Wahrscheinlichkeitsrechnung und Statistik

Für Studierende der Physik

Rudolf Frühwirth

Rudolf Frühwirth

Wahrscheinlichkeitsrechnung und Statistik für Studierende der Physik

Wahrscheinlichkeitsrechnung und Statistik für Studierende der Physik
1. Auflage

ISBN 978-87-403-0847-1

Inhalt

Vorwort

In der modernen Experimentalphysik ist die Anwendung von statistischen Methoden ein wichtiges, wenn nicht das wichtigste Werkzeug zur Analyse von experimentellen Daten. Die Datenanalyse benützt Elemente der beschreibenden oder deskriptiven Statistik ebenso wie das umfangreiche Arsenal der schließenden oder Inferenzstatistik. Während sich die deskriptive Statistik im Rahmen des Experiments meist auf graphische Präsentation und Berechnung von Kennzahlen von großen Datensätzen beschränkt, werden in der Inferenzstatistik sehr häufig, wenn auch keineswegs immer, stochastische Modelle verwendet, die die Zufälligkeit der experimentellen Resultate mit Begriffen der Wahrscheinlichkeitsrechnung erfassen und beschreiben.

Wir verwenden hier bewusst nicht den Ausdruck "Wahrscheinlichkeitstheorie", denn diese ist ohne die Grundlage der mathematischen Maßtheorie nicht denkbar. Um jedoch die Darstellung nicht mit mathematischen Begriffen und Sätzen zu überfrachten und Studierende der Physik nicht von Anfang an abzuschrecken, verzichten wir in diesem Buch gänzlich auf Maßtheorie. Grundkenntnisse der Differential- und Integralrechnung sowie der Vektor- und Matrizenrechnung werden jedoch vorausgesetzt.

Der Verzicht auf Maßtheorie hat zur Folge, dass grundlegende Begriffe wie der der Zufallsvariablen nicht streng mathematisch definiert, sondern lediglich anschaulich plausibel gemacht werden können. Glücklicherweise ist die statistische Behandlung von experimentellen Beobachtungen, sofern sie als (unabhängige) Stichproben vorliegen, auf dieser mathematisch unsicheren Grundlage trotzdem ohne weitere Schwierigkeiten durchzuführen. Hingegen ist die Analyse von zeitabhängigen Beobachtungen wie Zeitreihen oder stochastischen Prozessen, inbesonders in stetiger Zeit, ohne Maßtheorie nur mehr schwer bzw. gar nicht mehr möglich.

Eine Herausforderung taucht nach meiner Erfahrung in einer Vorlesung über Statistik für Studierende der Physik gleich zu Beginn auf. Bevor man nämlich mit der eigentlichen Statistik beginnen kann, müssen zuerst zahlreiche Begriffe und Resultate aus der Wahrscheinlichkeitsrechnung erarbeitet werden. Der erste Teil dieses Buches soll daher dazu dienen, den Studierenden eine leicht fassliche Einführung in jene Bereiche der Wahrscheinlichkeitsrechnung zu geben, die für die nachfolgende Darstellung der statistischen Methoden und die Untersuchung ihrer Eigenschaften von Bedeutung sind. Der zweite Teil bringt dann eine Auswahl von statistischen Methoden, die sich in der experimentellen Praxis bewährt haben oder von allgemeiner Bedeutung sind. Das Schätzen von physikalischen Größen, die Berechnung von Konfidenzintervallen und das Testen von Hypothesen sind hier die wichtigsten Themen. Angesichts der wachsenden Bedeutung von Methoden der Bayes-Statistik in der experimentellen Datenanalyse geben wir auch eine kurze Einführung in diesen wichtigen Zweig der Statistik. Selbstverständlich dürfen auch multivariate Methoden nicht zu kurz kommen. Aus diesem Grund beschließt ein Kapitel über Regression, lineare Modelle und Ausgleichsrechnung den zweiten Teil.

Am Ende von vielen Beispielen und in der Legende der meisten Abbildungen wird auf ein MATLAB-Programm hingewiesen. Dieses zeigt, wie das Beispiel gerechnet bzw. die Abbildung erzeugt wurde. Die Programme sind frei verfügbar, und ich empfehle allen Interessierten, sie herunterzuladen und wenn möglich auszuprobieren.

Abschließend möchte ich klarstellen, dass die Auswahl der statistischen Verfahren wie auch der Anwendungsbeispiele zu einem gewissen Teil von meiner praktischen Arbeit in der experimentellen Teilchenphysik geprägt ist. Dennoch hoffe ich, dass die Darstellung für Studierende in allen Bereichen der Experimentalphysik und Messtechnik von Nutzen ist.

Ich danke meinen Kollegen A. Beringer und O. Cencic sowie den Studierenden an der TU Wien V. Knünz, I. Krätschmer und T. Madlener für die sorgfältige Durchsicht des Manuskripts und für wertvolle didaktische Hinweise. Ich bitte alle Leserinnen und Leser, mich auf verbleibende Fehler und Unklarheiten aufmerksam zu machen. Für diese ist allein der Autor verantwortlich.

Wien, am 12. Jänner 2015

Anmerkungen zur Notation

Als Dezimaltrennungszeichen wird, wie international üblich, der Punkt verwendet: $\pi \approx 3.14159$

Die Eulersche Zahl wird mit e bezeichnet, die imaginäre Einheit mit i.

Mengen und Ereignisse sind aufrecht gesetzt: A, B, e_i, ...

Zufallsvariable sind in Großbuchstaben kursiv gesetzt: X, Y, ...

Zufallsvektoren sind in Großbuchstaben fett kursiv gesetzt: $\boldsymbol{X}$, $\boldsymbol{Y}$, ...

Vektoren sind in Kleinbuchstaben fett kursiv gesetzt: $\boldsymbol{x}$, $\boldsymbol{y}$, ...; der Nullvektor wird mit $\mathbf{0}$ bezeichnet.

Matrizen sind in Großbuchstaben fett aufrecht gesetzt: $\mathbf{A}$, $\mathbf{B}$, ...; die Einheitsmatrix wird mit $\mathbf{I}$, die Nullmatrix mit $\mathbf{O}$ bezeichnet.

Der transponierte Vektor von $\boldsymbol{x}$ ist $\boldsymbol{x}^{\mathrm{T}}$, die transponierte Matrix von $\mathbf{X}$ ist $\mathbf{X}^{\mathrm{T}}$.

Definitionen und Sätze sind *schräg* gesetzt. Definitionen werden mit ☞ gekennzeichnet, Sätze mit ☛.

Beispiele werden mit ▶ begonnen und mit ◀ abgeschlossen.

Auf MATLAB-Programme wird mit ➠ hingewiesen.

Teil I

Wahrscheinlichkeitsrechnung

Kapitel 1

Zufallsvariable und Ereignisse

1.1 Einleitung

Statistische Methoden werden zur Auswertung von Zufallsexperimenten im weitesten Sinn verwendet. Ein Zufallsexperiment kann eine Messung sein, es kann ein Zählvorgang sein, es kann aber auch eine Umfrage oder ein Glücksspiel sein. Das Charakteristikum eines Zufallsexperiments ist, dass das konkrete Ergebnis, der Ausgang des Experiments, nicht genau vorausgesagt werden kann. Dies kann mehrere Gründe haben:

- Mangelnde Kenntnis des Anfangszustandes und/oder der Dynamik des untersuchten Gegenstandes, wie z.B. beim Würfeln oder beim Roulettespiel.
- Die im Experiment beobachteten Objekte sind eine zufällige Auswahl (Stichprobe) aus einer größeren Grundgesamtheit, wie bei einer Umfrage.
- Der beobachtete Prozess ist prinzipiell indeterministisch, auf Grund von quantenmechanischen Effekten. Ein typisches Beispiel dafür ist der radioaktive Zerfall eines Atoms oder der Zerfall eines instabilen Elementarteilchens.
- Messfehler geben dem Ergebnis einen (teilweise) zufälligen Charakter. Dies trifft auf praktisch alle Messvorgänge zu.

Selbstverständlich kann auch eine Kombination mehrerer dieser Gründe vorliegen. Wird zum Beispiel die Lebensdauer eines Myons gemessen, treffen zwei Effekte zusammen: die Lebensdauer ist an sich eine zufällige Größe; dazu kommt noch der Messfehler der Zeitmessung.

Das konkrete Ergebnis eines Experiments wird, wenn es in Zahlen ausgedrückt wird, als die Realisierung einer *Zufallsvariablen* betrachtet. Die mathematisch exakte Definition einer Zufallsvariablen wird im Rahmen der mathematischen Maßtheorie formuliert. Wir begnügen uns hier mit der folgenden anschaulichen „Definition“:

☞ **Definition 1.1.1.** Zufallsvariable

Eine univariate Zufallsvariable X ordnet dem Ausgang eines Zufallsexperiments einen Zahlenwert zu. Wir sagen auch, dass die Zufallsvariable X das Experiment beschreibt oder repräsentiert.

☞ **Definition 1.1.2.** Zufallsvektor

Eine multivariate Zufallsvariable oder ein Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ der Dimension n ist ein Vektor von n univariaten Zufallsvariablen.

Eine Zufallsvariable sollte genau genommen von ihren Zahlenwerten, den *Realisierungen* der Zufallsvariablen, unterschieden werden. Jedoch verzichten wir der Einfachheit halber auf die Unterscheidung und schreiben X sowohl für die Zufallsvariable als auch für die Realisierung.

1.2 Merkmale

Eine univariate Zufallsvariable X wird auch als *Merkmal* bezeichnet. Die möglichen Realisierungen eines Merkmals werden dessen *Ausprägungen* genannt. Die Anzahl und Art der möglichen Ausprägungen gibt Anlass zur Klassifizierung von Merkmalen in verschiedene Merkmals- und Skalentypen.

1.2.1 Merkmalstypen

In der Praxis wird zwischen *qualitativen* Merkmalen und *quantitativen* Merkmalen unterschieden. Qualitative Merkmale sind beschreibender Natur, während quantitative Merkmale durch einen Zähl- oder Messvorgang realisiert werden. Der Übergang zwischen den beiden Typen ist fließend.

Binäre Merkmale haben nur zwei Ausprägungen. Ein typisches Beispiel ist das Merkmal "EU-Bürger", das auf eine Person zutreffen kann oder nicht. Binäre Merkmale sind qualitativ, die Kodierung durch 0 (trifft nicht zu) und 1 (trifft zu) ist üblich, aber nicht zwingend notwendig.

Kategoriale Merkmale haben mehrere, meist jedoch nur wenige Ausprägungen. Ein typisches Beispiel ist das Merkmal "Familienstand einer Person" mit den folgenden Ausprägungen: ledig=1/geschieden=2/verheiratet=3/verwitwet=4. Kategoriale Merkmale sind qualitativ, die Kodierung durch die Werte $1, \ldots, k$ ist üblich, aber nicht zwingend notwendig.

Ordinale Merkmale sind durch eine Ordnungsbeziehung zwischen den Ausprägungen gekennzeichnet. Diese können ebenfalls durch die numerischen Werte $1, \ldots, k$ kodiert werden. Ein Beispiel sind Prüfungsnoten, die die Ausprägungen $1, 2, 3, 4, 5$ haben. Die Kodierung ist konventionell, ihr ist jedoch eine Anordnung inhärent: 1 ist besser als 2, 2 ist besser als 3, usw. Ordinale Merkmale stehen am Übergang vom qualitativen zum quantitativen Typ. Prüfungsnoten etwa können durch eine Beurteilung zustande kommen (qualitativ), aber auch durch die Messung der erreichten Punktezahl (quantitativ).

Quantitative Merkmale beschreiben einen Zählvorgang oder einen Messvorgang. Im ersten Fall sind die Ausprägungen ganzzahlig; man nennt solche Merkmale auch *diskret*. Im Fall von gemessenen Merkmalen ist prinzipiell ein Kontinuum von Ausprägungen möglich; man nennt solche Merkmale daher *stetig* oder *kontinuierlich*. Die Unterscheidung ist insofern von Bedeutung, als die mathematische Behandlung der beiden Typen etwas unterschiedlich verläuft.

1.2.2 Skalentypen

Man kann den Merkmalstypen verschiedene Skalentypen zuordnen. Der Skalentypus gibt an, ob und in welchem Ausmaß arithmetische Operationen mit den Ausprägungen möglich und sinnvoll sind.

Der niedrigste Skalentypus ist die *Nominalskala*. Hier sind die Zahlenwerte nur Bezeichnung für einander ausschließende Kategorien, wie im Fall von binären oder allgemeiner kategorialen Merkmalen. Arithmetische Operationen mit diesen Zahlenwerten sind daher nicht sinnvoll.

Der nächsthöhere Skalentypus ist die *Ordinalskala*. Auf ihr ist die Ordnung der Zahlen wesentlich, wie es bei ordinalen Merkmalen der Fall ist. Differenzen von Zahlenwerten sind beschränkt interpretierbar, jedoch sind weitergehende numerische Operationen auf den Zahlenwerten in der Regel nicht sinnvoll.

Auf der nächsthöheren Skala, der *Intervallskala*, ist eine natürliche Anordnung vorhanden. Summen und Differenzen von Werten sind sinnvoll berechen- und interpretierbar. Der Nullpunkt ist auf einer solchen Skala jedoch willkürlich festgelegt, daher ist die Berechnung von Quotienten nicht sinnvoll.

Auf dem höchsten Skalentypus, der *Verhältnisskala*, können nicht nur Summen und Differenzen, sondern auch Quotienten oder Verhältnisse von Werten gebildet werden, da es einen natürlichen Nullpunkt gibt.

▶ Beispiel 1.2.1.
Einige Beispiele für die verschiedenen Skalentypen sind die folgenden:

1. Die Art des Zerfalls eines Z-Bosons wird durch Zahlen kodiert: 1=Myonen, 2=Elektronen, 3=Hadronen, 4=Neutrinos. Es liegt eine Nominalskala vor, da die Zuordnung der Zahlenwerte zu den Kategorien willkürlich ist.

2. Der Stand einer Mannschaft in der Meisterschaft wird durch den Rang in der Liga angegeben. Es liegt eine Ordinalskala vor.

3. Die Jahreszahlen (2013, 2008, ...) bilden eine Intervallskala, da der Nullpunkt willkürlich festgelegt ist. Differenzen von Jahreszahlen sind sinnvoll, Quotienten aber nicht.

4. Die Celsius-Skala der Temperatur ist eine Intervallskala, da der Nullpunkt willkürlich festgelegt ist.

5. Die Kelvin-Skala der Temperatur ist eine Verhältnisskala, da der Nullpunkt physikalisch festgelegt ist.

6. Die Größe einer Person wird in cm angegeben. Es liegt eine Verhältnisskala vor, da ein natürlicher Nullpunkt existiert. ◀

► **Beispiel 1.2.2.**
Von acht Personen werden drei Merkmale erhoben:

G... Geschlecht (1=W, 2=M);

A... Alter (in Jahren);

B... Ausbildung (1=Pflichtschule, 2=Höhere Schule, 3=Bachelor, 4=Master).

Die Resultate sind in der folgenden Datenmatrix zusammengestellt:

Nr	G	A	B
1	1	34	2
2	2	54	1
3	2	46	3
4	1	27	4
5	1	38	2
6	1	31	3
7	2	48	4
8	2	51	2

Das Merkmal G ist binär, es liegt eine Nominalskala vor. Das Merkmal A ist quantitativ auf einer Verhältnisskala. Das Merkmal B ist ordinal, es liegt daher eine Ordinalskala vor. ◀

1.3 Ereignisse

1.3.1 Ergebnismenge und Ereignisalgebra

Eine univariate Zufallsvariable X ist das numerische Resultat eines Zufallsexperiments (im weitesten Sinn). Die Menge aller möglichen Ausprägungen oder Realisierungen einer Zufallsvariablen heißt *Ergebnismenge* oder *Stichprobenraum* von X. Dieser wird mit $\Omega \subseteq \mathbf{R}$ bezeichnet. Die Ergebnismenge Ω kann *diskret*, also endlich bzw. abzählbar unendlich, oder *kontinuierlich*, also überabzählbar unendlich sein.

► **Beispiel 1.3.1.**

1. Beim Roulettespiel gibt es 37 mögliche Realisierungen der Zufallsvariablen X. Die Ergebnismenge $\Omega = \{0, 1, \ldots, 36\}$ ist diskret und endlich.

2. Wird eine radioaktive Quelle beobachtet, ist die Anzahl X der Zerfälle pro Sekunde im Prinzip unbeschränkt. Die Ergebnismenge $\Omega = \{0, 1, 2, \ldots\}$ ist diskret und abzählbar unendlich.

3. Die Wartezeit X zwischen zwei Zerfällen kann jeden beliebigen positiven Wert annehmen. Die Ergebnismenge $\Omega = \mathbf{R}_+$ ist kontinuierlich, also überabzählbar unendlich. ◀

☞ **Definition 1.3.2.** Ereignis

Es sei X eine Zufallsvariable mit der Ergebnismenge Ω. Ein Ereignis E *ist eine Teilmenge von Ω. Ist $x \in \Omega$ eine Realisierung von X, dann tritt* E *genau dann ein, wenn* E *das Ergebnis x enthält. Wir schreiben dafür auch einfach $X \in \mathrm{E}$.*

☞ **Definition 1.3.3.** Elementarereignis

Ein Elementarereignis e *ist eine einelementige Teilmenge der Ergebnismenge Ω.*

☞ **Definition 1.3.4.** Ereignisalgebra

Die Menge aller Ereignisse der Ergebnismenge Ω heißt die Ereignisalgebra $\mathfrak{S}(\Omega)$.

► **Beispiel 1.3.5.**

Die Zufallsvariable X, die den Wurf mit einem Würfel repräsentiert, hat die Ergebnismenge:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Die Ereignisalgebra besteht aus allen Teilmengen von Ω. Wie aus der elementaren Mengenlehre bekannt, gibt es insgesamt $2^6 = 64$ Ereignisse. Darunter sind sechs Elementarereignisse:

$$\mathrm{e}_1 = \{1\}, \mathrm{e}_2 = \{2\}, \mathrm{e}_3 = \{3\}, \mathrm{e}_4 = \{4\}, \mathrm{e}_5 = \{5\}, \mathrm{e}_6 = \{6\}$$

Das Ereignis G (gerade Zahl) ist die Teilmenge $\mathrm{G} = \{2, 4, 6\}$. G tritt ein, wenn eine gerade Zahl geworfen wird, also $X \in \mathrm{G}$. ◄

Allgemein kann im diskreten (endlichen oder abzählbar unendlichen) Fall *jede* Teilmenge als Ereignis betrachtet werden, und $\mathfrak{S}(\Omega)$ ist einfach die Potenzmenge* von Ω. Die Ereignisalgebra heißt ebenfalls *diskret*.

Im kontinuierlichen (überabzählbar unendlichen) Fall müssen gewisse pathologische (nicht messbare) Teilmengen ausgeschlossen werden. Diese sind jedoch ohne jegliche praktische Bedeutung. Die Ereignisalgebra heißt ebenfalls *kontinuierlich* oder *stetig*.

1.3.2 Verknüpfung von Ereignissen

Da Ereignisse nichts anderes als Teilmengen der Ergebnismenge sind, können auf sie die üblichen logischen oder Mengenoperationen angewendet werden. In Tabelle 1.1 sind die wichtigsten Operationen zusammengefasst. Mit diesen Operationen ist $\Sigma = \mathfrak{S}(\Omega)$ eine *Boole'sche Algebra*, also ein distributiver komplementärer Verband mit Null- und Einselement.† Das Nullelement $\emptyset$ ist das *unmögliche Ereignis*, da nie $X \in \emptyset$ gilt; das Einselement Ω ist das *sichere Ereignis*, da immer $X \in \Omega$ gilt.

Tabelle 1.1: Die wichtigsten Verknüpfungsoperationen.

Symbol	Logische Operation	Mengenoperation	Bedeutung
$A \cup B$	Disjunktion	Vereinigung	A oder B (oder beide)
$A \cap B$	Konjunktion	Durchschnitt	A und B (sowohl A als auch B)
A'	Negation	Komplement	nicht A (das Gegenteil von A)
$A \subseteq B$	Implikation	Teilmenge	aus A folgt B ($A' \cup B$)

▶ Beispiel 1.3.6.

Es seien $A, B, C \in \Sigma$ drei Ereignisse. Man kann mittels Verknüpfungen die folgenden Ereignisse formulieren:

1. $E_1 = A \cap B \cap C$.
   E_1 tritt genau dann ein, wenn alle drei Ereignisse eintreten.
2. $E_2 = A \cap B' \cap C$.
   E_2 tritt genau dann ein, wenn A und C eintreten, B aber nicht.
3. $E_3 = (A \cap B \cap C') \cup (A \cap B' \cap C) \cup (A' \cap B \cap C)$.
   E_3 tritt genau dann ein, wenn genau zwei der drei Ereignisse eintreten.
4. $E_4 = (A \cap B' \cap C') \cup (A' \cap B \cap C') \cup (A' \cap B' \cap C) \cup (A' \cap B' \cap C')$.
   E_4 tritt genau dann ein, wenn höchstens eines der Ereignisse eintritt. ◀

☞ **Definition 1.3.7.** Disjunkte Ereignisse

Zwei Ereignisse A *und* B *schließen einander aus, wenn* $A \cap B = \emptyset$.

Zwei Elementarereignisse schließen einander stets aus.

*http://de.wikipedia.org/wiki/Potenzmenge
†http://de.wikipedia.org/wiki/Boolesche_Algebra

1.3.3 Wiederholte Experimente und Häufigkeit

Wird ein Experiment n-mal wiederholt, so wird die Gesamtheit aller Experimente durch einen Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ repräsentiert. Jedes Ergebnis ist ein Vektor von Realisierungen $(x_1, \ldots, x_n)$. Die Ergebnismenge von $\boldsymbol{X}$ ist das n-fache kartesische Produkt $\Omega \times \ldots \times \Omega$.*

▶ **Beispiel 1.3.8.** Die Ereignisalgebra des Doppelwurfs

Der zweimalige Wurf mit einem Würfel wird durch den Zufallsvektor $\boldsymbol{X} = (X_1, X_2)$ beschrieben, wobei X_1 den ersten und X_2 den zweiten Wurf repräsentiert. Die Ergebnismenge von $\boldsymbol{X}$ ist das kartesische Produkt $\Omega \times \Omega$:

$$\Omega \times \Omega = \{(x_1, x_2) \,|\, x_1, x_2 \in \Omega\}$$

Das geordnete Paar (x_1, x_2) bedeutet: x_1 ist die Realisierung von X_1, x_2 ist die Realisierung von X_2. Die Ereignisalgebra $\mathfrak{S}(\Omega \times \Omega)$ hat folglich 36 Elementarereignisse $\mathrm{e}_{i,j}$:

$$\mathrm{e}_{i,j} = \{(i, j)\}, \;\; i, j = 1, \ldots, 6$$

Beispiele von Ereignissen in der Ereignisalgebra des Doppelwurfs sind:

6 beim ersten Wurf:	$\{(6,1), (6,2), \ldots, (6,6)\}$
6 beim zweiten Wurf:	$\{(1,6), (2,6), \ldots, (6,6)\}$
Beide Würfe gleich:	$\{(1,1), (2,2), \ldots, (6,6)\}$
Summe der Würfe gleich 7:	$\{(1,6), (2,5), \ldots, (6,1)\}$

Insgesamt enthält die Ereignisalgebra $2^{36} \approx 6.87 \cdot 10^{10}$ Ereignisse.

Analog zum Doppelwurf wird beim n-maligen Würfeln das Experiment durch den Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ repräsentiert. Die Ergebnismenge von $\boldsymbol{X}$ ist das n-fache kartesische Produkt $\Omega \times \ldots \times \Omega$. Jedes Elementarereignis enthält genau eine Folge der Länge n aus ganzen Zahlen von 1 bis 6. Es gibt daher 6^n Elementarereignisse. ◀

▶ **Beispiel 1.3.9.** Wiederholter Alternativversuch

Ein Experiment, das nur zwei mögliche Ergebnisse hat, heißt ein *Alternativversuch* oder *Bernoulliexperiment* (siehe Abschnitt 3.3.1). Die beiden Ergebnisse werden üblicherweise durch 0 und 1 kodiert. Wird ein Alternativversuch n-mal durchgeführt, ergibt sich eine Ergebnismenge mit 2^n Ergebnissen, nämlich den Folgen der Form $(i_1, \ldots, i_n)$ mit $i_j \in \{0, 1\}$. Jedes Ergebnis kann als Realisierung eines Zufallsvektors der Dimension n aufgefasst werden. Ein Beispiel ist das n-malige Werfen einer Münze. ◀

Oft wird angenommen, dass die Wiederholungen des Experiments *unabhängig* voneinander sind. Das heißt anschaulich, dass ein Versuch von den vorhergehenden nicht beeinflusst wird. Die exakte Definition wird im nächsten Kapitel gegeben werden (siehe Abschnitt 2.3.3).

