii) Quantile der $\mathcal{N}(a, \sigma^2)$ -Verteilung lassen sich leicht aus der $\mathcal{N}(0, 1)$ -Verteilung berechnen:

$$u_{p\%}^{\mathcal{N}(a, \sigma^2)} = a - u_{p\%}^{\mathcal{N}(0, 1)} \sigma.$$

- (iii) In manchen Büchern findet man die Quantile auch unter der Bezeichnung „Fraktile“.

Definition 7.3 (Modalwert). Jede Maximumstelle einer Z-Dichte oder R-Dichte f^X heißt **Modalwert** von X oder von P^X .

7.163

7.2 Erwartungswert

Definition 7.4 (Erwartungsvektor). Ist $X = (X_1, \dots, X_n)$ eine mehrdimensionale Zufallsvariable, dann wird der Erwartungswert von X als Vektor der Erwartungswerte der X_i definiert:

$$EX = (EX_1, \dots, EX_n).$$

7.164

7.3 Erwartungswert: diskrete Modelle

Definition 7.5 (Erwartungswert). $X : \Omega \rightarrow \Omega' \subset \mathbb{R}$ sei eine diskrete ZV mit $X \geq 0$ oder Ω' endlich. Dann heißt

$$EX := \sum_{k \in \Omega'} k \cdot P(X = k) = \sum_{k \in \Omega'} k \cdot f^X(k) \quad (7.2)$$

der **Erwartungswert** von X (oder P^X).

7.165

Folgerung 7.6 (Erwartungswerte wichtiger Verteilungen).

(a) Laplace-Verteilung	$L(\{1, \dots, N\})$,	$f^X(k) = 1/N$	$EX = \frac{N+1}{2}$
(b) Einpunktverteilung	ε_a ,	$f^X(a) = 1$	$EX = a$
(c) Bernoulli-Versuch	$B(p)$,	$f^{X_1}(1) = p$	$EX = p$
(d) Binomial-Verteilung	$B(n, p)$,	$f^{S_n}(k) = \binom{n}{k} p^k (1-p)^{n-k}$	$ES_n = np$
(e) Hypergeom. Vert.	$H(N, K, n)$,	$f^{Z_n}(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$	$EZ_n = n \frac{K}{N}$
(f) Poisson-Verteilung	$\pi(\lambda)$,	$f^X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	$EX = \lambda$
(g) Geo. Verteilung	$\text{Geo}^+(p)$,	$f^{W_1}(k) = p(1-p)^{k-1}$	$EW_1 = \frac{1}{p}$
	$\text{Geo}^0(p)$,	$f^{W'_1}(k) = p(1-p)^k$	$EW'_1 = \frac{1-p}{p}$
(h) Neg. Binomial-Vert.	$\text{Nb}^+(p)$,	$f^{W_r}(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$	$EW_r = \frac{r}{p}$
	$\text{Nb}^0(p)$,	$f^{W'_r}(k) = \binom{k+r-1}{r-1} p^r (1-p)^k$	$EW'_r = \frac{r(1-p)}{p}$

7.166

Definition 7.7 (Positiv-/Negativteil). (i) Der **Positivteil** einer reellen Zahl a ist

$$a^+ = \max(0, a) = \begin{cases} 0 & \text{für } a \leq 0, \\ a & \text{für } a \geq 0. \end{cases}$$

Der **Negativteil** einer reellen Zahl a ist

$$a^- = (-a)^+ = \max(0, -a) = \begin{cases} |a| & \text{für } a \leq 0, \\ 0 & \text{für } a \geq 0. \end{cases}$$

(ii) Entsprechend werden f^+ und f^- für die reellwertige Abbildung f definiert:

$$\begin{aligned} f^+(y) &= (f(y))^+ \\ f^-(y) &= (f(y))^- \end{aligned}$$

Für eine ZV $X : \Omega \rightarrow \Omega' \in \mathbb{R}$ ist der Positivteil X^+ und der Negativteil X^- erklärt. Es gilt

$$X = X^+ - X^-.$$

7.167

Definition 7.8 (Erwartungswert diskret). Es sei $X : \Omega \rightarrow \Omega' \in \mathbb{R}$ eine diskrete ZV mit Träger $T \subset \Omega'$ und $f^X(k)$ ($k \in T$) eine Z-Dichte. Dann heißt

$$EX := \sum_{k \in T} k P(X = k) = \sum_{k \in T} k f^X(k) \quad (7.3)$$

der **Erwartungswert von X** (oder von P^X), falls die positive oder die negative Teilsumme (oder beide) endlich ist, d.h. falls

$$EX^+ := \sum_{k \in T, k > 0} k f^X(k) < \infty \quad \text{oder} \quad EX^- := \sum_{k \in T, k < 0} |k| f^X(k) < \infty. \quad (7.4)$$

Anmerkungen

- $EX = EX^+ - EX^-$ ist unabhängig von der Summationsreihenfolge.
- Sind $EX^+ < \infty$ und $EX^- < \infty$, so heißt X „integrierbar“.

7.168

Satz 7.9 (Darstellungen des Erwartungswertes). *Die folgenden Gleichungen gelten unter der Voraussetzung, dass die Summen in den Gleichungen existieren. Die Existenz einer der beiden Seiten, zieht die Existenz der anderen Seite nach sich.*

- (i) $X : \Omega \rightarrow \Omega' \subset \mathbb{R}$ sei eine diskrete ZV, Ω, Ω' sind abzählbar und f eine Z-Dichte von P . Dann gilt

$$EX := \sum_{k \in \Omega'} k \cdot P(X = k) = \sum_{\omega \in \Omega} X(\omega) f(\omega). \quad (7.5)$$

- (ii) $X : \Omega \rightarrow \Omega' \subset \mathbb{R}$ sei eine diskrete ZV, $g : \Omega' \rightarrow \Omega'' \subset \mathbb{R}$, Ω', Ω'' abzählbar. Dann gilt

$$Eg(X) := \sum_{m \in \Omega''} m \cdot P(g(X) = m) = \sum_{k \in \Omega'} g(k) \cdot P(X = k). \quad (7.6)$$

- (iii) $X : \Omega \rightarrow \Omega'_1, Y : \Omega \rightarrow \Omega'_2$ sind diskrete ZV, $h : \Omega'_1 \times \Omega'_2 \rightarrow \Omega'' \subset \mathbb{R}$ eine Abbildung, $\Omega'_1, \Omega'_2, \Omega''$ sind abzählbar. Dann gilt

$$Eh(X, Y) := \sum_{m \in \Omega''} m \cdot P(h(X, Y) = m) = \sum_{k \in \Omega'_1} \sum_{l \in \Omega'_2} h(k, l) \cdot P(X = k, Y = l). \quad (7.7)$$

7.169

Satz 7.10 (Eigenschaften des Erwartungswertes). X, Y, X_1, \dots, X_n seien reellwertige ZV.

- (a) Gilt $P(X = a) = 1$, d.h. ist X („fast sicher“) konstant, dann besitzt X die Einpunktverteilung ϵ_a , und es ist $EX = a$.
- (b) Der Erwartungswert ist **monoton**, d.h.

$$X \leq Y \Rightarrow EX \leq EY, \text{ falls } EX, EY \text{ existieren.}$$

Speziell gilt:

$$a \leq X \leq b \Rightarrow a \leq EX \leq b.$$

- (c) Der Erwartungswert ist **linear**, d.h.

$$E(aX + b) = a \cdot EX + b. \quad (7.8)$$

- (d1) Existieren EX und EY und ist $EX + EY$ definiert (also nicht $\infty + (-\infty)$ oder $(-\infty) + \infty$), dann existiert auch $E(X + Y)$ und es gilt

$$E(X + Y) = EX + EY. \quad (7.9)$$

- (d2) Existieren EX_i und $EX_i \neq \pm\infty$ für alle i , dann gilt

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i. \quad (7.10)$$

(e) X, Y seien **stochastisch unabhängig**, EX und EY existieren und sind **endlich**. Dann existiert $EXY := E(XY)$ und es gilt

$$EXY = EX \cdot EY. \quad (7.11)$$

7.170

Folgerung 7.11. Ist $X : \Omega \rightarrow \Omega' \subset \mathbb{R}$ eine reellwertige ZV. Dann gilt für $|X| = X^+ + X^-$:

$$E|X| = EX^+ + EX^-, \quad (7.12)$$

$$EX \text{ existiert} \quad \Rightarrow \quad |EX| \leq E|X|, \quad (7.13)$$

$$X \text{ ist integrierbar} \quad \Rightarrow \quad E|X| < \infty. \quad (7.14)$$

7.171

7.4 Erwartungswert: stetige und gemischte Modelle

Definition 7.12. Es sei $X : \Omega \rightarrow \mathbb{R}$ eine reellwertige ZV mit R-Dichte f^X . Dann heißt

$$EX = \int_{-\infty}^{\infty} x f^X(x) dx \quad (7.15)$$

der **Erwartungswert** von X (oder von P^X), falls

$$EX^+ = \int_0^{\infty} x f^X(x) dx < \infty \quad (7.16)$$

oder

$$EX^- = \int_{-\infty}^0 |x| f^X(x) dx < \infty \quad (7.17)$$

Es gilt $EX = EX^+ - EX^-$.

Sprechweisen

Es heißt „ EX existiert“. Gilt $EX^+ < \infty$ und $EX^- < \infty$, dann heißt X „integrierbar“.

7.172

Definition 7.13. Es sei $X : \Omega \rightarrow \mathbb{R}$ eine ZV mit gemischter Verteilung $P^X = \alpha_d P_d^X + \alpha_s P_s^X$ mit $\alpha_d \in [0, 1]$ und $\alpha_s = 1 - \alpha_d$. Existieren die Erwartungswerte $E_d X$ von P_d^X und $E_s X$ von P_s^X , dann sei

$$EX = \alpha_d E_d X + \alpha_s E_s X \quad (7.18)$$

der **Erwartungswert von X** (bzw. von P^X).

Folgerung 7.14. Für eine reellwertige ZV $X : \Omega \rightarrow \mathbb{R}$ mit VF F^X lassen sich der Positiv- und der Negativteil des Erwartungswertes darstellen als

$$EX^+ = \int_0^{\infty} [1 - F^X(x)] dx \quad \text{und} \quad EX^- = \int_{-\infty}^0 F^X(x) dx. \quad (7.19)$$

7.173

Folgerung 7.15. (a) Die Eigenschaften aus Satz 7.10 gelten auch im allgemeinen Fall. Ebenso die Folgerung 7.11.

- (b1) Besitzt $X : \Omega \rightarrow \mathbb{R}$ die R-Dichte f^X und ist $g : \mathbb{R} \rightarrow \mathbb{R}$ eine R-integrierbare Abbildung, dann gilt

$$Eg(X) = \int_{-\infty}^{\infty} g(x)f^X(x) dx. \quad (7.20)$$

- (b2) Besitzen die ZV $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ die gemeinsame R-Dichte $f^{(X,Y)}$, ist h eine Abbildung $\mathbb{R}^2 \rightarrow \mathbb{R}$ und $h \cdot f^{(X,Y)}$ R-integrierbar, dann gilt

$$Eh(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f^{(X,Y)}(x,y) dx dy. \quad (7.21)$$

7.174

Folgerung 7.16 (Erwartungswerte wichtiger stetiger Verteilungen).

- | | | |
|--|---|----------------------------|
| (a) Rechteck-Verteilung $\mathcal{R}(a,b)$, | $f^X(x) = \frac{1}{b-a} 1_{a,b}(x)$ | $EX = \frac{a+b}{2}$ |
| (b) Exponential-Vert. $\text{Exp}(\alpha)$, | $f^X(x) = \alpha e^{-\alpha x} 1_{(0,\infty)}(x)$ | $EX = \frac{1}{\alpha}$ |
| (c) Gamma-Verteilung $\Gamma_{\alpha,\nu}$, | $f^X(x) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} 1_{(0,\infty)}(x)$ | $EX = \frac{\nu}{\alpha}$ |
| (d) Beta-Verteilung $Be(\mu, \nu)$, | $f^X(x) = \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\Gamma(\nu)} x^{\mu-1} (1-x)^{\nu-1}$ | $EX = \frac{\mu}{\mu+\nu}$ |
| (e) Normal-Verteilung $\mathcal{N}(a, \sigma^2)$, | $f^X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-a)^2}{\sigma^2}}$ | $EX = a$ |
| (f) Cauchy-Verteilung $C(\alpha)$, | $f^X(x) = \frac{1}{\pi} \frac{\alpha}{\alpha^2 + x^2}$ | EX ex. nicht |

7.175

7.5 Streuung und Varianz

Definition 7.17 (Varianz und Streuung). Ist $X : \Omega \rightarrow \Omega' \subset \mathbb{R}$ eine reellwertige ZV mit endlichem Erwartungswert, dann heißen

$$\text{Var} X := E(X - EX)^2 = EX^2 - (EX)^2 \quad (7.22)$$

und

$$\text{Str} X := \sqrt{E(X - EX)^2} = \sqrt{\text{Var} X} \quad (7.23)$$

die **Varianz** und die **Streuung** von X .

7.176

Satz 7.18. Es sei $a \in \mathbb{R}$.

- (a) Eine Verschiebung hat keinen Einfluss auf die Varianz und die Streuung:

$$\text{Var}(X+a) = \text{Var} X, \quad \text{Str}(X+a) = \text{Str} X. \quad (7.24)$$

- (b) Ein Faktor verändert die Varianz quadratisch, die Streuung proportional (mit dem Betrag des Faktors):

$$\text{Var}(aX) = a^2 \text{Var} X, \quad \text{Str}(aX) = |a| \text{Str} X \quad (7.25)$$

(c) Nützlich ist auch die folgende Formel

$$E(X - a)^2 = \text{Var} X + (EX - a)^2, \text{ speziell } EX^2 = \text{Var} X + (EX)^2. \quad (7.26)$$

(d) Konstante Zufallsvariablen besitzen die Streuung 0:

$$\text{Str} X = 0 \Leftrightarrow \text{Var} X = 0 \Leftrightarrow P(X = EX) = 1.$$

(e) Für stochastisch unabhängige Zufallsvariablen gilt „Varianz einer Summe = Summe der Varianzen“, d.h.

$$X, Y \text{ seien stoch.unabh.} \implies \text{Var}(X + Y) = \text{Var} X + \text{Var} Y. \quad (7.27)$$

7.177

Folgerung 7.19 (Varianz wichtiger diskreter Verteilungen).

(a) $L(\{1, \dots, N\})$,	$\text{Var} X = \frac{N^2 - 1}{12}$
(b) ε_a ,	$\text{Var} X = 0$
(c) $B(p)$,	$\text{Var} X_1 = p(1 - p)$
(d) $B(n, p)$,	$\text{Var} S_n = np(1 - p)$
(e) $H(N, K, n)$,	$\text{Var} Z_n = n \frac{K}{N} \frac{N - K}{N} \frac{N - n}{N - 1}$
(f) $\pi(\lambda)$,	$\text{Var} X = \lambda$
(g) $\text{Geo}^+(p)$,	$\text{Var} W_1 = \frac{1 - p}{p^2}$
$\text{Geo}^0(p)$,	$\text{Var}(W_1 - 1) = \text{Var} W_1$
(h) $\text{Nb}^+(p)$,	$\text{Var} W_r = \frac{r(1 - p)}{p^2}$
$\text{Nb}^0(p)$,	$\text{Var}(W_r - r) = \text{Var} W_r$

7.178

Folgerung 7.20 (Varianz wichtiger stetiger Verteilungen).

(a) $\mathcal{R}(a, b)$,	$\text{Var} X = \frac{(b - a)^2}{12}$
(b) $\text{Exp}(\alpha)$,	$\text{Var} X = \frac{1}{\alpha^2}$
(c) $\Gamma_{\alpha, \nu}$,	$\text{Var} X = \frac{\nu}{\alpha^2}$
(d) $\chi_n^2 = \Gamma_{\frac{1}{2}, \frac{n}{2}}$,	$\text{Var} X = 2n$
(e) $\mathcal{N}(0, 1)$,	$\text{Var} X = 1$
(f) $\mathcal{N}(a, \sigma^2)$,	$\text{Var} X = \sigma^2$

7.179

7.6 Kovarianz

Definition 7.21. Für die ZV $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ mit $EX^2 < \infty$ und $EY^2 < \infty$ heißt

$$\text{Kov}(X, Y) := EXY - EXEY = E[(X - EX)(Y - EY)] \quad (7.28)$$

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

$$\text{korr}(X, Y) := \frac{\text{Kov}(X, Y)}{\text{Str}X \text{Str}Y} \quad (7.29)$$

heißt **Korrelationskoeffizient von X und Y** , falls $\text{Str}X \neq 0$ und $\text{Str}Y \neq 0$. Anderenfalls sei $\text{korr}(X, Y) = 0$.

7.180

Folgerung 7.22. X, Y seien reellwertige ZV mit $EX^2 < \infty$ und $EY^2 < \infty$.

Dann gilt:

(i) (Kovarianz)

$$\text{Kov}(X, X) = \text{Var}(X), \quad (7.30)$$

$$\text{Kov}(X, Y) = \text{Kov}(Y, X). \quad (7.31)$$

(ii) (Varianz der Summe von Zufallsgrößen)

$$\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2\text{Kov}(X, Y). \quad (7.32)$$

Für $X_i : \Omega \rightarrow \mathbb{R}$ mit $EX_i^2 < \infty, i = 1, \dots, n$ gilt:

$$\text{Var} \sum_{i=1}^n X_i = \sum_{i=1}^n \text{Var}X_i + 2 \sum_{i < j} \text{Kov}X_i, X_j. \quad (7.33)$$

(iii) Sind X und Y stochastisch unabhängig, dann gilt $\text{Kov}(X, Y) = 0$.

7.181

Folgerung 7.23 (Eigenschaften von $\text{korr}(X, Y)$). Für die ZV $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ gelte $EX^2 < \infty$ und $EY^2 < \infty$, sowie $\text{Var}X \neq 0$ und $\text{Var}Y \neq 0$. Dann gilt:

a) Die mittlere quadratische Abweichung zwischen ZV Y und $a + bX$ ist minimal, wenn

$$b = \text{korr}(X, Y) \frac{\text{Str}Y}{\text{Str}X} \text{ und } a = EY - bEX. \quad (7.34)$$

Der Minimalwert ist dann

$$\text{Var}Y[1 - (\text{korr}(X, Y))^2].$$

b) Es gilt stets

$$-1 \leq \text{korr}(X, Y) \leq 1. \quad (7.35)$$

c) Es gilt

$$\text{korr}(X, Y) = \pm 1 \iff Y = a + bX.$$

Dann gilt auch $\text{sign}b = \text{sign}\text{korr}(X, Y)$.

7.182

7.7 Momente

Definition 7.24. X sei eine Zufallsvariable und $k \in \mathbb{N}$. Dann heißt der Erwartungswert der k -ten Potenz von X

$$m_k := E\left(X^k\right) \quad (7.36)$$

Moment der Ordnung k von X oder kürzer k -tes Moment von X (sofern der Erwartungswert existiert).

Der Erwartungswert der k -ten Potenz des Absolutbetrages $|X|$ von X

$$M_k := E\left(|X|^k\right) \quad (7.37)$$

heißt **k -tes absolutes Moment von X** .

Darstellung

Ist X eine reelle ZV mit der Dichte f^X und Verteilungsfunktion F^X , dann folgt aus der Definition des Erwartungswertes

$$m_k^X = \int_{-\infty}^{\infty} x^k dF^X(x) = \int_{-\infty}^{\infty} x^k f^X(x) dx. \quad (7.38)$$

7.183

Definition 7.25. X sei eine Zufallsvariable mit $\mu = EX$ und $k \in \mathbb{N}$. Dann heißt

$$\mu_k := E\left((X - \mu)^k\right) \quad (7.39)$$

zentrales Moment der Ordnung k von X und

$$\bar{\mu}_k := E\left(|X - \mu|^k\right) \quad (7.40)$$

heißt **k -tes absolutes zentrales Moment von X** .

Bemerkung

Das dritte zentrale Moment heißt **Schiefe**, das normierte vierte zentrale Moment $\frac{\mu_4(X)}{\sigma^2}$ heißt **Wölbung**.

7.184

Definition 7.26 (Momenterzeugende Funktion). Sei X eine ZV mit stetiger R-Dichte $f(x)$, dann ist die **momenterzeugende Funktion** durch

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2}{2!} x^2 + \dots\right) f(x) dx \\ &= 1 + tm_1^X + \frac{t^2}{2!} m_2^X + \dots \end{aligned}$$

gegeben, wobei m_i^X das i -te Moment von X ist.

Der Ausdruck $M_X(-t)$ ist die zweiseitige Laplacetransformation des durch X festgelegten Wahrscheinlichkeitsmaßes.

7.185

Satz 7.27 (Eigenschaften von M_X). *Die momenterzeugende Funktion hat folgende Eigenschaften:*

$$(i) \quad M_X(0) = 1, M'_X(0) = EX, M''_X(0) = EX^2.$$

$$(ii) \quad M_{aX+b}(t) = M_X(at)e^{tb}, \quad a > 0, b \in \mathbb{R}.$$

$$(iii) \quad M_{X+Y}(t) = M_X(t)M_Y(t), \text{ falls } X, Y \text{ stoch.unabh.}$$

$$(iv) \quad M_{X_n}(t) \rightarrow M_X(t) \Leftrightarrow X_n \xrightarrow{V} X \text{ für } n \rightarrow \infty.$$

$$(v) \quad Y \sim \mathcal{N}(0, 1) \Leftrightarrow M_Y(t) = e^{\frac{t^2}{2}}.$$

7.186

7.8 Mehrdimensionale Normalverteilung

Standardnormalverteilung im \mathbb{R}^n

Ist $X : \Omega \rightarrow \mathbb{R}^n$ standard-normalverteilt, dann besitzt X die Produkt-R-Dichte

$$f^X(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)} = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}\mathbf{x}^T \mathbf{x}} \quad (7.41)$$

mit $\mathbf{x} \in \mathbb{R}^n$.

7.187

Lineare Transformation

Gegeben sei die Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}^n, x \mapsto \mathbf{a} + \mathbf{A}\mathbf{x}$ mit $\mathbf{a} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{n \times n}$ und $\det \mathbf{A} \neq 0$. X sei eine standardnormalverteilte ZV, die mittels g in die ZV Y transformiert wird:

$$Y = g(X) = \mathbf{a} + \mathbf{A}X. \quad (7.42)$$

Es gilt:

$$(i) \quad EY_i = a_i.$$

(ii) Für die Kovarianz gilt mit $EX_i^2 = 1$ und $EX_k X_l = 0, (k \neq l)$ Folgendes:

$$\begin{aligned} \text{Kov}(Y_i, Y_j) &= E(Y_i - EY_i)(Y_j - EY_j) \\ &= E\left(\sum_{k=1}^n a_{ik}X_k\right)\left(\sum_{l=1}^n a_{jl}X_l\right) \\ &= \sum_{k=1}^n a_{ik}a_{jk} = \mathbf{A}\mathbf{A}^T. \end{aligned} \quad (7.43)$$

Notation

Mit $k_{ij} := \text{Kov}(Y_i, Y_j)$ und $k_{ii} := \text{Var } Y_i$ gilt

$$\mathbf{K} = \mathbf{A}\mathbf{A}^T.$$

Dichte für $Y = \mathbf{a} + \mathbf{A}X$

$$f^Y(y_1, \dots, y_n) = \frac{1}{|\det \mathbf{A}|} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}(\mathbf{y}-\mathbf{a})^T (\mathbf{A}^{-1})^T (\mathbf{A}^{-1})(\mathbf{y}-\mathbf{a})} \quad (7.44)$$

bzw.

$$f^Y(\mathbf{y}) = \frac{1}{\sqrt{|\det \mathbf{K}|}} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}(\mathbf{y}-\mathbf{a})^T (\mathbf{K}^{-1})(\mathbf{y}-\mathbf{a})}, \quad \mathbf{y} \in \mathbb{R}^n. \quad (7.45)$$

Definition 7.28 (n -dimensionale Normalverteilung). Das W-Maß über $(\mathbb{R}^n, \mathbb{B}^n)$, definiert mit $\mathbf{a} \in \mathbb{R}^n$ und $\mathbf{K} \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit durch die R-Dichte

$$f^Y(\mathbf{y}) = \frac{1}{\sqrt{|\det \mathbf{K}|}} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2}(\mathbf{y}-\mathbf{a})^T (\mathbf{K}^{-1})(\mathbf{y}-\mathbf{a})}, \quad \mathbf{y} \in \mathbb{R}^n, \quad (7.46)$$

heißt **n -dimensionale Normalverteilung** und wird mit $\mathcal{N}(\mathbf{a}, \mathbf{K})$ bezeichnet.

\mathbf{a} bezeichnet den **Erwartungsvektor** und \mathbf{K} die Kovarianzmatrix. Die n -dimensionale Standardnormalverteilung wird mit $\mathcal{N}(0, I_n)$ bezeichnet, wobei I_n die n -dimensionale Einheitsmatrix ist.

Folgerung 7.29. (a) Ist die ZV $X \mathcal{N}(0, I_n)$ -verteilt und ist $Y = \mathbf{a} + \mathbf{A}X$, $\mathbf{a} \in \mathbb{R}^n$ und $\mathbf{A} \in \mathbb{R}^{n \times n}$ regulär, dann ist $Y \mathcal{N}(\mathbf{a}, \mathbf{A}\mathbf{A}^T)$ -verteilt. Zu jeder $\mathcal{N}(\mathbf{a}, \mathbf{K})$ -verteilten ZV Y gibt es ein $\mathbf{a} \in \mathbb{R}^n$ und eine untere Dreiecksmatrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ mit $\mathbf{K} = \mathbf{A}\mathbf{A}^T$, so dass gilt:

$$Y = \mathbf{a} + \mathbf{A}X \text{ und } X \mathcal{N}(0, I_n) - \text{verteilt.}$$

(b) Ist $Y \mathcal{N}(\mathbf{a}, \mathbf{K})$ -verteilt und $Z = \mathbf{b} + \mathbf{B}Y$ mit $\mathbf{b} \in \mathbb{R}^n$ und $\mathbf{B} \in \mathbb{R}^{n \times n}$ regulär, dann ist $Z \mathcal{N}(\mathbf{b} + \mathbf{B}\mathbf{a}, \mathbf{B}\mathbf{K}\mathbf{B}^T)$ -verteilt.

(c) Alle Randverteilungen einer n -dimensionalen Normalverteilung sind wieder Normalverteilungen. Dies gilt auch für k -dimensionale Normalverteilungen mit $k < n$. Die Parameter a_i und k_{ij} bleiben unverändert. Für $Y = (Y_1, \dots, Y_n) \mathcal{N}(\mathbf{a}, \mathbf{K})$ -verteilt, sind die $Y_i \mathcal{N}(a_i, k_{ii})$ -verteilt.

(d) (b) gilt auch für $\mathbf{b} \in \mathbb{R}^m$ ($m < n$) und $\mathbf{B} \in \mathbb{R}^{m \times n}$ mit Vollrang. Z ist dann eine m -dimensionale ZV. Für $m = 1$, $\mathbf{b} = 0$ und $\mathbf{B} = (1, \dots, 1)$ gilt:

$$\sum_{i=1}^n Y_i \text{ ist } \mathcal{N}\left(\sum_{i=1}^n a_i, \sum_{i,j=1}^n k_{i,j}\right) - \text{verteilt.}$$

(e) Die ZVen Y_1, \dots, Y_n sind genau dann stochastisch unabhängig und $\mathcal{N}(a_i, \sigma_i^2)$ -verteilt, wenn $Y \mathcal{N}(\mathbf{a}, \mathbf{K})$ -verteilt ist mit $\mathbf{K} = \text{diag}(\sigma_i^2)$.

7.9 Zufällige Summen und bedingte Erwartungswerte

Definition 7.30 (Zufällige Summen). Es sei Y eine ZV mit Werten in \mathbb{N}_0 . X_1, X_2, \dots seien reellwertige ZV, identisch verteilt und stochastisch unabhängig, auch von Y . Dann heißt die ZV

$$S = \sum_{i=1}^Y X_i$$

mit zufälliger oberer Grenze eine **zufällige Summe**.

Satz 7.31. Für die zufällige Summe $S = \sum_{i=1}^Y X_i$ gilt, falls $EY < \infty$ und $EX_i < \infty$,

$$ES = EY \cdot EX_1, \quad (7.47)$$

$$\text{Var } S = EY \cdot \text{Var } X_1 + \text{Var } Y \cdot (EX_1)^2. \quad (7.48)$$

7.192

Bedingte Verteilung $P(X \in B | Y = y)$

Die Formel für die bedingte Wahrscheinlichkeit lautet:

$$P^{X|Y}(B|y) = P(X \in B | Y = y) = \frac{P[(X \in B) \cap (Y = y)]}{P(Y = y)}. \quad (7.49)$$

Sind X und Y Zufallsvariablen und $P^{X|Y}(B|y) = P(X \in B | Y = y)$ die bedingte Verteilung von X unter $Y = y$, so kann man natürlich den Mittelwert dieser Verteilung bilden.

Definition 7.32. Der Erwartungswert

$$E(X | Y = y) = \int x P^{X|Y}(dx|y)$$

der bedingten Verteilung von X unter Y bzw. die Zufallsvariable $E(X | Y)$

$$\omega \mapsto \int x P^{X|Y}(dx | Y(\omega))$$

heißen der **bedingte Erwartungswert** von X unter der Bedingung Y .

7.193

Definition 7.33. Sind $S : \Omega \rightarrow \Omega' \subset \mathbb{R}$ und $Y : \Omega \rightarrow \Omega''$ diskrete Zufallsvariablen und existiert der Erwartungswert ES , dann heißt

$$E(S | Y = n) := \sum_{k \in \Omega'} k \cdot P(S = k | Y = n) \quad (7.50)$$

der **bedingte Erwartungswert von S unter $Y = n$** . Es gilt die Formel vom **iterierten Erwartungswert**

$$ES = \sum_{n \in \Omega''} P(Y = n) E(S | Y = n). \quad (7.51)$$

7.194

7.10 Gesetze der großen Zahlen

Es seien Y und Y_1, Y_2, \dots ZV über (Ω, \mathcal{A}, P) mit Werten in \mathbb{R} .

Definition 7.34. (a) Y_n **konvergiert fast sicher** gegen Y , kurz $Y_n \xrightarrow{\text{f.s.}} Y$, wenn

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1,$$

d.h., wenn höchstens innerhalb einer Ausnahmemenge $N \in \mathcal{A}$ mit $P(N) = 0$ der Grenzwert $\lim_{n \rightarrow \infty} Y_n(\omega)$ nicht existiert oder $\neq Y(\omega)$ ist.

(b) Y_n **konvergiert stochastisch** gegen Y , kurz $Y_n \xrightarrow{\text{st}} Y$, wenn

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0,$$

d.h. für festes $\varepsilon > 0$ und für jedes n darf es eine Ausnahmemenge M_n geben, auf der $|Y_n - Y| \geq \varepsilon$ gilt, aber mit $P(M_n) \rightarrow 0$ für große n .

(c) Y_n **konvergiert im r -ten Mittel** gegen Y , kurz $\left(Y_n \xrightarrow{(r)} Y\right)$, wobei $1 \leq r < \infty$, wenn

$$E|Y_n - Y|^r \rightarrow 0.$$

Für $r = 1$ sagt man Y_n **konvergiert im Mittel**, für $r = 2$ **im quadratischen Mittel**.

(d) Y_n **konvergiert nach Verteilung** gegen Y , kurz $Y_n \xrightarrow{V} Y$, wenn

$$F^{Y_n}(x) \rightarrow F^Y(x) \quad \text{für alle } x \text{ mit „} F^Y \text{ stetig im Punkt } x \text{“}.$$

7.195

Satz 7.35 (Ungleichung von Chebychew-Markow). Für jede ZV $Y : \Omega \rightarrow \mathbb{R}$ und $r \geq 1$, $\varepsilon > 0$ gilt:

$$P(|Y| \geq \varepsilon) \leq \frac{1}{\varepsilon^r} E|Y|^r.$$

Existiert EY , so gilt für $r = 2$

$$P(|Y - EY| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var} Y.$$

Definition 7.36. Für ZV X_1, X_2, \dots mit $EX_i < \infty$ gilt das starke bzw. schwache Gesetz der großen Zahlen, wenn $\frac{1}{n} \sum_{i=1}^n (X_i - EX_i)$ fast sicher bzw. stochastisch gegen 0 konvergiert. Wenn die X_i identisch verteilt sind, heißt das

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{f.s.}} EX_1 \quad \text{bzw.} \quad \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{st}} EX_1.$$

7.196

Satz 7.37. Die ZV X_1, X_2, \dots seien identisch verteilt mit $\text{Var} X_i < \infty$.

(a) Sind die X_i auch stochastisch unabhängig, dann gilt das starke Gesetz der großen Zahlen.

(b) Sind die X_i nur paarweise unkorreliert, d.h. $\text{Kov}(X_i, X_j) = 0 \quad \forall i \neq j$, dann gilt das schwache Gesetz der großen Zahlen.

Satz 7.38 (Zentraler Grenzwertsatz). Sind die ZV X_1, X_2, \dots stochastisch unabhängig und identisch verteilt mit $\text{Str } X_i < \infty$, dann konvergieren die „standardisierten“ Teilsummen nach Verteilung gegen eine $\mathcal{N}(0, 1)$ -verteilte ZV Y , d.h.

$$\frac{S_n - \mathbb{E} S_n}{\text{Str } S_n} = \frac{\sum_{i=1}^n X_i - n \cdot \mathbb{E} X_1}{\sqrt{n} \text{Str } X_1} \xrightarrow{v} Y$$

mit $P^Y = \mathcal{N}(0, 1)$.

<http://134.100.146.165:3838/StatsWebApp/>

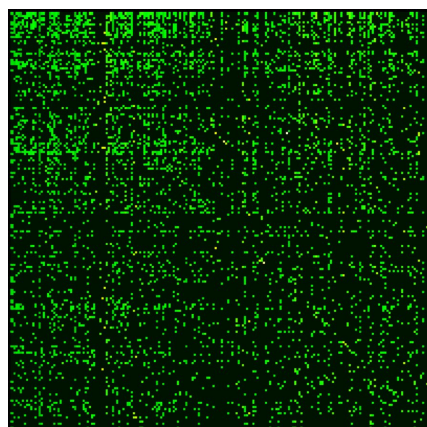
7.197

Kapitel 8

Markow-Ketten

8.1 Vorbemerkungen

Beispiel 8.1 (Google-Page-Rank). Die **Google-Matrix** ist eine quadratische Matrix, die bei der Konstruktion des **Page-Rank-Algorithmus** entsteht.



Zur Berechnung der **Page-Ranks** ist man insbesondere an der Existenz und Vielfachheit von **Linkseigenvektoren** der Matrix interessiert. **Warum ?**

Quelle: Wikipedia 2018

7.198

Definition 8.2 (Stochastischer Prozess). Gegeben sei ein W-Modell (Ω, \mathcal{A}, P) , ein Bildbereich Ω' (Zustandsraum, Zustandsmenge), ein Zeitbereich T (oft $T \subset \mathbb{R}$) und zu jedem $t \in T$ eine ZV $X_t : \Omega \rightarrow \Omega'$, die den Zustand zum Zeitpunkt t angibt. Dann heißt $(X_t) := (X_t, t \in T)$ ein **stochastischer Prozess**.

Satz 8.3 (Satz von Ionescu-Tulcea). Zu einer Folge von Ergebnismengen $\Omega_1, \Omega_2, \dots$, einer Dichte f_1 , und einer Folge von Übergangsdichten f_2^1, f_3^2, \dots gibt es ein W-Modell

$$(\Omega, \mathcal{A}, P) \quad \text{über} \quad \Omega = \bigtimes_{i=1}^{\infty} \Omega_i,$$

so dass die Wahrscheinlichkeiten für alle Ereignisse aus \mathcal{A} , die durch endlich viele Beobachtungen $\omega_1, \dots, \omega_n$ bestimmt sind, in dem entsprechenden n -stufigen Koppelungsmodell berechnet werden können.

7.199

8.2 Grundbegriffe

Definition 8.4 (Markow-Kette, Zustände $i \in I$). Eine **Markow-Kette** ist ein stochastischer Prozess. Die Folge der Beobachtungen X_0, X_1, X_2, \dots in einem unendlich-stufigen Versuch mit Markow-Koppelung und abzählbarer Zustandsmenge I . Die ZV $X_n : \Omega \rightarrow I$ beschreiben also den Zustand eines Systems zu den Zeitpunkten $n = 0, 1, 2, \dots$.

Definition 8.5 (Homogene Markowkette). Eine Markow-Kette (X_n) heißt **homogen**, falls die Übergangswahrscheinlichkeiten

$$f_n^{n-1}(i; j) = P(X_n = j | X_{n-1} = i) \quad n = 1, 2, \dots$$

für alle Zeitpunkte gleich sind.

In diesem Fall wird

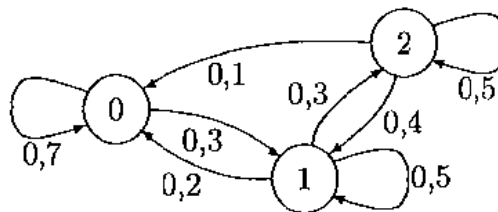
$$p_{ij} := f_n^{n-1}(i; j)$$

gesetzt. $\mathbf{P} := (p_{ij}), (i, j \in I)$ heißt **Übergangsmatrix**.

7.200

Definition 8.6 (Übergangsgraph). Ein **Übergangsgraph** einer homogenen Markow-Kette besteht aus allen möglichen Zuständen als Knoten des Graphs und den mit *positiver Wahrscheinlichkeit* möglichen Übergängen als Kanten. An der Kante von i nach j wird der Wert p_{ij} notiert.

Beispiel 8.7. Der Graph zum Beispiel der Belegung von Telefonleitungen eines Anschlusses:

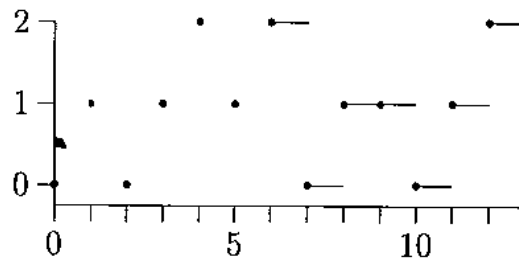


Quelle: Hübner, G.: Stochastik. 5. Aufl., Vieweg+Teubner, 2009

7.201

Definition 8.8 (Pfad einer Markow-Kette). Ein einzelner Verlauf einer Markow-Kette für einen festen Wert ω , also $(X_0(\omega), X_1(\omega), \dots)$ heißt **Pfad** der Markow-Kette.

Beispiel 8.9. Ein möglicher Pfad zum Beispiel der Belegung von Telefonleitungen eines Anschlusses:



Quelle: Hübner, G.: Stochastik. 5. Aufl., Vieweg+Teubner, 2009

7.202

Folgerung 8.10. Für eine homogene Markow-Kette mit der Übergangsmatrix $(p_{ij})_{i,j \in I}$ und der Startverteilung $(P(X_0 = i), i \in I)$ gilt

$$P(X_0 = i_0, \dots, X_n = i_n) = P(X_0 = i_0) p_{i_0 i_1} \cdots p_{i_{n-1} i_n}, \quad (8.1)$$

$$P(X_n = j) = \sum_{i \in I} P(X_{n-1} = i) p_{ij} \quad \text{bzw.} \quad \mathbf{p}_n = \mathbf{p}_{n-1} \mathbf{P}. \quad (8.2)$$

Bemerkung

Wahrscheinlichkeit zum Zeitpunkt n im Zustand j zu sein

ergibt sich aus den

Wahrscheinlichkeiten zum Zeitpunkt $n - 1$ in den Zuständen $i \in I$ zu sein

und der jeweiligen

Übergangswahrscheinlichkeit $i \rightarrow j$ gegeben durch p_{ij} .

7.203

Definition 8.11 (n -Schritt Übergangsmatrix). Für eine homogene Markow-Kette (X_n) ist die Matrix $\mathbf{P}^{(n)} = (p_{ij}^{(n)})$ mit $p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i)$ unabhängig von m und heißt **n -Schritt-Übergangsmatrix**. Es gilt mit $\mathbf{P}^{(1)} = \mathbf{P}$:

$$p_{ij}^{(n)} = \sum_{k \in I} p_{ik}^{(n-1)} p_{kj} \quad \text{bzw.} \quad \mathbf{P}^{(n)} = \mathbf{P}^{(n-1)} \mathbf{P} = \mathbf{P}^n. \quad (8.3)$$

Die Folgerung

$$p_{ij}^{(n+m)} = \sum_{k \in I} p_{ik}^{(n)} p_{kj}^{(m)} \quad \text{bzw.} \quad \mathbf{P}^{(n+m)} = \mathbf{P}^n \mathbf{P}^m \quad (8.4)$$

ist auch als **Chapman-Kolmogorov-Gleichung** bekannt.

7.204

Definition 8.12. Die Zustandsmenge I einer homogenen Markow-Kette wird nach folgender Vorschrift in disjunkte Klassen zerlegt:

Die Zustände i und j gehören zur selben Klasse, wenn

- (i) $i = j$ oder
- (ii) wenn sowohl j von i aus als auch i von j aus in endliche vielen Schritten erreicht werden kann.

[0pt] Jeder Zustand gehört zu genau einer Klasse K . [10pt] Eine homogene Markow-Kette heißt **irreduzibel**, falls alle Zustände zu einer Klasse gehören.

7.205

Definition 8.13. Eine Klasse K heißt **periodisch** mit der **Periode** d , wenn es $d(\geq 2)$ Teilmengen in K gibt, die der Reihe nach in d Schritten durchlaufen werden. Eine homogene Markow-Kette heißt **aperiodisch**, wenn es keine periodische Klasse gibt.

Bemerkung

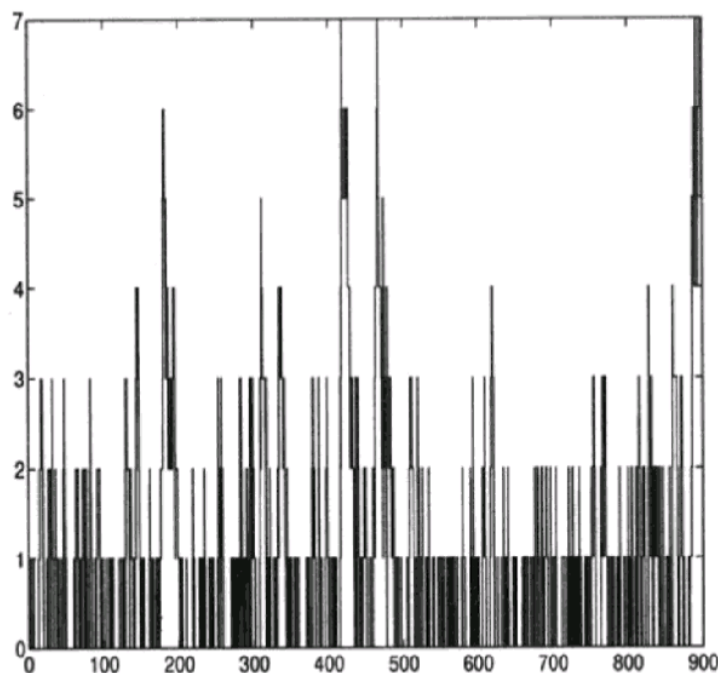
Startet man in einem Zustand mit Periode $d(i) > 1$, so kann das System höchstens zu den Zeitpunkten $nd(i)$ mit $n \in \mathbb{N}$ in diesen Zustand zurückkehren.

Bemerkung

In einer irreduziblen Markow-Kette besitzen alle Zustände dieselbe Periode. Die Markow-Kette ist also immer periodisch oder aperiodisch.

7.206

8.3 Gleichgewicht



Quelle: Hübner, G.: Stochastik. 5.Aufl., Vieweg+Teubner, 2009

7.207

Definition 8.14. (a) Eine homogene Markow-Kette (X_n) ist **im Gleichgewicht**, wenn für alle Zustände $i \in I$ die Wahrscheinlichkeiten $P(X_n = i)$ unabhängig vom Zeitpunkt n sind. Dann wird $\pi_i = P(X_n = i)$ bzw. $\pi = \mathbf{p}_n$ gesetzt. Die Dichte $\pi = (\pi_i, i \in I)$ heißt **Gleichgewichtsverteilung** der homogenen Markow-Kette (X_n) .

(b) Allgemein heißt π eine Gleichgewichtsverteilung (zur Übergangsmatrix \mathbf{P}), wenn eine aus (π_i) als Startverteilung und der Übergangsmatrix \mathbf{P} konstruierte homogene Markow-Kette im Gleichgewicht ist.

7.208

Satz 8.15. (a) Die homogene Markow-Kette (X_n) mit der Übergangsmatrix \mathbf{P} sei im Gleichgewicht, d.h. es gelte $P(X_n = j) = \pi_j$ bzw. $\mathbf{p} = \pi$ für alle $n = 0, 1, 2, \dots$ und $i \in I$. Wegen (8.2)

$$P(X_n = j) = \sum_{i \in I} P(X_{n-1} = i) \cdot p_{ij}$$

gelten für die Werte π_i ($i \in I$) die folgenden Gleichgewichtsbedingungen

$$\pi_j = \sum_{i \in I} \pi_i p_{ij} \quad \text{für alle } j \in I \quad \text{bzw.} \quad \pi = \pi P, \quad (\text{G})$$

$$\pi_j \geq 0 \quad \text{für alle } j \in I \quad \text{und} \quad \sum_{j \in I} \pi_j = 1. \quad (\text{N})$$

(b) Gelten die Gleichungen (G) und (N) für π , dann ist die homogene Markow-Kette mit Startverteilung $(\pi_j, j \in I)$ mit der Übergangsmatrix $(p_{ij})_{i,j \in I}$ im Gleichgewicht.

(c) Zur Übergangsmatrix $(p_{ij})_{i,j \in I}$ gibt es genau dann keine / genau eine / mehrere Gleichgewichtslösung(en), wenn die Bedingungen (G) und (N) keine / genau eine / mehrere Lösung(en) besitzen.

7.209

Satz 8.16. (a) Wird die Zustandsmenge I einer homogenen Markow-Kette in K und K^c zerlegt, dann folgt aus (G)

$$\sum_{i \in K} \sum_{j \in K^c} \pi_i p_{ij} = \sum_{i \in K} \sum_{j \in K^c} \pi_j p_{ji}. \quad (8.5)$$

(b) Zerteilt ein Schnitt zwischen den Zuständen i und j den Übergangsgraphen in zwei (disjunkte) Teile, dann folgt aus (G) die **lokale Gleichgewichtsbedingung** L_{ij}

$$\pi_i p_{ij} = \pi_j p_{ji}. \quad (8.6)$$

(c) Ist der Übergangsgraph einfach zusammenhängend, d.h. zerlegt ihn jeder Schnitt zwischen zwei verbundenen Zuständen in zwei Teile, dann ist (G) äquivalent zu

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j \in I. \quad (\text{L})$$

7.210

Beispiel 8.17 (Warteschlange in Router). Mit Wahrscheinlichkeit p kommt in einem Zeitintervall ein neues Datenpaket an. Falls in einem Zeitintervall kein Datenpaket ankommt, schickt der Router ein Datenpaket weiter. Mit Wahrscheinlichkeit q ist dies erfolgreich. Die Warteschlange kann $m - 1$ Datenpakete abspeichern.

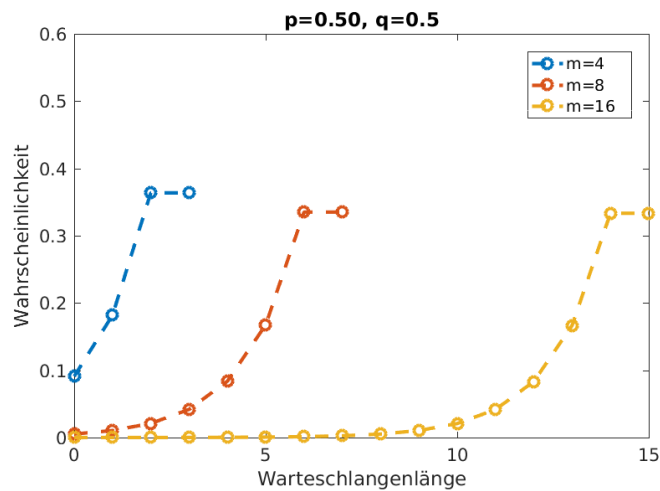
Dank an Lukas Pflug für die Aufbereitung des Beispiels.