*http://de.wikipedia.org/wiki/Kartesisches_Produkt

☞ **Definition 1.3.10.** Stichprobe

Wird ein Experiment, dessen Ausgang durch die Zufallsvariable X repräsentiert wird, n-mal unabhängig wiederholt, so heißt der Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ eine (zufällige) Stichprobe der Zufallsvariablen X.

☞ **Definition 1.3.11.** Absolute Häufigkeit

Ein Experiment werde durch die Zufallsvariable X mit Ergebnismenge Ω repräsentiert. Ist $\mathrm{E} \subseteq \Omega$ ein Ereignis, und tritt E in n Wiederholungen des Experiment k-mal ein, so heißt $k = h_n(\mathrm{E})$ die absolute Häufigkeit von E.

☞ **Definition 1.3.12.** Relative Häufigkeit

Ist $h_n(\mathrm{E})$ die absolute Häufigkeit von E in n Wiederholungen eines Experiments, so heißt $f_n(\mathrm{E}) = h_n(\mathrm{E})/n$ die relative Häufigkeit von E.

Kapitel 2

Wahrscheinlichkeit

2.1 Einleitung

Der konkrete Ausgang eines Zufallsexperiments kann im Allgemeinen nicht genau vorausgesagt werden. Die möglichen Ausgänge sind jedoch bekannt. Das Ziel der Wahrscheinlichkeitsrechnung ist es, den Ereignissen *im Kontext des Experiments* Wahrscheinlichkeiten zuzuweisen.

Es sind zwei Interpretationen der Wahrscheinlichkeit möglich. In der *Häufigkeitsinterpretation* ist die Wahrscheinlichkeit eines Ereignisses die relative Häufigkeit, die beobachtet wird, wenn das Experiment sehr oft unter den gleichen Bedingungen wiederholt wird. Dies ist natürlich nur sinnvoll, wenn das Experiment zumindest theoretisch beliebig oft wiederholt werden kann. Die auf dieser Interpretation basierende Statistik wird „frequentistisch“ genannt.

Die *subjektive Interpretation* ist auch auf Experimente anwendbar, wo eine Wiederholung unter den gleichen Bedingungen prinzipiell nicht möglich ist. Die Wahrscheinlichkeit eines Ereignisses ist dann eine Aussage über den Glauben oder die subjektive Einschätzung der Person, die die Wahrscheinlichkeit angibt. Die darauf basierende Statistik wird „bayesianisch“ oder „Bayes-Statistik“ genannt. Es ist jedoch zu beachten, dass die subjektive Einschätzung durchaus auf objektiven Tatsachen beruhen kann und meist auch beruht.

► **Beispiel 2.1.1.**

In der Teilchenphysik wird die Wahrscheinlichkeit des Zerfalls eines Z-Bosons in zwei Myonen ($Z \to \mu^+\mu^-$) interpretiert als der Grenzwert der relativen Häufigkeit für eine große Zahl von beobachteten Zerfällen. Die Beobachtung einer beliebig großen Zahl von solchen Zerfällen ist aus Zeit- und Geldgründen zwar nicht praktikabel, es bestehen aber keine prinzipiellen Hindernisse. ◄

► **Beispiel 2.1.2.**

Der Satz „Die Wahrscheinlichkeit, dass es morgen regnet, ist 40%“ quantifiziert die Einschätzung der Person, die diese Aussage tätigt. In der professionellen Wettervorhersage beruht diese Einschätzung sicherlich auf objektiven aktuellen und historischen Wetterdaten. Da aber die Bedingungen, unter denen die Voraussage zustandekommt, nicht wiederholbar sind, ist eine frequentistische Interpretation der Aussage wohl kaum sinnvoll. ◄

In der Praxis ist der Übergang zwischen den beiden Ansätzen oft fließend. Außerdem sind in vielen Fällen die Resultate annähernd identisch, und nur die Interpretation ist verschieden. Der bayesianische Ansatz ist allerdings umfassender und flexibler. Der frequentistische Ansatz ist oft einfacher in der Berechnung, aber beschränkter in der Anwendung.

2.2 Wahrscheinlichkeit

2.2.1 Wahrscheinlichkeitsmaße

Dia axiomatische Begründung des Wahrscheinlichkeitsbegriffes geht auf den russischen Mathematiker A. Kolmogorow* zurück.

☞ **Definition 2.2.1.** Wahrscheinlichkeitsmass

Es sei $\Sigma = \mathfrak{S}(\Omega)$ *eine Ereignisalgebra über der Ergebnismenge* Ω, A *und* B *Ereignisse in* Σ, *und* W *eine Abbildung von* Σ *in* $\mathbb{R}$. W *heißt ein Wahrscheinlichkeitsmaß, eine Wahrscheinlichkeitsverteilung oder kurz eine Wahrscheinlichkeit, wenn gilt:*

1. *Positivität:* $W(\mathrm{A}) \geq 0$, *für alle* $\mathrm{A} \in \Sigma$
2. *Additivität:* $\mathrm{A} \cap \mathrm{B} = \emptyset \implies W(\mathrm{A} \cup \mathrm{B}) = W(\mathrm{A}) + W(\mathrm{B})$, *für alle* $\mathrm{A}, \mathrm{B} \in \Sigma$
3. *Normierung:* $W(\Omega) = 1$

Enthält Σ *unendlich viele Ereignisse, verlangt man zusätzlich*

4. σ*-Additivität: Sind* A_n, $n \in \mathbb{N}$ *paarweise disjunkte Ereignisse, gilt*

$$W\left(\bigcup_{n\in\mathbb{N}} \mathrm{A}_n\right) = \sum_{n\in\mathbb{N}} W(\mathrm{A}_n)$$

☛ **Satz 2.2.2.** Rechenregeln für Wahrscheinlichkeitsmasse

Es sei Σ *eine Ereignisalgebra mit dem Wahrscheinlichkeitsmaß* W. *Dann gilt:*

1. $W(\emptyset) = 0$
2. $W(\mathrm{A}) \leq 1$, *für alle* $\mathrm{A} \in \Sigma$
3. $W(\mathrm{A}') = 1 - W(\mathrm{A})$, *für alle* $\mathrm{A} \in \Sigma$
4. $\mathrm{A} \subseteq \mathrm{B} \implies W(\mathrm{A}) \leq W(\mathrm{B})$, *für alle* $\mathrm{A}, \mathrm{B} \in \Sigma$
5. $W(\mathrm{A} \cup \mathrm{B}) = W(\mathrm{A}) + W(\mathrm{B}) - W(\mathrm{A} \cap \mathrm{B})$, *für alle* $\mathrm{A}, \mathrm{B} \in \Sigma$
6. *Hat* Σ *höchstens abzählbar viele Elementarereignisse* $\{\mathrm{e}_i \,|\, i \in I\}$, *so ist* $\sum_{i\in I} W(\mathrm{e}_i) = 1$.

In einer diskreten Ereignisalgebra Σ ist jedes Ereignis A die Vereinigung von höchstens abzählbar vielen Elementarereignissen. Die Wahrscheinlichkeit von A ist gleich der Summe der Wahrscheinlichkeiten dieser Elementarereignisse. Daher ist ein Wahrscheinlichkeitsmaß durch die Werte, die es den Elementarereignissen zuordnet, *eindeutig bestimmt*. Andererseits kann jede positive Funktion, die auf der Menge der Elementarereignisse definiert ist und Punkt 6 erfüllt, eindeutig zu einem Wahrscheinlichkeitsmaß fortgesetzt werden. Beispiele von wichtigen Wahrscheinlichkeitsmaßen auf diskreten Ereignisalgebren finden sich in Kapitel 3.

In einer kontinuierlichen Ereignisalgebra Σ ist die Wahrscheinlichkeit jedes Elementarereignisses gleich 0. Die Wahrscheinlichkeit eines Ereignisses kann daher nicht mehr durch Summation ermittelt werden. Statt dessen wird eine *Dichtefunktion* $f(x)$ angegeben, die jedem Ergebnis x einen nichtnegativen Wert $f(x)$ zuordnet.

*http://de.wikipedia.org/wiki/Andrei_Nikolajewitsch_Kolmogorow

☞ **Definition 2.2.3.** DICHTEFUNKTION

Eine nichtnegative Funktion $f(x)$, die integrierbar und normiert ist:

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$$

heißt Wahrscheinlichkeitsdichtefunktion, Dichtefunktion oder kurz Dichte.

Die Wahrscheinlichkeit eines Ereignisses $\mathrm{A} \in \Sigma$ wird im kontinuierlichen Fall durch *Integration* über die Dichtefunktion ermittelt:

$$W(\mathrm{A}) = \int_{\mathrm{A}} f(x)\,\mathrm{d}x$$

Die Dichtefunktion muss so beschaffen sein, dass das Integral für alle $\mathrm{A} \in \Sigma$ existiert. Beispiele von wichtigen Wahrscheinlichkeitsmaßen auf kontinuierlichen Ereignisalgebren finden sich in Kapitel 4.

2.2.2 Das Gesetz der großen Zahlen

Ein einfaches Zufallsexperiment ist der Wurf mit einer Münze. Es gibt zwei mögliche Ergebnisse, Kopf (K) und Zahl (Z). Unter der Annahme, dass die Münze symmetrisch

Tabelle 2.1: Häufigkeitstabelle des Wurfs mit einer Münze.

n	$h_n(\mathrm{K})$	$f_n(\mathrm{K})$	$\lvert f_n(\mathrm{K}) - 0.5\rvert$
10	6	0.6	0.1
100	51	0.51	0.01
500	252	0.504	0.004
1000	488	0.488	0.012
5000	2533	0.5066	0.0066

ist, sind K und Z gleichwahrscheinlich. Das Experiment wird n-mal wiederholt, und die Häufigkeit des Auftretens von K wird gezählt. Tabelle 2.1 und Abbildung 2.1 zeigen die Entwicklung der absoluten Häufigkeit $h_n(\mathrm{K})$ und der relativen Häufigkeit $f_n(\mathrm{K}) = h_n(\mathrm{K})/n$ von K als Funktion von n.

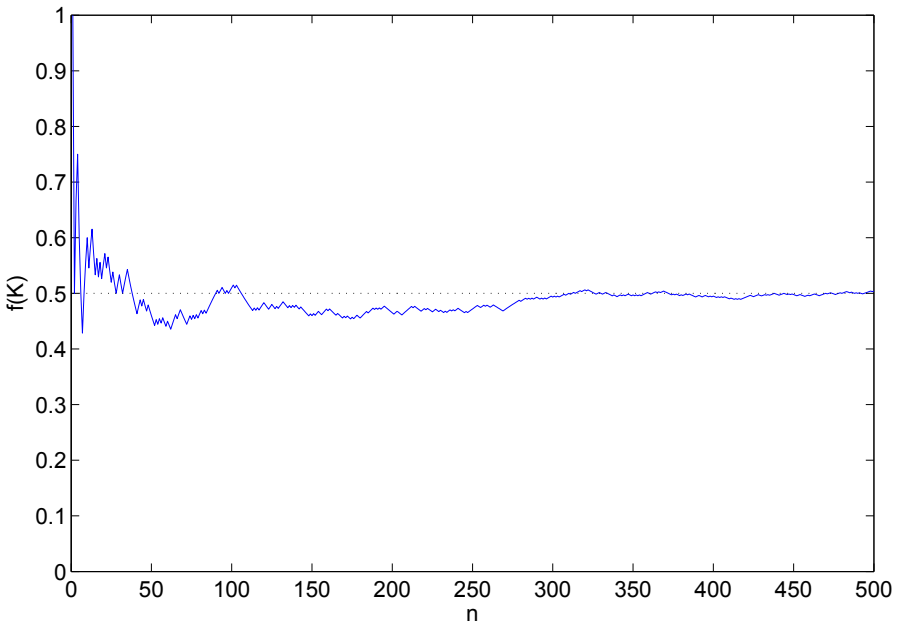

Abbildung 2.1: Entwicklung der relativen Häufigkeit $f_n(\mathrm{K})$ von K.
MATLAB: `make_coin`

Die relative Häufigkeit des Ereignisses K scheint gegen den Grenzwert 0.5 zu streben. Dieser Grenzwert wird als die *Wahrscheinlichkeit* $W(\mathrm{K})$ bezeichnet. Man spricht auch von einem *empirischen Gesetz der großen Zahlen*:

$$\lim_{n\to\infty} f_n(\mathrm{K}) = W(\mathrm{K})$$

Das mathematische Problem dieser „Definition“ liegt darin, dass die Existenz des Grenzwerts von vornherein nicht einzusehen ist und im klassisch analytischen Sinn tatsächlich nicht gegeben ist, sondern nur in einem speziellen statistischen Sinn.

2.2.3 Kombinatorik

Manchmal ist es auf Grund von Symmetrieüberlegungen möglich, alle Elementarereignisse als gleichwahrscheinlich anzusehen, wie etwa beim Würfeln oder beim Roulettespiel. Dies ist natürlich nur sinnvoll für endlich viele Elementarereignisse. Die Annahme der perfekten Symmetrie entspricht nicht immer der physikalischen Realität und muss im Zweifelsfall durch das Experiment überprüft werden.

Sind alle m Elementarereignisse gleichwahrscheinlich, gilt:

$$W(\mathrm{e}_1) = W(\mathrm{e}_2) = \ldots = W(\mathrm{e}_m) = \frac{1}{m}$$

Für ein Ereignis A, das sich aus g Elementarereignissen zusammensetzt, gilt dann die *Regel von Laplace*:

$$W(\mathrm{A}) = \frac{g}{m}$$

Die Wahrscheinlichkeit von A ist die Anzahl g der „günstigen“ durch die Anzahl m der „möglichen“ Fälle. Die Abzählung der günstigen und möglichen Fälle erfordert oft *kombinatorische Methoden*.

☞ **Definition 2.2.4.** Variation

Es sei M *eine Menge mit n Elementen. Eine geordnete Folge von k verschiedenen Elementen von* M *heißt eine Variation von n Elementen zur k-ten Klasse.*

☛ **Satz 2.2.5.** Anzahl der Variationen

Es gibt

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

Variationen von n Elementen zur k-ten Klasse.

Für den Spezialfall $k = n$ sieht man, dass sich die n Elemente der Menge M auf

$$n! = \prod_{i=1}^{n} i$$

verschiedene Weisen anordnen lassen. Diese werden *Permutationen* genannt.

☞ **Definition 2.2.6.** Kombination

Sei M *wieder eine Menge mit n Elementen. Eine k-elementige Teilmenge von* M *heißt eine Kombination von n Elementen zur k-ten Klasse.*

☛ **Satz 2.2.7.** Anzahl der Kombinationen

Es gibt

$$C_k^n = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Kombinationen von n Elementen zur k-ten Klasse.

Wie aus der Definition der Kombination folgt, ist die Summe aller $C_k^n,\ 0 \le k \le n$, für festes n gleich der Anzahl *aller* Teilmengen von M. Daher gilt:

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

► **Beispiel 2.2.8.**

Beim Roulette sind die Zahlen von 0 bis 36 als Ergebnis möglich, somit gibt es $n = 37$ Ergebnisse.

1. Wie groß ist die Wahrscheinlichkeit, dass sich in einer Serie von zehn Würfen keine Zahl wiederholt?
   Es gibt $g = V_k^n$ verschiedene Serien der Länge k, in denen sich keine Zahl wiederholt, und $m = n^k$ beliebige Serien. Die Wahrscheinlichkeit ist daher gleich

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

   Mit $n = 37$ und $k = 10$ erhält man eine Wahrscheinlichkeit von 0.2629.

2. Wie groß ist die Wahrscheinlichkeit, dass in einer Serie von 37 Würfen jede Zahl vorkommt?
   Es gibt $g = n!$ verschieden Serien der Länge n, in denen jede Zahl vorkommt, und $m = n^n$ beliebige Serien. Wahrscheinlichkeit ist daher gleich

$$\frac{g}{m} = \frac{n!}{n^n}$$

   Mit $n = 37$ erhält man eine Wahrscheinlichkeit von $1.3 \cdot 10^{-15}$. ◄

2.3 Bedingte Wahrscheinlichkeit

2.3.1 Kopplung und bedingte Wahrscheinlichkeit

Es seien A und B zwei Ereignisse, die bei einem Experiment eintreten können. Dann stellt sich die Frage, ob ein Zusammenhang zwischen den Ereignissen besteht. Ein solcher Zusammenhang wird *Kopplung* genannt. *Positive Kopplung* bedeutet: je öfter A eintritt, desto öfter tritt *tendenziell* auch B ein. *Negative Kopplung* bedeutet: je öfter A eintritt, desto seltener tritt tendenziell auch B ein.

Die absolute Häufigkeit des Eintretens von A und B in n Versuchen kann in einer *Vierfeldertafel* oder *Kontingenztafel* zusammengefasst werden. Eine Vierfeldertafel ist folgendermaßen aufgebaut:

	B	B′	
A	$h_n(\mathrm{A}\cap\mathrm{B})$	$h_n(\mathrm{A}\cap\mathrm{B}')$	$h_n(\mathrm{A})$
A′	$h_n(\mathrm{A}'\cap\mathrm{B})$	$h_n(\mathrm{A}'\cap\mathrm{B}')$	$h_n(\mathrm{A}')$
	$h_n(\mathrm{B})$	$h_n(\mathrm{B}')$	n

Die Zeilen- und Spaltensummen sind die Häufigkeiten der Ereignisse A, A' und B, B'.

► **Beispiel 2.3.1.** Vierfeldertafel

Bei einer Reihenuntersuchung von 1000 Personen werden die Merkmale „Geschlecht“ und „Diabeteserkrankung“ erhoben, mit den folgenden Ereignissen:

A: „weiblich“, A′: „männlich“

B: „hat Diabetes“, B′: „hat nicht Diabetes“

Eine Vierfeldertafel mit 1000 Einträgen kann dann so aussehen:

	B	B′	
A	19	526	545
A′	26	429	455
	45	955	1000

◄

Gewöhnliche relative Häufigkeiten von Ereignissen werden auf den Umfang n der gesamten Versuchsreihe bezogen (siehe Definition 1.3.12):

$$f_n(\mathrm{A}) = \frac{h_n(\mathrm{A})}{n}$$

Bedingte relative Häufigkeiten werden auf die Häufigkeit des Eintretens des anderen Ereignisses bezogen, d.h. es werden nur jene Versuche betrachtet, bei denen das andere Ereignis eingetreten ist.

☞ **Definition 2.3.2.** Bedingte relative Häufigkeit

Die bedingte relative Häufigkeit von A *unter der Bedingung* B *ist definiert durch:*

$$f_n(\mathrm{A}\,|\,\mathrm{B}) = \frac{f_n(\mathrm{A}\cap\mathrm{B})}{f_n(\mathrm{B})}$$

Die Vierfeldertafel in Beispiel 2.3.1 ergibt unter anderem folgende bedingte relative Häufigkeiten:

$$f_n(\mathrm{A}\,|\,\mathrm{B}) = \frac{19}{45} = 0.422, \quad f_n(\mathrm{A}\,|\,\mathrm{B}') = \frac{526}{955} = 0.551$$

Es ist somit zu vermuten, dass die beiden Merkmale gekoppelt sind. $f_n(\mathrm{A}\,|\,\mathrm{B}) > f_n(\mathrm{A})$ deutet auf eine positive Kopplung, $f_n(\mathrm{A}\,|\,\mathrm{B}) < f_n(\mathrm{A})$ auf eine negative Kopplung.

Stammen die Daten aus einem Zufallsexperiment, dann besitzen die Ereigniskonjunktionen auch Wahrscheinlichkeiten. Analog zur Vierfeldertafel kann dann eine Wahrscheinlichkeitstafel der folgenden Form erstellt werden:

	B	B′	
A	$W(\mathrm{A}\cap\mathrm{B})$	$W(\mathrm{A}\cap\mathrm{B}')$	$W(\mathrm{A})$
A′	$W(\mathrm{A}'\cap B)$	$W(\mathrm{A}'\cap B')$	$W(\mathrm{A}')$
	$W(\mathrm{B})$	$W(\mathrm{B}')$	1

Bedingte Wahrscheinlichkeiten werden nun analog zu den bedingten Häufigkeiten definiert.

☞ **Definition 2.3.3.** Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit von A *unter der Bedingung* B *ist definiert durch:*

$$W(\mathrm{A}\,|\,\mathrm{B}) = \frac{W(\mathrm{A}\cap\mathrm{B})}{W(\mathrm{B})}$$

sofern $W(\mathrm{B}) > 0$.

▶ **Beispiel 2.3.4.** Bedingte Wahrscheinlichkeit

Ist ein Würfel völlig symmetrisch, werden den Elementarereignissen $\mathrm{e}_i = \{i\}$ gleiche Wahrscheinlichkeiten zugeordnet:

$$W(\mathrm{e}_i) = \frac{1}{6},\ 1 \le i \le 6$$

Die Ereignisse U (ungerade Zahl) und G (gerade Zahl) sind definiert durch:

$$\mathrm{U} = \{1,3,5\},\ \mathrm{G} = \{2,4,6\}$$

Dann gilt zum Beispiel:

$$\begin{aligned}
W(\mathrm{e}_1\,|\,\mathrm{U}) &= \frac{W(\mathrm{e}_1\cap\mathrm{U})}{W(\mathrm{U})} = \frac{W(\mathrm{e}_1)}{W(\mathrm{U})} = \frac{1}{3}\\
W(\mathrm{e}_1\,|\,\mathrm{G}) &= \frac{W(\mathrm{e}_1\cap\mathrm{G})}{W(\mathrm{G})} = \frac{W(\emptyset)}{W(\mathrm{G})} = 0\\
W(\mathrm{U}\,|\,\mathrm{e}_1) &= \frac{W(\mathrm{e}_1\cap\mathrm{U})}{W(\mathrm{e}_1)} = \frac{W(\mathrm{e}_1)}{W(\mathrm{e}_1)} = 1\\
W(\mathrm{e}_1\cup\mathrm{e}_3\,|\,\mathrm{U}) &= \frac{W((\mathrm{e}_1\cup\mathrm{e}_3)\cap\mathrm{U})}{W(\mathrm{U})} = \frac{W(\mathrm{e}_1\cup\mathrm{e}_3)}{W(\mathrm{U})} = \frac{2}{3}\\
W(\mathrm{e}_1\cup\mathrm{e}_2\,|\,\mathrm{U}) &= \frac{W((\mathrm{e}_1\cup\mathrm{e}_2)\cap\mathrm{U})}{W(\mathrm{U})} = \frac{W(\mathrm{e}_1)}{W(\mathrm{U})} = \frac{1}{3}
\end{aligned}$$

◀

Aus der Definition der bedingten Wahrscheinlichkeit folgen sofort zwei wichtige Beziehungen.

☛ **Satz 2.3.5.** Produktformel und Inverse Wahrscheinlichkeit

Sind A *und* B *zwei Ereignisse, gilt:*

1. *Produktformel:*

$$W(\mathrm{A} \cap \mathrm{B}) = W(\mathrm{A}|\mathrm{B})W(\mathrm{B}) = W(\mathrm{B}|\mathrm{A})W(\mathrm{A})$$

2. *Formel für die inverse Wahrscheinlichkeit:*

$$W(\mathrm{B}|\mathrm{A}) = \frac{W(\mathrm{A}|\mathrm{B})W(\mathrm{B})}{W(\mathrm{A})}$$

Beide Aussagen gelten auch für relative Häufigkeiten.

2.3.2 Satz von Bayes

Der Satz von Bayes ist eine einfache Folgerung aus der Definition der bedingten Wahrscheinlichkeit, hat jedoch weitreichende Konsequenzen, da er die Grundlage der gesamten Bayes-Statistik ist. Er ist nach dem englischen Mathematiker und Pfarrer T. Bayes* benannt.