Übergangsmatrix für $m = 4$:

$$\mathbf{P} = \begin{pmatrix} (1-p) & p & 0 & 0 \\ (1-p)q & (1-p)(1-q) & p & 0 \\ 0 & (1-p)q & (1-p)(1-q) & p \\ 0 & 0 & q & (1-q) \end{pmatrix}$$

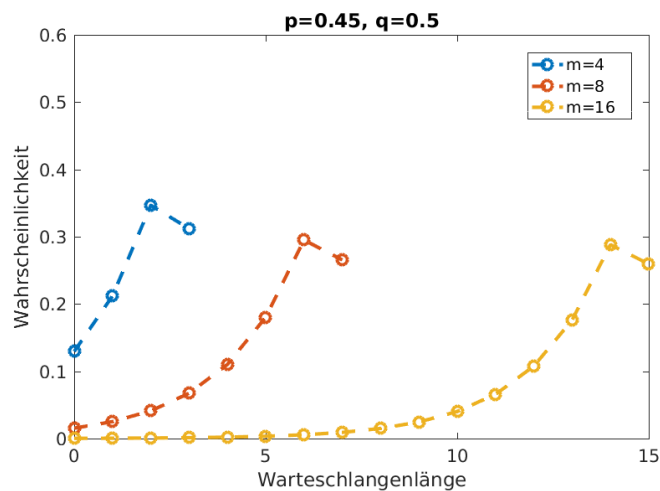
7.211

Beispiel 8.18 (Warteschlange in Router).



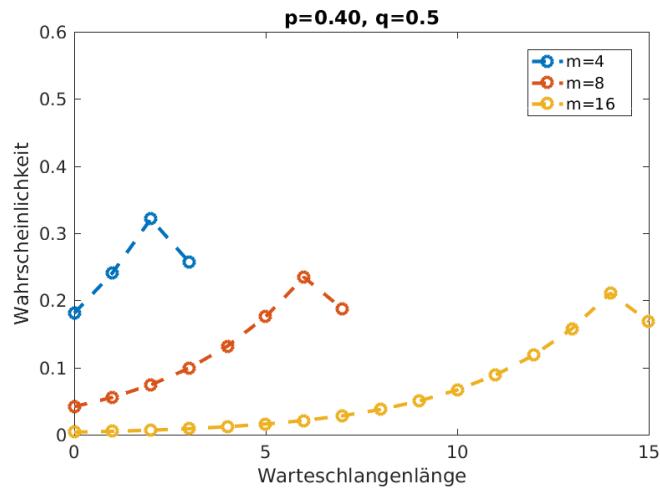
7.212

Beispiel 8.19 (Warteschlange in Router).



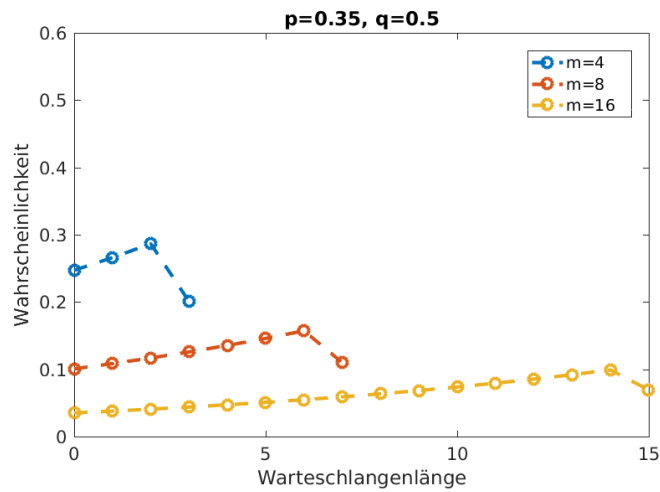
7.213

Beispiel 8.20 (Warteschlange in Router).



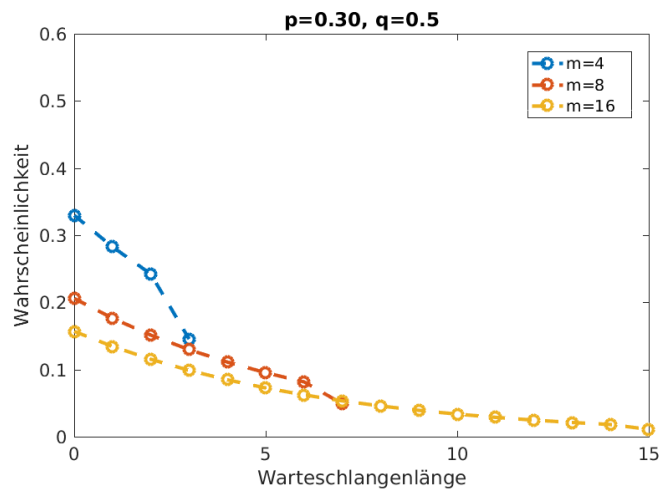
7.214

Beispiel 8.21 (Warteschlange in Router).



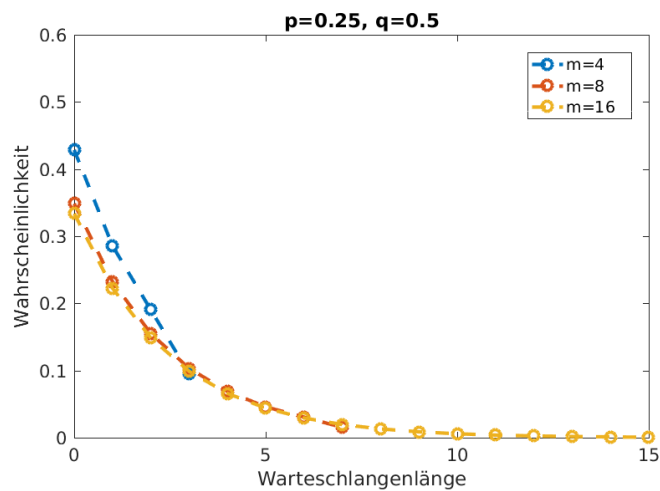
7.215

Beispiel 8.22 (Warteschlange in Router).



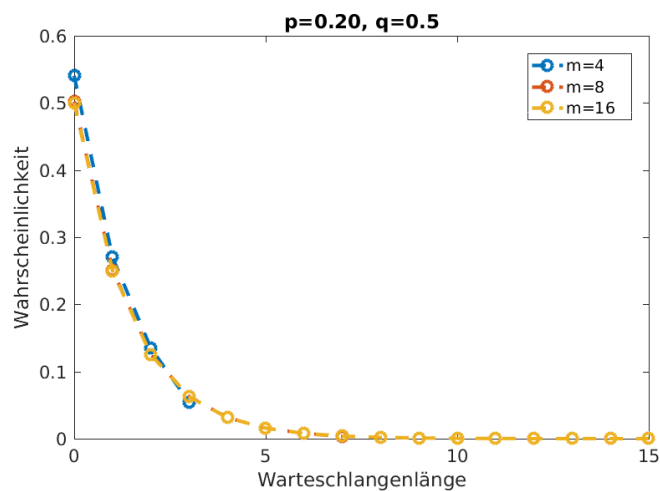
7.216

Beispiel 8.23 (Warteschlange in Router).



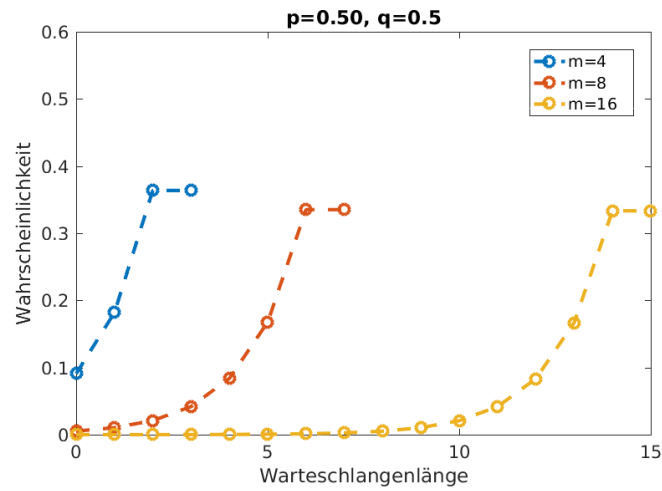
7.217

Beispiel 8.24 (Warteschlange in Router).



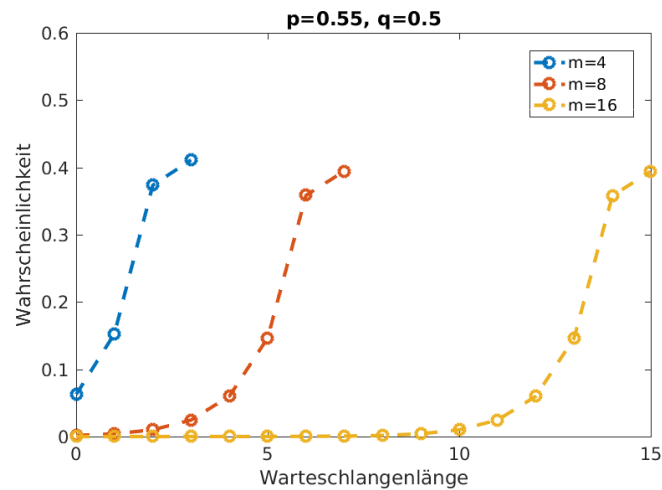
7.218

Beispiel 8.25 (Warteschlange in Router).



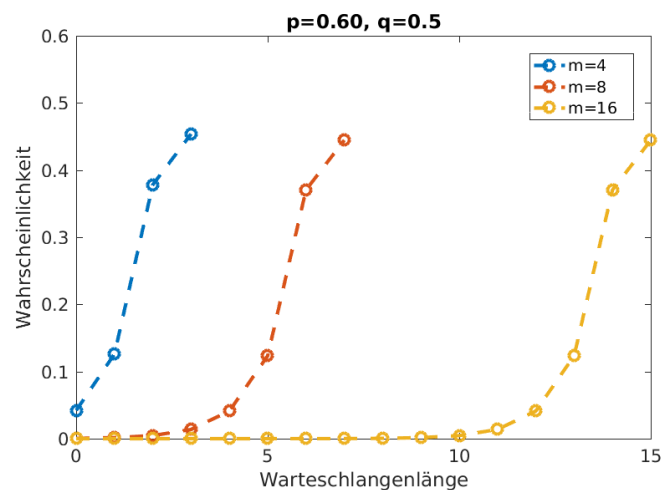
7.219

Beispiel 8.26 (Warteschlange in Router).



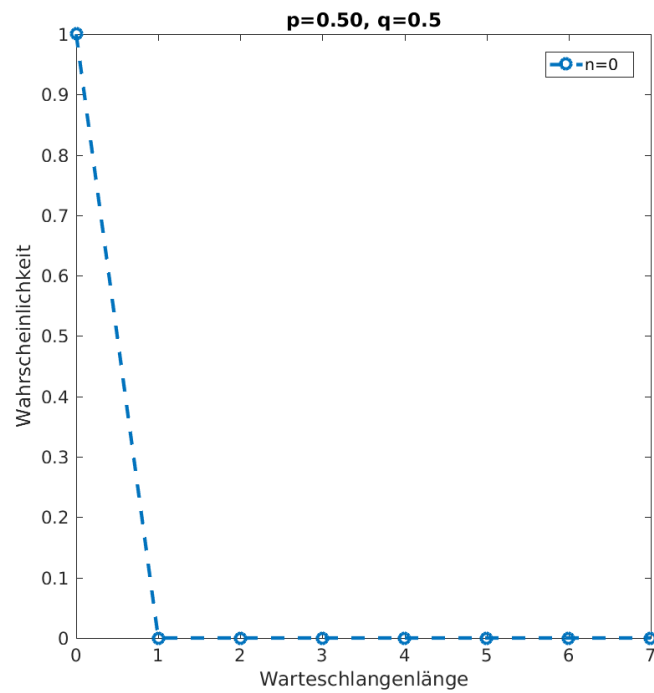
7.220

Beispiel 8.27 (Warteschlange in Router).



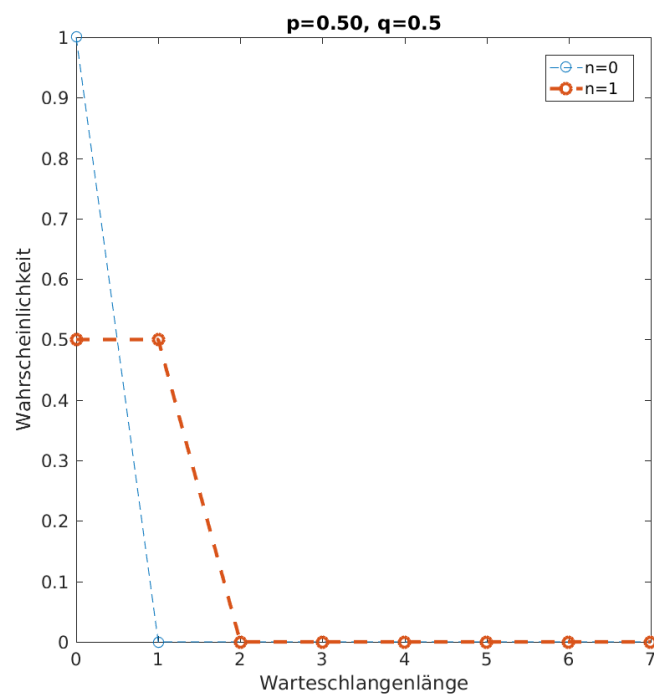
7.221

Beispiel 8.28 (Warteschlange in Router).



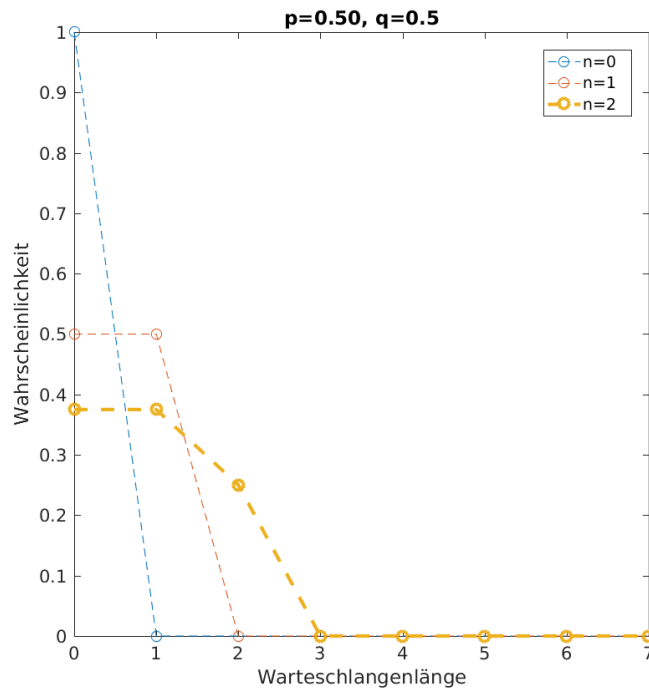
7.222

Beispiel 8.29 (Warteschlange in Router).



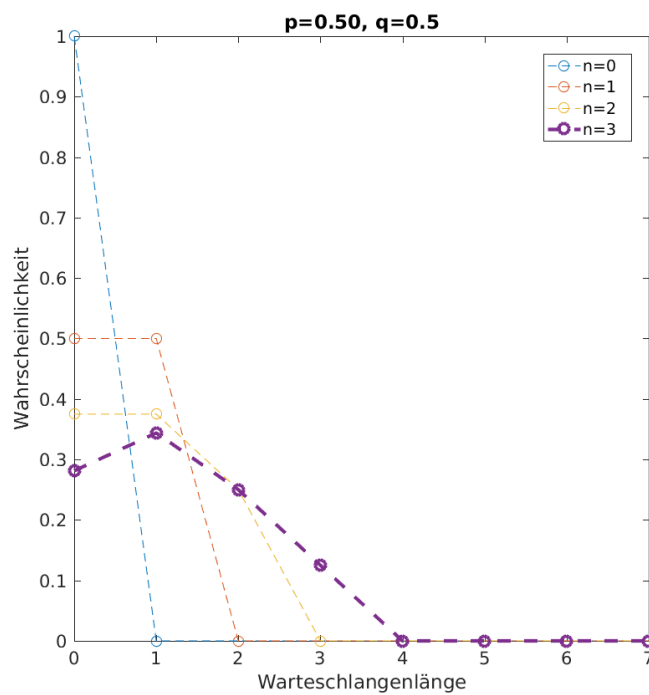
7.223

Beispiel 8.30 (Warteschlange in Router).



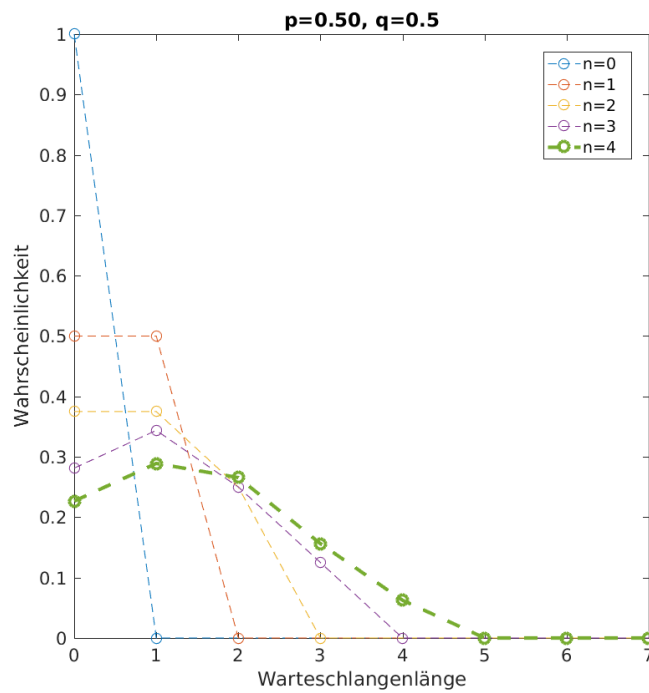
7.224

Beispiel 8.31 (Warteschlange in Router).



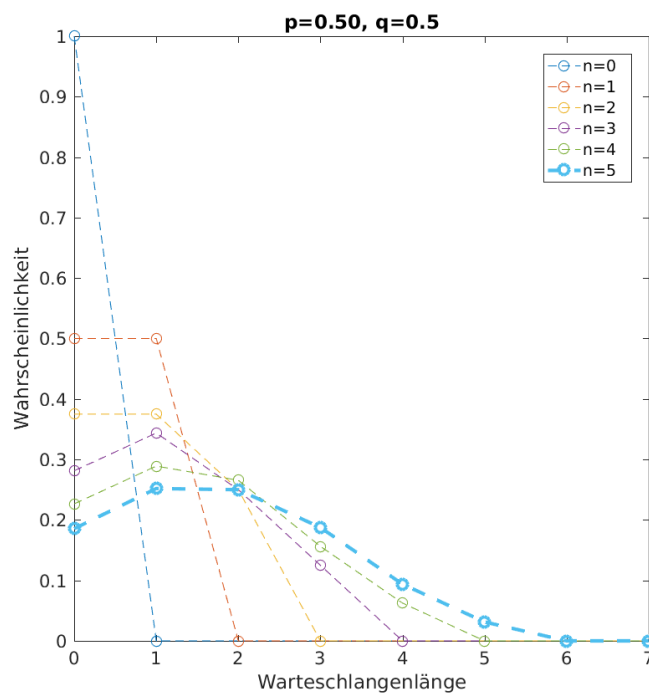
7.225

Beispiel 8.32 (Warteschlange in Router).



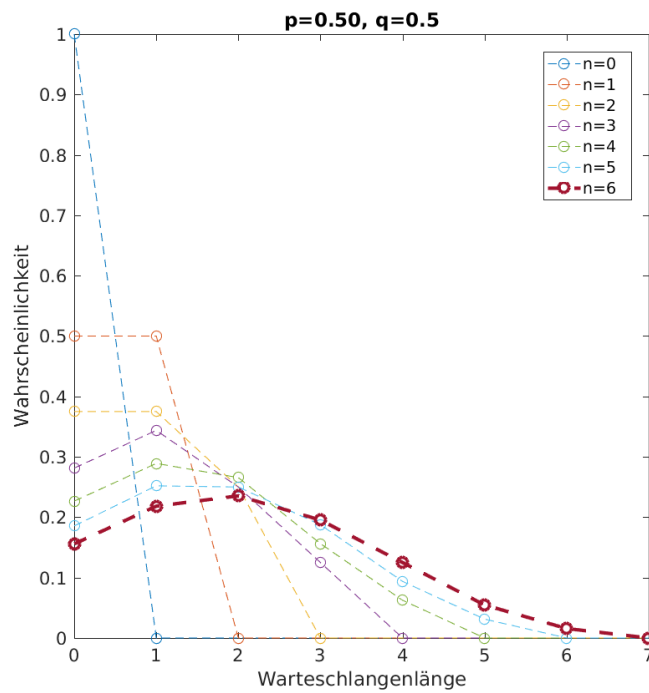
7.226

Beispiel 8.33 (Warteschlange in Router).



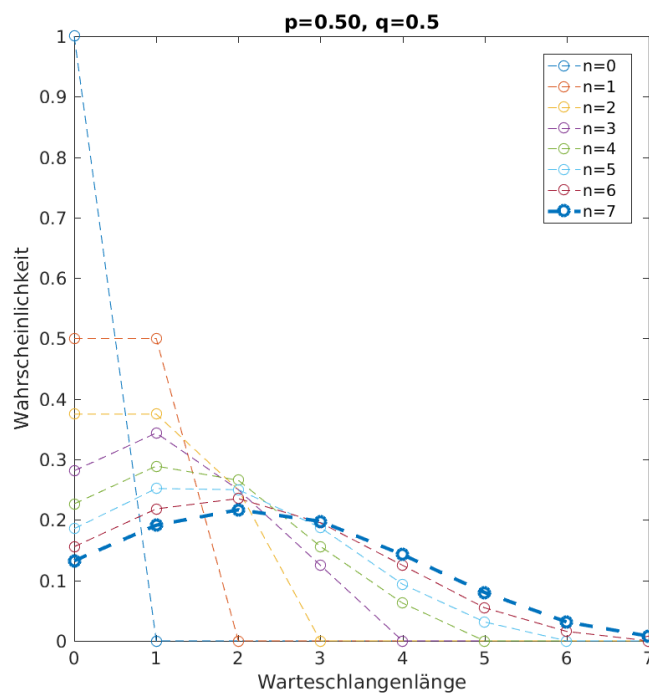
7.227

Beispiel 8.34 (Warteschlange in Router).



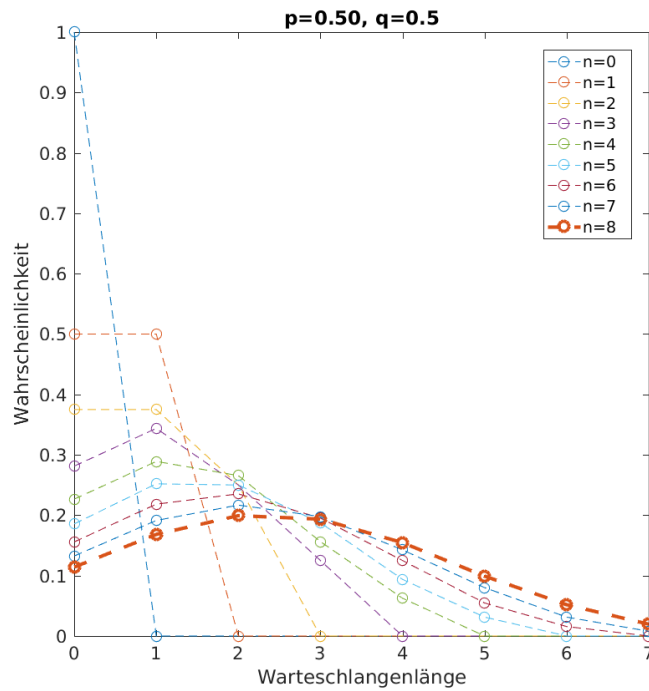
7.228

Beispiel 8.35 (Warteschlange in Router).



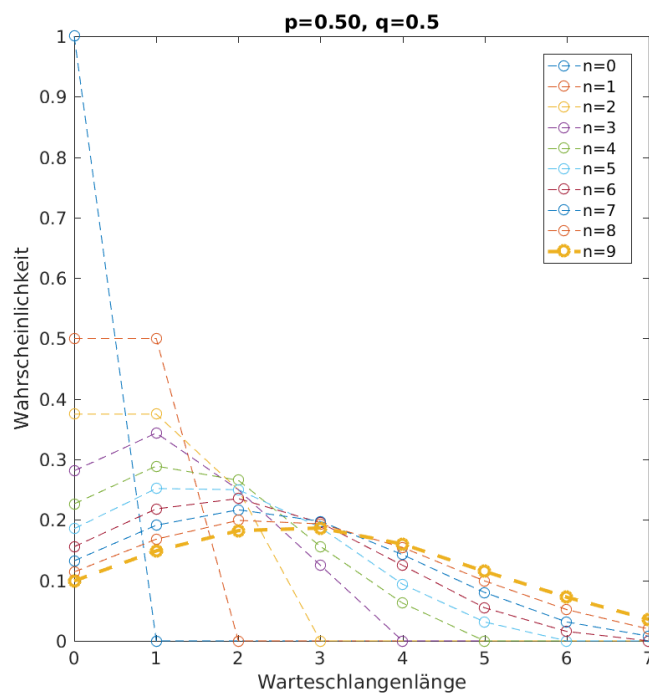
7.229

Beispiel 8.36 (Warteschlange in Router).



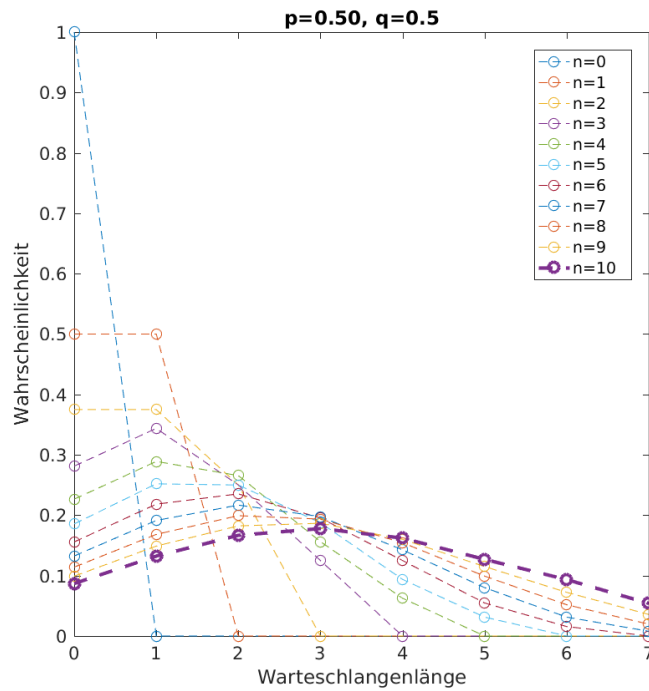
7.230

Beispiel 8.37 (Warteschlange in Router).



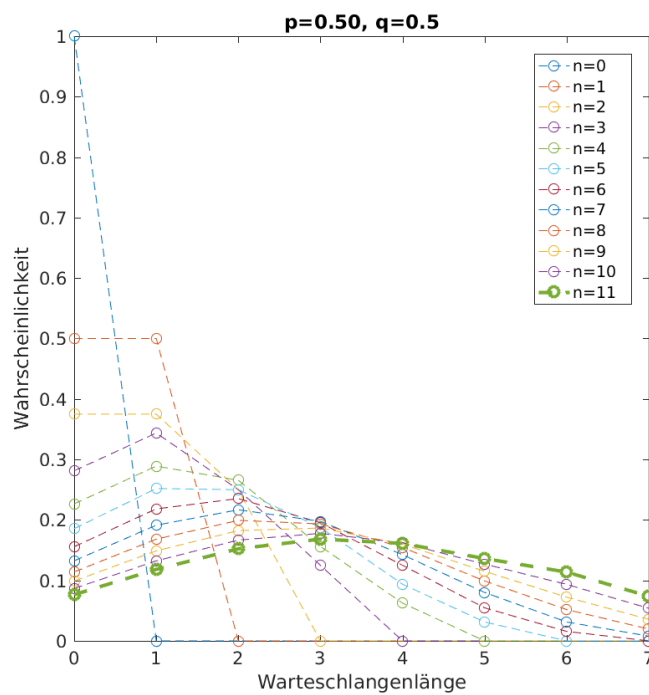
7.231

Beispiel 8.38 (Warteschlange in Router).



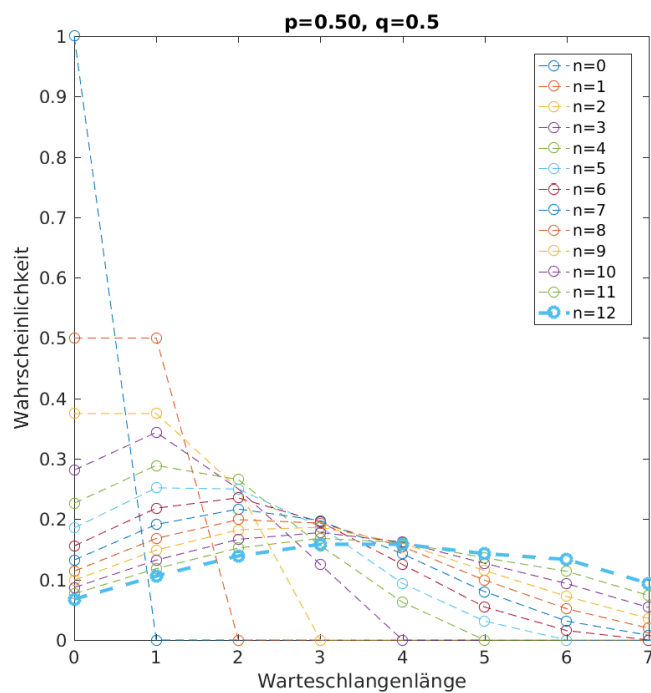
7.232

Beispiel 8.39 (Warteschlange in Router).



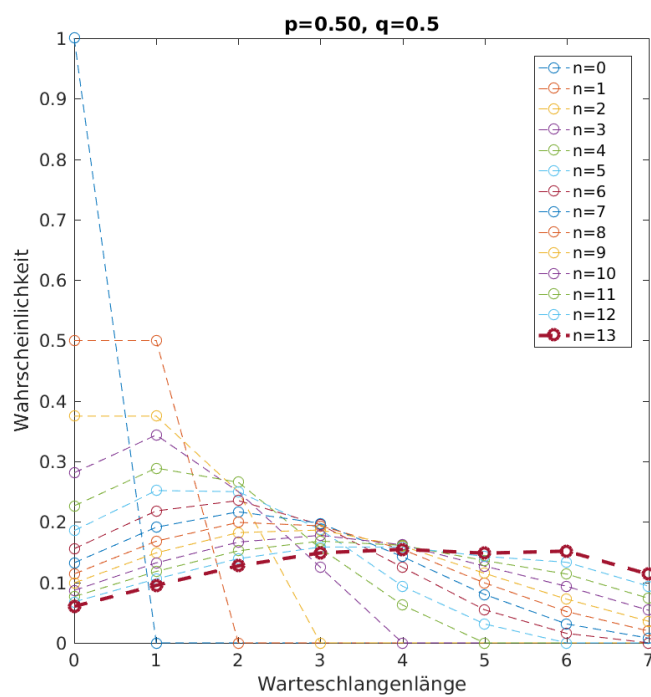
7.233

Beispiel 8.40 (Warteschlange in Router).



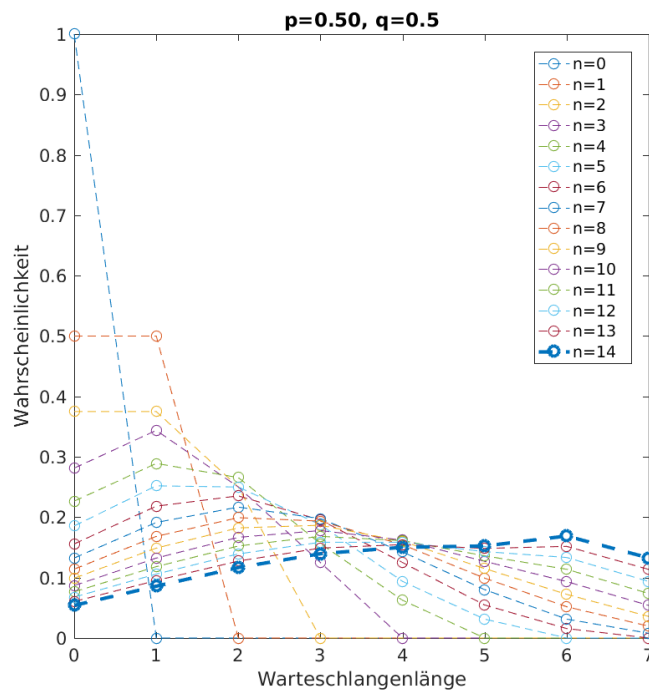
7.234

Beispiel 8.41 (Warteschlange in Router).



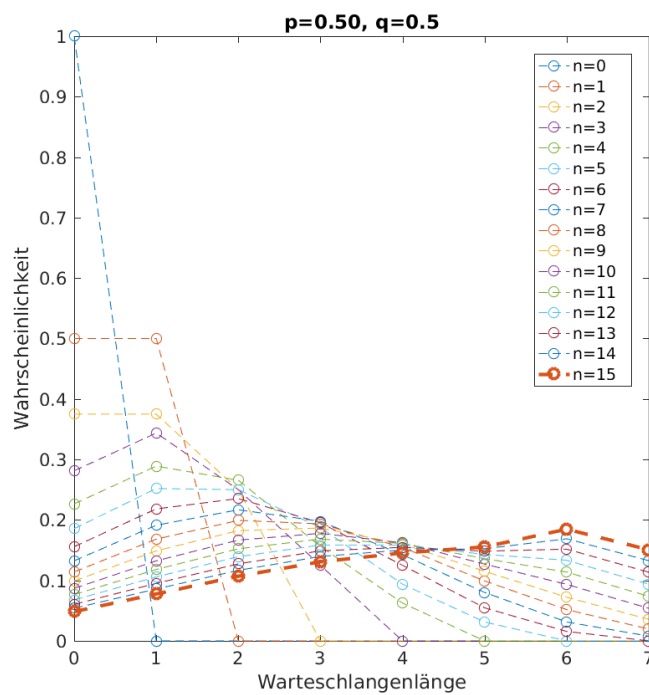
7.235

Beispiel 8.42 (Warteschlange in Router).



7.236

Beispiel 8.43 (Warteschlange in Router).



7.237

Bemerkung

Irreduzible Markow-Kette mit endlichem Zustandsraum besitzen immer mindestens eine stationäre Verteilung. Diese entspricht genau dem normierten Linkseigenvektor zum Eigenwert 1 der Übergangsmatrix \mathbf{P}

Bemerkung

7.238

Satz 8.44 (Grenzwertsatz). *Ist die homogene Markow-Kette (X_n) mit der Übergangsmatrix $(p_{ij})_{i,j \in I}$ **irreduzibel** und **aperiodisch**, dann konvergiert $P(X_n = i)$ für alle $i \in I$ unabhängig von der Startverteilung gegen einen Wert π_i mit $0 \leq \pi_i \leq 1$. Dabei sind*

- ***entweder** alle $\pi_i = 0$ und es gibt keine Gleichgewichtsverteilung zu (p_{ij}) ,*
- ***oder** es sind alle $\pi_i > 0$ und (π_i) ist die einzige Gleichgewichtsverteilung zu (p_{ij}) . Die Lösung wird aus (G) und (N) bestimmt.*

7.239

Bemerkung

Ein hinreichendes Kriterium für die Irreduzibilität und Aperiodizität von (X_n) entspricht der Tatsache, dass es eine Potenz $m \in \mathbb{N}$ gibt, so dass die Übergangsmatrix \mathbf{P}^m nur strikt positive Einträge enthält. Dieses Kriterium ist für die praktische Anwendung (Übungsaufgaben) sehr hilfreich.

Beispiel 8.45 (Google-Page-Rank). Durch die Berechnung der Linkseigenvektoren des Eigenwertes 1 wird der Gleichgewichtszustand der durch die Google-Matrix beschriebene Markov-Kette berechnet

7.240

Kapitel 9

Grundgesamtheit und Stichprobe

9.1 Rückblick: Beschreibende Statistik

Beispiel 9.1. Anhand von Stichproben sollen die Eigenschaften einer Betonsorte ermittelt werden. Zur Bestimmung der Druckfestigkeit wurden 20 Proben gefertigt und unter gleichen Bedingungen geprüft. Die Messungen der Druckfestigkeit [10^{-1} MPa] ergab folgende Ergebnisse:

183	181	183	180	182	182	185	182	184	179
183	184	180	181	179	180	182	180	181	183

Begriffe

Merkmale	Eigenschaften des Untersuchungsobjekts, X, Y, \dots
Merkmalswerte, Messwerte	Messergebnisse der einzelnen Merkmale, x_i, y_k, \dots
Urliste	Menge der unbearbeiteten Messergebnisse

Erste Auswertung

- Ordnungsstatistik, $x_{[1]}, \dots, x_{[n]}$
- Spannweite, Variationsbreite $R := x_{[n]} - x_{[1]}$
- absolute, relative Häufigkeiten
- Histogramme

9.241

Kenngrößen

- arithmetisches Mittel: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- Modalwert

- geometrisches Mittel: $G := \sqrt[n]{\prod_{i=1}^n x_i}$
 $\log G := \frac{1}{n} \sum_{i=1}^n \log x_i$
- empirische Varianz $s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- Variationskoeffizient $v := \frac{s_x}{\bar{x}} 100\%$.

Beschreibende Statistik bei zwei Merkmalen

Neben den o.g. Kenngrößen werden betrachtet:

- empirische Kovarianz: $s_{XY} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
- empirischer Korrelationskoeffizient: $r_{XY} := \frac{s_{XY}}{s_X s_Y}$

9.242

9.2 Grundgesamtheit und Stichprobe

Definition 9.2. Eine Zufallsvariable X (mit einer bestimmten Wahrscheinlichkeitsverteilung), durch die ein Merkmal X beschrieben wird, heißt **Grundgesamtheit**.

Definition 9.3. Eine Menge von n Realisierungen $\{x_1, x_2, \dots, x_n\}$ einer Grundgesamtheit X heißt **Stichprobe vom Umfang n** . Jede einzelne Realisierung heißt **Element der Stichprobe** oder **Merkmalswert**.

Bezeichnung

Eine Stichprobe wird auch mit $\mathbf{x} = (x_1, x_2, \dots, x_n)$ bezeichnet.

Definition 9.4. Eine Stichprobe $\mathbf{x} = (x_1, x_2, \dots, x_n)$ heißt **unabhängig**, wenn die ZVen X_1, X_2, \dots, X_n stoch.unabh. sind, d.h. wenn für beliebige reelle Zahlen gilt:

$$P(X_1 \leq c_1, \dots, X_n \leq c_n) = P(X_1 \leq c_1) \cdot \dots \cdot P(X_n \leq c_n).$$

9.243

Definition 9.5. Die Stichprobe heißt **einfach**, wenn die ZVen $X_i, (i = 1, \dots, n)$ stoch.unabh. sind und alle dieselbe Verteilungsfunktion F besitzen.

Definition 9.6. Eine n -dimensionale ZV $\mathbf{X} = (X_1, X_2, \dots, X_n)$ heißt **mathematische Stichprobe**, wenn die ZV $X_i (i = 1, \dots, n)$ untereinander unabhängig und identisch entsprechend der Grundgesamtheit X verteilt sind.

Bemerkung

In diesem Kontext ist eine *einfache* Stichprobe eine Realisierung einer *mathematischen* Stichprobe.

9.244

Fragen der mathematischen Statistik

- (i) Schätzen der Parameter der Grundgesamtheit
- (ii) Prüfen von Hypothesen

Ausgangspunkt

Der Ausgangspunkt aller Auswertungen mit Methoden der mathematischen Statistik ist eine konkrete Stichprobe und das mit Hilfe der beschreibenden Statistik aufbereitete Zahlenmaterial. Dabei fordern wir von unserer Stichprobe, dass sie die folgenden Bedingungen erfüllt:

- Sie soll repräsentativ für die Grundgesamtheit sein.
- Die Auswahl von Einheiten des Untersuchungsobjekts soll jeweils gleichwahrscheinlich erfolgen. („Zufallsstichprobe“)

9.245

Auswahlvarianten

- Lotterie
- Nutzen von Zufallszahlen

9.246

Verteilungsfunktion

Die Grundgesamtheit eines Merkmales X ist eine ZV mit einer Verteilungsfunktion $F^X(t)$. Diese ist unbekannt und wird auch *theoretische Verteilungsfunktion* genannt. Zu einer konkreten Stichprobe x_1, x_2, \dots, x_n vom Umfang n wird die *empirische Verteilungsfunktion* bestimmt.

Definition 9.7 (Empirische Verteilungsfunktion). Die Funktion $\hat{F}_n : \mathbb{R} \rightarrow [0, 1]$,

$$t \mapsto \hat{F}_n(t) := \text{Summe der relativen Häufigkeiten aller Merkmalswerte, die kleiner oder gleich } t \text{ sind.}$$

heißt **empirische Verteilungsfunktion** der Stichprobe $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

Formulierung mit Zufallsvariablen

Zur Stichprobe (X_1, X_2, \dots, X_n) wird jedem X_i die ZV

$$Y_i = \begin{cases} 1, & \text{falls } \{X_i < t\} \text{ eintritt} \\ 0, & \text{falls } \{X_i \geq t\} \text{ eintritt} \end{cases} \quad (9.1)$$

zugeordnet. $F_n(t)$ mit der Darstellung

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i$$

ist wieder eine ZV. \hat{F}_n ist eine Realisierung von F_n .

9.247

Satz 9.8 (Satz von Gliwenko). Ist $F_n(t)$ die empirische Verteilungsfunktion der einfachen Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n und $F^X(t)$ die Verteilungsfunktion der Grundgesamtheit X , dann gilt

$$F_n(t) \xrightarrow{fs} F.$$

Rückgriff auf Funktionen von Zufallsvariablen

Die Kenngrößen aus der deskriptiven Statistik (Kapitel 2 und 9.1) lassen sich als Realisierungen von ZV auffassen. Die ZV X_1, X_2, \dots, X_n bilden eine einfache Stichprobe, die als Argumente für *Stichprobenfunktionen* dienen. Beispiele sind:

9.248

Stichprobenfunktionen

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{arithmetisches Mittel} \quad (9.2)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{empirische Varianz} \quad (9.3)$$

9.249

χ^2 -Verteilung mit n Freiheitsgraden

Die Komponenten des Zufallsvektors (X_1, \dots, X_n) seien standardnormalverteilt, d.h. $(X_i) \sim \mathcal{N}(0, 1)$. Dann besitzt

$$Y_n = \sum_{i=1}^n X_i^2 \quad (9.4)$$

die Dichtefunktion

$$f^{Y_n}(t) = \begin{cases} \frac{t^{\frac{n}{2}-1} e^{-\frac{t}{2}}}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})}, & \text{für } t > 0, \\ 0, & \text{für } t \leq 0. \end{cases} \quad (9.5)$$

und Erwartungswert und Varianz

$$\mathbb{E} Y_n = n, \quad \text{Var}(Y_n) = 2n.$$

n ist dabei die Anzahl der unabhängigen Summanden, also die *Anzahl der Freiheitsgrade*.

9.250

Definition 9.9. Y_n mit der Dichtefunktion (9.5) unterliegt einer **Chi-Quadrat-Verteilung mit n Freiheitsgraden**, kurz $\chi^2(n)$.

Normierte Summen

Sind die ZV X_1, X_2, \dots, X_n unabhängig und $\mathcal{N}(\mu, \sigma^2)$ -verteilt, so gilt für die ZV

$$Y_{n-1}^* := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1). \quad (9.6)$$

\bar{X} ist wie in (9.2) definiert.

9.251

Student-t-Verteilung, t-Verteilung

Für zwei stoch.unabh. ZV gelte $X \sim \mathcal{N}(0, 1)$, $Y_n \sim \chi^2(n)$. Dann besitzt die ZV

$$Z_n := \frac{X}{\sqrt{\frac{Y_n}{n}}} \quad (9.7)$$

die Dichtefunktion

$$f^{Z_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}. \quad (9.8)$$

Definition 9.10. Eine stetige ZV Z_n mit der Dichtefunktion (9.8) unterliegt einer **Student-t-Verteilung mit n Freiheitsgraden**.

9.252

Normierte Summen

Sind die ZV X_1, X_2, \dots, X_n unabhängig und $\mathcal{N}(\mu, \sigma^2)$ -verteilt, so gilt für die ZV

$$Z_{n-1}^* := \frac{1}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} (\bar{X} - \mu) \sqrt{n} = \frac{(\bar{X} - \mu)}{S} \sqrt{n} \quad (9.9)$$

Z_{n-1}^* unterliegt einer Student-Verteilung mit $(n-1)$ Freiheitsgrade.

9.253

Kapitel 10

Schätzverfahren

10.1 Einleitung

Beispiel 10.1. Langskript: Maßabweichung bei der Produktion eines Maschinenteils

Beispiel 10.2. Langskript: Thema eintragen

Die Grundgesamtheit X ist mit einer Verteilung verteilt, die nicht immer bekannt ist. Nun kann ein Verteilungstyp als bekannt angenommen werden, der Parameter der Verteilung sei allerdings unbekannt.

9.254

10.2 Punktschätzung

10.2.1 Begriffe

Aufgabe

Aufgrund einer stoch.unabh. und identisch verteilten Stichprobe X_1, \dots, X_n soll ein Parameter Θ einer unbekannten Verteilung P^X geschätzt werden.

Begriffe

$\hat{\Theta}(X_1, X_2, \dots, X_n)$ **Schätzfunktion**, Schätzer

$\hat{\theta}(x_1, x_2, \dots, x_n)$ **Schätzwert**

9.255

Veranschaulichung

$$\hat{\Theta}_1(X_1, X_2, \dots, X_n) = \bar{X}(X_1, X_2, \dots, X_n) = \bar{X}$$

$$\hat{\Theta}_2(X_1, X_2, \dots, X_n) = \tilde{X}(X_1, X_2, \dots, X_n) = \tilde{X}$$

sind Schätzfunktionen von $\Theta = EX$ der Grundgesamtheit X .

$$\begin{aligned}\hat{\theta}_1(x_1, x_2, \dots, x_n) &= \bar{x}(x_1, x_2, \dots, x_n) = \bar{x} \\ \hat{\theta}_2(x_1, x_2, \dots, x_n) &= \tilde{x}(x_1, x_2, \dots, x_n) = \tilde{x}\end{aligned}$$

sind Schätzwerte von $\Theta = EX$ der Grundgesamtheit X .

9.256

Definition 10.3. Eine Schätzfunktion $\hat{\Theta}(X_1, X_2, \dots, X_n)$ eines Parameters Θ heißt **erwartungstreu (unverzerrt)**, wenn der $E(\hat{\Theta}) = \Theta$ ist.

Eine Schätzfunktion $\hat{\Theta}$ eines Parameters Θ heißt **asymptotisch erwartungstreu**, falls für wachsenden Stichprobenumfang der Grenzwert des Erwartungswertes von $\hat{\Theta}$ gleich dem Parameter Θ ist, d.h. wenn gilt:

$$\lim_{n \rightarrow \infty} E(\hat{\Theta}(X_1, \dots, X_n)) = \Theta.$$

Beispiel 10.4. Siehe Vorlesung.