☞ **Definition 2.3.6.** Zerlegung

Die Ereignisse $\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_m$ *bilden eine Zerlegung der Ergebnismenge* Ω*, wenn gilt:*

1. *Unvereinbarkeit:* $\mathrm{B}_i \cap \mathrm{B}_j = \emptyset,\ i, j = 1, \ldots, m,\ j \neq i$
2. *Vollständigkeit:* $\mathrm{B}_1 \cup \mathrm{B}_2 \cup \ldots \cup \mathrm{B}_m = \Omega$

*http://de.wikipedia.org/wiki/Thomas_Bayes

☛ **Satz 2.3.7.**

Bilden die Ereignisse $\mathrm{B}_1, \mathrm{B}_2, \ldots, \mathrm{B}_m$ *eine Zerlegung der Ergebnismenge* Ω*, dann gilt:*

$$W(\mathrm{B}_1) + W(\mathrm{B}_2) + \ldots + W(\mathrm{B}_m) = W(\Omega) = 1$$

☛ **Satz 2.3.8.** Satz von der totalen Wahrscheinlichkeit

Es sei $\mathrm{B}_1, \ldots, \mathrm{B}_m$ *eine Zerlegung. Dann gilt:*

$$W(\mathrm{A}) = W(\mathrm{A}\,|\,\mathrm{B}_1)W(\mathrm{B}_1) + \ldots + W(\mathrm{A}\,|\,\mathrm{B}_m)W(\mathrm{B}_m)$$

▶ **Beispiel 2.3.9.**

Ein Betrieb erzeugt Halogenleuchten mit 100 W (35% der Produktion), 200 W (45%) und 300 W (20%). Nach einem Jahr sind noch 98% der 100 W-Leuchten funktionsfähig, 96% der 200 W-Leuchten, und 93% der 300 W-Leuchten. Der Anteil an allen Leuchten, die nach einem Jahr noch funktionsfähig (Ereignis F) sind, berechnet sich folgendermaßen:

Anteile der Produktion:
$W(100) = 0.35, W(200) = 0.45, W(300) = 0.2$

Bedingte Wahrscheinlichkeiten von F:
$W(\mathrm{F}\,|\,100) = 0.98, W(\mathrm{F}\,|\,200) = 0.96, W(\mathrm{F}\,|\,300) = 0.93$

Totale Wahrscheinlichkeit von F:
$W(\mathrm{F}) = 0.98 \cdot 0.35 + 0.96 \cdot 0.45 + 0.93 \cdot 0.2 = 0.961$ ◀

☛ **Satz 2.3.10.** Satz von Bayes

Es sei $\mathrm{B}_1, \ldots, \mathrm{B}_m$ *eine Zerlegung. Dann gilt:*

$$W(\mathrm{B}_i\,|\,\mathrm{A}) = \frac{W(\mathrm{A}\,|\,\mathrm{B}_i)W(\mathrm{B}_i)}{W(\mathrm{A})} = \frac{W(\mathrm{A}\,|\,\mathrm{B}_i)W(\mathrm{B}_i)}{W(\mathrm{A}\,|\,\mathrm{B}_1)W(\mathrm{B}_1) + \ldots + W(\mathrm{A}\,|\,\mathrm{B}_m)W(\mathrm{B}_m)}$$

$W(\mathrm{B}_i)$ wird die *a-priori* Wahrscheinlichkeit von B_i genannt, $W(\mathrm{B}_i\,|\,\mathrm{A})$ die *a-posteriori* Wahrscheinlichkeit. Die a-posteriori Wahrscheinlichkeit enthält die zusätzliche Information, dass A eingetreten ist.

▶ **Beispiel 2.3.11.**

Ein Experiment kauft Bauteile von zwei Anbietern (1 und 2), wobei der Anteil des ersten 65% beträgt. Erfahrungsgemäß ist der Ausschussanteil (Ereignis A) bei Anbieter 1 gleich 3% und bei Anbieter 2 gleich 4%.

1. Der totale Ausschussanteil ist gleich:

$$W(\mathrm{A}) = W(\mathrm{A}\,|\,1)W(1) + W(\mathrm{A}\,|\,2)W(2) = 0.03 \cdot 0.65 + 0.04 \cdot 0.35 = 0.0335$$

2. Die Wahrscheinlichkeit, dass ein einwandfreier Bauteil von Anbieter 2 kommt, ist gleich:

$$W(2\,|\,\mathrm{A}') = \frac{W(\mathrm{A}'\,|\,2)W(2)}{W(\mathrm{A}')} = \frac{0.96 \cdot 0.35}{1 - 0.0335} = 0.3476$$

3. Die Wahrscheinlichkeit, dass ein mangelhafter Bauteil von Anbieter 1 kommt, ist gleich:

$$W(1\,|\,\mathrm{A}) = \frac{W(\mathrm{A}\,|\,1)W(1)}{W(\mathrm{A})} = \frac{0.03 \cdot 0.65}{0.0335} = 0.5821$$

Es gilt $W(1|\mathrm{A}) < W(1)$, d.h. die Information dass der Bauteil mangelhaft ist, lässt die Wahrscheinlichkeit, dass er von Anbieter 1 kommt, sinken. Anbieter 1 liefert ja die bessere Qualität. ◀

2.3.3 Unabhängigkeit

Zwei Ereignisse sind *positiv gekoppelt*, wenn

$$W(\mathrm{A}|\mathrm{B}) > W(\mathrm{A}) \quad \text{oder} \quad W(\mathrm{A} \cap \mathrm{B}) > W(\mathrm{A})W(\mathrm{B})$$

Zwei Ereignisse sind *negativ gekoppelt*, wenn

$$W(\mathrm{A}|\mathrm{B}) < W(\mathrm{A}) \quad \text{oder} \quad W(\mathrm{A} \cap \mathrm{B}) < W(\mathrm{A})W(\mathrm{B})$$

Liegt weder positive noch negative Kopppelung vor, sind A und B *unabhängig*.

☞ **Definition 2.3.12.** Unabhängigkeit
Zwei Ereignisse A *und* B *heißen (stochastisch) unabhängig, wenn gilt:*

$$W(\mathrm{A}|\mathrm{B}) = W(\mathrm{A}) \quad \textit{oder} \quad W(\mathrm{A} \cap \mathrm{B}) = W(\mathrm{A})W(\mathrm{B})$$

Die Ereignisse $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$ *heißen unabhängig, wenn gilt:*

$$W(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_n) = W(\mathrm{A}_1) \cdot \ldots \cdot W(\mathrm{A}_n).$$

Es ist zu beachten, dass es zur Unabhängigkeit von n Ereignissen nicht hinreichend ist, dass je zwei Ereignisse A_i und A_j paarweise unabhängig sind, wie das folgende Beispiel zeigt.

► **Beispiel 2.3.13.**
Ein Experiment besteht im zweimaligen Wurf einer Münze (Kopf/Zahl). Die Ergebnismenge ist $\Omega = \{\mathrm{KK}, \mathrm{KZ}, \mathrm{ZK}, \mathrm{ZZ}\}$. Ferner werden folgende Ereignisse definiert:

$$\begin{aligned}
\mathrm{E}_1 &= \{\mathrm{KK}, \mathrm{KZ}\} \ldots \text{Kopf beim ersten Wurf} \\
\mathrm{E}_2 &= \{\mathrm{KK}, \mathrm{ZK}\} \ldots \text{Kopf beim zweiten Wurf} \\
\mathrm{E}_3 &= \{\mathrm{KK}, \mathrm{ZZ}\} \ldots \text{Gerade Zahl von Köpfen}
\end{aligned}$$

Dann gilt für alle $i \neq j$:

$$W(\mathrm{E}_i \cap \mathrm{E}_j) = \frac{1}{4} = W(\mathrm{E}_i) \cdot W(\mathrm{E}_j)$$

aber auch:

$$W(\mathrm{E}_1 \cap \mathrm{E}_2 \cap \mathrm{E}_3) = \frac{1}{4} \neq \frac{1}{8} = W(\mathrm{E}_1) \cdot W(\mathrm{E}_2) \cdot W(\mathrm{E}_3)$$ ◀

Sind A und B unabhängig, gilt $W(\mathrm{A}|\mathrm{B}) = W(\mathrm{A})$ und $W(\mathrm{B}|\mathrm{A}) = W(\mathrm{B})$. Die Wahrscheinlichkeitstafel für zwei unabhängige Ereignisse ist daher folgendermaßen aufgebaut:

	B	B′	
A	$W(\mathrm{A})W(\mathrm{B})$	$W(\mathrm{A})W(\mathrm{B}')$	$W(\mathrm{A})$
A′	$W(\mathrm{A}')W(\mathrm{B})$	$W(\mathrm{A}')W(\mathrm{B}')$	$W(\mathrm{A}')$
	$W(\mathrm{B})$	$W(\mathrm{B}')$	1

Die Kopplung zwischen A und B kann unter anderem durch die *Vierfelderkorrelation* gemessen werden.

☞ **Definition 2.3.14.** Vierfelderkorrelation

Die Vierfelderkorrelation $\rho(\mathrm{A},\mathrm{B})$ *ist definiert durch:*

$$\rho(\mathrm{A},\mathrm{B}) = \frac{W(\mathrm{A}\cap\mathrm{B}) - W(\mathrm{A})W(\mathrm{B})}{\sqrt{W(\mathrm{A})W(\mathrm{A}')W(\mathrm{B})W(\mathrm{B}')}}$$

☛ **Satz 2.3.15.** Eigenschaften der Vierfelderkorrelation

Die Vierfelderkorrelation hat die folgenden Eigenschaften:

1. $-1 \leq \rho(\mathrm{A},\mathrm{B}) \leq 1$
2. $\rho(\mathrm{A},\mathrm{B}) = 0 \iff$ *A und* B *sind unabhängig*
3. $\rho(\mathrm{A},\mathrm{B}) > 0 \iff$ *A und* B *sind positiv gekoppelt*
4. $\rho(\mathrm{A},\mathrm{B}) < 0 \iff$ *A und* B *sind negativ gekoppelt*

Das Vorzeichen von $\rho(\mathrm{A},\mathrm{B})$ gibt die *Richtung* der Kopplung an. Der Betrag von $\rho(\mathrm{A},\mathrm{B})$ gibt die *Stärke* der Kopplung an. Speziell gilt:

$$\mathrm{A} = \mathrm{B} \implies \rho(\mathrm{A},\mathrm{B}) = 1$$
$$\mathrm{A} = \mathrm{B}' \implies \rho(\mathrm{A},\mathrm{B}) = -1$$

Eine bestehende Kopplung ist *kein Beweis* für einen kausalen Zusammenhang. Die Kopplung kann auch durch eine gemeinsame Ursache für beide Ereignisse entstehen.

Zwei physikalische Ereignisse können als unabhängig postuliert werden, wenn zwischen ihnen keine wie immer geartete Verbindung besteht, da dann das Eintreten des einen Ereignisses die Wahrscheinlichkeit des anderen nicht beeinflussen kann. Zwei Elementarereignisse sind niemals unabhängig, weil das Eintreten des einen das Eintreten des anderen mit Sicherheit ausschließt.

☛ **Satz 2.3.16.**

Sind E_1 *und* E_2 *zwei unabhängige Ereignisse, so sind auch* E_1 *und* E_2', E_1' *und* E_2, *sowie* E_1' *und* E_2' *unabhängig.*

▶ **Beispiel 2.3.17.** Wiederholung eines Alternativversuchs

Die Ereignisalgebra hat 2^n Elementarereignisse, nämlich die Folgen der Form $(i_1,\ldots,i_n)$, $i_j \in \{0,1\}$ (siehe Beispiel 1.3.9). Sind die Wiederholungen unabhängig, und bezeichnet p die Wahrscheinlichkeit des Eintretens von 1, ist die Wahrscheinlichkeit einer Folge gegeben durch:

$$W(\{(i_1,\ldots,i_n)\}) = p^k(1-p)^{n-k}$$

wo k bzw. $n-k$ die Häufigkeit von 1 bzw. 0 angibt. ◀

Kapitel 3

Diskrete Verteilungen

3.1 Grundbegriffe

3.1.1 Diskrete Verteilungen

Diskrete Zufallsvariable sind entweder kategoriale Merkmale oder das Resultat von Zählexperimenten. In der experimentellen Praxis kommen beide Fälle häufig vor.

► **Beispiel 3.1.1.**
Ein instabiles Elementarteilchen kann in der Regel auf verschiedene Arten zerfallen. Wird jedem Zerfallsmodus eine natürliche Zahl zugeordnet, ist dadurch eine diskrete Zufallsvariable definiert. Wird in einer Stichprobe von n Zerfällen die Häufigkeit eines bestimmten Zerfallskanals abgezählt, ist dadurch ebenfalls eine diskrete Zufallsvariable definiert. ◄

► **Beispiel 3.1.2.**
Wird die Anzahl der Zerfälle einer radioaktiven Quelle pro Zeiteinheit gezählt, ist dadurch eine diskrete Zufallsvariable definiert. ◄

Im Folgenden wird oft angenommen, dass die Werte der diskreten Zufallsvariablen X nichtnegative ganze Zahlen $k \in \mathbb{N}_0$ sind. Die Ereignisalgebra besteht dann aus allen Teilmengen von $\mathbb{N}_0$. Die Wahrscheinlichkeit, mit der X das Ergebnis k liefert, wird mit $f_X(k)$ bezeichnet.

☞ **Definition 3.1.3.** Wahrscheinlichkeitsfunktion

Die Funktion $f_X(k) = W(X = k)$ heißt Wahrscheinlichkeitsfunktion (WKF) der Zufallsvariablen X. Alternativ wird auch die Bezeichnung Dichtefunktion oder Dichte verwendet.

☞ **Definition 3.1.4.** Verteilung einer Zufallsvariablen

Das Wahrscheinlichkeitsmaß W_X, das durch

$$W_X(E) = W(X \in E)$$

definiert ist, heißt die Wahrscheinlichkeitsverteilung oder einfach die Verteilung von X.

Die Wahrscheinlichkeit $W_X(\mathrm{E})$ eines Ereignisses E lässt sich bequem mit Hilfe der Wahrscheinlichkeitsfunktion berechnen:

$$W_X(\mathrm{E}) = \sum_{k \in \mathrm{E}} f_X(k)$$

☞ **Definition 3.1.5.** Diskrete Verteilungsfunktion

Ist X eine diskrete Zufallsvariable, so ist die Verteilungsfunktion F_X von X definiert durch:

$$F_X(x) = W(X \leq x) = \sum_{k \leq x} f_X(k)$$

☛ **Satz 3.1.6.** Eigenschaften einer diskreten Verteilungsfunktion

Es sei $F = F_X$ eine diskrete Verteilungsfunktion. Dann gilt:

1. *F hat Sprungstellen in allen Punkten der Ergebnismenge.*
2. *Die Sprunghöhe im Punkt k ist gleich $f_X(k)$.*
3. $0 \leq F(x) \leq 1$, *für alle* $x \in \mathbb{R}$
4. $x \leq y \implies F(x) \leq F(y)$, *für alle* $x, y \in \mathbb{R}$
5. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$
6. *Die Wahrscheinlichkeit, dass X in das Intervall $(a, b]$ fällt, ist gleich $F(b) - F(a)$:*

$$W(X \leq a) + W(a < X \leq b) = W(X \leq b) \implies W(a < X \leq b) = F(b) - F(a)$$

Abbildung 3.1 zeigt ein Beispiel einer diskreten Wahrscheinlichkeitsfunktion und der zugehörigen Verteilungsfunktion.

3.1.2 Mittelwert und Varianz

Zwei wichtige Kennzahlen einer Wahrscheinlichkeitsverteilung sind *Mittelwert* und *Varianz*.

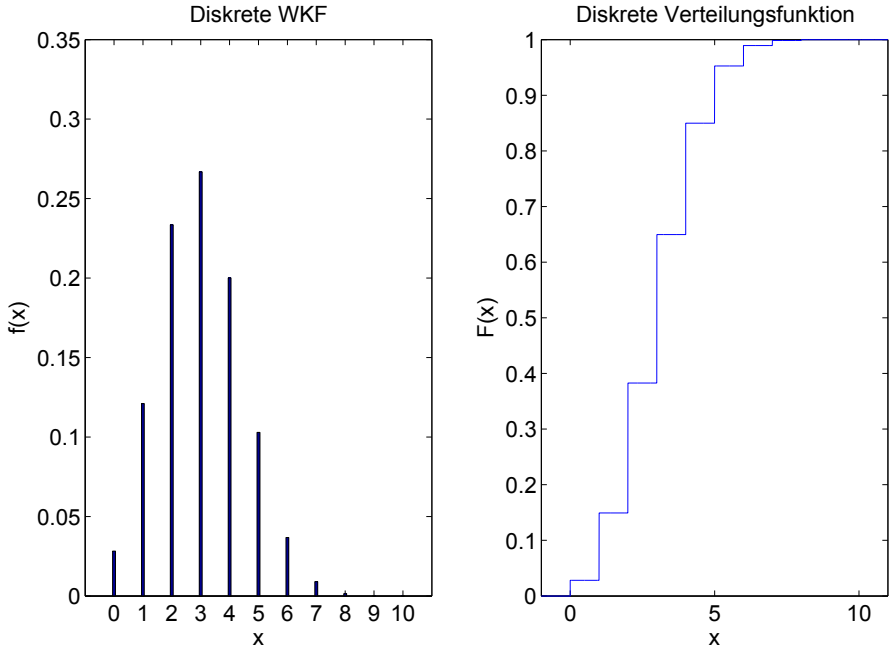

Abbildung 3.1: Beispiel einer diskreten Wahrscheinlichkeitsfunktion und der zugehörigen Verteilungsfunktion. MATLAB: `make_diskret_cdf_pdf`

☞ **Definition 3.1.7.** ERWARTUNGSWERT, MITTELWERT

Es sei X eine diskrete Zufallsvariable mit der Wahrscheinlichkeitsfunktion $f_X(x)$. Ferner sei g eine beliebige Funktion. Die Erwartung von $g(X)$ ist definiert durch:

$$\mathsf{E}[g(X)] = \sum_{k \in \mathbb{N}_0} g(k) f_X(k)$$

Ist $g(x) = x$, so heißt $\mathsf{E}[g(X)] = \mathsf{E}[X]$ der Erwartungswert von X oder der Mittelwert der Verteilung von X.

Erwartungswert und Mittelwert sind synonym; im Folgenden wird jedoch vorzugsweise der Begriff Erwartungswert verwendet, wenn es sich um eine Zufallsvariable handelt, und der Begriff Mittelwert meist dann, wenn es sich um die Verteilung der Zufallsvariablen handelt.

Aus der Definition folgt unmittelbar:

$$\mathsf{E}[X] = \sum_{k \in \mathbb{N}_0} k \, f_X(k)$$

Der Erwartungswert ist der *Schwerpunkt* der durch die Wahrscheinlichkeitsfunktion definierten Massenverteilung. Er hat ferner die folgenden Eigenschaften, die ihn zu einem *linearen Operator*[*] machen.

[*]Ein Operator ist hier eine Funktion, die einer Zufallsvariablen eine relle Zahl zuordnet. Das Argument eines Operators wird in eckige Klammern gesetzt.

☛ **Satz 3.1.8.** Eigenschaften des Erwartungswerts

Es seien X und Y Zufallsvariable und $a, b \in \mathbb{R}$. Dann gilt:

$$\mathsf{E}[aX+b] = a\,\mathsf{E}[X] + b$$
$$\mathsf{E}[X+Y] = \mathsf{E}[X] + \mathsf{E}[Y]$$

Der Erwartungswert von X ist der wichtigste *Lageparameter* der Verteilung von X. Ein Lageparameter ist ein Operator L mit den folgenden Eigenschaften (siehe auch Abschnitt 5.2.6.1):

$$\mathrm{L}[aX+b] = a\,\mathrm{L}[X] + b,\ a, b \in \mathbb{R} \quad \text{und} \quad \min(X) \le \mathrm{L}[X] \le \max(X)$$

☞ **Definition 3.1.9.** Varianz

Die Varianz von X, bezeichnet mit $\mathsf{var}[X]$, ist definiert durch:

$$\mathsf{var}[X] = \mathsf{E}[(X - \mathsf{E}[X])^2] = \sum_{k \in \mathbb{N}_0} (k - \mathsf{E}[X])^2 \, f_X(k)$$

Die Wurzel aus der Varianz heißt die Standardabweichung von X, bezeichnet mit $\sigma[X]$.

Die Standardabweichung hat die gleiche Dimension wie X. Sie ist ein *Skalenparameter*, der die Breite der Verteilung beschreibt. Ein Skalenparameter ist ein Operator S mit der folgenden Eigenschaft (siehe auch Abschnitt 5.2.6.2):

$$\mathrm{S}[aX+b] = |a| \cdot \mathrm{S}[X],\ a, b \in \mathbb{R} \quad \text{und} \quad \mathrm{S}[X] \ge 0$$

☛ **Satz 3.1.10.** Verschiebungssatz
Die Varianz einer Zufallsvariablen X kann auch folgendermaßen berechnet werden:

$$\mathsf{var}[X] = \mathsf{E}[X^2] - \mathsf{E}[X]^2$$

☛ **Satz 3.1.11.** Eigenschaften der Varianz
Es sei X eine Zufallsvariable und $a, b \in \mathbb{R}$. Dann gilt:

$$\mathsf{var}[aX+b] = a^2\mathsf{var}[X]$$
$$\sigma[aX+b] = |a|\,\sigma[X]$$

Varianz und Standardabweichung sind invariant gegen Translationen (Verschiebungen) von X.

3.2 Ziehungsexperimente

Viele Experimente können so interpretiert werden, dass aus einer Grundgesamtheit von Objekten eine Stichprobe zufällig gezogen wird. Es wird dann abgezählt, wieviele der gezogenen Objekte eine gewisse Eigenschaft haben. Ein solches Experiment wird ein *Ziehungsexperiment* genannt. Einmal gezogene Objekte können zurückgelegt werden, d.h. bei der nächsten Ziehung wiederverwendet werden, oder nicht.

3.2.1 Ziehen mit Zurücklegen

Der Ausgangspunkt ist eine Grundgesamtheit von beliebiger Größe. Der Anteil der Merkmalsträger, d.h. der Objekte mit einer gewissen Eigenschaft E, ist gleich p. Es werden n Objekte zufällig gezogen, wobei ein gezogenes Objekte wieder *zurückgelegt* wird. Es wird vorausgesetzt, dass die Ziehungen unabhängig sind, sodass bei jeder Ziehung jedes Objekt die gleiche Wahrscheinlichkeit hat, gezogen zu werden. Die Anzahl der gezogenen Objekt mit der Eigenschaft E ist dann eine diskrete Zufallsvariable X. Die Verteilung von X wird *Binomialverteilung* genannt und mit $\mathrm{Bi}(n,p)$ bezeichnet. Die Schreibweise dafür ist $X \sim \mathrm{Bi}(n,p)$.

☛ **Satz 3.2.1.** Wahrscheinlichkeitsfunktion der Binomialverteilung
Die Wahrscheinlichkeitsfunktion der Binomialverteilung $\mathrm{Bi}(n,p)$ *lautet:*

$$f(k\,|\,n,p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \le k \le n$$

Die Wahrscheinlichkeitsfunktionen einiger Binomialverteilungen sind in Abbildung 3.2 dargestellt.