9.257

Definition 10.5. Eine Schätzfunktion $\hat{\Theta}(X_1, X_2, \dots, X_n)$ eines Parameters Θ heißt **konsistent (passend)**, wenn mit wachsendem n die Schätzfunktion $\hat{\Theta}$ stochastisch gegen Θ konvergiert, d.h. wenn für beliebiges $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|\hat{\Theta}(X_1, X_2, \dots, X_n) - \Theta| < \varepsilon) = 1.$$

Konsistenzkriterium

Eine Schätzfunktion $\hat{\Theta}$ ist konsistent, falls

$$\text{Var}(\hat{\Theta}(X_1, \dots, X_n)) \xrightarrow{n \rightarrow \infty} 0.$$

Es gilt dann:

$$P(|\hat{\Theta}(X_1, \dots, X_n) - \Theta| > \varepsilon) \leq \frac{\text{Var}(\hat{\Theta})}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Beispiel 10.6. Siehe Vorlesung.

9.258

Definition 10.7. Die erwartungstreue Schätzfunktion $\hat{\Theta}_1$ des Parameters Θ einer Grundgesamtheit heißt **effizienter (wirksamer)** als die erwartungstreue Schätzfunktion $\hat{\Theta}_2$ desselben Parameters, wenn für die Varianzen

$$\text{Var}(\hat{\Theta}_1) = E((\hat{\Theta}_1 - \Theta)^2) \text{ und } \text{Var}(\hat{\Theta}_2) = E((\hat{\Theta}_2 - \Theta)^2)$$

gilt:

$$\text{Var}(\hat{\Theta}_1) < \text{Var}(\hat{\Theta}_2).$$

Das Verhältnis

$$\eta = \frac{\text{Var}(\hat{\Theta}_1)}{\text{Var}(\hat{\Theta}_2)} \quad (10.1)$$

heißt **relative Wirksamkeit (Wirkungsgrad)** von $\hat{\Theta}_2$ in Bezug auf $\hat{\Theta}_1$. Die Schätzfunktion für einen Parameter einer Grundgesamtheit X mit der kleinsten Varianz heißt **effektive (wirksamste) Schätzung**.

9.259

Beispiel 10.8. Siehe Vorlesung.

9.260

10.2.2 Maximum-Likelihood-Methode

Ausgangspunkt

Es sei eine Stichprobe (x_1, x_2, \dots, x_n) vom Umfang n aus einer Grundgesamtheit X gegeben. Der Verteilungstyp sei bekannt, die Parameter Θ_i , $(i = 1, \dots, m)$ seien unbekannt und sollen geschätzt werden. Es wird nur ein Parameter betrachtet.

Definition 10.9. (x_1, x_2, \dots, x_n) sei eine Stichprobe vom Umfang n aus einer Grundgesamtheit, die durch eine stetige ZV X mit Dichte $f^X(x; \Theta)$, $x \in \mathbb{R}$, Θ unbekannt, beschrieben wird. Die Funktion

$$L(x_1, x_2, \dots, x_n; \Theta) := \prod_{i=1}^n f^X(x_i; \Theta) \quad (10.2)$$

heißt dann **Likelihood-Funktion** der Stichprobe.

9.261

Definition 10.10. (x_1, x_2, \dots, x_n) ist eine Stichprobe vom Umfang n aus einer Grundgesamtheit, die durch eine diskrete ZV X beschrieben wird. Die Einzelwahrscheinlichkeiten sind $P(X = x_i; \Theta)$, wobei Θ unbekannt ist. Die Funktion

$$L(x_1, x_2, \dots, x_n; \Theta) := \prod_{i=1}^n P(X = x_i; \Theta) \quad (10.3)$$

heißt dann **Likelihood-Funktion** der Stichprobe.

Prinzip der Maximum-Likelihood-Methode

- (i) Die Likelihood-Funktion ist für jede Stichprobe eine Funktion in Θ .
- (ii) Als Schätzwert $\hat{\theta}$ wird derjenige Wert gewählt, für den

$$\max_{\theta} L(x_1, x_2, \dots, x_n; \Theta)$$

gilt. (L kann mehrere Maxima besitzen.)

9.262

Differenzierbarkeit

Anwenden der notwendigen Bedingung, falls L differenzierbar ist

$$\frac{dL}{d\Theta} = 0 \quad (10.4)$$

Enthält die Likelihood-Funktion einen Exponentialausdruck, so ist es günstiger $\ln L$ statt L zu verwenden und von der Gleichung

$$\frac{d \ln L}{d\Theta} = 0 \quad (10.5)$$

auszugehen.

Die Lösung $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n; \Theta)$ ist eine Realisierung (ein Punktschätzwert) der Schätzfunktion $\hat{\Theta} = \hat{\Theta}(x_1, x_2, \dots, x_n; \Theta)$. $\hat{\Theta}$ heißt **Maximum-Likelihood-Schätzung** für Θ .

9.263

Bemerkung

Der Maximum-Likelihood-Schätzer ist unter bestimmten Voraussetzungen konsistent, wenigstens asymptotisch erwartungstreu und asymptotisch normalverteilt.

Beispiel 10.11. Siehe Vorlesung.

Beispiel 10.12. Siehe Vorlesung.

9.264

10.3 Konfidenzschätzung

10.3.1 Begriffe

Warum Konfidenzschätzung?

Bisher wurde nur der Parameter Θ geschätzt. Daraus lässt sich aber keine Aussage über die Genauigkeit einer solchen Schätzung ableiten. Die Abweichungen vom Wert des Parameters Θ können erheblich sein, insbesondere bei kleinen Stichproben.

Idee

Gesucht wird ein Intervall $I = (G_1, G_2)$ derart, dass gilt

$$P(G_1 < \Theta < G_2) = 1 - \alpha. \quad (10.6)$$

Begriffe

G_1, G_2	Konfidenzgrenzen
(G_1, G_2)	Konfidenzintervall
$1 - \alpha$	Konfidenzniveau
α	Irrtumswahrscheinlichkeit

9.265

Konfidenzgrenzen als Stichprobenfunktionen

Es sind nun $G_1 = G_1(X_1, X_2, \dots, X_n)$ und $G_2 = G_2(X_1, X_2, \dots, X_n)$ Stichprobenfunktionen der Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n . Das Intervall (G_1, G_2) ist ein *Zufallsintervall*. Für eine konkrete Stichprobe (x_1, x_2, \dots, x_n) (Realisierung von X) erhält man Realisierungen g_1, g_2 der Konfidenzgrenzen G_1, G_2 . (g_1, g_2) wird als **konkrete Konfidenzschätzung** bezeichnet.

9.266

Ermittlung des Konfidenzintervalls

Es wird $\hat{\Theta}$ verwendet und mittels den Größen δ_1 und δ_2 formulieren wir

$$G_1 = \hat{\Theta} - \delta_1 \quad \text{und} \quad G_2 = \hat{\Theta} + \delta_2 \quad (10.7)$$

bzw.

$$G_1 = \hat{\Theta} \cdot \delta_1 \quad \text{und} \quad G_2 = \hat{\Theta} \cdot \delta_2. \quad (10.8)$$

Damit ergibt sich

$$P(\hat{\Theta} - \delta_1 < \Theta < \hat{\Theta} + \delta_2) = 1 - \alpha, \quad (10.9)$$

$$P(\hat{\Theta} \cdot \delta_1 < \Theta < \hat{\Theta} \cdot \delta_2) = 1 - \alpha. \quad (10.10)$$

9.267**10.3.2 Konfidenzschätzung für den Erwartungswert**

Es wird die Konfidenzschätzung für den Erwartungswert μ einer normalverteilten Grundgesamtheit X betrachtet, wobei σ^2 bekannt ist.

Die Schätzfunktion für $\Theta = \mu$ lautet $\hat{\Theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

In (10.9) eingesetzt, erhalten wir

$$P(\bar{X} - \delta < \mu < \bar{X} + \delta) = P(|\bar{X} - \mu| < \delta) = 1 - \alpha. \quad (10.11)$$

9.268**Bestimmung δ (Schätzung von μ bei bekannter Varianz σ^2)**

Es ist $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$; $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ ist $Z \sim \mathcal{N}(0, 1)$. Damit gilt

$$P\left(\left|\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| < \frac{\delta\sqrt{n}}{\sigma}\right) = 1 - \alpha \quad (10.12)$$

bzw.

$$P\left(|Z| < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \quad \text{mit } z_{1-\frac{\alpha}{2}} = \frac{\delta\sqrt{n}}{\sigma}. \quad (10.13)$$

$z_{1-\frac{\alpha}{2}}$ ist ein Perzentil der Normalverteilung und kann aus entsprechenden Tabellen abgelesen werden.

9.269**Ergebnis**

Die Konfidenzschätzung für $\Theta = \mu$ ist

$$\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}. \quad (10.14)$$

Jede Realisierung liefert eine Realisierung des Konfidenzintervalls

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (10.15)$$

mit der Länge des Konfidenzintervalls

$$2\delta = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}. \quad (10.16)$$

$\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$ überdeckt $\Theta = \mu$ mit Wahrscheinlichkeit $(1 - \alpha)$.

Beispiel 10.13. Siehe Vorlesung.

9.270

Schätzung von μ bei unbekannter Varianz σ^2

Es wird die Konfidenzschätzung für den Erwartungswert μ einer normalverteilten Grundgesamtheit X betrachtet, wobei σ^2 nicht bekannt ist.

Die Schätzfunktion für $\Theta_1 = \mu$ lautet $\hat{\Theta}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Die Schätzfunktion für $\Theta_2 = \sigma^2$ lautet $\hat{\Theta}_2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Ansatz zur Berechnung von δ_1 und δ_2

$$Z_{n-1}^* = \frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1}, \quad (10.17)$$

wobei Z Student-t-verteilt mit $m = n - 1$ Freiheitsgraden ist.

Wegen der Symmetrie gilt: $\delta = \delta_1 = \delta_2$.

9.271

Bestimmung δ (Schätzung von μ bei unbekannter Varianz σ^2)

Der Wert $t_{\frac{\alpha}{2}; n-1}$ wird aus der Tabelle für die Student-t-Verteilung abgelesen, da gelten muss:

$$P\left(|Z_{n-1}^*| < t_{1-\frac{\alpha}{2}; n-1}\right) = 1 - \alpha. \quad (10.18)$$

Einsetzen von Z_m^* liefert:

$$P\left(\left|\frac{\bar{X} - \mu}{S} \sqrt{n}\right| < t_{1-\frac{\alpha}{2}; n-1}\right) = 1 - \alpha \quad (10.19)$$

bzw.

$$P\left(-t_{1-\frac{\alpha}{2}; n-1} < \frac{\bar{X} - \mu}{S} \sqrt{n} < t_{1-\frac{\alpha}{2}; n-1}\right) = 1 - \alpha. \quad (10.20)$$

Daraus ergibt sich die Konfidenzschätzung

$$\bar{X} - t_{1-\frac{\alpha}{2}; m} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2}; m} \frac{S}{\sqrt{n}}.$$

9.272

Für eine konkrete Stichprobe \mathbf{x} ergibt sich dann

$$\bar{x} - t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}}. \quad (10.21)$$

Beispiel 10.14. Hektarerträge von Versuchsflächen einer neuen Getreidesorte [dt]:

35,6; 33,7; 37,8; 31,2; 37,2; 34,1; 35,8; 36,6; 37,1; 34,9; 35,6; 34,0

9.273

Es wird die Konfidenzschätzung für die Varianz σ^2 einer normalverteilten Grundgesamtheit X betrachtet.

Die Schätzfunktion für $\Theta = \sigma^2$ lautet $\hat{\Theta} = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Grundlage zur Berechnung des Konfidenzintervalls

$$Y_{n-1}^* = \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2, \quad (10.22)$$

wobei Y_{n-1}^* χ^2 -verteilt mit $m = n - 1$ Freiheitsgraden ist.

9.274

Bestimmen des Konfidenzintervalls

Gesucht sind $c_1 < c_2$ mit

$$P(c_1 < Y^* < c_2) = \int_{c_1}^{c_2} f_{\chi^2}(x) dx = 1 - \alpha. \quad (10.23)$$

Da die χ^2 -Verteilung nicht symmetrisch ist, wird folgender Ansatz verwendet:

$$P(Y^* \geq c_1) = \int_{c_1}^{\infty} f_{\chi^2}(x) dx = 1 - \frac{\alpha}{2}, \quad (10.24)$$

$$P(Y^* \geq c_2) = \int_{c_2}^{\infty} f_{\chi^2}(x) dx = \frac{\alpha}{2}. \quad (10.25)$$

Es wird gesetzt:

$$c_1 = \chi_{\frac{\alpha}{2};m}^2 \text{ und } c_2 = \chi_{1-\frac{\alpha}{2};m}^2. \quad (10.26)$$

Damit gilt:

$$P\left(\chi_{\frac{\alpha}{2};m}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\frac{\alpha}{2};m}^2\right) = 1 - \alpha. \quad (10.27)$$

9.275

Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$

$$\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2};m}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2};m}^2}. \quad (10.28)$$

Die Realisierung zu einer konkreten Stichprobe ist demnach

$$\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2};m}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2};m}^2}. \quad (10.29)$$

Beispiel 10.15. Siehe Vorlesung

9.276

Konfidenzschätzung für p von $B(p)$

Schätzfunktion

Für eine Grundgesamtheit X mit $X \sim B(p)$ soll ein Konfidenzintervall für p bestimmt werden.

Als Schätzfunktion wählen wir $\hat{\Theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Ansatz

Es gilt $n\bar{X} = \sum_{i=1}^n X_i \sim B(n, p)$.

- Für kleine n kann das Konfidenzintervall mittels der Binomialverteilung bestimmt werden.
- Für große n ist \bar{X} annähernd $\mathcal{N}(p; \frac{pq}{n})$ -verteilt. Demnach ist $Z = \frac{\bar{X} - p}{\sqrt{\frac{pq}{n}}}$ annähernd $\mathcal{N}(0, 1)$ verteilt.

9.277

Konfidenzintervall

Aus (10.9) folgt mit $\Theta = p$, $\hat{\Theta} = \bar{X}$ und $\delta_1 = \delta_2 = \delta$

$$P(|\bar{X} - p| < \delta) = 1 - \alpha, \quad (10.30)$$

$$P\left(\left|\frac{\bar{X} - p}{\sqrt{\frac{pq}{n}}}\right| < \frac{\delta\sqrt{n}}{\sqrt{pq}}\right) = 1 - \alpha \quad (10.31)$$

und damit

$$P(|Z| < z_{1-\frac{\alpha}{2}}) \quad (10.32)$$

mit $z_{1-\frac{\alpha}{2}} = \frac{\delta\sqrt{n}}{\sqrt{pq}}$.

9.278

Damit folgt das Konfidenzintervall für das Konfidenzniveau $1 - \alpha$ aus

$$\left|\frac{\bar{X} - p}{\sqrt{\frac{pq}{n}}}\right| < z_{1-\frac{\alpha}{2}}, \quad (10.33)$$

$$\frac{n}{n+z_{1-\frac{\alpha}{2}}^2} \left(\bar{X} + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \left(\frac{z_{1-\frac{\alpha}{2}}}{2n}\right)^2} \right) < p < \left(\bar{X} + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \left(\frac{z_{1-\frac{\alpha}{2}}}{2n}\right)^2} \right).$$

9.279

Kapitel 11

Testen von Hypothesen

11.1 Grundbegriffe

Exemplarische Fragestellungen

Beispiel 11.1. Langskript:

11.280

Hypothesentests

Eine **Hypothese** H wird mittels **statistischem Test**, (auch statische Prüfverfahren, Tests) überprüft. Die Aufgabe ist, die Hypothese aufgrund einer konkreten Stichprobe anzunehmen oder abzulehnen.

Die Hypothese H heißt **Nullhypothese** H_0 , wenn noch andere Hypothesen oder **Alternativhypothesen** aufgestellt werden können.

11.281

Beispiel 11.2. Langskript:

Folgerungen aus dem Beispiel

Mögliche Entscheidungen

- + H_0 wird nicht abgelehnt: Die Abweichung des Mittelwertes ist gering und wird als zufällig angenommen.
- H_0 wird abgelehnt: Die Abweichung des Mittelwertes ist so groß, dass die konkrete Stichprobe anscheinend nicht zur Grundgesamtheit X gehört. Die auftretende Abweichung wird dann als **signifikant** oder **statistisch gesichert** bezeichnet.

Frage

Wie wird eine Schranke c für die Entscheidung gefunden?

11.282

Idee zum Wählen von c

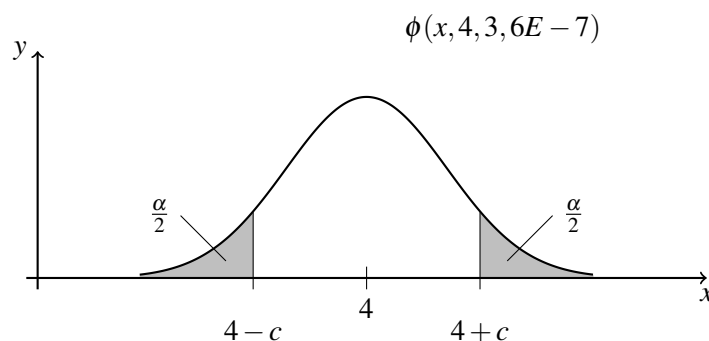
- Ein Wert α ($0 < \alpha < 1$) wird gewählt.
- c wird so bestimmt, dass die Wahrscheinlichkeit für den Betrag des Abweichens der ZV \bar{X} vom Nennmaß um mindestens c gerade die Wahrscheinlichkeit α beträgt, wenn H_0 richtig ist. D.h.:

$$P(|\bar{X} - \mu| \geq c | H_0) = \alpha. \quad (11.1)$$

α wird als **Irrtumswahrscheinlichkeit** oder auch **Signifikanzniveau** bezeichnet. Üblich ist die Wahl $\alpha = 0,05; 0,01; 0,001$. Aus α wird die Schranke c bestimmt und somit der **Ablehnungsbereich (kritischer Bereich)** K für die Nullhypothese H_0 gewonnen.

11.283

Beispiel 11.3 (Fortsetzung Beispiel 282). Siehe Vorlesung



11.284

Fehlerarten

1. Art: Die Nullhypothese H_0 wird abgelehnt, obwohl sie richtig ist.
2. Art: Die Nullhypothese H_0 wird nicht abgelehnt, obwohl sie falsch ist.

11.285

Fehler 1. Art

Betrachten wir die Realisierung u der Prüfgröße U , so gilt

$$P(u \in K | H_0) = \alpha. \quad (11.2)$$

α ist die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen.

Schwierigkeit

Wird α klein gewählt, dann ist auch die Wahrscheinlichkeit für den Fehler 1. Art klein. Aber je kleiner α ist, um so schwieriger ist es, die Falschheit einer Hypothese zu zeigen. (K wird für U kleiner.)

Folge: Die Wahrscheinlichkeit für den Fehler 2. Art steigt.

11.286

Vorgehen in der allgemeinen Testtheorie

Aufstellen der Hypothesen

$$\begin{aligned}H_0: & E(X) = \mu_0 = 4 \\H_1: & E(X) = \mu, (\mu \neq 4)\end{aligned}$$

Idee

Bestimme K derart, dass die Wahrscheinlichkeit für die Ablehnung falscher H_0 möglichst groß ist. Dies entspricht dann der Wahrscheinlichkeit, H_1 anzunehmen unter der Voraussetzung, dass H_1 richtig ist.

$$P(U \in K | H_1) = 1 - \beta. \quad (11.3)$$

β heißt **Güte** oder **Trennschärfe** des Tests oder Prüfverfahrens.

11.287**Wahl von K**

Ziel ist es, K derart zu wählen, dass

- der Fehler 1. Art durch ein **vorgegebenes** möglichst kleines α und
- die Wahrscheinlichkeit, die Nullhypothese H_0 richtigerweise abzulehnen, durch ein **vorgegebenes** möglichst großes $1 - \beta$

begrenzt ist.

Fehlerarten

	nicht abgelehnt	abgelehnt
H_0 richtig	richtige Entscheidung $p_1 = 1 - \alpha$	Fehler 1. Art $p_2 = \alpha$
H_0 falsch	Fehler 2. Art $p_3 = \beta$	richtige Entscheidung $p_1 = 1 - \beta$

Bemerkung

Wird nur α vorgegeben und auf die Berücksichtigung der Fehler 2. Art verzichtet, dann wird von **Signifikanztests** gesprochen.

11.288**Im Auge behalten:**

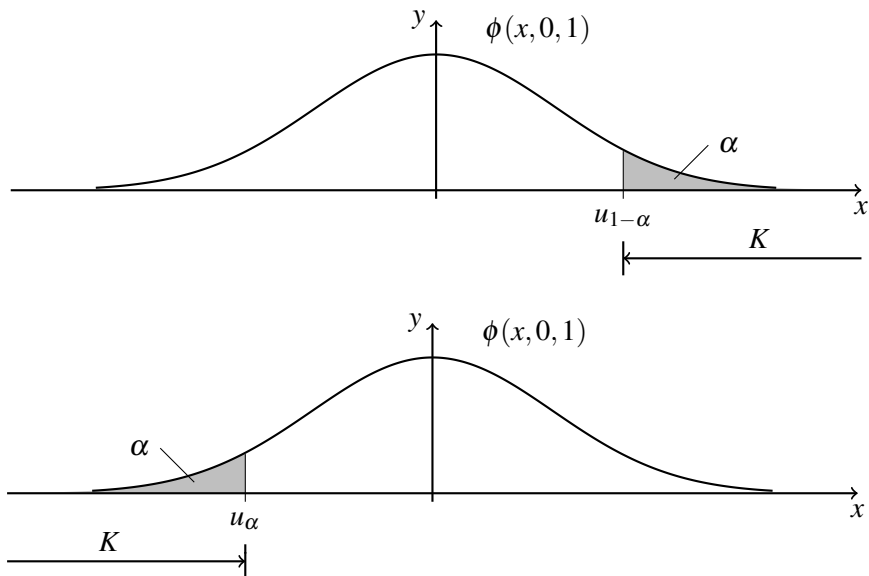
Fehler 2. Art können immer auftreten.

Einseitige Fragestellungen

U ist symmetrisch verteilt und es wird

$$U \geq u_{1-\alpha} \text{ oder } U \leq -u_{1-\alpha}$$

gewählt.

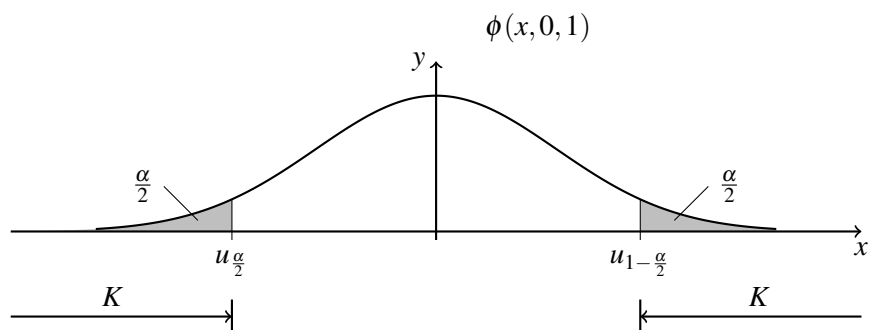


11.289

Zweiseitige Fragestellungen

Für die Prüfgröße U und die gegebene Irrtumswahrscheinlichkeit α wählen wir

$$P\left(|U| \geq u_{1-\frac{\alpha}{2}} | H_0\right) = \alpha.$$



Beispiel 11.4. Siehe Vorlesung

11.290

Vorgehen bei Tests

1. Aufstellen der Nullhypothese H_0 .
2. Vorgabe der Irrtumswahrscheinlichkeit α .
3. Wahl einer geeigneten Prüfgröße $U = U(X_1, X_2, \dots, X_n)$. Die Verteilungsfunktion sei bekannt.

4. Ermittlung von K aus $P(U \in K|H_0) = \alpha$.
5. Berechnung einer Realisierung u von U mit Hilfe einer konkreten Stichprobe (x_1, x_2, \dots, x_n) vom Umfang n .
6. Falls $u \in K$, wird H_0 abgelehnt; falls $u \notin K$, wird H_0 nicht abgelehnt.

11.291

11.2 Test für $\mathcal{N}(\mu, \sigma^2)$, $H_0 = \{\mathcal{N}(\mu, \sigma^2), \mu = \mu_0\}$

Aufgabe

Der Erwartungswert einer Grundgesamtheit X mit $X \sim \mathcal{N}(\mu, \sigma^2)$ soll bei bekannter Varianz σ^2 geprüft werden.

Stichworte: Sollwert, statistische Qualitätssicherung

Vorgehen

1. $H_0: EX = \mu_0$
2. Vorgabe der Irrtumswahrscheinlichkeit α
3. Zu einer mathematischen Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n aus $X \sim \mathcal{N}(\mu, \sigma^2)$ wird

$$U = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \quad (11.4)$$

gewählt.

$$U \sim \mathcal{N}(0, 1)$$

11.292

4. Bestimmung von K

(i) Zweiseitige Fragestellung

Ausgehend von

$$P(|U| \geq z_{1-\frac{\alpha}{2}} | H_0) = \alpha$$

ergibt sich

$$\bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \quad \text{und} \quad \bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}. \quad (11.5)$$

(ii) Einseitige Fragestellung

Ausgehend von

$$P(U \geq z_{1-\alpha} | H_0) = \alpha \quad \text{bzw.} \quad P(U \leq z_{1-\alpha} | H_0) = \alpha$$

ergibt sich

$$\bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \quad \text{bzw.} \quad \bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}. \quad (11.6)$$

5. Aus einer konkreten Stichprobe (x_1, x_2, \dots, x_n) wird eine Realisierung u von U berechnet

$$u = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}. \quad (11.7)$$

6. Entscheidung

- (i) *Zweiseitige Fragestellung*: H_0 wird abgelehnt, falls $|u| \geq z_{1-\frac{\alpha}{2}}$. H_0 wird nicht abgelehnt, falls $|u| < z_{1-\frac{\alpha}{2}}$.
- (ii) *Einseitige Fragestellung*: H_0 wird abgelehnt, falls $u \geq z_{1-\alpha}$ bzw. $u \leq -z_{1-\alpha}$. H_0 wird nicht abgelehnt, falls $u < z_{1-\alpha}$ bzw. $u > -z_{1-\alpha}$.

Beispiel 11.5. Siehe Vorlesung.

11.3 Test für $\mathcal{N}(\mu, \sigma^2)$, $H_0 = \{\mathcal{N}(\mu, \sigma^2), \mu = \mu_0\}$, Varianz unbekannt

Aufgabe

Der Erwartungswert einer Grundgesamtheit X mit $X \sim \mathcal{N}(\mu, \sigma^2)$ soll bei unbekannter Varianz σ^2 geprüft werden.

Stichworte: Sollwert, statistische Qualitätssicherung

Vorgehen

1. $H_0: EX = \mu_0$.
2. Vorgabe der Irrtumswahrscheinlichkeit α .
3. Zu einer mathematischen Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n aus $X \sim \mathcal{N}(\mu, \sigma^2)$ wird

$$U = \frac{\bar{X} - \mu_0}{S} \sqrt{n} \sim t_{n-1} \quad (11.8)$$

gewählt. U ist Student- t -verteilt mit $m = n - 1$ Freiheitsgraden.

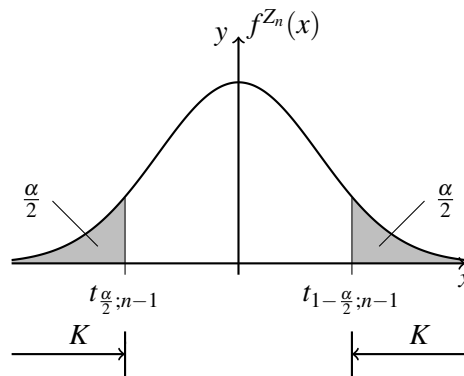
4. Bestimmung von K

- (i) *Zweiseitige Fragestellung*
Ausgehend von

$$P(|U| \geq t_{1-\frac{\alpha}{2}; n-1} | H_0) = \alpha$$

ergibt sich

$$\bar{X} \geq \mu_0 + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}; n-1} \quad \text{und} \quad \bar{X} \leq \mu_0 - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}; n-1}. \quad (11.9)$$



11.296

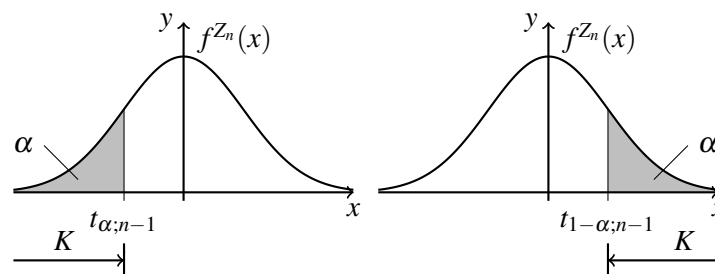
4. Bestimmung K (ii) *Einseitige Fragestellung*

Ausgehend von

$$P(U \leq t_{1-\alpha; n-1} | H_0) = \alpha \quad \text{bzw.} \quad P(U \geq t_{1-\alpha; n-1} | H_0) = \alpha$$

ergibt sich

$$\bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} t_{1-\alpha; n-1} \quad \text{bzw.} \quad \bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} t_{1-\alpha; n-1} \quad (11.10)$$



11.297

5. Aus einer konkreten Stichprobe (x_1, x_2, \dots, x_n) wird eine Realisierung u von U berechnet:

$$u = \frac{\bar{x} - \mu_0}{s} \sqrt{n}. \quad (11.11)$$

6. Entscheidung

(i) *Zweiseitige Fragestellung*: H_0 wird abgelehnt, falls $|u| \geq t_{1-\frac{\alpha}{2}; n-1}$. H_0 wird nicht abgelehnt, falls $|u| < t_{1-\frac{\alpha}{2}; n-1}$.

(ii) *Einseitige Fragestellung*: H_0 wird abgelehnt, falls

$$u \geq t_{1-\alpha; n-1} \quad \text{bzw.} \quad u \leq -t_{1-\alpha; n-1}.$$

H_0 wird nicht abgelehnt, falls

$$u < t_{1-\alpha; n-1} \quad \text{bzw.} \quad u > -t_{1-\alpha; n-1}.$$

Beispiel 11.6. Siehe Vorlesung.

11.298

11.4 Prüfung der Gleichheit

Fragestellung

$$\begin{array}{cc} \text{Werk 1} & \text{Werk 2} \\ X \sim \mathcal{N}(\mu_X, \sigma_X^2) & Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \\ EX = \mu_X & \stackrel{?}{=} EX = \mu_Y \end{array}$$

Voraussetzungen: X, Y stoch. unabh.

Vorgehen

Voraussetzung: $\text{Var}X = \text{Var}Y$.

1. $H_0: EX = EY$.
2. Vorgabe der Irrtumswahrscheinlichkeit α .

11.299

3. Zu einer mathematischen Stichprobe $(X_1, X_2, \dots, X_{n_1})$ vom Umfang n_1 und einer mathematischen Stichprobe $(Y_1, Y_2, \dots, Y_{n_2})$ vom Umfang n_2 wird die Testgröße

$$U = \frac{\bar{X} - \bar{Y}}{\sqrt{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \sim t_m \quad (11.12)$$

gewählt. U ist Student- t -verteilt mit $m = n_1 + n_2 - 2$ Freiheitsgraden.

4. Bestimmung von K

- (i) *Zweiseitige Fragestellung*
Ausgehend von

$$P(|U| \geq t_{1-\frac{\alpha}{2};m} | H_0) = \alpha$$

ergibt sich

$$|\bar{X} - \bar{Y}| \leq t_{1-\frac{\alpha}{2};m} \sqrt{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2} \sqrt{\frac{n_1 + n_2}{n_1 n_2 (n_1 + n_2 - 2)}}. \quad (11.13)$$

- (ii) *Einseitige Fragestellung* Wird hier nicht behandelt.

11.300

5. Aus einer konkreten Stichprobe (x_1, x_2, \dots, x_n) wird eine Realisierung u von U berechnet:

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}. \quad (11.14)$$

6. Entscheidung der *zweiseitigen Fragestellung*: H_0 wird abgelehnt, falls $|u| \geq t_{1-\frac{\alpha}{2};m}$.
 H_0 wird nicht abgelehnt, falls $|u| < t_{1-\frac{\alpha}{2};m}$.

Beispiel 11.7. Siehe Vorlesung

11.301

11.5 Nicht vorgestellte Tests

- Prüfung der Varianz einer normalverteilten Grundgesamtheit.
- Prüfung der Gleichheit der Varianzen zweier unabhängiger normalverteilter Grundgesamtheiten.
- Prüfung p der $B(p)$ -Verteilung.

11.302

11.6 χ^2 -Anpassungstest

1. $H_0: F^X(t) \equiv F_0(t)$
2. Vorgabe der Irrtumswahrscheinlichkeit α
3. Zur mathematischen Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n wird die Prüfgröße

$$U = \sum_{m=1}^k \frac{(H_m - np_m)^2}{np_m} \sim \chi_{k-r-1}^2 \quad (11.15)$$

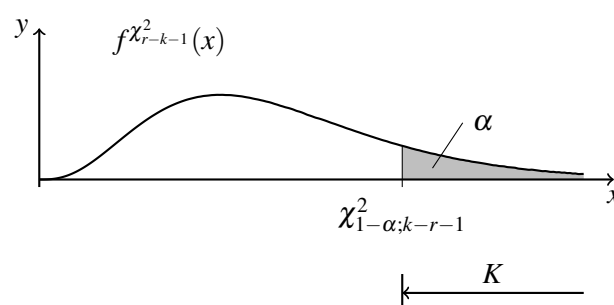
betrachtet. Es gilt: $U \sim \chi^2(k-r-1)$ und

- die Elemente der Stichprobe wurden in k Klassen eingeteilt,
- H_m ist die absolute Häufigkeit (ZV) von Elementen in der m -ten Klasse, ($m = 1, \dots, k$)
- p_m ist die mittels $F_0(t)$ errechnete Wahrscheinlichkeit, dass ein Wert der Grundgesamtheit X in der m -ten Klasse liegt,
- np_m ist die entsprechende absolute Häufigkeit von n Elementen in der m -ten Klasse,
- r ist die Anzahl der zu schätzenden Parameter in der VF $F_0(t)$.

11.303

4. Kritischer Bereich K (hier ist nur die einseitige Fragestellung von Interesse)

$$P(U \geq \chi_{1-\alpha; k-r-1}^2 | H_0) = \alpha. \quad (11.16)$$



5. Aus konkreter Stichprobe (x_1, x_2, \dots, x_n) vom Umfang n wird eine Realisierung u von U bestimmt:

$$u = \sum_{m=1}^k \frac{(h_m - np_m)^2}{np_m}. \quad (11.17)$$

h_m ist Realisierung von H_m .

6. Entscheidung:

H_0 wird abgelehnt, falls $u \geq \chi^2_{1-\alpha; k-r-1}$. H_0 wird nicht abgelehnt, falls $u < \chi^2_{1-\alpha; k-r-1}$.

Beispiel 11.8.

m	h_m	p_m	np_m	$ h_m - np_m $	$(h_m - np_m)^2$	$\frac{(h_m - np_m)^2}{np_m}$
0	57	0,021	54,8	2,2	4,84	0,088
1	203	0,081	211,2	8,2	67,24	0,318
2	383	0,156	406,8	23,8	566,44	1,392
3	525	0,201	524,2	0,8	0,64	0,001
4	532	0,195	508,6	23,4	547,56	1,007
5	408	0,151	393,8	14,2	201,64	0,512
6	273	0,097	253,0	20,0	400,00	1,581
7	139	0,054	140,8	1,8	3,24	0,023
8	45	0,026	67,8	22,8	519,84	7,667
9	27	0,011	28,7	1,7	2,89	0,101
≥ 10	16	0,007	18,3	2,3	5,29	0,289
	2608	1,000				$u = 13,049$

Beyer et al.: Wahrscheinlichkeitsrechnung und mathematische Statistik. Teubner Leipzig, 1980

Kapitel 12

Regressions- und Korrelationsanalyse

12.1 Einführung

Beispiel 12.1. Siehe Vorlesung

Aufgabe

Aus Beobachtungen (Realisierungen) des 2-dimensionalen Zufallsvektors (X, Y) soll auf einen möglichen Zusammenhang zwischen X und Y geschlossen werden.

Bisher wurden betrachtet:

- empirische Kovarianz s_{xy} ,
- empirischer Korrelationskoeffizient r_{xy} .

12.308

12.2 Regressionsanalyse

Beobachtungen

Ein Wert von X kann verschiedene Werte von Y besitzen bzw. annehmen. Es wird dann von einer **stochastischen Abhängigkeit** gesprochen.

X heißt **Einflussgröße** und Y heißt **Zielgröße**.

Aufgabe der Regressionsanalyse

Untersuchung der Art des Zusammenhangs von X und Y und die Angabe von Prüf- und Schätzverfahren.

Annahme

Für eine feste Realisierung x von X sei Y normalverteilt mit

$$E(Y|X=x) = \eta(x) = \beta_1 + \beta_2 x \quad (12.1)$$

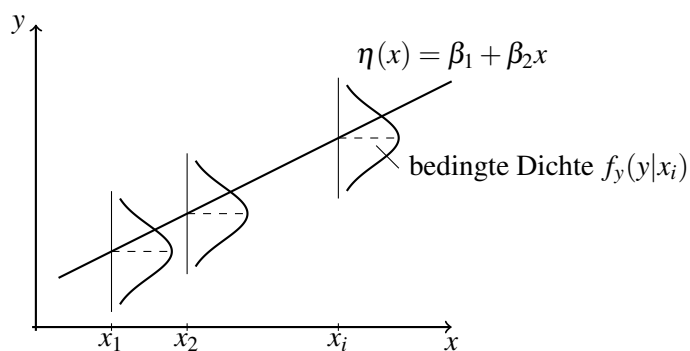
und

$$\text{Var}(Y|X=x) = \sigma^2 \equiv \text{const.} \quad (12.2)$$

12.309

Definition 12.2. $\eta(x)$ (aus (12.1)) heißt **(theoretische) Regressionsgerade** und β_2 (aus (12.1)) **(theoretischer) Regressionskoeffizient**.

(12.2) heißt die **Varianz um die Regressionsgerade**



12.310

12.3 Schätzung von β_1 , β_2 und σ^2

Ausgangspunkt

Es liegt eine konkrete Stichprobe $\mathbf{x} = (x_1, x_2, \dots, x_n)$ vor. Davon ausgehend wird eine Realisierung b_1 von β_1 und eine Realisierung b_2 von β_2 berechnet.

Daraus ergibt sich (die Realisierung von $\eta(x)$):

$$y(x) = b_1 + b_2 x. \quad (12.3)$$

12.311

Berechnung der Realisierung

Die Berechnung erfolgt mittels der Methode der kleinsten Quadrate

$$\min \left\| A \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} - y \right\|^2 \quad (12.4)$$

mit

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Die Lösung von (12.4) ist die Lösung der Normalengleichung

$$A^T A \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = A^T \mathbf{y} \quad (12.5)$$

mit

$$A^T A = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}. \quad (12.6)$$

12.312

Rückblick

Aus Satz 2.21

wissen wir bereits:

$$b_1 = \bar{y} - b_2 \bar{x}, \quad (12.7)$$

$$b_2 = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x} \quad (12.8)$$

mit

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (12.9)$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad (12.10)$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y}. \quad (12.11)$$

12.313

Ergebnis

Damit ergibt sich

$$y(x) = (\bar{y} - b_2 \bar{x}) + b_2 x = \bar{y} + b_2 (x - \bar{x}), \quad (12.12)$$

dies ist die empirische Regressionsgerade „von Y in Bezug auf X “.

12.314

Beispiel 12.3. Siehe Vorlesung

Bemerkung

(i) In gleicher Weise kann die Regression für X in Bezug auf Y ermittelt werden:

$$\begin{aligned} x(y) &= c_1 + c_2 y, \\ c_1 &= \bar{x} - c_2 \bar{y}, \end{aligned} \quad (12.13)$$

$$c_2 = r_{xy} \frac{s_x}{s_y}. \quad (12.14)$$

- (ii) Bei der orthogonalen Regression wird die Quadratsumme der orthogonalen Abstände minimiert.

12.315

Es liegt eine Stichprobe $(x_i, y_i), i = 1, \dots, n$, vor.

Die Punktschätzung $\hat{\sigma}^2$ ist die Schätzfunktion der Restvarianz σ^2 . $\hat{\sigma}^2$ heißt **empirische Restvarianz**.

Die Realisierung von $\hat{\sigma}^2$ wird mit s_R^2 bezeichnet und nennt sich **konkrete empirische Restvarianz**.

Berechnung

$$s_R^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - y(x_i))^2. \quad (12.15)$$

Für die numerische Auswertung ist die Darstellung

$$s_R^2 = \frac{1}{n-2} \left[\sum_{i=1}^n y_i - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 - b_2 \left(\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) \right]$$

günstiger.

12.316

12.4 Prüfen der Parameter β_1, β_2

Im Folgenden wird die Frage beantwortet, ob die ermittelten Werte b_1 und b_2 mit vorgegebenen Werten β_{20} und β_{10} vereinbar sind. Die Prüfung erfolgt mittels Hypothesentest.

12.317

Vorgehen für den Signifikanztest $\beta_2 = \beta_{20}$

1. $H_0 : \beta_2 = \beta_{20}$.
2. Vorgabe der Irrtumswahrscheinlichkeit α .
3. Zu einer mathematischen Stichprobe (X_i, Y_i) vom Umfang n wird

$$U = \frac{\hat{\beta}_2 - \beta_{20}}{S_{\beta_2}} \quad (12.16)$$

gewählt. Es gilt:

$$U \sim t_m$$

mit $m = n - 2$ und

$$S_{\beta_2}^2 = \frac{\hat{\sigma}_2}{(n-1)S_X^2}, \quad (12.17)$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (12.18)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (12.19)$$

4. Bestimmung von K : Ausgehend von

$$P(|U| \geq t_{1-\frac{\alpha}{2};m} | H_0) = \alpha$$

ergibt sich

$$\hat{\beta}_2 \geq \beta_{20} + t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_2} \quad \text{und} \quad \hat{\beta}_2 \leq \beta_{20} - t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_2}. \quad (12.20)$$

5. Für eine konkrete Stichprobe (x_i, y_i) , $i = 1, \dots, n$, vom Umfang n ergibt sich

$$u = \frac{b_2 - \beta_{20}}{s_{\beta_2}} \quad (12.21)$$

mit

$$s_{\beta_2}^2 = \frac{s_R^2}{(n-1)s_x^2}. \quad (12.22)$$

6. Entscheidung: H_0 wird abgelehnt, falls $|u| \geq t_{1-\frac{\alpha}{2};m}$.

H_0 wird nicht abgelehnt, falls $|u| < t_{1-\frac{\alpha}{2};m}$.

Beispiel 12.4. Siehe Vorlesung

für den Signifikanztest $\beta_1 = \beta_{10}$

1. $H_0 : \beta_1 = \beta_{10}$.
2. Vorgabe der Irrtumswahrscheinlichkeit α .
3. Zu einer mathematischen Stichprobe (X_i, Y_i) vom Umfang n wird

$$U = \frac{\hat{\beta}_1 - \beta_{10}}{S_{\hat{\beta}_1}} \quad (12.23)$$

gewählt. Es gilt:

$$U \sim t_m$$

mit $m = n - 2$ und

$$S_{\hat{\beta}_1}^2 = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)S_X^2} \right), \quad (12.24)$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (12.25)$$

4. Bestimmung von K Ausgehend von

$$P(|U| \geq t_{1-\frac{\alpha}{2};m} | H_0) = \alpha$$

ergibt sich

$$\hat{\beta}_1 \geq \beta_{10} + t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_1} \quad \text{und} \quad \hat{\beta}_1 \leq \beta_{10} - t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_1} \quad (12.26)$$

5. Für eine konkrete Stichprobe (X_i, Y_i) , $i = 1, \dots, n$, vom Umfang n ergibt sich

$$u = \frac{b_1 - \beta_{10}}{s_{\beta_1}} \quad (12.27)$$

mit

$$s_{\beta_1}^2 = s_R^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2} \right). \quad (12.28)$$

6. Entscheidung: H_0 wird abgelehnt, falls $|u| \geq t_{1-\frac{\alpha}{2};m}$. H_0 wird nicht abgelehnt, falls $|u| < t_{1-\frac{\alpha}{2};m}$.

Beispiel 12.5. Siehe Vorlesung

12.321

Stand der Dinge

Mit den Rezepten aus dem bisherigen Teil des Kapitels kann für eine konkrete Stichprobe (\mathbf{x}, \mathbf{y}) eine Realisierung

$$y(x) = b_1 + b_2 x$$

ermittelt werden. Dies ist eine Realisierung der Schätzung

$$\hat{\eta}(x) = \hat{\beta}_1 + \hat{\beta}_2 x$$

von $\eta(x) = \beta_1 + \beta_2 x$.

12.322

Konfidenzintervall für β_2

Ausgehen von der Schätzfunktion $\hat{\beta}_2$ für den Parameter β_2 ergibt sich zur Bestimmung des Konfidenzintervalls

$$P \left(\left| \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}} \right| < t_{1-\frac{\alpha}{2};m} \right) = 1 - \alpha \quad (12.29)$$

mit $m = n - 2$.

Das Konfidenzintervall lautet für die Schätzfunktion

$$\hat{\beta}_2 - t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_2} < \beta_2 < \hat{\beta}_2 + t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_2} \quad (12.30)$$

und für die Realisierung

$$b_2 - t_{1-\frac{\alpha}{2};m} s_{\hat{\beta}_2} < \beta_2 < b_2 + t_{1-\frac{\alpha}{2};m} s_{\hat{\beta}_2}. \quad (12.31)$$

Beispiel 12.6. Siehe Vorlesung

12.323

Konfidenzintervall für β_1

Ausgehend von der Schätzfunktion $\hat{\beta}_1$ für den Parameter β_1 ergibt sich zur Bestimmung des Konfidenzintervalls:

$$P \left(\left| \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \right| < t_{1-\frac{\alpha}{2};m} \right) = 1 - \alpha. \quad (12.32)$$

Das Konfidenzintervall lautet für die Schätzfunktion

$$\hat{\beta}_1 - t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_1} < \beta_1 < \hat{\beta}_1 + t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_1} \quad (12.33)$$

und für die Realisierung

$$b_1 - t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_1} < \beta_1 < b_1 + t_{1-\frac{\alpha}{2};m} S_{\hat{\beta}_1}. \quad (12.34)$$

Beispiel 12.7. Siehe Vorlesung

12.324

Ebenfalls ohne Herleitung wird das Konfidenzintervall für $\eta(x)$ angegeben.

$$\hat{\eta} - t_{1-\frac{\alpha}{2};m} S_{\hat{\eta}} < \eta(x) < \hat{\eta} + t_{1-\frac{\alpha}{2};m} S_{\hat{\eta}} \quad (12.35)$$

mit

$$S_{\hat{\eta}}^2 = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x - \bar{X})^2}{(n-1)S_X^2} \right). \quad (12.36)$$

Für eine Realisierung ergibt sich damit

$$y(x) - t_{1-\frac{\alpha}{2};m} S_{\hat{\eta}} < \eta(x) < y(x) + t_{1-\frac{\alpha}{2};m} S_{\hat{\eta}} \quad (12.37)$$

mit

$$s_{\hat{\eta}}(x) = \sqrt{s_R^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2} \right)}. \quad (12.38)$$

Beispiel 12.8. Siehe Vorlesung

12.325

12.326

12.5 Korrelationsanalyse

Ausgangspunkt

Es liegt eine Grundgesamtheit (X, Y) vor, die einer mehrdimensionalen Normalverteilung unterliegt. Es sei

$$\begin{aligned} E(X) &= \mu_X, & \text{Var}(X) &= \sigma_X^2, \\ E(Y) &= \mu_Y, & \text{Var}(Y) &= \sigma_Y^2. \end{aligned}$$

Die zweidimensionale Normalverteilung ist mit Kenntnis von $\rho_{XY} = \text{korr}(X, Y)$ vollständig bestimmt.