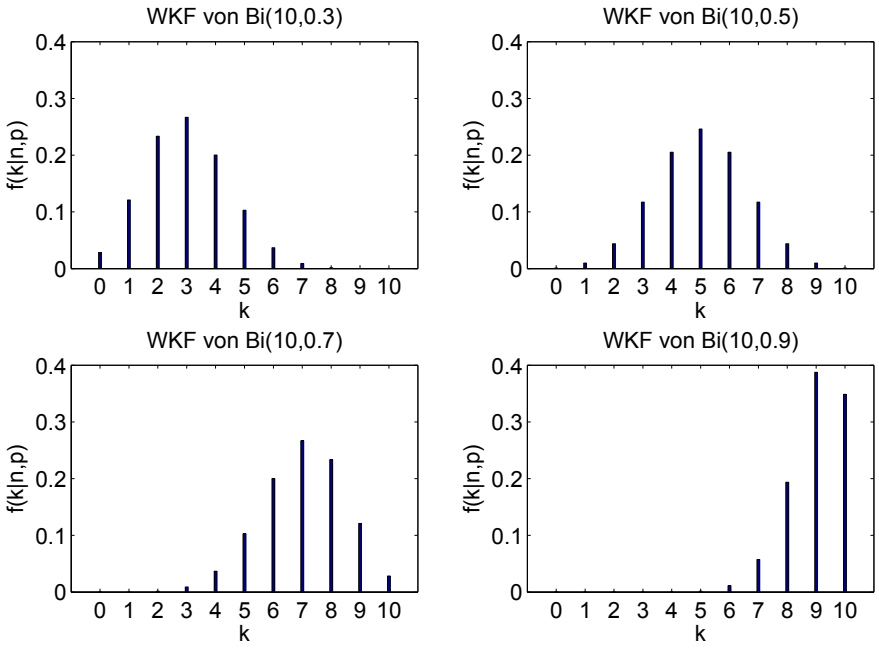

Abbildung 3.2: Die Wahrscheinlichkeitsfunktionen einiger Binomialverteilungen. MATLAB: `make_binomial`

☛ **Satz 3.2.2.** MITTELWERT UND VARIANZ DER BINOMIALVERTEILUNG
Es sei $X \sim \mathrm{Bi}(n,p)$. *Dann gilt*

$$\mathsf{E}[X] = np \text{ und } \mathsf{var}[X] = np(1-p)$$

► **Beispiel 3.2.3.**
Aus einer Lieferung von Bauteilen wird eine Stichprobe vom Umfang $n = 50$ mit Zurücklegen gezogen. Aus Erfahrung ist bekannt, dass 1.5% der Bauteile fehlerhaft sind. Die Wahrscheinlichkeit, in der Stichprobe höchstens einen fehlerhaften Bauteil zu finden, beträgt dann:

$$\begin{aligned} W(X \leq 1) &= f_X(0) + f_X(1) = \\ &= 1 \cdot 0.015^0 \cdot 0.985^{50} + 50 \cdot 0.015^1 \cdot 0.985^{49} \approx \\ &\approx 0.8273 \end{aligned}$$

Die erwartete Anzahl von fehlerhaften Bauteilen in der Stichprobe ist gleich 0.75. ◄

Ist der Anteil der Merkmalsträger p unbekannt, kann eine Beobachtung von X dazu verwendet werden, um p zu schätzen (siehe Abschnitt 6.2.2).

3.2.2 Ziehen ohne Zurücklegen

Der Ausgangspunkt ist eine endliche Grundgesamtheit der Größe N. Von diesen haben M Objekte, die wieder Merkmalsträger genannt werden, eine gewisse Eigenschaft E. Es werden n Objekte zufällig gezogen, wobei einmal gezogene Objekte *nicht zurückgelegt* werden. Bei jeder Ziehung hat jedes Objekt die gleiche Wahrscheinlichkeit, gezogen zu werden. Die Anzahl der gezogenen Merkmalsträger ist eine diskrete Zufallsvariable X. Die Verteilung von X wird *hypergeometrische Verteilung* $\mathrm{Hy}(N, M, n)$ genannt.

☛ **Satz 3.2.4.** Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung
Die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung $\mathrm{Hy}(N, M, n)$ *lautet:*

$$f(m|N, M, n) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}, \quad 0 \leq m \leq \min(n, M)$$

Die Wahrscheinlichkeitsfunktionen von zwei hypergeometrischen Verteilungen sind in Abbildung 3.3 dargestellt. Zum Vergleich sind auch die Wahrscheinlichkeitsfunktionen der korrespondierenden Binomialverteilungen mit gleicher Versuchszahl und gleichem Anteil von Merkmalsträgern gezeigt. Man sieht, dass die beiden Verteilungen mit wachsender Größe der Grundgesamtheit einander ähnlicher werden, da es bei einer großen Grundgesamtheit weniger Unterschied macht, ob zurückgelegt wird oder nicht.

☛ **Satz 3.2.5.** Mittelwert und Varianz der hypergeometrischen Verteilung
Es sei $X \sim \mathrm{Hy}(N, M, n)$. *Dann gilt*

$$\mathsf{E}[X] = \frac{nM}{N} \text{ und } \mathsf{var}[X] = \frac{nM(N-n)(N-M)}{N^2(N-1)}$$

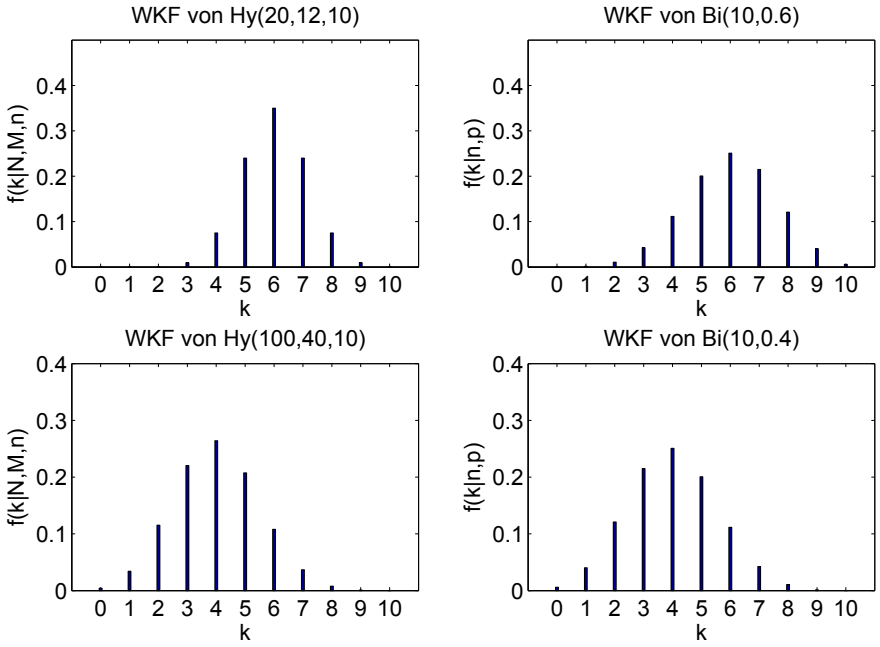

Abbildung 3.3: Die Wahrscheinlichkeitsfunktionen von zwei hypergeometrischen Verteilungen (links) und die der korrespondierenden Binomialverteilungen (rechts).
MATLAB: `make_hypgeom`

► Beispiel 3.2.6.

In einer Lieferung von $N = 1000$ Widerständen sind $M = 20$ Stück defekt. Die Wahrscheinlichkeit, in einer Stichprobe vom Umfang $n = 50$ mehr als 2 defekte Stücke vorzufinden, wenn ohne Zurücklegen gezogen wird, beträgt:

$$\begin{aligned} W(X>2) &= 1 - W(X \le 2) = 1 - (f_X(0) + f_X(1) + f_X(2)) = \\ &= 1 - \frac{\binom{20}{0}\binom{980}{50}}{\binom{1000}{50}} - \frac{\binom{20}{1}\binom{980}{49}}{\binom{1000}{50}} - \frac{\binom{20}{2}\binom{980}{48}}{\binom{1000}{50}} = \\ &= 0.0736 \end{aligned}$$

Die erwartete Anzahl von defekten Widerständen in der Stichprobe ist gleich 1.

► Beispiel 3.2.7.

Beim Lottospiel werden $n = 6$ Zahlen ohne Zurücklegen aus einer Grundgesamtheit von $N = 45$ Zahlen gezogen. Es gibt auf jedem Lottoschein $M = 6$ Merkmalsträger, nämlich die Zahlen, auf die gesetzt wird. Die Wahrscheinlichkeit eines „Vierers“ beträgt:

$$W(X=4) = \frac{\binom{6}{4}\binom{39}{2}}{\binom{45}{6}} = 0.0014 \qquad \blacktriangleleft$$

Ist entweder N bekannt und M unbekannt, oder M bekannt und N unbekannt, kann eine Beobachtung von X dazu verwendet werden, um M bzw. N zu schätzen. Dies wird in Abschnitt 6.3.2.1 näher erläutert.

3.3 Zählexperimente

3.3.1 Der Alternativversuch

Ein Versuch mit nur zwei möglichen Ausgängen heißt ein *Alternativversuch* oder *Bernoulliexperiment*, benannt nach dem Schweizer Mathematiker Jakob I. Bernoulli.* Er wird durch eine Zufallsvariable X beschrieben, die nur zwei mögliche Ergebnisse hat: 0 (Misserfolg) oder 1 (Erfolg), wobei die Wahrscheinlichkeit eines Erfolgs gleich $p \in [0,1]$ ist. Die Verteilung von X heißt *Alternativverteilung* $\mathrm{Al}(p)$.

☞ **Definition 3.3.1.** Wahrscheinlichkeitsfunktion der Alternativverteilung

Die Wahrscheinlichkeitsfunktion der Alternativverteilung lautet:

$$f(k\,|\,p) = p^k(1-p)^{1-k}, \quad k \in \{0,1\}$$

☛ **Satz 3.3.2.** Mittelwert und Varianz der Alternativverteilung

Es sei $X \sim \mathrm{Al}(p)$. *Dann gilt*

$$\mathsf{E}[X] = p \text{ und } \mathsf{var}[X] = p(1-p)$$

*http://de.wikipedia.org/wiki/Jakob_I._Bernoulli

3.3.2 Der wiederholte Alternativversuch

Wird ein Alternativversuch n-mal unabhängig wiederholt, und bezeichnet X die *Anzahl der Erfolge in n Versuchen*, dann ist X *binomialverteilt* gemäß $\mathrm{Bi}(n,p)$ (siehe Abschnitt 3.2.1).

3.3.3 Die Poissonverteilung

Die *Poissonverteilung* $\mathrm{Po}(\lambda)$ entsteht aus der Binomialverteilung $\mathrm{Bi}(n,p)$ durch den Grenzübergang $n \to \infty$ unter der Bedingung $n \cdot p = \lambda$. Das klassische Beispiel einer poissonverteilten Zufallsvariablen ist die Anzahl X der Zerfälle pro Zeiteinheit in einer radioaktiven Quelle. Die Poissonverteilung ist nach dem französischen Physiker und Mathematiker S. Poisson* benannt. Abbildung 3.4 zeigt einige Beispiele.

☛ **Satz 3.3.3.** Wahrscheinlichkeitsfunktion der Poissonverteilung
Die Wahrscheinlichkeitsfunktion der Poissonverteilung $\mathrm{Po}(\lambda)$ *lautet:*

$$f(k\,|\,\lambda) = \frac{\lambda^k}{k!} \cdot \mathrm{e}^{-\lambda}, \quad k \in \mathbb{N}_0$$

☛ **Satz 3.3.4.** Mittelwert und Varianz der Poissonverteilung
Es sei $X \sim \mathrm{Po}(\lambda)$. *Dann gilt*

$$\mathsf{E}[X] = \lambda \text{ und } \mathsf{var}[X] = \lambda$$

Die Poissonverteilung hat die Besonderheit, dass Mittelwert und Varianz identisch sind.

Abbildung 3.5 zeigt, dass für großes n und kleines p die Binomialverteilung $\mathrm{Bi}(n,p)$ sehr gut durch die Poissonverteilung $\mathrm{Po}(np)$ angenähert werden kann. Dies ist eine in der Praxis häufig benützte Vereinfachung, da mit der Poissonverteilung leichter zu rechnen ist als mit der Binomialverteilung.

Ist der Mittelwert λ unbekannt, können unabhängige Beobachtungen von X dazu verwendet werden, um λ zu schätzen (siehe Abschnitt 6.2.2).

*http://de.wikipedia.org/wiki/Simeon_Denis_Poisson

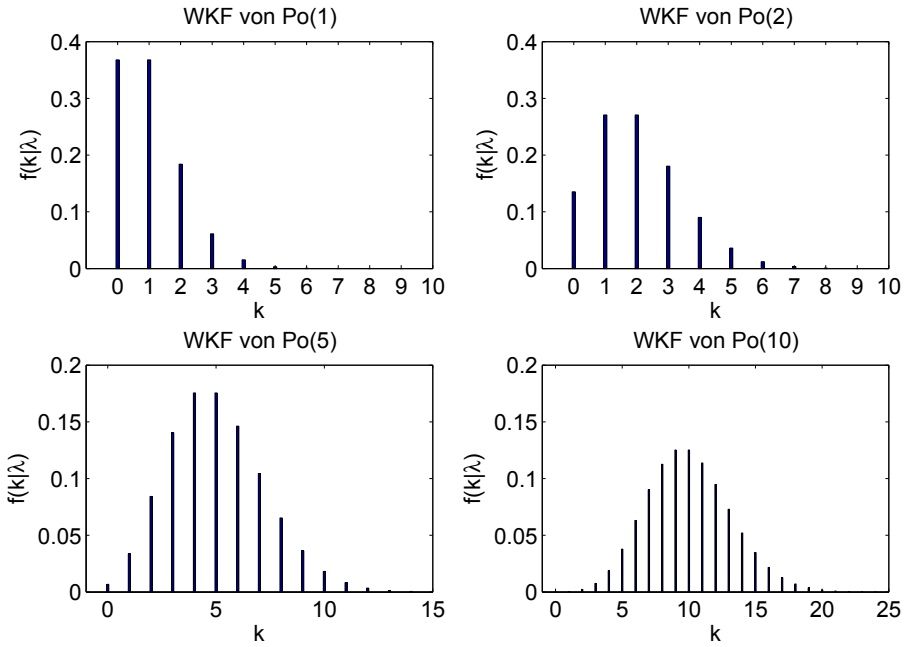

Abbildung 3.4: Die Wahrscheinlichkeitsfunktionen einiger Poissonverteilungen. MATLAB: `make_poisspdf`

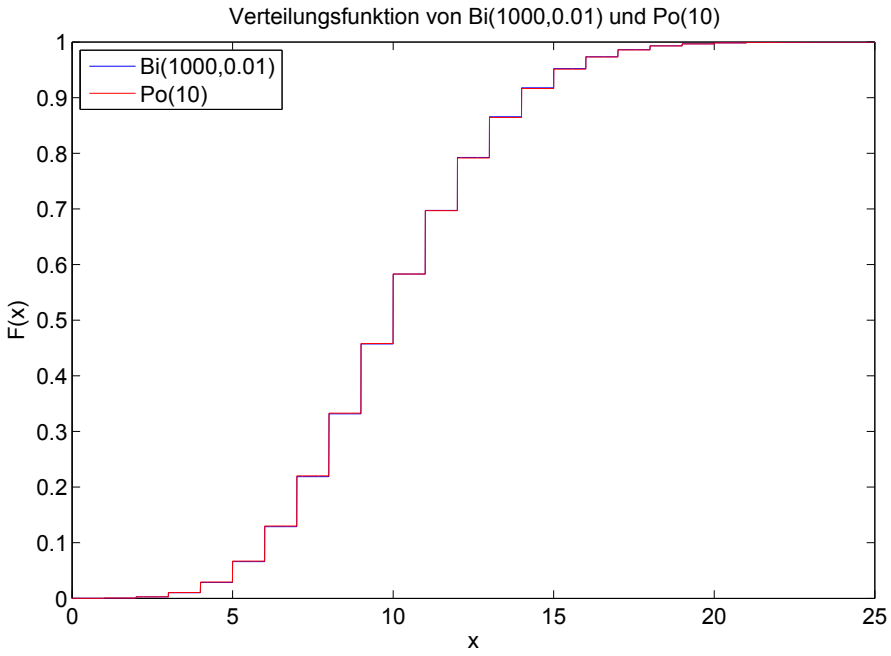

Abbildung 3.5: Näherung von $\mathrm{Bi}(1000, 0.01)$ durch $\mathrm{Po}(10)$. MATLAB: `make_binolimit_poisson`

3.4 Diskrete Zufallsvektoren

3.4.1 Grundbegriffe

☞ **Definition 3.4.1.** Diskreter Zufallsvektor

Eine multivariate Zufallsvariable, deren n Komponenten diskrete Zufallsvariable sind, heißt ein n-dimensionaler diskreter Zufallsvektor.

▶ **Beispiel 3.4.2.**

Der Ausgang des Wurfs mit zwei Würfeln wird durch einen 2-dimensionalen Zufallsvektor $\boldsymbol{X} = (X_1, X_2)$ beschrieben. ◀

▶ **Beispiel 3.4.3.**

Die Häufigkeiten in einer gruppierten Häufigkeitstabelle mit n Gruppen bilden einen diskreten Zufallsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$. ◀

☞ **Definition 3.4.4.** Wahrscheinlichkeitsfunktion eines diskreten Zufallsvektors

Die Funktion $f_{\boldsymbol{X}}(k_1, \ldots, k_n)$, die dem Ergebnis $(k_1, \ldots, k_n)$ seine Wahrscheinlichkeit unter der Verteilung von $\boldsymbol{X}$ zuordnet, wird als Wahrscheinlichkeitsfunktion oder als Dichtefunktion von $\boldsymbol{X}$ bezeichnet:

$$f_{\boldsymbol{X}}(k_1, \ldots, k_n) = W(X_1 = k_1 \cap \ldots \cap X_n = k_n)$$

☞ **Definition 3.4.5.** Randverteilung

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein diskreter Zufallsvektor. Die Verteilung einer Komponente X_i heißt die Randverteilung von $\boldsymbol{X}$ bzgl. X_i. Die Wahrscheinlichkeitsfunktion von X_i wird als Randdichte bezeichnet.

☛ **Satz 3.4.6.** Berechnung der Randverteilung

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein diskreter Zufallsvektor. Die Randdichte von $\boldsymbol{X}$ bzgl. X_i ist gegeben durch:

$$f_{X_i}(k_i) = W(X_i = k_i) = \sum_{j \neq i} \sum_{k_j \in \mathbb{N}_0} f_{\boldsymbol{X}}(k_1, \ldots, k_n)$$

Beispiel 3.4.14 zeigt die Berechnung der Randdichte anhand einer konkreten Verteilung.

☞ **Definition 3.4.7.** Unabhängigkeit

Es sei $\boldsymbol{X} = (X_1, X_2)$ eine diskreter Zufallsvektor mit der Wahrscheinlichkeitsfunktion $f_{\boldsymbol{X}}(k_1, k_2)$. X_1 und X_2 heißen unabhängig, wenn für alle Paare (k_1, k_2) gilt:

$$\begin{aligned} f_{\boldsymbol{X}}(k_1, k_2) &= W(X_1 = k_1 \cap X_2 = k_2) \\ &= W(X_1 = k_1) \cdot W(X_2 = k_2) \\ &= f_{X_1}(k_1) \cdot f_{X_2}(k_2) \end{aligned}$$

Allgemein sind die Komponenten eines Zufallsvektors unabhängig, wenn ihre gemeinsame Wahrscheinlichkeitsfunktion das Produkt der Randdichten ist.

☛ **Satz 3.4.8.** Erwartung eines Produkts

Sind X_1 und X_2 unabhängig, gilt:

$$\mathsf{E}[X_1 \cdot X_2] = \mathsf{E}[X_1] \cdot \mathsf{E}[X_2]$$

☞ **Definition 3.4.9.** Kovarianz

Die Kovarianz von X_1 und X_2, bezeichnet mit $\mathsf{cov}[X_1, X_2]$, ist definiert durch:

$$\begin{aligned}\mathsf{cov}[X_1, X_2] &= \mathsf{E}[(X_1 - \mathsf{E}[X_1])(X_2 - \mathsf{E}[X_2])] \\ &= \mathsf{E}[X_1 X_2] - \mathsf{E}[X_1]\mathsf{E}[X_2] \\ &= \mathsf{cov}[X_2, X_1]\end{aligned}$$

X_1 und X_2 heißen unkorreliert, wenn ihre Kovarianz gleich 0 ist.

Aus Satz 3.4.8 folgt, dass unabhängige Zufallsvariable stets unkorreliert sind. Die Umkehrung gilt im Allgemeinen jedoch nicht: unkorrelierte Zufallsvariable müssen keineswegs auch unabhängig sein.

☞ **Definition 3.4.10.** Kovarianzmatrix

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein Zufallsvektor. Die Matrix $\mathbf{C}$, definiert durch:

$$\mathbf{C}_{ij} = \mathsf{cov}[X_i, X_j]$$

heißt die Varianz-Kovarianzmatrix oder kurz Kovarianzmatrix von $\boldsymbol{X}$. Sie wird mit $\mathsf{Cov}[\boldsymbol{X}]$ bezeichnet.

Die Kovarianzmatrix ist symmetrisch und positiv (semi-)definit.* Sie enthält in der Diagonale die Varianzen und neben der Diagonale die paarweisen Kovarianzen aller Komponenten von $\boldsymbol{X}$.

☞ **Definition 3.4.11.** Korrelationsmatrix

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein Zufallsvektor und $\mathbf{C} = \mathsf{Cov}[\boldsymbol{X}]$ die Kovarianzmatrix von $\boldsymbol{X}$. Die Matrix $\mathbf{R}$ mit:

$$\mathbf{R}_{ij} = \frac{\mathbf{C}_{ij}}{\sqrt{\mathbf{C}_{ii}\mathbf{C}_{jj}}}$$

heißt die Korrelationsmatrix von $\boldsymbol{X}$.

Die Korrelationsmatrix ist ebenfalls symmetrisch und positiv (semi-)definit. Sie enthält in der Diagonale 1 und neben der Diagonale die paarweisen Korrelationskoeffizienten ρ_{ij} aller Komponenten von $\boldsymbol{X}$.

☛ **Satz 3.4.12.**

Der Korrelationskoeffizient erfüllt stets die Ungleichung

$$-1 \leq \rho_{ij} \leq 1$$

Ist ρ_{ij} positiv, heißen X_i und X_j positiv korreliert; ist ρ_{ij} negativ, heißen sie negativ korreliert.

☞ **Definition 3.4.13.** Bedingte Dichte

Es sei $\boldsymbol{X} = (X_1, X_2)$ ein 2-dimensionaler diskrete Zufallsvektor mit der Wahrscheinlichkeitsfunktion $f(k_1, k_2)$ und den Randdichten $f_1(k_1)$ bzw. $f_2(k_2)$. Die Funktion $f_{12}(k_1 \,|\, k_2)$, definiert durch:

$$f_{12}(k_1 \,|\, k_2) = \frac{f(k_1, k_2)}{f_2(k_2)}$$

heißt die durch X_2 bedingte Dichte von X_1.

Die bedingte Dichte ist für festes k_2 die Wahrscheinlichkeitsfunktion der durch $X_2 = k_2$ bedingten Verteilung von X_1.

► **Beispiel 3.4.14.** Randdichten und bedingte Dichten

Es sei (X_1, X_2) eine bivariate diskrete Zufallsvariable. Ihre gemeinsame Wahrscheinlichkeitsfunktion ist durch die folgende Kontingenztafel definiert:

*Das heißt, dass sie keine negativen Eigenwerte hat.

$f(k_1,k_2)$	$k_2=1$	$k_2=2$	$k_2=3$	$k_2=4$	$f_1(k_1)$
$k_1=1$	0.080	0.070	0.094	0.090	0.334
$k_1=2$	0.102	0.096	0.100	0.094	0.392
$k_1=3$	0.078	0.066	0.058	0.072	0.274
$f_2(k_2)$	0.260	0.232	0.252	0.256	1.000

Die Randdichte von X_1 ist in der rechten Randspalte zu sehen, die Randdichte von X_2 in der unteren Randzeile. Daher stammt übrigens die Bezeichnung „Randdichte“.

Die bedingte Dichte $f_{12}(k_1\,|\,k_2=2)$ von X_1 bedingt durch $X_2=2$ ist gleich der zweiten Spalte dividiert durch $f_2(2)=0.232$. Es gilt also:

$$f_{12}(k_1=1\,|\,k_2=2)=0.3017, f_{12}(k_1=2\,|\,k_2=2)=0.4138, f_{12}(k_1=3\,|\,2)=0.2845$$

Die bedingte Dichte $f_{21}(k_2 \mid k_1=3)$ von X_2 bedingt durch $X_1=3$ ist gleich der dritten Zeile dividiert durch $f_1(3)=0.274$. Es gilt also:

$$f_{21}(k_2=1\,|\,k_1=3)=0.2847, f_{21}(k_2=2\,|\,k_1=3)=0.2409,$$
$$f_{21}(k_2=3\,|\,k_1=3)=0.2117, f_{21}(k_2=4\,|\,k_1=3)=0.2628 \quad \blacktriangleleft$$

3.4.2 Die Multinomialverteilung

Der Alternativversuch kann dahingehend verallgemeinert werden, dass man nicht nur zwei, sondern k Ausgänge $1,\ldots,k$ zulässt, denen die Wahrscheinlichkeiten $p_1,\ldots,p_k$ zugeordnet werden. Diese müssen die Gleichung

$$\sum_{i=1}^{k} p_i = 1$$

erfüllen. Führt man den verallgemeinerten Alternativversuch n-mal durch, so ist jedes Ergebnis eine Folge der Form:

$$(i_1,\ldots,i_n),\ i_j\in\{1,\ldots,k\},\ j=1,\ldots,n$$

Sind die Versuche unabhängig, ist die Wahrscheinlichkeit einer solchen Folge gleich:

$$W(\{(i_1,\ldots,i_n)\})=\prod_{j=1}^{n} p_{i_j}$$

In der Regel ist man jedoch nur daran interessiert, wie oft im Lauf der n Versuche jeder der k Ausgänge eingetreten ist. Bezeichnet man die Häufigkeit des Ergebnisses i mit X_i, erhält man einen k-dimensionalen diskreten Zufallsvektor $\boldsymbol{X}=(X_1,\ldots,X_k)$. Die Verteilung von $\boldsymbol{X}$ wird als *Multinomialverteilung* $\mathrm{Mu}(n,p_1,\ldots,p_k)$ mit den Parametern n und $p_1,\ldots,p_k$ bezeichnet. Klarerweise muss gelten:

$$\sum_{i=1}^{k} X_i = n$$

☛ **Satz 3.4.15.** Wahrscheinlichkeitsfunktion der Multinomialverteilung

Die Wahrscheinlichkeitsfunktion der Multinomialverteilung $\mathrm{Mu}(n,p_1,\ldots,p_k)$ *lautet:*

$$f(n_1,\ldots,n_k\,|\,n,p_1,\ldots,p_k)=\frac{n!}{n_1!\ldots n_k!}\prod_{i=1}^{k} p_i^{n_i},\ \sum_{i=1}^{k} n_i=n,\ \sum_{i=1}^{k} p_i=1$$

Ist $k = 2$, wird aus der Multinomialverteilung eine Binomialverteilung. Ein häufiges Beispiel eines multinomialverteilten Zufallsvektors ist das Histogramm, das zur graphischen Darstellung der *experimentellen gruppierten Häufigkeit* verwendet wird (Abbildung 3.6). X_i ist die Anzahl der Fälle, in denen die Zufallsvariable Y, das experimentelle Ergebnis, in Gruppe i fällt. Die Wahrscheinlichkeit, dass Y in Gruppe i fällt, sei gleich p_i. Werden in das Histogramm n Ergebnisse eingefüllt, so sind die Gruppeninhalte $(X_1, \ldots, X_k)$ multinomial gemäß $\mathrm{Mu}(n, p_1, \ldots, p_k)$ verteilt.

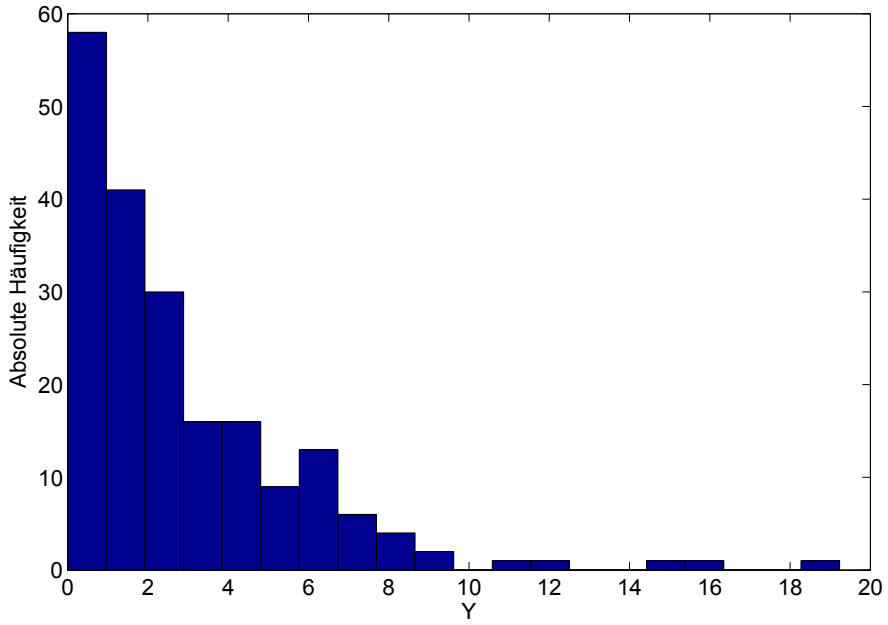

Abbildung 3.6: Ein Histogramm.

Für jede Gruppe i ist die Komponente X_i *binomialverteilt* mit den Parametern n und p_i. Diese Verteilung ist also die *Randverteilung* von $\boldsymbol{X}$ bzgl. X_i. Die Komponenten X_i sind *nicht unabhängig* voneinander, da ihre Summe gleich n sein muss. Die Wahrscheinlichkeitsfunktion von $\boldsymbol{X}$ ist daher *nicht* das Produkt ihrer Randdichten.

☛ **Satz 3.4.16.** Kovarianzmatrix der Multinomialverteilung

Die Kovarianzmatrix $\mathbf{C}$ *der Multinomialverteilung* $\mathrm{Mu}(n, p_1, \ldots, p_k)$ *hat die folgende Form:*

$$\mathbf{C}_{ij} = \begin{cases} np_i(1-p_i), & i = j \\ -np_ip_j, & i \neq j \end{cases}$$

Da die Summe aller Komponenten von $\boldsymbol{X}$ immer gleich n ist, sind die Komponenten stets negativ korreliert. Der Rang von $\mathbf{C}$ ist $k-1$, $\mathbf{C}$ ist daher positiv semi-definit.

3.5 Funktionen von diskreten Zufallsvariablen

Es sei X eine diskrete Zufallsvariable und $h : \mathbb{N}_0 \to \mathbb{R}$ eine beliebige Funktion. Dann ist $Y = h(X)$ wieder eine diskrete Zufallsvariable mit einer diskreten Ergebnismenge Ω. Der folgende Satz zeigt, wie die Verteilung bzw. die Wahrscheinlichkeitsfunktion von Y berechnet wird.

☛ **Satz 3.5.1.** Transformation von diskreten Zufallsvariablen
Es sei X eine diskrete Zufallsvariable und $Y = h(X)$. Dann gilt für $x \in \Omega$:

$$f_Y(y) = W_X(h^{-1}(y)) = \sum_{h(j)=y} f_x(j)$$

Hier bezeichnet $h^{-1}(y)$ das *Urbild* von y unter h, d.h. die Menge aller j, die von der Funktion h auf y abgebildet werden.

► **Beispiel 3.5.2.**
Die geraden Augenzahlen des Würfels werden auf die Zahl 0 abgebildet, die ungeraden auf die Zahl 1:

$$Y = h(X) = \mathrm{mod}(X, 2)$$

Die Wahrscheinlichkeitsfunktion von Y ist dann gegeben durch:

$$f_Y(0) = W_X(h^{-1}(0)) = W_X(\{2,4,6\}) = \tfrac{1}{2}$$
$$f_Y(1) = W_X(h^{-1}(1)) = W_X(\{1,3,5\}) = \tfrac{1}{2}$$ ◄

Mit dem gleichen Verfahren kann auch die Verteilung von Funktionen von diskreten Zufallsvektoren berechnet werden.

► **Beispiel 3.5.3.**

Dem Ergebnis $\boldsymbol{X} = (X_1, X_2)$ eines Doppelwurfs wird die Summe der Augenzahlen zugeordnet:

$$Y = h(X_1, X_2) = X_1 + X_2$$

Sind die Würfe unabhängig, ist die Wahrscheinlichkeit jedes Ergebnisses des Doppelwurfs gleich 1/36. Die Werte von Y sind die natürlichen Zahlen von 2 bis 12. Die Verteilung von Y kann folgendermaßen bestimmt werden. Für jeden möglichen Wert k werden alle Paare (i, j) aufgelistet, deren Summe gleich k ist. Die Wahrscheinlichkeitsfunktion $f_Y(k)$ ist dann die Summe der Wahrscheinlichkeiten dieser Paare und damit gleich der Anzahl der Paare mal 1/36. Die Anzahl ist gleich $k-1$ für $k \leq 7$ und gleich $13-k$ für $k > 7$. Zusammenfassend ergibt sich:

$$f_Y(k) = W_{\boldsymbol{X}}(h^{-1}(k)) = \sum_{i+j=k} f_{\boldsymbol{X}}((i,j)) = \begin{cases} \dfrac{k-1}{36}, & k \leq 7 \\ \dfrac{13-k}{36}, & k \geq 7 \end{cases}$$

Abbildung 3.7 zeigt die Wahrscheinlichkeitsfunktion und die Verteilungsfunktion der Zufallsvariablen $Y = X_1 + X_2$.

➠MATLAB: `make_doppelwurf` ◄

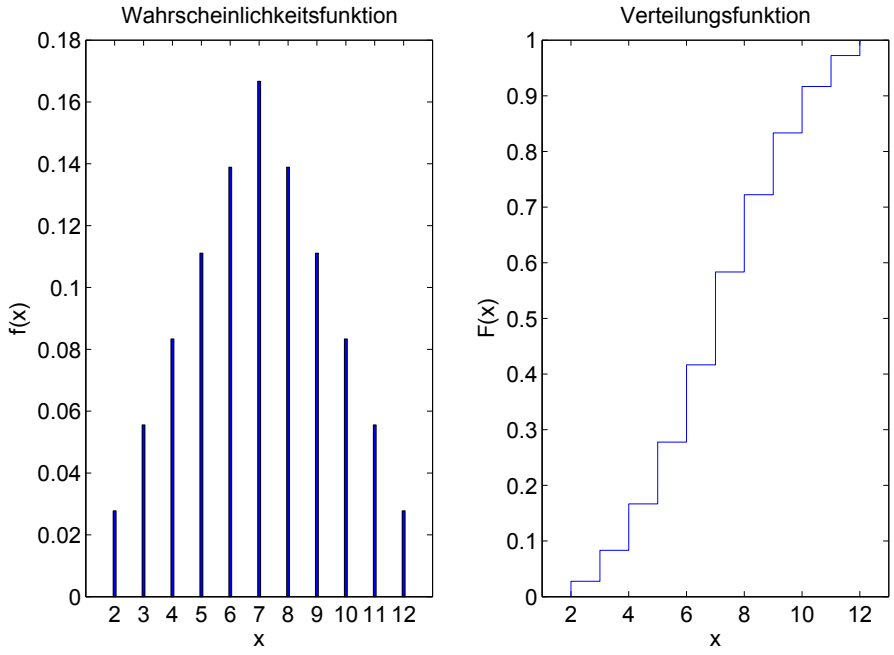

Abbildung 3.7: Wahrscheinlichkeitsfunktion (links) und Verteilungsfunktion (rechts) der Summe eines Doppelwurfs. MATLAB: `make_doppelwurf`

Kapitel 4

Stetige Verteilungen

4.1 Grundbegriffe

4.1.1 Stetige Verteilungen

Stetige Zufallsvariable sind meist das Resultat von *Messungen*. Der zufällige Charakter der Messungen kann verschiedene Ursachen haben: die fundamentale Zufälligkeit von quantenmechanischen Prozessen, Unkenntnis des genauen Anfangszustands oder der Dynamik eines Systems, Messfehler der Messapparatur, und andere mehr. Oft wirken mehrere Ursachen zusammen.

▶ **Beispiel 4.1.1.** LEBENSDAUERMESSUNG

Die Lebensdauer eines instabilen Teilchens ist prinzipiell zufällig, und zwar exponentialverteilt; dazu kommt noch der Messfehler, der durch die beschränkte Auflösung der Messapparatur verursacht wird. ◀

Eine stetige Zufallsvariable X kann im Prinzip ein Kontinuum von Werten annehmen. Als Ereignisse kommen vor allem Intervalle von reellen Zahlen in Betracht, sowie die Mengen, die man durch die Bildung von (abzählbar vielen) Durchschnitten und Vereinigungen aus Intervallen gewinnen kann. Im Folgenden ist die Ereignisalgebra Σ die Menge aller solcher Teilmengen von $\mathbb{R}$. Die Ereignisalgebra hat überabzählbar viele Elementarereignisse.

Es ist in diesem Fall nicht möglich, die Verteilung von X durch die Wahrscheinlichkeiten der Elementarereignisse zu spezifizieren. Die Verteilung muss vielmehr durch die Angabe einer (stetigen) Funktion definiert werden. Dafür sind drei Möglichkeiten im Gebrauch: die Dichtefunktion, die Verteilungsfunktion, d.h. eine Stammfunktion der Dichtefunktion, oder die charakteristische Funktion, d.h. die Fouriertransformierte der Dichtefunktion. Die anschaulichste Darstellung ist jene durch die Dichtefunktion.

☞ **Definition 4.1.2.** Stetige Dichtefunktion

Eine stetige Funktion $f(x) : \mathbb{R} \to \mathbb{R}$ ist eine Wahrscheinlichkeitsdichtefunktion oder kurz Dichtefunktion, wenn sie nichtnegativ und auf 1 normiert ist:

$$f(x) \geq 0, \quad \textit{für alle } x \in \mathbb{R} \quad \textit{und} \quad \int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$$

Manche Dichtefunktionen haben eine Polstelle, an der sie gegen ∞ streben. Dies ist zulässig, solange das Integral über die Funktion existiert und gleich 1 ist.

Die Verteilung W_X von X lässt sich leicht mit Hilfe der Dichtefunktion f_X von X angeben:

$$W_X(\mathrm{A}) = \int_{\mathrm{A}} f_X(x)\,\mathrm{d}x$$

Die Wahrscheinlichkeit eines Elementarereignisses ist immer gleich 0:

$$W_X(\{x\}) = \int_x^x f_X(x)\,\mathrm{d}x = 0$$

Daher gilt auch:

$$W_X((x_1, x_2]) = W_X((x_1, x_2)) = W_X([x_1, x_2])$$

Da jedes Elementarereignis Wahrscheinlichkeit 0 hat, ist es sinnvoller nach der Wahrscheinlichkeit zu fragen, mit der X in ein bestimmtes Intervall fällt. Ist das Intervall $[x, x + \mathrm{d}x]$ infinitesimal klein, ist die Wahrscheinlichkeit gleich $f_X(x)\,\mathrm{d}x$.

Die Wahrscheinlichkeit, dass X einen Wert x nicht übersteigt, ist dann gleich

$$W(X \leq x) = \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x = F_X(x)$$

$F_X(x)$ ist offenbar die Stammfunktion von $f_X(x)$ mit $\lim_{x\to\infty} = 1$. Sie wird die *Verteilungsfunktion* von X genannt.

☞ **Definition 4.1.3.** Stetige Verteilungsfunktion

Es sei X eine stetige Zufallsvariable. Die Funktion F_X, definiert durch:

$$F_X(x) = W(X \leq x) = \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x$$

heißt die Verteilungsfunktion von X.

☛ **Satz 4.1.4.**

Die Wahrscheinlichkeit, dass X in ein Intervall $(x_1, x_2]$ fällt, ist gleich:

$$W(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$$

Abbildung 4.1 zeigt ein Beispiel einer stetigen Dichtefunktion und der zugehörigen Verteilungsfunktion.

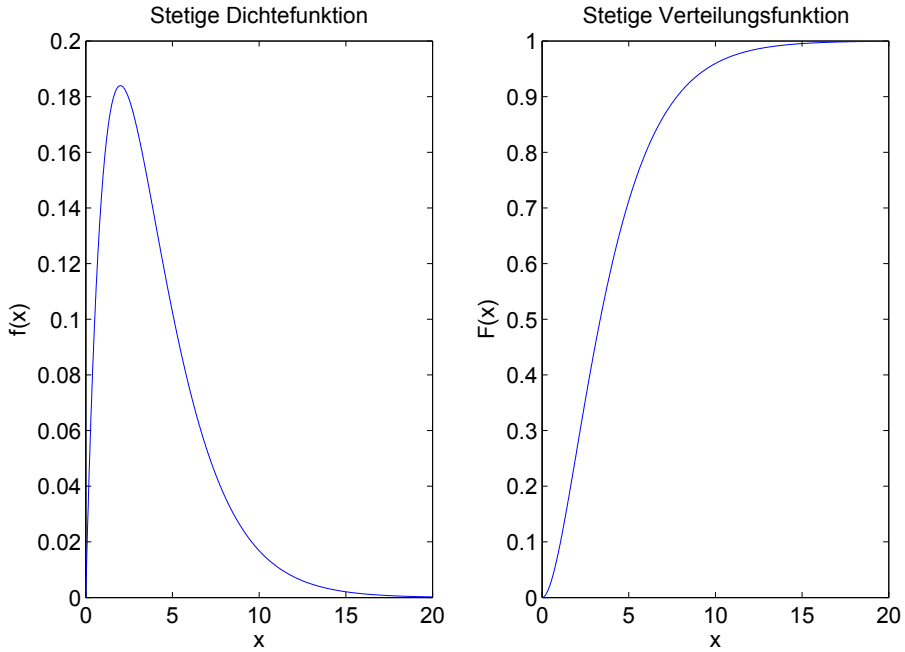

Abbildung 4.1: Beispiel einer stetigen Dichtefunktion (links) und der zugehörigen Verteilungsfunktion (rechts). MATLAB: `make_stetig_pdf_cdf`

☛ **Satz 4.1.5.** EIGENSCHAFTEN EINER STETIGEN VERTEILUNGSFUNKTION

Eine stetige Verteilungsfunktion $F(x)$ hat die folgenden Eigenschaften:

1. $0 \leq F(x) \leq 1$, *für alle* $x \in \mathbb{R}$
2. $x_1 \leq x_2 \implies F(x_1) \leq F(x_2)$, *für alle* $x_1, x_2 \in \mathbb{R}$
3. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$

☞ **Definition 4.1.6.** QUANTILSFUNKTION

Es sei $F_X(x)$ eine stetige und umkehrbare Verteilungsfunktion. Der Wert x_α, für den

$$F_X(x_\alpha) = \alpha, \quad 0 < \alpha < 1$$

gilt, heißt das α-Quantil der Verteilung von X. Die Funktion

$$x = F_X^{-1}(\alpha), \quad 0 < \alpha < 1$$

heißt die Quantilsfunktion von X.

☞ **Definition 4.1.7.** QUARTIL

Die Quantile zu den Niveaus $\alpha = 0.25, 0.5, 0.75$ heißen Quartile. Das Quantil zum Niveau $\alpha = 0.5$ heißt Median der Verteilung.

Quantile können auch für diskrete Verteilungen definiert werden, jedoch sind sie dann nicht immer eindeutig. Abbildung 4.2 zeigt ein Beispiel einer stetigen Verteilungsfunktion und der zugehörigen Quantilsfunktion.

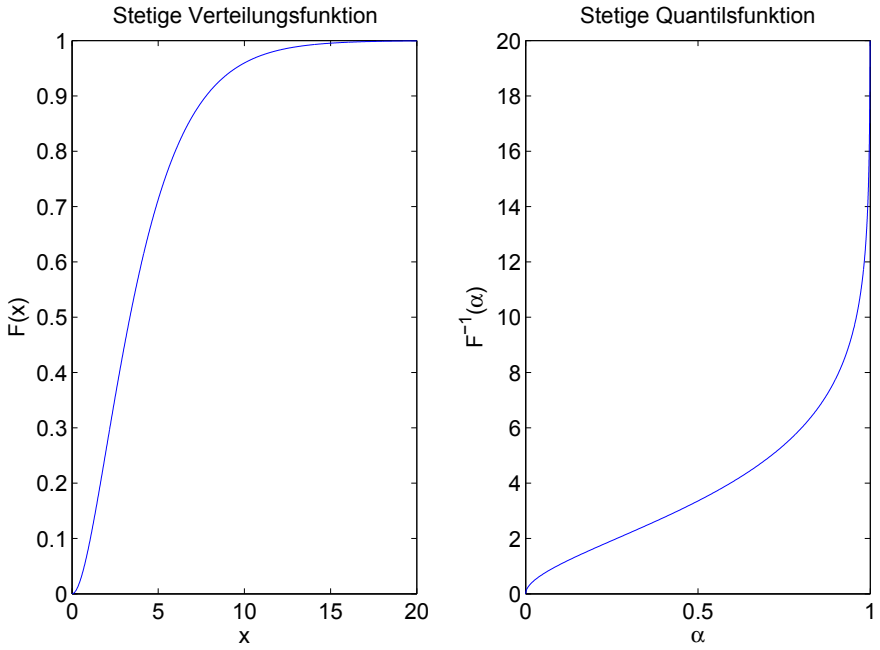

Abbildung 4.2: Beispiel einer stetigen Verteilungsfunktion (links) und der zugehörigen Quantilsfunktion (rechts). MATLAB: `make_stetig_cdf_pdf`

☞ **Definition 4.1.8.** MODUS

Ein Modus ist eine Stelle, an der die Dichtefunktion ein lokales Maximum hat. Verteilungen mit genau einem Modus heißen unimodal.

Die im Folgenden verwendeten Verteilungen sind sämtlich unimodal, jedoch können in den Anwendungen auch Verteilungen mit mehreren Maxima auftreten.

4.1.2 Erwartungswert und Varianz

☞ **Definition 4.1.9.** ERWARTUNG

Es sei X eine stetige Zufallsvariable mit der Dichtefunktion $f_X(x)$. Ferner sei g eine beliebige stetige reelle oder komplexe Funktion. Man definiert $\mathsf{E}[g(X)]$ durch:

$$\mathsf{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\,\mathrm{d}x$$

$\mathsf{E}[g(X)]$ heißt die Erwartung von $g(X)$. Ist $g(x) = x$, so heißt $\mathsf{E}[g(X)] = \mathsf{E}[X]$ der Erwartungswert von X oder der Mittelwert der Verteilung von X. Es gilt:

$$\mathsf{E}[X] = \int_{\mathbb{R}} x\, f_X(x)\,\mathrm{d}x$$

Der Mittelwert einer Verteilung braucht nicht zu existieren. Ein Beispiel ist die Cauchy-Verteilung (T-Verteilung mit einem Freiheitsgrad, siehe Abschnitt 4.2.8) mit der Dichtefunktion:

$$f(x) = \frac{1}{\pi(1+x^2)},\ x \in \mathbb{R}$$

Der Mittelwert ist wie im diskreten Fall ein Lageparameter. Er ist die x-Koordinate des *Schwerpunkts* der von der x-Achse und dem Graphen der Dichtefunktion eingeschlossenen Fläche.

☞ **Definition 4.1.10.** VARIANZ

Die Varianz von X, bezeichnet mit $\mathsf{var}[X]$, ist definiert durch:

$$\mathsf{var}[X] = \mathsf{E}[(X - \mathsf{E}[X])^2] = \int_{\mathbb{R}} (x - \mathsf{E}[X])^2\, f_X(x)\,\mathrm{d}x$$

Die Wurzel aus der Varianz heißt die Standardabweichung von X, bezeichnet mit $\sigma[X]$.

Die Standardabweichung ist ein Skalenparameter, der die Breite der Verteilung beschreibt. Sie hat die gleiche Dimension wie X. Varianz und Standardabweichung sind invariant gegen Translationen. Die Sätze 3.1.8 und 3.1.11 können unverändert vom diskreten auf den stetigen Fall übertragen werden.

Erwartungswert und Varianz sind Spezialfälle eines allgemeineren Begriffs, nämlich der des *Moments* einer Verteilung.

☞ **Definition 4.1.11.** MOMENTE

Sei X eine Zufallsvariable. Die Erwartung von $g(x) = (x-a)^k$ mit $k \in \mathbb{N}, a \in \mathbb{R}$, sofern sie existiert, heißt k-tes Moment von X um a. Das k-te Moment um 0 wird mit μ'_k bezeichnet. Ist $a = \mathsf{E}[X]$, wird das k-te Moment als zentrales Moment μ_k bezeichnet.