Ziel der Korrelationsanalyse

Da $(X, Y) \sim \mathcal{N}(\mathbf{a}, \mathbf{K})$, sind X und Y genau dann stoch. unabh., wenn für den Korrelationskoeffizient gilt: $\text{korr}(X, Y) = 0$. Anhand einer Schätzfunktion $\hat{\rho}_{XY}$ (gewonnen aus der Stichprobe $(X_i, Y_i), i = 1, \dots, n$) wird aus einer Stichprobe $(x_i, y_i), i = 1, \dots, n$, ein r_{XY} berechnet. r_{XY} ist eine Realisierung von ρ_{XY} .

Mit Hilfe von r_{XY} wird dann entschieden, ob X und Y stoch. unabh. sind.

12.327**Vorgehen**

1. $H_0 : \rho_{XY} = 0$.
2. Vorgabe der Irrtumswahrscheinlichkeit α .
3. Zu einer mathematischen Stichprobe (X_i, Y_i) vom Umfang n einer normalverteilten Grundgesamtheit (X, Y) wird die Prüfgröße

$$U = \frac{\hat{\rho}_{XY} \sqrt{n-2}}{\sqrt{1 - \hat{\rho}_{XY}^2}} \sim t_m \quad (12.39)$$

mit $m = n - 2$ gewählt.

12.328

4. Bestimmung von K : Ausgehend von

$$P(|U| \geq t_{1-\frac{\alpha}{2};m} | H_0) = \alpha$$

ergibt sich

$$\hat{\rho}_{XY} \geq t_{1-\frac{\alpha}{2};m} \frac{1}{\sqrt{n-2 + t_{1-\frac{\alpha}{2};m}^2}} \quad \text{und} \quad \hat{\rho}_{XY} \leq -t_{1-\frac{\alpha}{2};m} \frac{1}{\sqrt{n-2 + t_{1-\frac{\alpha}{2};m}^2}}. \quad (12.40)$$

5. Für eine konkrete Stichprobe $(X_i, Y_i), i = 1, \dots, n$, vom Umfang n ergibt sich

$$u = \frac{r_{XY} \sqrt{n-2}}{\sqrt{1 - r_{XY}^2}}. \quad (12.41)$$

6. Entscheidung:

H_0 wird abgelehnt, falls $|u| \geq t_{1-\frac{\alpha}{2};m}$. H_0 wird nicht abgelehnt, falls $|u| < t_{1-\frac{\alpha}{2};m}$.

Beispiel 12.9. Siehe Vorlesung

12.329**Vorgehen**

1. $H_0 : \rho_{XY} = \rho$.

2. Vorgabe der Irrtumswahrscheinlichkeit α .
3. Zu einer mathematischen Stichprobe (X_i, Y_i) vom Umfang n einer normalverteilten Grundgesamtheit (X, Y) wird die Transformation

$$W = \frac{1}{2} \ln \left(\frac{1 + \hat{\rho}_{XY}}{1 - \hat{\rho}_{XY}} \right) \quad (12.42)$$

verwendet. Es gilt

$$E(W) = \mu_W = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) + \frac{\rho}{2(n-1)}, \quad (12.43)$$

$$\text{Var}(W) = \sigma_W^2 = \frac{1}{n-3}. \quad (12.44)$$

Als Prüfgröße wird

$$U = \frac{W - \mu_W}{\sigma_W} \sim \mathcal{N}(0, 1) \quad (12.45)$$

gewählt.

12.330

4. Bestimmung von K : Ausgehend von

$$P(|U| \geq z_{1-\frac{\alpha}{2}} | H_0) = \alpha$$

ergibt sich:

$$W \geq \mu_W + \sigma_W z_{1-\frac{\alpha}{2}} \quad \text{und} \quad W \leq \mu_W + \sigma_W z_{1-\frac{\alpha}{2}}. \quad (12.46)$$

5. Für eine konkrete Stichprobe (X_i, Y_i) , $i = 1, \dots, n$, vom Umfang n ergibt sich

$$u = \frac{w - \mu_w}{\sigma_w} \quad (12.47)$$

6. Entscheidung:

H_0 wird abgelehnt, falls $|u| \geq z_{1-\frac{\alpha}{2}}$. H_0 wird nicht abgelehnt, falls $|u| < z_{1-\frac{\alpha}{2}}$.

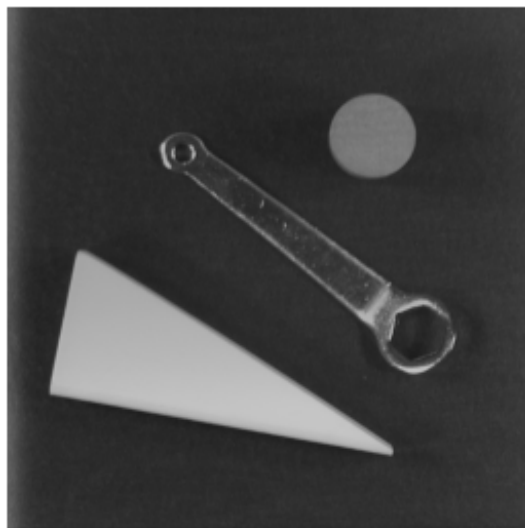
Beispiel 12.10. Siehe Vorlesung

12.331

Anhang A

Anwendungen

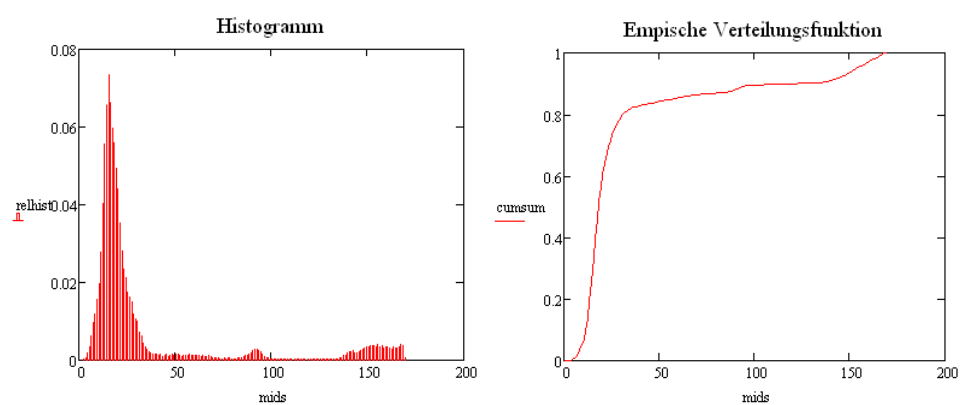
A.1 Kontrastanhebung



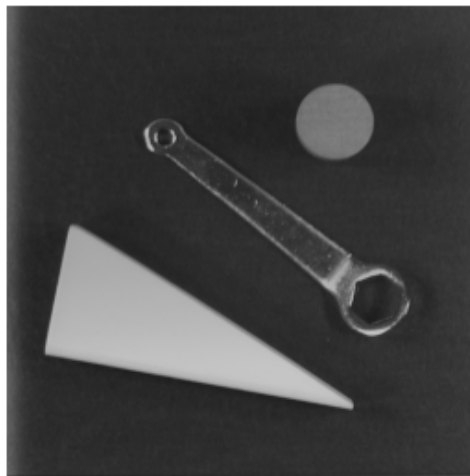
Beispiel A.1. test

Quelle: PTC: Image-Prozessing Extension Pack für Mathcad 15

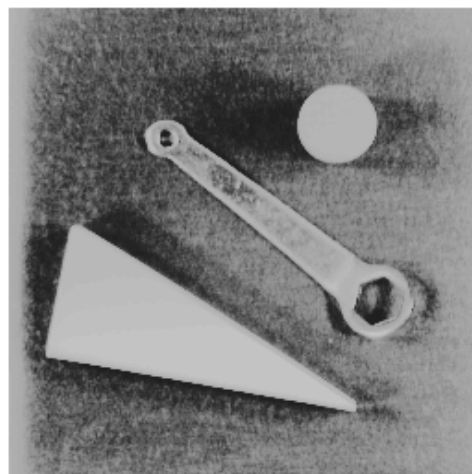
13.332



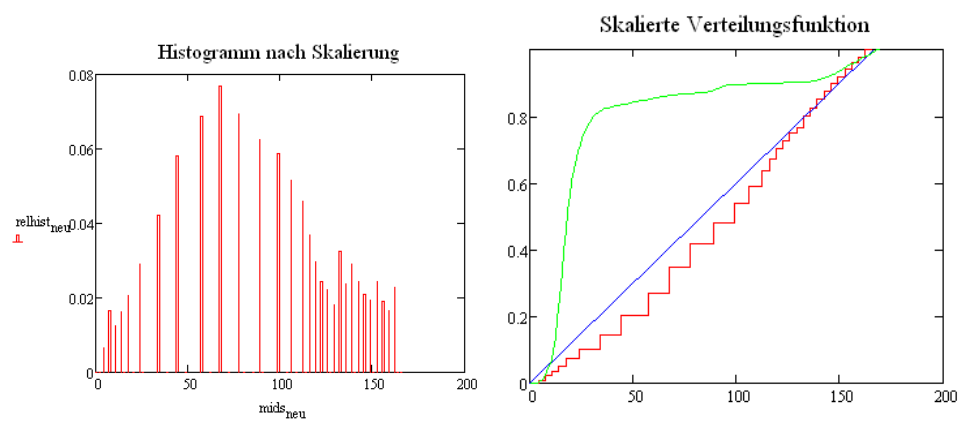
Beispiel A.2.



Beispiel A.3. test



testneu + min(test)



Beispiel A.4.

Anhang B

Integration im \mathbb{R}^n

B.1 Parameterintegrale

$I \subseteq \mathbb{R}$ sei ein Intervall und $f : [a, b] \times I \rightarrow \mathbb{R}$ eine Funktion $f(t, x)$ von zwei Variablen mit dem Definitionsbereich

$$[a, b] \times I = \{(t, x) \in \mathbb{R}^2 : t \in [a, b], x \in I\}$$

Definition B.1. Eine Funktion $F : I \rightarrow \mathbb{R}$ der Gestalt

$$F(x) = \underbrace{\int_a^b f(t, x) \, dt}_{\text{Parameterintegral}}$$

heißt eine **Integralfunktion** und das Integral ein **Parameterintegral**.

B.336

B.1.1 Eigentliche Parameterintegrale

Beispiel B.2.

$$F(x) = \int_1^2 \frac{e^{tx}}{t} \, dt$$

für $x \in \mathbb{R}$

Beispiel B.3. Die Bessel-Funktionen erster Gattung der Ordnung n .

$$\begin{aligned} J_{2n}(x) &= \frac{1}{\pi} \int_0^\pi \cos(x \sin t) \cos(2nt) \, dt \\ J_{2n+1}(x) &= \frac{1}{\pi} \int_0^\pi \sin(x \sin t) \sin((2n+1)t) \, dt \end{aligned}$$

für $n \in \mathbb{Z}$ und $x \in \mathbb{R}$.

B.337

B.1.2 Uneigentliche Parameterintegrale

Beispiel B.4. Die Gamma-Funktion

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt$$

für $x \in (0, \infty)$.

Beispiel B.5. Gesucht ist eine Parabel p mit Scheitel im Ursprung, für die die integrierte quadratische Abweichung von $q(x) = x^4$ im Intervall $[-1, 1]$ minimal wird.

B.338

Frage

Kann man die Ableitung der Funktion

$$F(x) = \int_a^b f(t, x) dt$$

dadurch berechnen, dass man den Integranden nach der Variablen x differenziert?

B.339

B.2 Stetigkeit und Differenzierbarkeit von Funktionen zweier Variablen

Sei $D \subseteq \mathbb{R}^2$ und $f : D \rightarrow \mathbb{R}$ eine Funktion.

Für die Elemente des \mathbb{R}^2 schreiben wir \mathbf{x} oder (x, y) und für die Funktionswerte $f(\mathbf{x})$ oder $f(x, y)$.

Für $\mathbf{x} = (x, y)$ bezeichnet $\|\mathbf{x}\| = \sqrt{x^2 + y^2}$ die euklidische Norm.

Die Stetigkeit einer Funktion $f(x, y)$ wird wie bei Funktionen einer reellen Variablen definiert, wobei der Abstand zweier Punkte mit der euklidischen Norm gemessen wird.

Definition B.6. Eine Funktion $f : D \rightarrow \mathbb{R}$ heißt an der Stelle $\mathbf{x}_0 \in D$ **stetig**, wenn es zu jedem $\varepsilon > 0$ ein $\delta = \delta(\varepsilon) > 0$ gibt, so dass für alle $\mathbf{x} \in D$ mit $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ gilt:

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| < \varepsilon$$

Definition B.7. Eine Folge (\mathbf{x}_n) von Elementen des \mathbb{R}^2 konvergiert gegen $\mathbf{x}_0 \in \mathbb{R}^2$ wenn

$$\lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{x}_0\| = 0$$

Man schreibt dann wieder $\mathbf{x}_0 = \lim_{n \rightarrow \infty} \mathbf{x}_n$.

Satz B.8. Die Folge der $\mathbf{x}_n = (x_{1n}, x_{2n})$ konvergiert genau dann gegen $\mathbf{x}_0 = (x_0, y_0)$, wenn

$$\lim_{n \rightarrow \infty} x_n = x_0 \quad \text{und} \quad \lim_{n \rightarrow \infty} y_n = y_0$$

Satz B.9. Eine Funktion $f : D \rightarrow \mathbb{R}$ auf einer Teilmenge $D \subseteq \mathbb{R}^2$ ist im Punkt $\mathbf{x}_0 \in D$ genau dann stetig, wenn für alle Folgen von Elementen $\mathbf{x}_n \in D$ mit $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}_0$ gilt

$$\lim_{n \rightarrow \infty} f(\mathbf{x}_n) = f(\mathbf{x}_0)$$

Bezeichnung:

Eine Funktion f heißt auf einer Menge D stetig, wenn sie in jedem Punkt $\mathbf{x} \in D$ stetig ist.

Partielle Ableitungen

I, J seien Intervalle,

$$D := I \times J = \{(x, y) \in \mathbb{R}^2 : x \in I \text{ und } y \in J\}$$

und $f : D \rightarrow \mathbb{R}$ eine Funktion in den Variablen x und y .

Definition B.10. Die Funktion f heißt im Punkt (x_0, y_0) **partiell nach x differenzierbar**, wenn die Funktion einer reellen Variablen

$$x \mapsto f(x, y_0)$$

(bei festem y_0) an der Stelle x_0 differenzierbar ist.

Die Ableitung dieser Funktion an der Stelle x_0 wird mit

$$f_{,x}(x_0, y_0) \quad \text{oder} \quad \frac{\partial f}{\partial x}(x_0, y_0)$$

bezeichnet und heißt die **partielle Ableitung von f nach der Variablen x** an der Stelle (x_0, y_0) .

Analog definiert man die partielle Ableitung von f nach der Variablen y ($f_{,y}(x_0, y_0)$).

Beispiel B.11. Langskript:

Für $x \in \mathbb{R}$ und $y \in (0, \infty)$ sei

$$f(x, y) = \frac{e^{xy}}{y}$$

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \left(\frac{\partial}{\partial x} e^{xy} \right) \frac{1}{y} \\ &= (y e^{xy}) \frac{1}{y} \\ &= e^{xy} \end{aligned}$$

Eigenschaften eigentlicher Parameterintegrale

Satz B.12. $f : [a, b] \times I \longrightarrow \mathbb{R}$ sei eine gegebene Funktion, und $F : I \rightarrow \mathbb{R}$ bezeichne die zugehörige Integralfunktion. Dann gilt:

- (i) Ist f auf $[a, b] \times I$ stetig, so ist F auf I stetig.
- (ii) Ist f auf $[a, b] \times I$ stetig, existiert auf $[a, b] \times I$ die partielle Ableitung $\frac{\partial f}{\partial x}$ und ist diese dort stetig, so ist F auf I differenzierbar und es gilt:

$$F'(x) = \frac{dF}{dx}(x) = \int_a^b \frac{\partial f(t, x)}{\partial x} dt$$

Beispiel B.13. $F(x) = \int_1^2 \frac{e^{tx}}{t} dt$ für $x \neq 0$

(F läßt sich nicht analytisch integrieren!)

$$\begin{aligned} F'(x) &= \int_1^2 \frac{\partial}{\partial x} \frac{e^{tx}}{t} dt = \int_1^2 \left(\frac{\partial}{\partial x} e^{tx} \right) \frac{1}{t} dt \\ &= \int_1^2 (te^{tx}) \frac{1}{t} dt = \int_1^2 e^{tx} dt \\ &= \left[\frac{1}{x} e^{tx} \right]_{t=1}^{t=2} = \frac{1}{x} (e^{2x} - e^x) \end{aligned}$$

B.340

Schließlich untersuchen wir auch Parameterintegrale, bei denen die Integrationsgrenzen variabel sind.

Satz B.14. $I \subset \mathbb{R}$ sei ein gegebenes Intervall,

$\varphi : I \rightarrow \mathbb{R}$ und $\psi : I \rightarrow \mathbb{R}$ seien auf I stetig differenzierbare Funktionen,

und f sei eine gegebene reellwertige Funktion, die auf der Menge

$$M := \{(t, x) \in \mathbb{R} \times I : t \in [\varphi(x), \psi(x)] \text{ oder } t \in [\psi(x), \varphi(x)]\}$$

stetig ist und dort eine stetige partielle Ableitung $\frac{\partial f}{\partial x}$ besitzt.

Dann ist die Integralfunktion $F : I \rightarrow \mathbb{R}$ mit

$$F(x) = \int_{\varphi(x)}^{\psi(x)} f(t, x) dt$$

differenzierbar auf I und es gilt

$$F'(x) = \int_{\varphi(x)}^{\psi(x)} \frac{\partial f(t,x)}{\partial x} dt + \psi'(x) \cdot f(\psi(x), x) - \varphi'(x) \cdot f(\varphi(x), x)$$

B.341

Beispiel B.15. **Langskript:**

$$F(x) = \int_0^{x^2} \cos(t^2 x) dt$$

$$\begin{aligned} F'(x) &= \int_0^{x^2} \frac{\partial}{\partial x} \cos(t^2 x) dt + 2x \cos(x^4 x) - 0 \\ &= - \int_0^{x^2} t^2 \sin(t^2 x) dt + 2x \cos(x^5) \end{aligned}$$

B.342

B.2.1 Uneigentliche Parameterintegrale

Bei uneigentlichen Parameterintegralen

$$F(x) = \int_0^\infty f(t, x) dt$$

muss man aufpassen! Zum Beispiel ist

$$f(x, t) = \frac{x}{1 + (xt)^2}$$

für alle x und t stetig und stetig partiell differenzierbar, trotzdem ist

$$F(x) = \int_0^\infty \frac{x}{1 + (xt)^2} dt = \arctan(xt) \Big|_{t=0}^{t=\infty} = \begin{cases} -\frac{\pi}{2} & , \quad x < 0 \\ 0 & , \quad x = 0 \\ \frac{\pi}{2} & , \quad x > 0 \end{cases}$$

unstetig an der Stelle 0 und damit auch nicht differenzierbar.

B.343

Voraussetzung dafür, dass man bei uneigentlichen Integralen die Ableitung und die Integration vertauschen darf, ist die **gleichmäßige Konvergenz** der Integrale. Wir definieren diese Eigenschaft nicht genauer, sondern geben eine hinreichende Bedingung an, unter der sie erfüllt ist. Außerdem beschränken wir uns bei der Formulierung auf uneigentliche Integrale auf dem Intervall $[0, \infty)$. Für andere Intervalle gilt der folgende Satz entsprechend.

Satz B.16. Die Funktion $f(t, x)$ sei auf der Menge

$$M := \{(t, x) : 0 \leq t < \infty, a \leq x \leq b\}$$

stetig und stetig partiell nach x differenzierbar.

Es gebe Funktionen $g(t)$ und $h(t)$, die in $[0, \infty)$ uneigentlich integrierbar sind, so dass für alle $(t, x) \in M$ gilt

$$|f(t, x)| \leq g(t) \quad \text{und} \quad |f_{,x}(t, x)| \leq h(t).$$

Dann konvergiert für alle $x \in [a, b]$ das uneigentliche Integral

$$F(x) = \int_0^\infty f(t, x) \, dt$$

und die dadurch definierte Funktion $F(x)$ ist differenzierbar mit der Ableitung

$$F'(x) = \int_0^\infty f_{,x}(t, x) \, dt$$

B.344

Beispiel B.17. Langskript:

$$F(x) = \int_0^\infty e^{-t^2} \cos(xt) \, dt$$

Schritt 1: g und h bestimmen Wegen $|\cos(xt)| \leq 1$ und $|\sin(xt)| \leq 1$ ist für $f(t, x) = e^{-t^2} \cos(xt)$

$$|f(t, x)| \leq e^{-t^2} =: g(t)$$

und für $f_{,x}(t, x) = e^{-t^2}(-t \sin(xt))$

$$|f_{,x}(t, x)| \leq t e^{-t^2} =: h(t)$$

mit uneigentlich integrierbaren Funktionen $g(t)$ und $h(t)$.

Schritt 2: Berechnung $F'(x)$ Man kann also den Satz B.16 anwenden. Unter Verwendung partieller Integration erhält man

$$\begin{aligned} F'(x) &= \int_0^\infty (-t e^{-t^2}) \sin(xt) \, dt \\ &= \int_0^\infty \left(\frac{\partial}{\partial t} \frac{1}{2} e^{-t^2} \right) \sin(xt) \, dt \\ &= \left[\frac{1}{2} e^{-t^2} \sin(xt) \right]_{t=0}^{t=\infty} - \int_0^\infty \frac{1}{2} e^{-t^2} \frac{\partial}{\partial t} \sin(xt) \, dt \\ &= 0 - \int_0^\infty \frac{1}{2} e^{-t^2} (x \cos(xt)) \, dt \\ &= -\frac{x}{2} \int_0^\infty e^{-t^2} \cos(xt) \, dt \\ &= -\frac{x}{2} F(x) \end{aligned}$$

Die Integralfunktion $F(x)$ ist also Lösung der *Differentialgleichung*

$$F'(x) = -\frac{x}{2}F(x)$$

Schritt 3: Bestimmung von $F'(x)$ Unter der Annahme, dass $F(x) \neq 0$ ist, ergibt dies

$$\begin{aligned}\frac{F'(x)}{F(x)} &= -\frac{x}{2} \\ (\ln |F(x)|)' &= -\frac{x}{2} \\ \ln |F(x)| &= -\frac{x^2}{4} + c \\ |F(x)| &= e^{-\frac{x^2}{4} + c} = e^{-\frac{x^2}{4}} e^c \\ F(x) &= C e^{-\frac{x^2}{4}}\end{aligned}$$

Es gilt:

$$F(0) = \int_0^\infty e^{-t^2} dt = \sqrt{\pi}$$

woraus sich der Wert der noch offenen Konstanten C ergibt:

$$F(x) = \sqrt{\pi} e^{-\frac{x^2}{4}}$$

B.345

B.3 Integration von Funktionen mehrerer Variablen

Es stellt sich zunächst die Frage, ob Integrale mehrerer Veränderlicher in den Anwendungen verwendet werden. Eine kleine Liste von Anwendungen ist zumindest ein Indiz dafür, dass wir uns damit beschäftigen sollten.

Beispiel für mehrdimensionale Integrale

Anwendungen Integration über 2 Veränderliche

- Flächeninhalte
- Schwerpunkte
- Volumen, Durchflussmengen
- Flächenträgheitsmomente
- Energien
- ...

B.3.1 Zweidimensionale Integrale

Idee

Im \mathbb{R}^2 betrachten wir ein Rechteck der Form

$$I := \left\{ (x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d \right\}.$$

$f : I \rightarrow \mathbb{R}$ sei eine beschränkte Funktion. Zur Integration von f auf I gehen wir wie folgt vor:

1. Wir bilden eine Zerlegung Z_1 von $[a, b]$ in $n + 1$ Teilpunkte

$$a = x_0 < x_1 < \dots < x_n = b$$

und eine Zerlegung Z_2 von $[c, d]$ in $m + 1$ Teilpunkte

$$c = y_0 < y_1 < \dots < y_m = d.$$

Wir bezeichnen

$$\|Z_1\| := \max_{1 \leq i \leq n} (x_i - x_{i-1}), \quad \|Z_2\| := \max_{1 \leq j \leq m} (y_j - y_{j-1})$$

als Normen dieser Zerlegungen.