Der Erwartungswert $\mathsf{E}[X]$ ist das erste Moment um 0, die Varianz $\mathsf{var}[X]$ ist das zweite zentrale Moment $\mu_2 = \mathsf{E}[(X - \mathsf{E}[X])^2]$. Existieren die zentralen Momente $\mu_1, \ldots, \mu_k$, so können sie aus den Momenten um 0 $\mu'_1, \ldots, \mu'_k$ berechnet werden, und umgekehrt. Ist die Dichtefunktion von X außerhalb eines endlichen Intervalls gleich 0, existieren Momente für alle $k \in \mathbb{N}$.

4.2 Wichtige Verteilungen

4.2.1 Verteilungsfamilien

Die im Folgenden beschriebenen Verteilungen, die alle in den Anwendungen häufig vorkommen, hängen von einem oder mehreren Parametern ab. Es handelt sich also um *Verteilungsfamilien*. Manche Familien, wie etwa die Familie der Normalverteilungen, sind invariant unter affinen Transformationen der Form $Y = m + sX$, d.h. Verschiebung und Skalierung. Sie werden daher *Lagen-Skalen-Familien* genannt. In diesem Fall ist einer der beiden Parameter ein Lageparameter $m \in \mathbb{R}$, der andere ein Skalenparameter $s \in \mathbb{R}_+$ (siehe auch Abschnitt 3.1.2). Es sei $f_X(x\,|\,0,1)$ die Dichtefunktion der Zufallsvariablen X aus der Lagen-Skalen-Familie $\mathfrak{F}$ mit $m = 0$ und $s = 1$. Diese Dichtefunktion wird als *standardisiert* bezeichnet. Jede Verteilung in der Familie $\mathfrak{F}$ entsteht durch eine affine Transformation der Form $Y = m + sX$. Die Dichtefunktion von Y ist dann gegeben durch (siehe Satz 4.3.1):

$$f_Y(y\,|\,m,s) = \frac{1}{s}\, f_X\left(\frac{y-m}{s}\,\middle|\,0,1\right)$$

Nicht alle im Folgenden besprochenen Familien sind Lagen-Skalen-Familien; zum Beispiel ist die Familie der Gammaverteilungen (siehe Abschnitt 4.2.5) keine Lagen-Skalen-Familie.

4.2.2 Die Gleichverteilung

Die *Gleichverteilungsfamilie* $\mathrm{Un}(a,b)$ ist durch konstante Dichtefunktionen auf endlichen Intervallen charakterisiert.

☞ **Definition 4.2.1.** Dichtefunktion der Gleichverteilung

Die Dichtefunktion der Gleichverteilung $\mathrm{Un}(a,b)$ *lautet:*

$$f(x\,|\,a,b) = \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x)$$

Die Funktion $\mathbb{1}_{[a,b]}(x)$ ist die *Indikatorfunktion* auf dem Intervall $[a,b]$:

$$\mathbb{1}_{[a,b]}(x) = \begin{cases} 1, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$$

☛ **Satz 4.2.2.** Verteilungsfunktion der Gleichverteilung

Die Verteilungsfunktion der Gleichverteilung $\mathrm{Un}(a,b)$ *lautet:*

$$F(x\,|\,a,b) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a \le x \le b \\ 1, & x > b \end{cases}$$

Abbildung 4.3 zeigt Dichtefunktion (links) und Verteilungsfunktion (rechts) der Gleichverteilung $\mathrm{Un}(0,1)$.

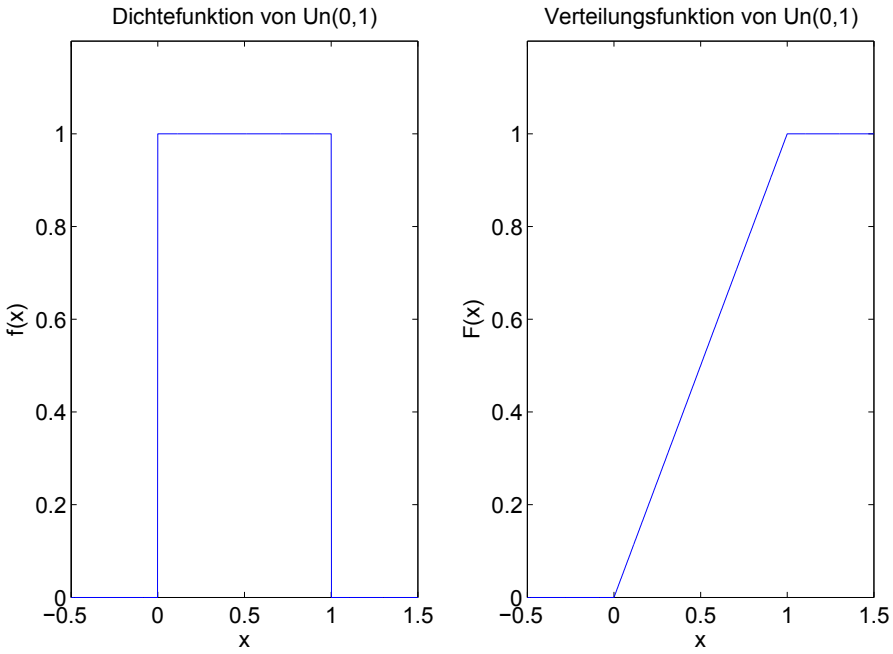

Abbildung 4.3: Dichtefunktion und Verteilungsfunktion der Gleichverteilung Un(0, 1). MATLAB: `make_unifcdf_pdf`

☛ **Satz 4.2.3.** Mittelwert und Varianz der Gleichverteilung
Es sei $X \sim \mathrm{Un}(a, b)$. *Dann gilt:*

$$\mathsf{E}[X] = \frac{a+b}{2} \quad \textit{und} \quad \mathsf{var}[X] = \frac{(b-a)^2}{12}$$

☛ **Satz 4.2.4.** Median und Quantile der Gleichverteilung
Es sei $X \sim \mathrm{Un}(a, b)$. *Dann ist der Median gleich dem Mittelwert* $(a+b)/2$ *und das* α*-Quantil ist gleich* $a + \alpha(b-a)$.

Die Gleichverteilung kann auch als Lagen-Skalen-Familie mit dem Lageparameter $m = (a+b)/2$ und dem Skalenparameter $s = (b-a)/2$ parametrisiert werden. Umgekehrt gilt dann $a = m - s$ und $b = m + s$.

▶ **Beispiel 4.2.5.** Gleichverteilter Messfehler
Wird eine kontinuierliche gemessene Größe durch Abrunden auf einen ganzzahligen Wert diskretisiert, ist der dadurch entstehende Messfehler gleichverteilt im Intervall $[-1, 0]$, mit Mittelwert -0.5 und Standardabweichung $1/\sqrt{12} \approx 0.289$. Ein Beispiel dafür ist ein Analog-Digital-Konverter (ADC), der ganzzahlige Werte ausgibt. ◀

4.2.3 Die Normalverteilung

4.2.3.1 Dichtefunktion und Verteilungsfunktion

Die Dichtefunktion der *Normalverteilungsfamilie* hängt von zwei Parametern μ und σ^2 ab. Sie ist eine Lagen-Skalen-Familie, wobei μ der Lageparameter und σ der Skalenparameter ist. Die Familie wird mit $\mathrm{No}(\mu, \sigma^2)$ bezeichnet. Die Normalverteilung wird auch als Gaußverteilung bezeichnet, nach dem deutschen Mathematiker, Astronomen, Geodäten und Physiker C. F. Gauß.*

☞ **Definition 4.2.6.** Dichtefunktion der Normalverteilung
Die Dichtefunktion der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$ *lautet:*

$$f(x\,|\,\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad \mu \in \mathbb{R}, \sigma > 0$$

☛ **Satz 4.2.7.** Mittelwert und Varianz der Normalverteilung
Es sei $X \sim \mathrm{No}(\mu, \sigma^2)$. *Dann gilt:*

$$\mathsf{E}[X] = \mu \quad \textit{und} \quad \mathsf{var}[X] = \sigma^2$$

Der Median und der Modus der Verteilung sind gleich dem Mittelwert.

Die Wendepunkte der Dichtefunktion sind bei $x = \mu \pm \sigma$. Die halbe Breite auf halber Höhe (HWHM) ist gleich $\sigma\sqrt{2\ln 2} \approx 1,177\sigma$ (Abbildung 4.4). Ist $X \sim \mathrm{No}(\mu, \sigma^2)$, gilt:

$$\begin{aligned} W(|X-\mu| \le \sigma) &= 68.3\% \\ W(|X-\mu| \le 2\sigma) &= 95.4\% \\ W(|X-\mu| \le 3\sigma) &= 99.7\% \end{aligned}$$

*http://de.wikipedia.org/wiki/Gauss

4.2.3.2 Die Standardnormalverteilung

☞ **Definition 4.2.8.** STANDARDNORMALVERTEILUNG
Die Normalverteilung $\mathrm{No}(0,1)$ *mit* $\mu = 0$ *und* $\sigma^2 = 1$ *wird als Standardnormalverteilung bezeichnet. Ihre Dichtefunktion lautet:*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

☞ **Definition 4.2.9.** STANDARDSCORE
Ist $X \sim \mathrm{No}(\mu, \sigma^2)$*, so wird*

$$Z = \frac{X - \mu}{\sigma}$$

als das Standardscore von X *bezeichnet.*

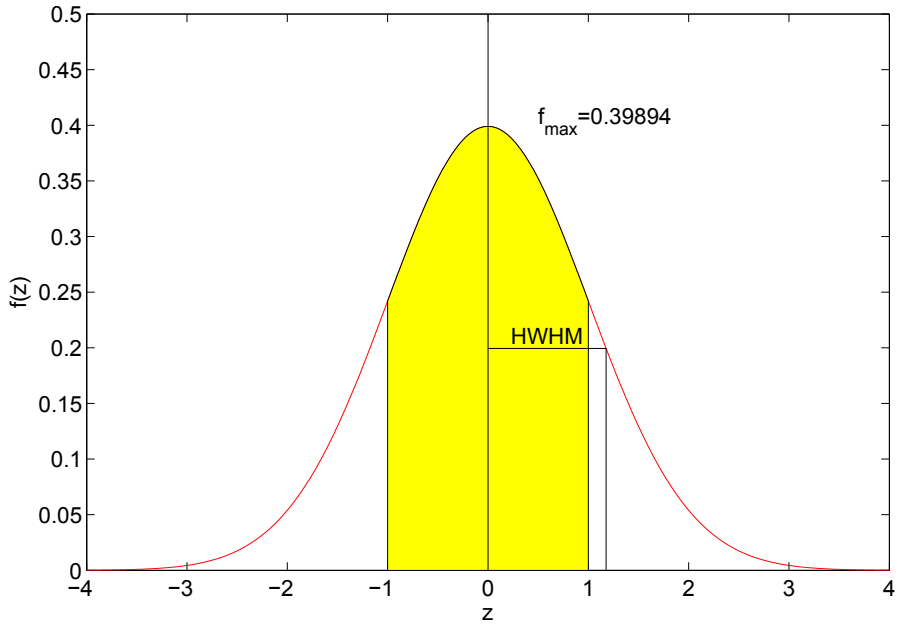

Abbildung 4.4: Dichtefunktion von $Z = (X - \mu)/\sigma$. Der gelbe Bereich enthält 68.2% der Gesamtfläche. MATLAB: `make_normpdf`

☛ **Satz 4.2.10.**
Ist $X \sim \mathrm{No}(\mu, \sigma^2)$*, so ist das Standardscore*

$$Z = \frac{X - \mu}{\sigma}$$

standardnormalverteilt.

Die Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung ist definiert durch:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u)\,\mathrm{d}u$$

Sie kann nicht durch elementare Funktionen ausgedrückt werden. Abbildung 4.5 zeigt Dichtefunktion und Verteilungsfunktion. Das α-Quantil wird mit z_α bezeichnet. Tabellen von $\Phi(x)$ und z_α sind im Anhang T.1 bzw. T.2 zu finden. Wegen der Symmetrie der Verteilung gilt $\Phi(-x) = 1 - \Phi(x)$ und $z_\alpha = -z_{1-\alpha}$.

☛ **Satz 4.2.11.**
Die Verteilungsfunktion $\Phi(x\,|\,\mu,\sigma^2)$ *der Normalverteilung* $\mathrm{No}(\mu,\sigma^2)$ *ist gleich:*

$$\Phi(x\,|\,\mu,\sigma^2) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Das α*-Quantil von* $\mathrm{No}(\mu,\sigma^2)$ *ist gleich* $\mu + \sigma z_\alpha$.

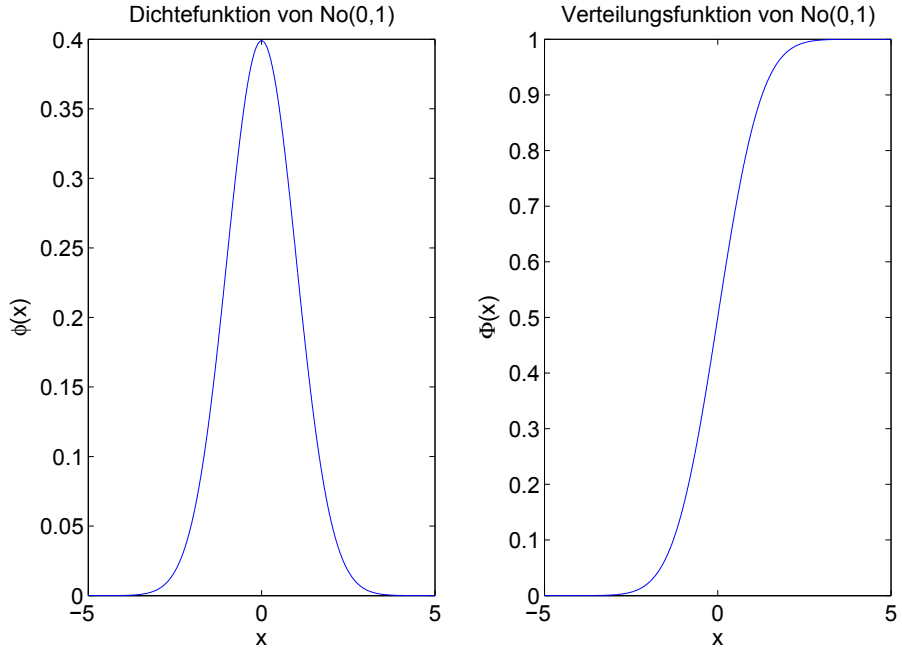

Abbildung 4.5: Dichtefunktion (links) und Verteilungsfunktion (rechts) der Standardnormalverteilung. MATLAB: `make_normcdf_pdf`

4.2.4 Die Exponentialverteilung

4.2.4.1 Dichtefunktion und Momente

Die *Exponentialverteilung* $\mathrm{Ex}(\tau)$ ist die Wartezeitverteilung des radioaktiven Zerfalls von Atomen und allgemein des Zerfalls von Elementarteilchen.

☞ **Definition 4.2.12.** DICHTEFUNKTION DER EXPONENTIALVERTEILUNG
Die Dichtefunktion der Exponentialverteilung $\mathrm{Ex}(\tau)$ *lautet:*

$$f(x\,|\,\tau) = \frac{1}{\tau}\,\mathrm{e}^{-x/\tau}\cdot\mathbb{1}_{[0,\infty)}(x), \quad \tau > 0$$

☛ **Satz 4.2.13.** VERTEILUNGSFUNKTION DER EXPONENTIALVERTEILUNG
Die Verteilungsfunktion der Exponentialverteilung $\mathrm{Ex}(\tau)$ *lautet:*

$$F(x\,|\,\tau) = \left(1-\mathrm{e}^{-x/\tau}\right)\cdot\mathbb{1}_{[0,\infty)}(x)$$

Abbildung 4.6 zeigt Dichtefunktion und Verteilungsfunktion der Exponentialverteilung $\mathrm{Ex}(2)$.

☛ **Satz 4.2.14.** Mittelwert und Varianz der Exponentialverteilung
Es sei $X \sim \mathrm{Ex}(\tau)$. *Dann gilt:*

$$\mathsf{E}[X] = \tau \quad \textit{und} \quad \mathsf{var}[X] = \tau^2$$

☛ **Satz 4.2.15.** Median und Quantile der Exponentialverteilung
Der Median der Verteilung ist gleich $\tau \ln 2$. *Er wird auch als Halbwertszeit bezeichnet. Das* α*-Quantil ist gleich* $-\tau \ln(1-\alpha)$.

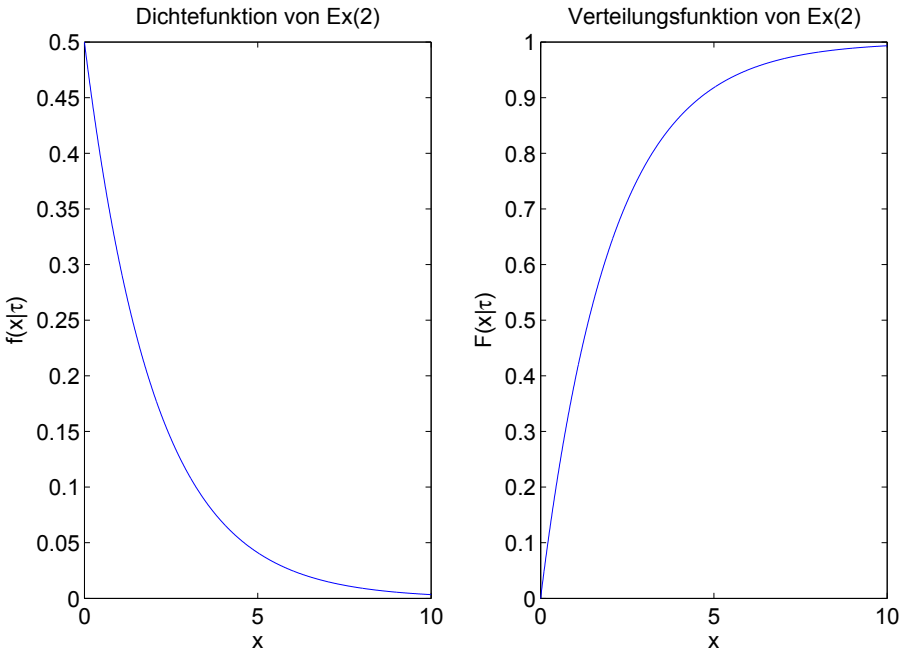

Abbildung 4.6: Dichtefunktion (links) und Verteilungsfunktion (rechts) der Exponentialverteilung Ex(2). MATLAB: `make_expcdf_pdf`

4.2.4.2 Der Poissonprozess

Es wird ein Prozess beobachtet, bei dem gewisse Ereignisse zu zufälligen Zeitpunkten eintreten. Ist die Anzahl der Ereignisse pro Zeiteinheit unabhängig und poissonverteilt gemäß $\mathrm{Po}(\lambda)$, nennt man den Prozess einen *Poissonprozess* mit Intensität oder Rate λ.

☛ **Satz 4.2.16.** Eigenschaften eines Poissonprozesses
Ein Poissonprozess mit Rate λ *hat die folgenden Eigenschaften:*

1. *Die Anzahl der Ereignisse in einem Zeitintervall der Länge* t *ist poissonverteilt gemäß* $\mathrm{Po}(\lambda t)$.
2. *Die Wartezeit zwischen zwei aufeinanderfolgenden Ereignissen ist exponentialverteilt gemäß* $\mathrm{Ex}(1/\lambda)$.
3. *Sind die Wartezeiten eines Prozesses unabhängig und exponentialverteilt gemäß* $\mathrm{Ex}(\tau)$, *so ist der Prozess ein Poissonprozess mit Rate* $\lambda = 1/\tau$.

Ein gutes Beispiel eines Poissonprozesses liefern die Zerfälle einer radioaktiven Quelle, die so groß ist, dass die Intensität im Beobachtungszeitraum nicht messbar abnimmt.

Die Exponentialverteilung kann auch mittels der Rate $\lambda = 1/\tau$ parametrisiert werden. Die Dichtefunktion ist dann gleich $f(x|\lambda) = \lambda e^{-\lambda x}$, die Verteilungsfunktion ist gleich $F(x|\lambda) = 1 - e^{-\lambda x}$, der Mittelwert ist $1/\lambda$, und die Varianz ist $1/\lambda^2$. Diese Parametrisierung ist in der Lebensdaueranalyse und in der Bayes-Statistik gebräuchlich.

4.2.5 Die Gammaverteilung

Die Exponentialverteilung ist ein Spezialfall einer allgemeineren Familie von Verteilungen, der *Gammaverteilungsfamilie*. Diese ist keine Lagen-Skalen-Familie, da sie nicht invariant unter affinen Transformationen ist; sie ist aber invariant unter Streckungen, d.h. Transformationen der Form $y = c \cdot x$ mit $c > 0$.

☞ **Definition 4.2.17.** Dichtefunktion der Gammaverteilung
Die Dichtefunktion der Gammaverteilung $\mathrm{Ga}(a, b)$ *lautet:*

$$f(x|a,b) = \frac{x^{a-1}e^{-x/b}}{b^a \Gamma(a)} \cdot \mathbb{1}_{[0,\infty)}(x), \quad a, b > 0$$

wo $\Gamma(a)$ *die Eulersche Gammafunktion ist.*

Die Eulersche Gammafunktion* erfüllt die Rekursionsformel:

$$\Gamma(x) = (x-1)\Gamma(x-1),\ x > 0$$

Für ganzzahlige Argumente $n \in \mathbb{N}$ gilt daher $\Gamma(n) = (n-1)!$. Für halbzahlige Argumente kann die Gammafunktion mit der Rekursionsformel und dem Wert $\Gamma(1/2) = \sqrt{\pi}$ berechnet werden.

Abbildung 4.7 zeigt die Dichtefunktionen einiger Gammaverteilungen. Die Exponentialverteilung $\mathrm{Ex}(\tau)$ ist die Gammaverteilung $\mathrm{Ga}(1, \tau)$.

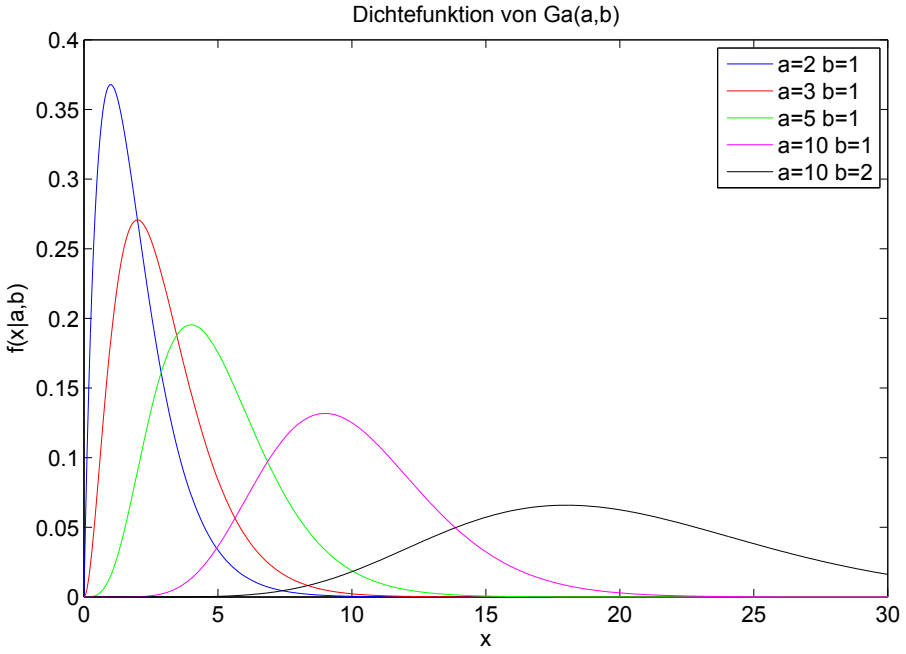

Abbildung 4.7: Dichtefunktionen einiger Gammaverteilungen.
MATLAB: `make_gampdf`

☛ **Satz 4.2.18.** VERTEILUNGSFUNKTION DER GAMMAVERTEILUNG

Die Verteilungsfunktion der Gammaverteilung $\mathrm{Ga}(a, b)$ *ist die folgende Stammfunktion der Dichtefunktion:*

$$F(x\,|\,a, b) = \int_0^x \frac{t^{a-1}\mathrm{e}^{-t/b}}{b^a\Gamma(a)}\,\mathrm{d}t$$

☛ **Satz 4.2.19.** MITTELWERT UND VARIANZ DER GAMMAVERTEILUNG

Es sei $X \sim \mathrm{Ga}(a, b)$. *Dann gilt:*

$$\mathsf{E}[X] = ab \quad \textit{und} \quad \mathsf{var}[X] = ab^2$$

Der Median ist außer dem Spezialfall $a = 1$ *nicht in geschlossener Form darstellbar.*

a wird als der Formparameter der Familie bezeichnet, weil er die Form der Dichtefunktion bestimmt (siehe Abbildung 4.7). b ist hingegen ein reiner Skalenparameter. Das α-Quantil der $\mathrm{Ga}(a, b)$-Verteilung wird mit $\gamma_{\alpha|a,b}$ bezeichnet. Die Quantile $\gamma_{\alpha|n,1}$ sind für $n \in \mathbb{N}$ im Anhang T.3 tabelliert.

*http://de.wikipedia.org/wiki/Gammafunktion

☛ **Satz 4.2.20.**
Ist $X \sim \mathrm{Ga}(a,b)$ *und* $c > 0$*, so ist* $Y = c \cdot X \sim \mathrm{Ga}(a, bc)$*. Daraus folgt:*

$$\gamma_{\alpha|a,b} = b \cdot \gamma_{\alpha|a,1}$$

Das α*-Quantil der Exponentialverteilung* $\mathrm{Ex}(\tau)$ *ist gleich* $\tau \cdot \gamma_{\alpha|1,1}$.

Analog zur Exponentialverteilung kann die Gammaverteilung auch mit dem Formparameter a und der Rate $\beta = 1/b$ parametrisiert werden. Die Dichtefunktion lautet dann:

$$f(x\,|\,a,\beta) = \frac{\beta^a x^{a-1} \mathrm{e}^{-\beta x}}{\Gamma(a)} \cdot \mathbb{1}_{[0,\infty)}(x), \quad a,\beta > 0$$

Diese Parametrisierung ist in der Bayes-Statistik gebräuchlich.

4.2.6 Die χ^2-Verteilung

☞ **Definition 4.2.21.** Dichtefunktion der χ^2-Verteilung
Die $\chi^2(n)$*-Verteilung mit* n *Freiheitsgraden ist die Gammaverteilung* $\mathrm{Ga}(n/2, 2)$*. Ihre Dichtefunktion lautet:*

$$f(x\,|\,n) = \frac{x^{n/2-1} \mathrm{e}^{-x/2}}{2^{n/2}\Gamma(n/2)} \cdot \mathbb{1}_{[0,\infty)}(x), \quad n \in \mathbb{N}$$

☛ **Satz 4.2.22.** Mittelwert und Varianz der χ^2-Verteilung
Es sei X $\chi^2(n)$*-verteilt mit* n *Freiheitsgraden. Dann gilt:*

$$\mathsf{E}[X] = n \quad \textit{und} \quad \mathsf{var}[X] = 2n$$

Abbildung 4.8 zeigt die Dichtefunktionen einiger χ^2-Verteilungen. Das α-Quantil der $\chi^2(n)$-Verteilung wird mit $\chi^2_{\alpha|n}$ bezeichnet. Eine Tabelle von $\chi^2_{\alpha|n}$ ist im Anhang T.4.

Die Bedeutung der χ^2-Verteilung beruht auf dem folgenden Satz (siehe Beispiele 4.3.9 und 4.3.23).

☛ **Satz 4.2.23.**
Ist X standardnormalverteilt, so ist $Y = X^2$ $\chi^2(1)$-verteilt mit einem Freiheitsgrad. Sind $X_1, \ldots, X_k$ unabhängig und standardnormalverteilt, so ist

$$Y = \sum_{i=1}^{k} {X_i}^2$$

$\chi^2(k)$-verteilt mit k Freiheitsgraden.

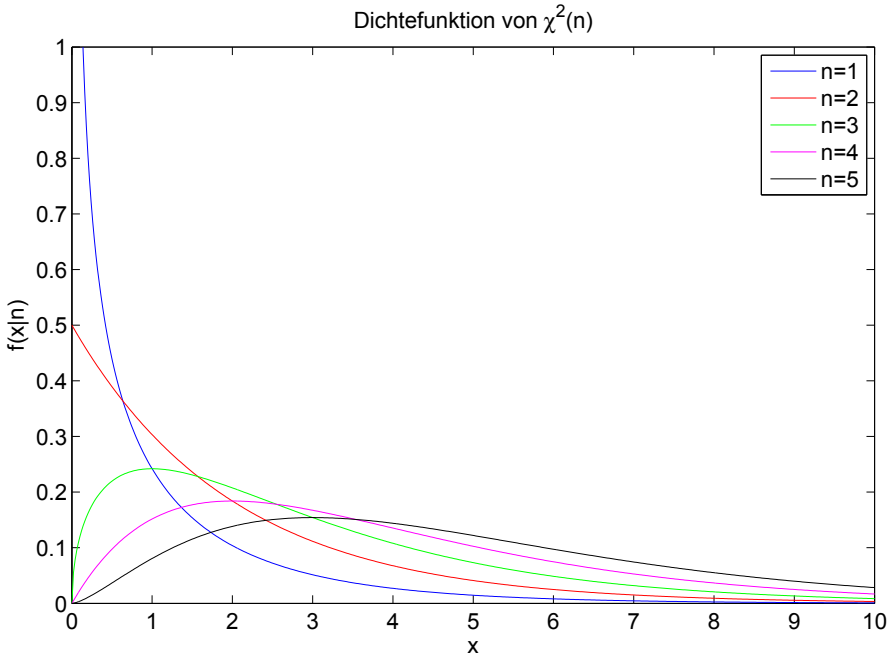

Abbildung 4.8: Dichtefunktionen einiger χ^2-Verteilungen.
MATLAB: `make_chi2pdf`

4.2.7 Die Betaverteilung

☞ **Definition 4.2.24.** DICHTEFUNKTION DER BETAVERTEILUNG
Die Dichtefunktion der Betaverteilung $\mathrm{Be}(a, b)$ *lautet:*

$$f(x\,|\,a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)} \cdot \mathbb{1}_{[0,1]}(x), \quad a, b > 0$$

wo $B(a, b)$ die Eulersche Betafunktion ist.

Die Eulersche Betafunktion* kann definiert werden durch:

$$B(a, b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$$

*http://de.wikipedia.org/wiki/Eulersche_Betafunktion

☛ **Satz 4.2.25.** Verteilungsfunktion der Betaverteilung

Die Verteilungsfunktion der Betaverteilung ist die folgende Stammfunktion der Dichtefunktion:

$$F(x\,|\,a,b) = \int_0^x \frac{t^{a-1}(1-t)^{b-1}}{B(a,b)}\,\mathrm{d}t$$

☛ **Satz 4.2.26.** Mittelwert und Varianz der Betaverteilung

Es sei $X \sim \mathrm{Be}(a,b)$. *Dann gilt:*

$$\mathsf{E}[X] = \frac{a}{a+b} \quad \textit{und} \quad \mathsf{var}[X] = \frac{ab}{(a+b)^2(a+b+1)}$$

Der Median ist außer in Spezialfällen ($a = b$, $a = 0$, $b = 0$) *nicht in geschlossener Form darstellbar.*

Das α-Quantil der $\mathrm{Be}(a,b)$-Verteilung wird mit $\beta_{\alpha|a,b}$ bezeichnet. Abbildung 4.9 zeigt die Dichtefunktionen einiger Betaverteilungen.

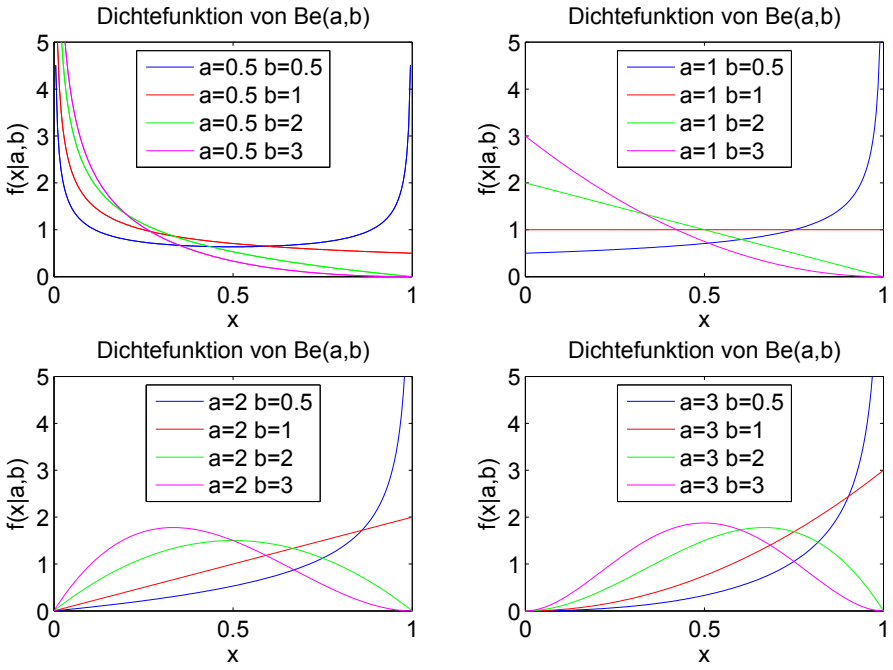

Abbildung 4.9: Dichtefunktionen einiger Betaverteilungen. MATLAB: `make_betapdf`

4.2.8 Die T-Verteilung

☞ **Definition 4.2.27.**

Die Dichtefunktion der $\mathrm{T}(n)$*-Verteilung mit* n *Freiheitsgraden lautet:*

$$f(x\,|\,n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \quad n \in \mathbb{N}$$

Die Verteilung wird auch als Student's T-Verteilung bezeichnet. „Student" war das Pseudonym des englischen Statistikers W. Gosset.*

☛ **Satz 4.2.28.** Mittelwert und Varianz der T-Verteilung

Es sei $X \sim \mathrm{T}(n)$. *Dann gilt:*

$$\mathsf{E}[X] = 0,\ n > 1 \quad \textit{und} \quad \mathsf{var}[X] = \frac{n}{n-2},\ n > 2$$

Der Median und der Modus sind gleich 0. Das k-te Moment der Verteilung existiert für $n > k$.

Das α-Quantil der T(n)-Verteilung wird mit $t_{a|n}$ bezeichnet. Eine Tabelle von $t_{a|n}$ ist im Anhang T.5. Wegen der Symmetrie der Verteilung gilt $t_{a|n} = -t_{1-a|n}$.

Die T(1)-Verteilung wird auch als Cauchy-Verteilung oder, in der Physik, als Breit-Wigner-Verteilung bezeichnet. Sie ist das bekannteste Beispiel einer Verteilung, von der kein einziges endliches Moment existiert, insbesondere weder Mittelwert noch Varianz. Abbildung 4.10 zeigt die Dichtefunktionen einiger T-Verteilungen. Für $n \to \infty$ strebt die T(n)-Verteilung gegen die Standardnormalverteilung.

*http://de.wikipedia.org/wiki/William_Sealy_Gosset

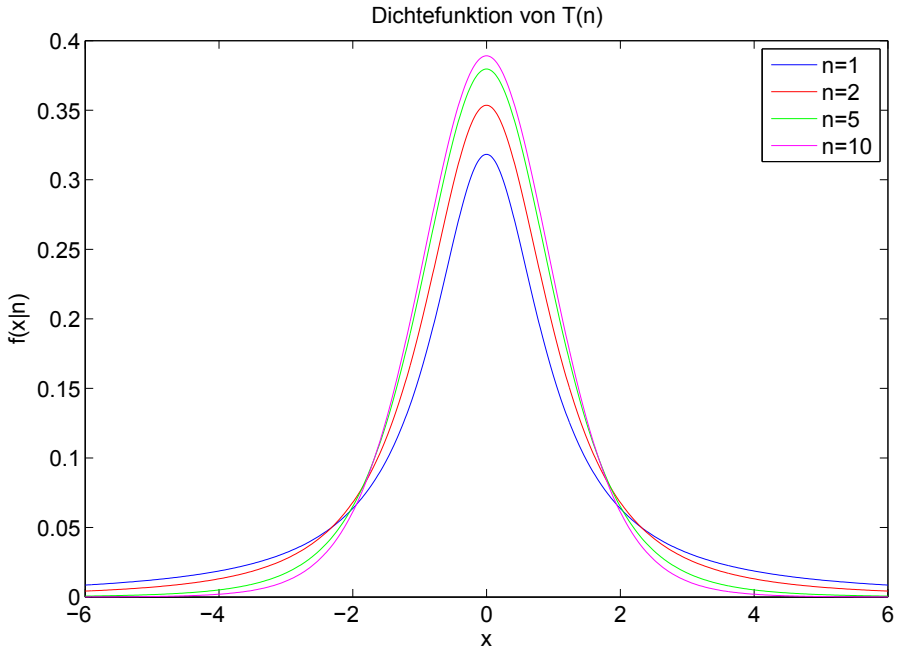

Abbildung 4.10: Dichtefunktionen einiger T-Verteilungen. MATLAB: `make_studpdf`

Aus der T-Verteilung kann für jeden Freiheitsgrad n eine Lagen-Skalen-Familie erzeugt werden. Ist X cauchyverteilt $(n = 1)$, so führt die affine Transformation $Y = sX + m$ auf eine Verteilung mit der Dichtefunktion:

$$f_Y(y\,|\,m, s) = \frac{1}{\pi s(1 + (y - m)^2/s^2)}$$

Hier ist m der Median, der Lageparameter der Verteilung, und s der Skalenparameter. s ist gerade die halbe Breite der Dichtefunktion bei halber maximaler Höhe:

$$f_Y(m \pm s\,|\,m, s) = \tfrac{1}{2} f_Y(m\,|\,m, s)$$

Die Verteilung wird mit $\mathrm{T}(1\,|\,m, s)$ bezeichnet.

4.3 Rechnen mit Verteilungen

4.3.1 Funktionen von Zufallsvariablen

Es sei X eine stetige Zufallsvariable und $h : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion. Dann ist $Y = h(X)$ wieder eine stetige Zufallsvariable. Der folgende Satz zeigt, wie die Verteilung bzw. die Dichtefunktion von Y berechnet wird, soferne h gewisse Voraussetzungen erfüllt.

☛ **Satz 4.3.1.** TRANSFORMATION VON STETIGEN ZUFALLSVARIABLEN

Es sei X eine stetige Zufallsvariable und $Y = h(X)$, wobei h eine umkehrbare und differenzierbare Funktion ist. Dann gilt:

$$f_Y(y)\,\mathrm{d}y = f_X(x)\,\mathrm{d}x \quad \textit{oder} \quad f_Y(y) = f_X(h^{-1}(y)) \cdot \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right|$$

▶ **Beispiel 4.3.2.** Affine Transformation

Es sei X eine stetige Zufallsvariable mit der Dichtefunktion $f_X(x)$ und $Y = aX + b$, mit $a \neq 0$. Dann ist die Dichtefunktion von Y gegeben durch:

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|}$$ ◀

▶ **Beispiel 4.3.3.**

Es sei $U \sim \mathrm{Un}(0,1)$, $F(x)$ eine stetige und umkehrbare Verteilungsfunktion und $X = F^{-1}(U)$. Dann gilt:

$$f_X(x) = 1 \cdot \left|\frac{\mathrm{d}F(x)}{\mathrm{d}x}\right| = F'(x)$$

Die Verteilungsfunktion von X ist also gleich F. Mit dieser Transformation können aus gleichverteilten Zufallszahlen Zufallszahlen mit vorgegebener Verteilung F erzeugt werden. ◀

☛ **Satz 4.3.4.** Transformation auf vorgegebene Verteilung

Ist $U \sim \mathrm{Un}(0,1)$ und $F(x)$ eine stetige und umkehrbare Verteilungsfunktion, dann hat $X = F^{-1}(U)$ die Verteilungsfunktion $F(x)$.

▶ **Beispiel 4.3.5.**

Es sei $U \sim \mathrm{Un}(0,1)$ und $F(x) = 1 - \exp(-x/\tau)$ die Verteilungsfunktion der Expo- nentialverteilung $\mathrm{Ex}(\tau)$. Da die Umkehrfunktion von F gleich $F^{-1}(x) = -\tau \ln(1-x)$ ist, ist $X = -\tau \ln(1-U)$ exponentialverteilt mit Mittel τ. Dann ist aber auch $X = -\tau \ln(U) \sim \mathrm{Ex}(\tau)$. ◀

▶ **Beispiel 4.3.6.**

Es sei $X \in [0, \pi]$ eine Zufallsvariable mit der Dichtefunktion

$$f_X(x) = \tfrac{1}{2}\sin x, \quad 0 \leq x \leq \pi$$

und $Y = \cos X$. Dann gilt:

$$f_Y(y) = \tfrac{1}{2}\sin(\arccos y)\frac{1}{\sqrt{1-y^2}} = \tfrac{1}{2}, \quad -1 \leq y \leq 1$$

Y ist also gleichverteilt im Intervall $[-1, 1]$. ◀

▶ **Beispiel 4.3.7.**

Es sei X eine exponentialverteilte Zufallsvariable mit der Dichtefunktion

$$f_X(x) = \frac{1}{\tau}\exp\left(-x/\tau\right)$$

und $Y = \sqrt{X}$. Dann ist mit $\lambda = 1/\tau$

$$f_Y(y) = 2\lambda y \exp\left(-\lambda y^2\right)$$

Das ist die Maxwell-Boltzmann-Verteilung* in 2 Dimensionen, auch Rayleigh-Verteilung genannt. ◀

*http://de.wikipedia.org/wiki/Maxwell-Boltzmann-Verteilung

Ist h nicht eindeutig umkehrbar, muss die Rechnung etwas modifiziert werden.

► **Beispiel 4.3.8.**
Es sei X eine stetige Zufallsvariable mit der Dichtefunktion $f_X(x)$ und $Y = X^2$. Ist f_X symmetrisch um 0, also $f_X(x) = f_X(-x)$, gilt:

$$f_Y(y) = 2f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{y}} \cdot f_X(\sqrt{y})$$

Der Faktor 2 entsteht dadurch, dass sowohl x als auch $-x$ auf $y = x^2$ abgebildet werden. ◄

► **Beispiel 4.3.9.**
Es sei X eine standardnormalverteilte Zufallsvariable mit der Dichtefunktion

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

und $Y = X^2$. Dann gilt:

$$f_Y(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left(-y/2\right) = \frac{y^{-\frac{1}{2}} \mathrm{e}^{-y/2}}{\sqrt{2}\,\Gamma(1/2)}$$

Das ist gerade die Dichtefunktion der $\chi^2(1)$-Verteilung. ◄

4.3.2 Fehlerfortpflanzung

Es soll nun untersucht werden, wie sich Mittelwert und Varianz einer Zufallsvariablen X unter nichtlinearen Transformationen verhalten. Es sei also $Y = h(X)$ und h eine glatte (mindestens zweimal stetig differenzierbare) Funktion. Approximiert man h an der Stelle $\mathsf{E}[X]$ durch die Taylorentwicklung 1. Ordnung, gilt:

$$h(x) \approx h(\mathsf{E}[X]) + h'(\mathsf{E}[X])(x - \mathsf{E}[X])$$

Unter Benützung der Eigenschaften von Erwartungswert und Varianz (Satz 3.1.8 und Satz 3.1.11) ergibt sich das folgende Resultat.

☛ **Satz 4.3.10.** Lineare Fehlerfortpflanzung

$$\mathsf{E}[Y] \approx h(\mathsf{E}[X]) \quad \textit{und} \quad \mathsf{var}[Y] \approx \left(\frac{\mathrm{d}h}{\mathrm{d}x}\right)^2 \cdot \mathsf{var}[X]$$

Ist der Mittelwert, wie es in den Anwendungen meist der Fall ist, nicht bekannt, wird die Ableitung am beobachteten Wert berechnet.

Lineare Fehlerfortpflanzung versagt, wenn die Funktion $h(x)$ in der Umgebung von $\mathsf{E}[X]$ nicht ausreichend gut durch ihre Taylorentwicklung 1. Ordnung approximiert werden kann. In diesem Fall kann der Näherungswert von $\mathsf{E}[Y]$ verbessert werden, indem h bis zur 2. Ordnung entwickelt wird:

$$h(x) \approx h(\mathsf{E}[X]) + h'(\mathsf{E}[X])(x - \mathsf{E}[X]) + \tfrac{1}{2}h''(\mathsf{E}[X])(x - \mathsf{E}[X])^2$$

Daraus folgt dann die verbesserte Näherung des Erwartungswerts.

☛ **Satz 4.3.11.**

$$\mathsf{E}[Y] \approx h(\mathsf{E}[X]) + \tfrac{1}{2}h''(\mathsf{E}[X]) \cdot \mathsf{var}[X]$$

► **Beispiel 4.3.12.**
In einem Teilchendetektor wird der inverse Impuls $q = 1/p$ eines Teilchens gemessen. Lineare Fehlerfortpflanzung ergibt dann:

$$\mathsf{E}[p] \approx 1/\mathsf{E}[q], \quad \mathsf{var}[p] \approx \mathsf{var}[q]/q^4$$

Die Näherung 2. Ordnung ergibt

$$\mathsf{E}[p] \approx 1/\mathsf{E}[q] + \mathsf{var}[q]/q^3$$

Es sei q normalverteilt ist mit $\mu = 2$ und $\sigma = 0.2$. Mittelwert und Standardabweichung von $p = 1/q$ können anhand einer sehr großen simulierten Stichprobe von q exakt bestimmt werden. Sie sind gleich 0.5051 bzw. 0.0521. Mit linearer Fehlerfortpflanzung erhält man:

$$\mathsf{E}[p] \approx 0.5, \quad \sigma[p] \approx 0.05$$

In 2. Ordnung gilt dagegen

$$\mathsf{E}[p] \approx 0.5050$$

Dies enspricht fast genau dem exakten Wert. ◄

► **Beispiel 4.3.13.**

Es sei X standardnormalverteilt und $Y = X^2$. Aus Beispiel 4.3.9 ist bekannt, dass Y $\chi^2(1)$-verteilt ist, and daher $\mathsf{E}[Y] = 1$ und $\mathsf{var}[Y] = 2$ gilt. Mit linearer Fehlerfortpflanzung ergibt sich jedoch $\mathsf{E}[Y] \approx 0$ und $\mathsf{var}[Y] \approx 0$. Das liegt natürlich daran, dass die Funktion $h(x) = x^2$ an der Stelle $x = \mathsf{E}[X] = 0$ eine Taylorentwicklung 1. Ordnung hat, die identisch gleich 0 ist. Die Taylorentwicklung 2. Ordnung, die in diesem Fall identisch mit der Funktion ist, ergibt zumindest für den Mittelwert von Y den korrekten Wert von 1. Das Beispiel zeigt, dass es in der Praxis unbedingt ratsam ist, sich über die Genauigkeit der linearen Näherung in der Umgebung der Beobachtung klarzuwerden. ◄

4.3.3 Faltung und Messfehler

In vielen Fällen müssen Zufallsvariable addiert werden, zum Beispiel um einen Messfehler zu berücksichtigen. Sind die Summanden unabhängig, spricht man von *Faltung*. Für die exakte Definition der Unabhängigkeit siehe Definition 4.4.6. Mittelwert und Varianz, sofern sie existieren, werden bei der Faltung addiert (siehe Satz 4.4.24).

☞ **Definition 4.3.14.** Faltung

Es seien X_1 und X_2 zwei unabhängige Zufallsvariable. Die Summe $X = X_1 + X_2$ heißt die Faltung von X_1 und X_2.

☛ **Satz 4.3.15.** Berechnung der Faltung

Sind X_1 und X_2 zwei unabhängige Zufallsvariable mit der gemeinsamen Dichtefunktion $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$, so hat ihre Summe $X = X_1 + X_2$ die Dichtefunktion:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_1(x - x_2) \cdot f_2(x_2)\,\mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(x - x_1)\,\mathrm{d}x_1 \end{aligned}$$

Die Dichtefunktion f_X wird als *Faltungsprodukt* von f_1 und f_2 bezeichnet: $f_X = f_1 * f_2$.