2. Wir wählen ein $\bar{x}_i \in [x_{i-1}, x_i]$ und ein $\bar{y}_j \in [y_{j-1}, y_j]$ und bilden die Riemann'sche Summe

$$S_{Z_1, Z_2} = \sum_{i=1}^n \sum_{j=1}^m f(\bar{x}_i, \bar{y}_j) \cdot (x_i - x_{i-1}) \cdot (y_j - y_{j-1}).$$

$f(\bar{x}_i, \bar{y}_j) \cdot (x_i - x_{i-1}) \cdot (y_j - y_{j-1})$ beschreibt das Volumen des Quaders mit den Grundseiten $x_i - x_{i-1}$, $y_j - y_{j-1}$ und der Höhe $f(\bar{x}_i, \bar{y}_j)$.

3. Wir gehen zu immer feineren Zerlegungen über, d.h. $\|Z_1\| \rightarrow 0$ und $\|Z_2\| \rightarrow 0$.

B.346

Definition B.18 (Norm der Zerlegung). Z ist eine Zerlegung eines Intervalls $[a, b]$ mit $a = x_0 < x_1 < \dots < x_n = b$. Dann heißt

$$\|Z\| := \max_{1 \leq i \leq n} (x_i - x_{i-1})$$

die **Norm der Zerlegung** Z .

B.347

Definition B.19 (Integrierbarkeit, bestimmtes Integral). f sei eine auf I beschränkte reellwertige Funktion. Streben die Riemann'schen Summen S_{Z_1, Z_2} für alle Zerlegungen Z_1 von $[a, b]$ mit $\|Z_1\| \rightarrow 0$ und alle Zerlegungen Z_2 von $[c, d]$ mit $\|Z_2\| \rightarrow 0$ sowie jede Wahl der Zwischenstellen (\bar{x}_i, \bar{y}_j) einem gemeinsamen Grenzwert S zu, so heißt f **integrierbar** auf I . Die Zahl S heißt **bestimmtes Integral** über f auf I . Wir schreiben

$$\begin{aligned} S &= \lim_{\substack{\|Z_1\| \rightarrow 0 \\ \|Z_2\| \rightarrow 0}} \sum_{i=1}^n \sum_{j=1}^m f(\bar{x}_i, \bar{y}_j) \cdot (x_i - x_{i-1}) \cdot (y_j - y_{j-1}) \\ &= \iint_I f(x, y) d(x, y). \end{aligned}$$

B.348

Definition B.20. Eine nichtleere Teilmenge G des \mathbb{R}^2 heißt **beschränkt**, falls es ein Rechteck

$$I := \left\{ (x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d \right\}$$

gibt mit $G \subset I$.

B.349

Definition B.21. $G \subset \mathbb{R}^2$ sei eine beschränkte Menge (mit $G \subset I$), und $f : G \rightarrow \mathbb{R}$ sei eine beschränkte Funktion. Existiert für die Erweiterungsfunktion f_I von f auf I mit

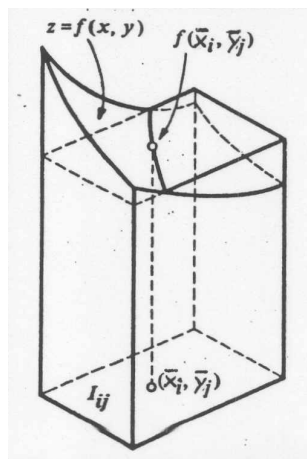
$$f_I(x, y) := \begin{cases} f(x, y) & , \text{ falls } (x, y) \in G \\ 0 & , \text{ falls } (x, y) \in I \setminus G \end{cases}$$

das Integral

$$S = \iint_I f_I(x, y) d(x, y),$$

so heißt f **integrierbar** auf G , und wir setzen

$$S = \iint_G f(x, y) d(x, y).$$



Quelle: Meyberg, K., Vachenaue, P.: Höhere Mathematik 2. 4. Aufl., 2003

Bemerkung

Gilt für die Funktion $f : G \rightarrow \mathbb{R}$

$$f(x, y) \geq 0 \quad \forall (x, y) \in G,$$

so beschreibt das Integral $\iint_G f(x, y) d(x, y)$ das Volumen des Raumstücks über G und unter der Fläche $z = f(x, y)$.

Schreibweise

$$S = \iint_G f(x, y) d(x, y) = \iint_G f(x, y) dG.$$

Satz B.22 (Linearität des Integrals). f und g seien integrierbar auf $G \subset \mathbb{R}^2$. Dann ist für alle $\alpha, \beta \in \mathbb{R}$ die Funktion $\alpha f + \beta g$ ebenfalls integrierbar auf G , und es gilt

$$\iint_G [\alpha f(x, y) + \beta g(x, y)] d(x, y) = \alpha \iint_G f(x, y) d(x, y) + \beta \iint_G g(x, y) d(x, y).$$

Definition B.23 (Maß). Ist $G \subset \mathbb{R}^2$ eine beschränkte Menge und ist die Funktion $f(x, y) \equiv 1$ integrierbar auf G , so heißt G **messbar**. Der Wert des Integrals

$$\mu(G) = \iint_G d(x, y)$$

heißt der zweidimensionale **Inhalt** oder das **Maß** von G .

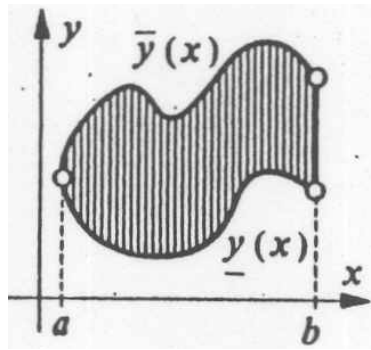
Wir wenden uns nun der Berechnung zweidimensionaler Integrale zu und benötigen zunächst folgende Definition.

Definition B.24 (y-Projizierbarkeit). G sei eine nichtleere Teilmenge des \mathbb{R}^2 . G heißt **y-projizierbar**, wenn es auf einem Intervall $[a, b]$ der x -Achse stetige Funktionen $\underline{y}(x)$ und $\bar{y}(x)$ mit

$$\underline{y}(x) \leq \bar{y}(x) \quad \forall x \in [a, b]$$

so gibt, dass gilt

$$G = \left\{ (x, y) \mid x \in [a, b], \underline{y}(x) \leq y \leq \bar{y}(x) \right\}.$$



Quelle: Meyberg, K., Vachenaue, P.: Höhere Mathematik 2. 4. Aufl., 2003

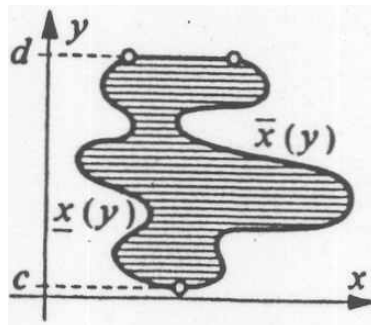
B.354

Definition B.25 (x -Projizierbarkeit). G sei eine nichtleere Teilmenge des \mathbb{R}^2 . G heißt **x -projizierbar**, wenn es auf einem Intervall $[c, d]$ der y -Achse stetige Funktionen $\underline{x}(y)$ und $\bar{x}(y)$ mit

$$\underline{x}(y) \leq \bar{x}(y) \quad \forall y \in [c, d]$$

so gibt, dass gilt

$$G = \left\{ (x, y) \mid y \in [c, d], \underline{x}(y) \leq x \leq \bar{x}(y) \right\}.$$



Quelle: Meyberg, K., Vachenaue, P.: Höhere Mathematik 2. 4. Aufl., 2003

B.355

Definition B.26 (projizierbar). G sei eine nichtleere Teilmenge des \mathbb{R}^2 . G heißt **projizierbar**, falls G y -projizierbar oder x -projizierbar ist.

Definition B.27 (Standardmenge). G sei eine nichtleere Teilmenge des \mathbb{R}^2 . G heißt **Standardmenge** im \mathbb{R}^2 , falls G sowohl y -projizierbar als auch x -projizierbar ist.

Bemerkung

Es gibt Mengen $G \subset \mathbb{R}^2$, die weder y - noch x -projizierbar sind.

B.356

Satz B.28. $G \subset \mathbb{R}^2$ sei eine projizierbare Menge, und $f : G \rightarrow \mathbb{R}$ sei eine stetige Funktion. Dann existiert das Integral $\iint_G f(x,y) d(x,y)$, und es gilt,

(i) falls G **y-projizierbar** ist:

$$\iint_G f(x,y) d(x,y) = \int_a^b \left[\int_{\underline{y}(x)}^{\bar{y}(x)} f(x,y) dy \right] dx;$$

(ii) falls G **x-projizierbar** ist:

$$\iint_G f(x,y) d(x,y) = \int_c^d \left[\int_{\underline{x}(y)}^{\bar{x}(y)} f(x,y) dx \right] dy.$$

Feinheit bei der Definition: projizierbare Menge impliziert Stetigkeit von f auf dem abg. Intervall

B.357

Beispiel B.29. Langskript:

(i) Bestimmung des Flächeninhaltes und des Schwerpunktes des Dreiecks

$$D := \left\{ (x,y) \in \mathbb{R}^2 : x \in [0,b], 0 \leq y \leq b - \frac{b}{a}x \right\}.$$

(ii) Berechne $\iint_G (x^2 + y^2) d(x,y)$ für

$$G := \{ (x,y) \in \mathbb{R}^2 : x \in [0,2], x^2 \leq y \leq 4 \}.$$

(iii) Berechne Volumen I , das durch das Paraboloid $f(x,y) = 1 - x^2 - y^2$ und die (x,y) -Ebene begrenzt wird. G ist der Einheitskreis in der (x,y) -Ebene.

Nichtprojizierbare Mengen

Ist $G \subset \mathbb{R}^2$ nicht projizierbar, aber zerlegbar in N projizierbare Mengen G_1, \dots, G_N , so gilt

$$\iint_G f(x,y) d(x,y) = \iint_{G_1} f(x,y) d(x,y) + \dots + \iint_{G_N} f(x,y) d(x,y).$$

B.358

B.3.2 Substitutionsregel für zweidimensionale Integrale

Wir wenden uns nun der **Substitutionsregel** für zweidimensionale Integrale zu. Zur Bestimmung von

$$\iint_G f(x,y) d(x,y)$$

substituiert man $x = x(u,v)$ und $y = y(u,v)$. Dabei seien x und y stetige Funktionen, die auf einer Menge H der (u,v) -Ebene stetige partielle Ableitungen erster Ordnung besitzen.

Ziel

Bestimme

$$\iint_G f(x,y) d(x,y)$$

mit $x = x(u,v)$ und $y = y(u,v)$.

Nötige Eigenschaften von x,y

- $T : H \rightarrow G$ wobei $H, G \subset \mathbb{R}^2$ gilt, $\begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} x(u,v) \\ y(u,v) \end{pmatrix}$,
- x,y sind stetig in u und v ,
- alle partiellen Ableitungen erster Ordnung von x und y existieren und sind stetig auf H in der (u,v) -Ebene,
- T ist bijektiv zwischen H und G .

B.359

Satz B.30 (Transformationssatz). *Jedem $(u_0, v_0) \in H$ wird genau ein $(x(u_0, v_0), y(u_0, v_0)) \in G$ zugeordnet. Für die Jacobi'sche Funktionalmatrix*

$$J_T(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \begin{pmatrix} \frac{\partial x(u,v)}{\partial u} & \frac{\partial x(u,v)}{\partial v} \\ \frac{\partial y(u,v)}{\partial u} & \frac{\partial y(u,v)}{\partial v} \end{pmatrix} = \begin{pmatrix} x_{,u}(u,v) & x_{,v}(u,v) \\ y_{,u}(u,v) & y_{,v}(u,v) \end{pmatrix}$$

gelte $\det \left(\frac{\partial(x,y)}{\partial(u,v)} \right) \neq 0$ für alle $(u,v) \in H$. Dann gilt

$$\boxed{\iint_G f(x,y) d(x,y) = \iint_H f(x(u,v), y(u,v)) \cdot \left| \det \left(\frac{\partial(x,y)}{\partial(u,v)} \right) \right| d(u,v).}$$

Beispiel B.31. Langskript: Polarkoordinaten: Berechne $\iint_G e^{x^2+y^2} d(x,y)$.

$$G := \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 2\}$$

$$H := \{(r, \varphi) \in \mathbb{R}^2 : 1 \leq r \leq 2, \varphi \in [0, 2\pi]\}$$

B.360

B.3.3 Dreidimensionale Integrale

Wie für zweidimensionale Integrale verwenden wir folgende

Schreibweise

$$S = \iiint_G f(x,y,z) d(x,y,z)$$

Bemerkung

Zur Berechnung werden zunächst projizierbare Mengen im \mathbb{R}^3 eingeführt.

B.361

Integrale über Funktionen $f(x,y,z)$ mit drei Variablen werden ähnlich wie im zweidimensionalen Fall definiert. Zur Berechnung von $\iiint_G f(x,y,z) d(x,y,z)$ führen wir zunächst projizierbare Mengen im \mathbb{R}^3 ein.

Definition B.32. G sei eine nichtleere Teilmenge des \mathbb{R}^3 .

- a) G heißt **z-projizierbar**, wenn es eine projizierbare Menge G_z in der (x,y) -Ebene und auf G_z stetige Funktionen $\underline{z}(x,y)$ und $\bar{z}(x,y)$ mit

$$\underline{z}(x,y) \leq \bar{z}(x,y) \quad \forall (x,y) \in G_z$$

so gibt, dass gilt

$$G = \{(x,y,z) \in \mathbb{R}^3 \mid (x,y) \in G_z, \underline{z}(x,y) \leq z \leq \bar{z}(x,y)\}.$$

- b) G heißt **y-projizierbar**, wenn es eine projizierbare Menge G_y in der (x,z) -Ebene und auf G_y stetige Funktionen $\underline{y}(x,z)$ und $\bar{y}(x,z)$ mit

$$\underline{y}(x,z) \leq \bar{y}(x,z) \quad \forall (x,z) \in G_y$$

so gibt, dass gilt

$$G = \{(x,y,z) \in \mathbb{R}^3 \mid (x,z) \in G_y, \underline{y}(x,z) \leq y \leq \bar{y}(x,z)\}.$$

- c) G heißt **x-projizierbar**, wenn es eine projizierbare Menge G_x in der (y, z) -Ebene und auf G_x stetige Funktionen $\underline{x}(y, z)$ und $\bar{x}(y, z)$ mit

$$\underline{x}(y, z) \leq \bar{x}(y, z) \quad \forall (y, z) \in G_x$$

so gibt, dass gilt

$$G = \left\{ (x, y, z) \in \mathbb{R}^3 \mid (y, z) \in G_x, \underline{x}(y, z) \leq x \leq \bar{x}(y, z) \right\}.$$

- d) G heißt **projizierbar**, falls G z -projizierbar oder y -projizierbar oder x -projizierbar ist.
- e) G heißt **Standardmenge** im \mathbb{R}^3 , falls G sowohl z -projizierbar als auch y -projizierbar und x -projizierbar ist.

Satz B.33. $G \subset \mathbb{R}^3$ sei eine projizierbare Menge, und $f : G \rightarrow \mathbb{R}$ sei eine stetige Funktion. Dann existiert das Integral $\iiint_G f(x, y, z) d(x, y, z)$ und es gilt,

(i) falls G **z -projizierbar** ist:

$$\iiint_G f(x, y, z) d(x, y, z) = \iint_{G_z} \left[\int_{\underline{z}(x, y)}^{\bar{z}(x, y)} f(x, y, z) dz \right] d(x, y);$$

(ii) falls G **y -projizierbar** ist:

$$\iiint_G f(x, y, z) d(x, y, z) = \iint_{G_y} \left[\int_{\underline{y}(x, z)}^{\bar{y}(x, z)} f(x, y, z) dy \right] d(x, z);$$

(iii) falls G **x -projizierbar** ist:

$$\iiint_G f(x, y, z) d(x, y, z) = \iint_{G_x} \left[\int_{\underline{x}(y, z)}^{\bar{x}(y, z)} f(x, y, z) dx \right] d(y, z).$$

B.362

Berechnung der Integrale

Man beachte, dass die Integrale in eckigen Klammern eindimensional und die äußeren Integrale zweidimensional sind.

Ist z.B. G **z -projizierbar** und G_z **y -projizierbar**, so folgt

$$\iiint_G f(x, y, z) d(x, y, z) = \int_a^b \left(\int_{\underline{y}(x)}^{\bar{y}(x)} \left[\int_{\underline{z}(x, y)}^{\bar{z}(x, y)} f(x, y, z) dz \right] dy \right) dx$$

Beispiel B.34. Langskript: Berechne das Volumen des Tetraeders

$$G = \left\{ (x, y, z) \in \mathbb{R}^3 \mid x, y, z \geq 0, \frac{x}{a} + \frac{y}{b} + \frac{z}{c} \leq 1 \right\}.$$

Schwerpunkt: $\iiint_G x \, d(x, y, z)$

$$G_z = \left\{ (x, y) \in \mathbb{R}^2 : x, y \geq 0, \frac{x}{a} + \frac{y}{b} \leq 1 \right\}$$

B.363

Definition B.35. Ist $G \subset \mathbb{R}^3$ eine beschränkte Menge (d.h., G ist Teilmenge eines Quaders im \mathbb{R}^3) und ist die Funktion $f(x, y, z) \equiv 1$ integrierbar auf G , so heißt G **messbar**. Der Wert des Integrals

$$\mu(G) := \iiint_G d(x, y, z)$$

heißt der dreidimensionale **Inhalt** oder das **Maß** von G .

Bemerkung

$\mu(G)$ gibt das Volumen von G an

Beispiel B.36. Langskript: Gesucht ist das Volumen der Menge

$$G = \{(x, y, z) \in \mathbb{R}^3 \mid 0 \leq x \leq 1, 0 \leq y \leq x, -y^2 \leq z \leq x^2\}.$$

B.364

B.3.4 Substitutionsregel für dreidimensionale Integrale

Transformationssatz im \mathbb{R}^3

Für zwei gegebene Mengen H und G , die bijektiv aufeinander abgebildet werden können, ist es zum Teil leichter

$$\iiint_G f(x, y, z) \, d(x, y, z)$$

zu bestimmen, als

$$\iiint_H f(u, v, w) \, d(u, v, w)$$

zu bestimmen.

B.365

Wir wenden uns nun der **Substitutionsregel** für dreidimensionale Integrale zu. Zur Bestimmung von

$$\iiint_G f(x, y, z) \, d(x, y, z)$$

substituiert man $x = x(u, v, w)$, $y = y(u, v, w)$ und $z = z(u, v, w)$. Dabei seien x, y und z stetige Funktionen, die auf einer Menge $H \subset \mathbb{R}^3$ stetige partielle Ableitungen erster Ordnung besitzen. Diese Funktionen seien so beschaffen, dass jedem $(u_0, v_0, w_0) \in H$ genau

ein $T((u_0, v_0, w_0)) = (x(u_0, v_0, w_0), y(u_0, v_0, w_0), z(u_0, v_0, w_0))^T \in G$ zugeordnet wird und umgekehrt. Für die Jacobi'sche Funktionalmatrix

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{pmatrix} = \begin{pmatrix} x_{,u} & x_{,v} & x_{,w} \\ y_{,u} & y_{,v} & y_{,w} \\ z_{,u} & z_{,v} & z_{,w} \end{pmatrix}$$

gelte $\det \left(\frac{\partial(x, y, z)}{\partial(u, v, w)} \right) \neq 0$ für alle $(u, v, w) \in H$. Dann gilt der

Satz B.37. $G \subset \mathbb{R}^3$ und $H \subset \mathbb{R}^3$ sind zwei Mengen im \mathbb{R}^3 , $T : H \rightarrow G$ bildet H bijektiv auf G ab. Es ist

$$T(u, v, w) := \begin{pmatrix} x(u, v, w) \\ y(u, v, w) \\ z(u, v, w) \end{pmatrix}.$$

T ist einmal stetig partiell differenzierbar und $\det(J_T(u, v, w)) \neq 0$ auf H .

Dann

$$\iiint_G f(x, y, z) \, d(x, y, z) = \iiint_H f(x(u, v, w), y(u, v, w), z(u, v, w)) \cdot \left| \det \left(\frac{\partial(x, y, z)}{\partial(u, v, w)} \right) \right| d(u, v, w).$$

Beispiel B.38. Langskript: Zylinderkoordinaten

$$\left. \begin{aligned} x &= r \cos \varphi \\ y &= r \sin \varphi \\ z &= t \end{aligned} \right\} \quad \begin{aligned} 0 &\leq \varphi \leq 2\pi, \, r \geq 0 \\ -\infty &< t < \infty \end{aligned}$$

$$\Rightarrow \det \left(\frac{\partial(x, y, z)}{\partial(r, \varphi, t)} \right) = \det \begin{pmatrix} \cos \varphi & -r \sin \varphi & 0 \\ \sin \varphi & r \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} = r \cos^2 \varphi + r \sin^2 \varphi = r.$$

Beispiel B.39. Langskript: Kugelkoordinaten

$$\left. \begin{aligned} x &= r \sin \varphi \cos \vartheta \\ y &= r \sin \varphi \sin \vartheta \\ z &= r \cos \varphi \end{aligned} \right\} \quad 0 \leq \varphi \leq \pi, \, 0 \leq \vartheta \leq 2\pi, \, r \geq 0$$

$$\begin{aligned}\Rightarrow \det\left(\frac{\partial(x,y,z)}{\partial(r,\varphi,\vartheta)}\right) &= \det\begin{pmatrix} \sin\varphi\cos\vartheta & r\cos\varphi\cos\vartheta & -r\sin\varphi\sin\vartheta \\ \sin\varphi\sin\vartheta & r\cos\varphi\sin\vartheta & r\sin\varphi\cos\vartheta \\ \cos\varphi & -r\sin\varphi & 0 \end{pmatrix} \\ &= \cos\varphi(r^2\cos\varphi\cos^2\vartheta\sin\varphi + r^2\cos\varphi\sin^2\vartheta\sin\varphi) \\ &\quad + r\sin\varphi(r\sin^2\varphi\cos^2\vartheta + r\sin^2\varphi\sin^2\vartheta) \\ &= r^2\cos^2\varphi\sin\varphi(\cos^2\vartheta + \sin^2\vartheta) + r^2\sin^3\varphi(\cos^2\vartheta + \sin^2\vartheta) \\ &= r^2\cos^2\varphi\sin\varphi + r^2\sin^3\varphi = r^2\sin\varphi(\cos^2\varphi + \sin^2\varphi) \\ &= r^2\sin\varphi.\end{aligned}$$

Index

p -Quantil, 20

Abbildung
messbare, 51

Bindung, 17
Binomial-Verteilung, 53
Binomialverteilung
negative, 55

Datensatz, 16
Maximum, 17
Minimum, 17
Durchschnitt, 19

Ereignis, 30
Ereignissystem, 30

Funktion
messbare, B-10

Gebiet
projizierbares, B-11
 x -projizierbares, B-11
 y -projizierbares, B-10
Grenzwertsatz
zentraler, 54

Häufigkeit
relative, 10
Häufigkeitsverteilung, 18
Histogramm, 18

Integral
bestimmtes, B-9
integrierbar, B-9
Integrierbarkeit, B-9
Integral
Linearität, B-10

Klassen, 18
Klassenmitte, 18
Korrelationskoeffizient
empirische, 24
Kovarianz

empirische, 24

Maß
einer Menge, B-16
Maß, B-10
Messraum, 51
Mittel, 19
arithmetisches, 19
geometrisches, 19
gewichtetes, 20
gewichtetes arithmetisches, 20
gewichtetes geometrisches, 20
gewichtetes harmonisches, 20
harmonisches, 19

Norm
der Zerlegung, B-8

Ordnungstatistik, 17

Perzentil, 21
projizierbar, B-11

Quantil, 20
Quantile-Quantile-Plot, 23
Quartil, 20

Rangwertreihe, 17
Regressionsgerade, 25

Spannweite, 19
Standardabweichung, 22
Standardmenge, B-11
Stichprobe, 16
Streuung, 22
empirische, 22

Transformationssatz, B-13

Urliste, 16

Varianz, 22
empirische, 22, 24

Verteilung
hypergeometrische, 53

negative Binomial-, 55
Verteilungsfunktion
empirische, 19
x-projizierbar, B-11
y-projizierbar, B-10