► **Beispiel 4.3.16.** Faltung von zwei Exponentialverteilungen

Es seien X_1 und X_2 exponentialverteilt gemäß $\mathrm{Ex}(\tau)$. Die Summe $X = X_1 + X_2$ hat die folgende Dichtefunktion:

$$f_X(x) = \int_{-\infty}^{\infty} f_1(x - x_2) f_2(x_2)\,\mathrm{d}x_2 = \int_0^x \frac{1}{\tau^2} \mathrm{e}^{(x_2 - x)/\tau} \mathrm{e}^{-x_2/\tau}\,\mathrm{d}x_2 = \frac{1}{\tau^2}\, x\mathrm{e}^{-x/\tau}$$

Das ist die Dichtefunktion der Gammaverteilung $\mathrm{Ga}(2, \tau)$. ◄

Die Faltung von zwei Zufallsvariablen X_1 und X_2 kann auch mit Hilfe ihrer *charakteristischen Funktionen* berechnet werden.

☞ **Definition 4.3.17.** Charakteristische Funktion

Es sei X eine diskrete oder stetige Zufallsvariable. Die charakteristische Funktion von X ist definiert durch:

$$\phi_X(t) = \mathsf{E}[\,\exp(\mathrm{i}tX)\,],\ t \in \mathbb{R}$$

wo i *die imaginäre Einheit ist.*

Die charakteristische Funktion ist die *Fouriertransformierte* der Dichtefunktion. Sie ist genau dann reellwertig, wenn die Dichtefunktion symmetrisch um 0 ist.

☛ **Satz 4.3.18.** Transformation der charakteristischen Funktion

Es sei $\phi_X(t)$ die charakteristische Funktion von X und $Y = m + sX$. Dann ist die charakteristische Funktion von Y gegeben durch:

$$\phi_Y(t) = \mathrm{e}^{\mathrm{i}tm} \phi_X(st)$$

☛ **Satz 4.3.19.** Charakteristische Funktion der Faltung

Ist $X = X_1 + X_2$ die Faltung von X_1 und X_2, so gilt:

$$\phi_X(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t)$$

► **Beispiel 4.3.20.** Faltung von zwei Poissonverteilungen

Es seien $X_1 \sim \mathrm{Po}(\lambda_1)$ und $X_2 \sim \mathrm{Po}(\lambda_2)$. Die charakteristische Funktion von X_i lautet:

$$\phi_{X_i}(t) = \sum_{k=0}^{\infty} \frac{\mathrm{e}^{\mathrm{i}tk} \lambda_i^k \mathrm{e}^{-\lambda_i}}{k!} = \mathrm{e}^{-\lambda_i} \sum_{k=0}^{\infty} \frac{(\mathrm{e}^{\mathrm{i}t} \lambda_i)^k}{k!} = \exp(\mathrm{e}^{\mathrm{i}t} \lambda_i - \lambda_i) = \exp[\lambda_i(\mathrm{e}^{\mathrm{i}t} - 1)]$$

Die charakteristische Funktion von $X = X_1 + X_2$ ist daher gleich

$$\phi_X(t) = \exp[\lambda_1(\mathrm{e}^{\mathrm{i}t} - 1)] \exp[\lambda_2(\mathrm{e}^{\mathrm{i}t} - 1)] = \exp[(\lambda_1 + \lambda_2)(\mathrm{e}^{\mathrm{i}t} - 1)]$$

X ist poissonverteilt gemäß $\mathrm{Po}(\lambda)$ mit $\lambda = \lambda_1 + \lambda_2$. Der Parameter λ ist sowohl Mittelwert als auch Varianz und daher additiv unter Faltung. ◄

► **Beispiel 4.3.21.** Faltung von zwei Normalverteilungen

Es seien $X_1 \sim \mathrm{No}(\mu_1, \sigma_1^2)$ und $X_2 \sim \mathrm{No}(\mu_2, \sigma_2^2)$. Die charakteristische Funktion von X_i lautet:

$$\phi_{X_i}(t) = \exp\left(\mathrm{i}t\mu_i - t^2\sigma_i^2/2\right)$$

Die charakteristische Funktion von $X = X_1 + X_2$ ist daher gleich

$$\phi_X(t) = \exp\left[\mathrm{i}t(\mu_1 + \mu_2) - t^2(\sigma_1^2 + \sigma_2^2)/2\right]$$

X ist normalverteilt gemäß $\mathrm{No}(\mu, \sigma^2)$ mit $\mu = \mu_1 + \mu_2$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Die Lageparameter und die Quadrate der Skalenparameter sind additiv unter Faltung. ◄

► **Beispiel 4.3.22.** Faltung von zwei Gammaverteilungen

Es seien $X_1 \sim \mathrm{Ga}(a_1, b)$ und $X_2 \sim \mathrm{Ga}(a_2, b)$. Die charakteristische Funktion von X_i lautet:

$$\phi_{X_i}(t) = (1 - \mathrm{i}tb)^{-a_i}$$

Die charakteristische Funktion von $X = X_1 + X_2$ ist daher gleich

$$\phi_X(t) = (1 - \mathrm{i}tb)^{-(a_1+a_2)}$$

X ist gammaverteilt gemäß $\mathrm{Ga}(a_1+a_2, b)$. Die Formparameter sind bei gleichen Skalenparametern additiv unter Faltung. Sind die Skalenparameter verschieden, ist X nicht gammaverteilt. ◄

► **Beispiel 4.3.23.** Faltung von zwei χ^2-Verteilungen

Es seien $X_1 \sim \chi^2(n_1)$ und $X_2 \sim \chi^2(n_2)$. Die charakteristische Funktion von X_i lautet:

$$\phi_{X_i}(t) = (1 - 2\mathrm{i}t)^{-n_i/2}$$

Die charakteristische Funktion von $X = X_1 + X_2$ ist daher gleich

$$\phi_X(t) = (1 - 2\mathrm{i}t)^{-n_1/2}(1 - 2\mathrm{i}t)^{-n_2/2} = (1 - 2\mathrm{i}t)^{-(n_1+n_2)/2}$$

X ist χ^2-verteilt mit $n = n_1 + n_2$ Freiheitsgraden. Die Freiheitsgrade sind additiv unter Faltung. ◄

► **Beispiel 4.3.24.** Faltung von zwei Cauchy-Verteilungen

Es seien X_1 und X_2 Cauchyverteilt mit Lageparametern m_1, m_2 und Skalenparametern s_1, s_2. Die charakteristische Funktion von X_i lautet (siehe Beispiel 4.3.26 und Satz 4.3.18):

$$\phi_{X_i}(t) = \exp\left(\mathrm{i}tm_i - s_i\,|t|\right)$$

Die charakteristische Funktion von $X = X_1 + X_2$ ist daher gleich

$$\phi_X(t) = \exp\left[\mathrm{i}t(m_1 + m2) - (s_1 + s_2)\,|t|\right]$$

X ist Cauchyverteilt mit Lageparameter $m = m_1 + m_2$ und Skalenparameter $s = s_1 + s_2$. Sowohl die Lageparameter als auch die Skalenparameter sind additiv unter Faltung. ◄

Die charakteristische Funktion kann auch zur Berechnung der Momente der Verteilung verwendet werden, wie der folgende Satz zeigt.

☛ **Satz 4.3.25.**

Ist $\phi_X(t)$ die charakteristische Funktion der Zufallsvariablen X und $\phi_X^{(n)}(t)$ ihre n-te Ableitung, gilt:

$$\phi_X^{(n)}(0) = \mathrm{i}^n \cdot \mathsf{E}[X^n]$$

Die n-te Ableitung von $\phi_X(t)$ an der Stelle $t = 0$ existiert genau dann, wenn das n-te Moment von X existiert.

► **Beispiel 4.3.26.** Die charakteristische Funktion der Cauchy-Verteilung

Ist X Cauchyverteilt, so ist die charakteristische Funktion von X gegeben durch:

$$\phi_X(t) = \mathrm{e}^{-|t|}$$

Da $\phi_X(t)$ an der Stelle $t = 0$ nicht differenzierbar ist, existiert keine einziges Moment der Verteilung. ◄

4.3.4 Systematische Fehler

Kann der Messfehler durch eine Zufallsvariable mit Mittelwert 0 beschrieben werden, hat die Messung nur einen *statistischen Fehler*. Die Messung kann jedoch durch eine falsche Kalibration, z.B. einen *Skalenfehler* oder einen *Nullpunktfehler* des Messgeräts verfälscht sein. Solche Fehler werden *systematische Fehler* genannt. Systematische Fehler werden durch Vergrößerung der Stichprobe *nicht kleiner*, und auch das Gesetz der Fehlerfortpflanzung gilt nicht für systematische Fehler.

Die Korrektur von systematischen Fehlern erfordert sorgfältige Kalibration der Messaparatur, Überprüfung von theoretischen Annahmen, etc. In der Experimentalphysik ist das Studium der Quellen von systematischen Fehlern eine der wichtigsten Aufgaben der Experimentatoren. Das folgende einfache Beispiel zeigt, dass die Behandlung der systematischen Fehler anderen Regeln gehorcht als die Behandlung der statistischen Fehler.

▶ **Beispiel 4.3.27.**

Zwei Spannungen X_1, X_2 werden mit dem gleichen Messgerät gemessen. Durch fehlerhafte Kalibration hat das Gerät neben dem statistischen Fehler auch einen Skalenfehler und einen Nullpunktfehler. Es misst also statt der wahren Spannung U die Spannung $X_\mathrm{m} = aU + b + \varepsilon$, mit $a = 0.99, b = 0.05$ und $\sigma[\varepsilon] = 0.03\,\mathrm{V}$. Der Mittelwert $\overline{X}$ der beiden Spannungen hat dann einen statistischen Fehler von $0.03/\sqrt{2} \approx 0.02\,\mathrm{V}$. Der systematische Fehler des Mittelwerts wird beschrieben durch $\overline{X}_\mathrm{m} = a\overline{X} + b$, ist also gleich dem der Einzelmessung. Der systematische Fehler wird also durch Mittelung nicht kleiner. Der systematische Fehler der Differenz ΔX wird hingegen beschrieben durch $\Delta X_\mathrm{m} = a\Delta X$. Der Nullpunktfehler ist also verschwunden, der Skalenfehler bleibt aber. Wird das Verhältnis der beiden gemessenen Spannungen berechnet, und ist der Nullpunktfehler klein im Vergleich zu den gemessenen Werten, was wohl anzunehmen ist, gilt näherungsweise:

$$\frac{aU_1 + b}{aU_2 + b} \approx \frac{U_1}{U_2} + \frac{b}{a} \cdot \frac{U_2 - U_1}{U_2^2}$$

Der Skalenfehler ist in dieser Näherung verschwunden, und der Nullpunktfehler des Quotienten ist deutlich kleiner geworden. ◀

4.3.5 Grenzverteilungssätze

Die Verteilung der Summe von unabhängigen Zufallsvariablen strebt unter recht allgemeinen Bedingungen gegen eine Normalverteilung. Speziell gelten die folgenden Versionen des *zentralen Grenzwertsatzes*.

☛ **Satz 4.3.28.** Zentraler Grenzwertsatz für identisch verteilte Folgen von Zufallsvariablen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen, die alle die gleiche Verteilung besitzen, mit endlichem Mittelwert μ und endlicher Varianz σ^2. Definiert man S_n und U_n durch:

$$S_n = \sum_{i=1}^{n} X_i,\ U_n = \frac{S_n - \mathsf{E}[S_n]}{\sigma[S_n]} = \frac{S_n - n\mu}{\sqrt{n}\cdot\sigma}$$

so konvergiert die Verteilungsfunktion von U_n für $n \to \infty$ gegen die Verteilungsfunktion der Standardnormalverteilung.

☛ **Satz 4.3.29.** Zentraler Grenzwertsatz für beliebig verteilte Folgen von Zufallsvariablen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit beliebigen Verteilungen, die jedoch für alle $i \in \mathbb{N}$ endlichen Mittelwert $\mu_i = \mathsf{E}[X_i]$ und endliche Varianz $\sigma_i^2 = \mathsf{var}[X_i]$ besitzen. Definiert man U_i und Y_n durch:

$$U_i = \frac{X_i - \mu_i}{\sqrt{n}\sigma_i},\ Y_n = \sum_{i=1}^{n} U_i$$

so konvergiert die Verteilungsfunktion von Y_n für $n \to \infty$ gegen die Verteilungsfunktion der Standardnormalverteilung.

Der zentrale Grenzwertsatz erklärt, warum die Normalverteilung in der Natur eine so bedeutende Rolle spielt, etwa bei der Verteilung der Impulskomponenten von Gasmolekülen, die das Ergebnis von zahlreichen Stößen ist, oder bei der Verteilung eines Messfehlers, der sich aus vielen unabhängigen Komponenten zusammensetzt.

Auch bei relativ kleinem n ist die Normalverteilung oft eine gute Näherung für die Summe von Zufallsvariablen.

▶ **Beispiel 4.3.30.** Binomialverteilung für grosses n

Da eine gemäß $\mathrm{Bi}(n,p)$ verteilte Zufallsvariable als Summe von n alternativverteilten Zufallsvariablen dargestellt werden kann, muss die Binomialverteilung für $n \to \infty$ und festes p gegen eine Normalverteilung streben. Abbildung 4.11 zeigt die Verteilungsfunktion der Binomialverteilung $\mathrm{Bi}(n,p)$ mit $n = 200$ und $p = 0.1$, sowie die Verteilungsfunktion der Normalverteilung $\mathrm{No}(\mu,\sigma^2)$ mit $\mu = np = 20$ und $\sigma^2 = np(1-p) = 18$.

➠MATLAB: `make_binolimit` ◀

▶ **Beispiel 4.3.31.** Poissonverteilung für grosses λ

Da eine gemäß $\mathrm{Po}(\lambda)$ verteilte Zufallsvariable für $\lambda \in \mathbb{N}$ als Summe von λ $\mathrm{Po}(1)$-verteilten Zufallsvariablen dargestellt werden kann, muss die Poissonverteilung für $\lambda \to \infty$ gegen eine Normalverteilung streben. Abbildung 4.12 zeigt die Verteilungsfunktion der Poissonverteilung $\mathrm{Po}(\lambda)$ mit $\lambda = 25$, sowie die Verteilungsfunktion der Normalverteilung $\mathrm{No}(\mu,\sigma^2)$ mit $\mu = \lambda = 25$ und $\sigma^2 = \lambda = 25$.

➠MATLAB: `make_poisslimit` ◀

Die Näherung von Binomial- und Poissonverteilung durch eine entsprechende Normalverteilung hat gewisse praktische Vorteile; zum Beispiel wird die Berechnung der Quantile der Verteilung wesentlich einfacher (siehe Abschnitte 7.2.3 und 7.2.4).

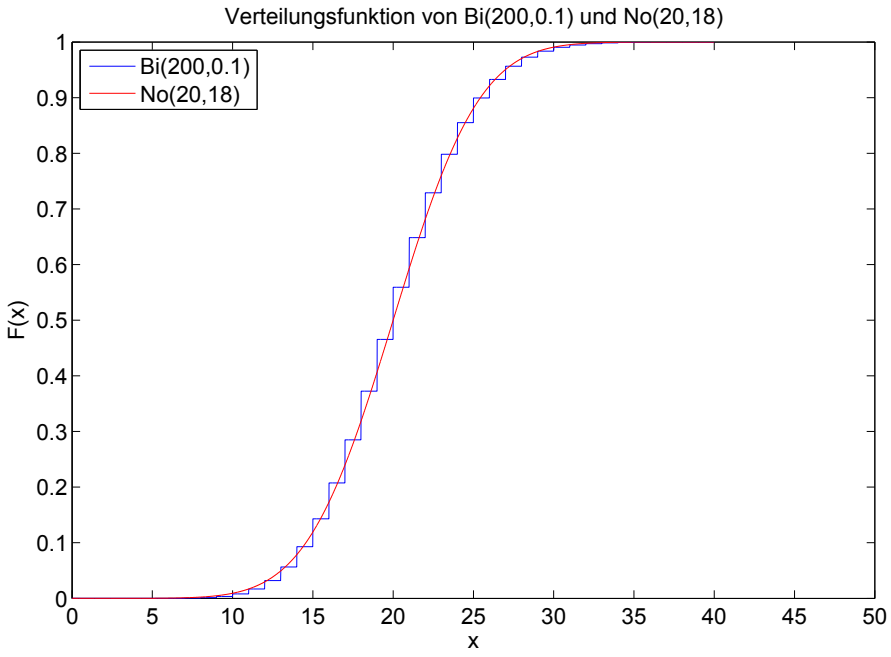

Abbildung 4.11: Näherung von $\mathrm{Bi}(200, 0.1)$ durch $\mathrm{No}(20, 18)$.
MATLAB: `make_binolimit`

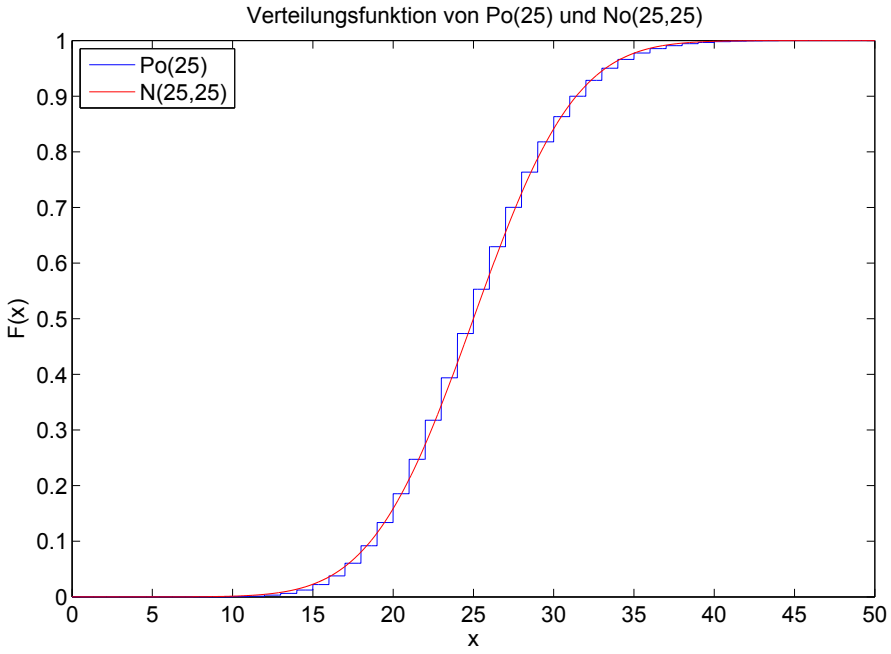

Abbildung 4.12: Näherung von Po(25) durch No(25, 25).
MATLAB: `make_poisslimit`

4.4 Stetige Zufallsvektoren

4.4.1 Grundbegriffe

☞ **Definition 4.4.1.** Stetiger Zufallsvektor

Eine multivariate Zufallsvariable, deren n Komponenten stetige Zufallsvariable sind, heißt ein n-dimensionaler stetiger Zufallsvektor.

► **Beispiel 4.4.2.**

Die gleichzeitige Messung von Impuls und Geschwindigkeit eines geladenen Teilchens wird durch einen 2-dimensionalen stetigen Zufallsvektor $\boldsymbol{X} = (X_1, X_2)$ beschrieben. ◄

☞ **Definition 4.4.3.** Dichtefunktion eines stetigen Zufallsvektors

Jede stetige nichtnegative Funktion $f_{\boldsymbol{X}}(x_1, \ldots, x_n)$, die auf 1 normiert ist, ist die Wahrscheinlichkeitsdichtefunktion oder kurz Dichtefunktion eines Zufallsvektors $\boldsymbol{X}$. Die Wahrscheinlichkeit eines Ereignisses $\mathrm{A} \subseteq \mathbb{R}^n$ unter der Verteilung von $\boldsymbol{X}$ ist dann gegeben durch das Integral:

$$W_{\boldsymbol{X}}(\mathrm{A}) = \int_{\mathrm{A}} f_{\boldsymbol{X}}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

☞ **Definition 4.4.4.** Randverteilung

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein stetiger Zufallsvektor. Die Verteilung einer Komponente X_i heißt die Randverteilung von $\boldsymbol{X}$ bzgl. X_i. Die Dichtefunktion von X_i wird als Randdichte bezeichnet.

☛ **Satz 4.4.5.** Berechnung der Randverteilung

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein stetiger Zufallsvektor mit der Dichtefunktion $f_{\boldsymbol{X}}(x_1, \ldots, x_n)$. Die Randdichte von X_i wird berechnet durch Integration über alle anderen Komponenten:

$$f_i(x_i) = \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} f_{\boldsymbol{X}}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \ldots \mathrm{d}x_n$$

☞ **Definition 4.4.6.** Unabhängigkeit

Es sei $\boldsymbol{X} = (X_1, X_2)$ ein stetiger Zufallsvektor mit der Dichtefunktion $f_{\boldsymbol{X}}(x_1, x_2)$. X_1 und X_2 heißen unabhängig, wenn gilt:

$$f_{\boldsymbol{X}}(x_1, x_2) = f_1(x_1) \cdot f_2(x_2).$$

Die Komponenten eines Zufallsvektors sind also unabhängig, wenn ihre gemeinsame Dichtefunktion das Produkt der Randdichten ist.

☛ **Satz 4.4.7.** Erwartungswert eines Produkts

Sind X_1 und X_2 unabhängig, gilt wieder:

$$\mathsf{E}[X_1 \cdot X_2] = \mathsf{E}[X_1] \cdot \mathsf{E}[X_2].$$

Die Definition der Kovarianz, der Korrelation und der Kovarianzmatrix können unverändert von diskreten auf stetige Zufallsvektoren übertragen werden (siehe Abschnitt 3.4).

☞ **Definition 4.4.8.** Bedingte Dichte

Es sei $\boldsymbol{X} = (X_1, X_2)$ ein stetiger Zufallsvektor mit der Dichtefunktion $f_{\boldsymbol{X}}(x_1, x_2)$ und den Randdichten $f_1(x_1)$ bzw. $f_2(x_2)$. Die Funktion $f_{12}(x_1 \,|\, x_2)$, definiert durch:

$$f_{12}(x_1 \,|\, x_2) = \frac{f_{\boldsymbol{X}}(x_1, x_2)}{f_2(x_2)}.$$

heißt die durch X_2 bedingte Dichte von X_1.

Die bedingte Dichte ist für festes x_2 die Dichtefunktion der durch $X_2 = x_2$ bedingten Verteilung von X_1. Aus der Kombination der Sätze 4.4.5 und 4.4.8 erhält man die stetige Version des Satzes von der totalen Wahrscheinlichkeit (Satz 2.3.8).

☛ **Satz 4.4.9.**

$$f_1(x_1) = \int_{\mathbb{R}} f_{12}(x_1 \,|\, x_2) f_2(x_2) \,\mathrm{d}x_2$$

▶ **Beispiel 4.4.10.** Der Messfehler

Es soll eine Größe X, etwas ein Streuwinkel oder eine Lebensdauer, gemessen werden. Durch den unvermeidlichen Messfehler wird jedoch tatsächlich Y statt X beobachtet. Der Messfehler $Y - X$ wird durch eine bedingte Dichte $g(y \,|\, x)$ beschrieben, die die Wahrscheinlichkeit angibt, dass der Wert y von Y beobachtet wird, wenn X den Wert x annimmt. Für die beobachtete Verteilung von Y gilt dann:

$$f_Y(y) = \int_{-\infty}^{\infty} g(y \,|\, x) f_X(x) \,\mathrm{d}x = \int_{-\infty}^{\infty} f(y, x) \,\mathrm{d}x$$

Im einfachsten Fall hängt g nur von der Differenz $y - x$ ab, oder eine weitere explizite Abhängigkeit von x ist vernachlässigbar. Dann wird aus dem Integral ein Faltungsintegral:

$$f_Y(y) = \int_{-\infty}^{\infty} g(y - x) f_X(x) \,\mathrm{d}x$$

Dies ist genau dann der Fall, wenn der Messfehler $Y - X$ und der zu messende Wert X unabhängig sind. ◀

Aus Satz 4.4.9 folgt wiederum der Satz von Bayes in der Version für stetige Dichtefunktionen.

☛ **Satz 4.4.11.** Satz von Bayes für stetige Dichtefunktionen

$$f_{21}(x_2 \,|\, x_1) = \frac{f_{12}(x_1 \,|\, x_2) f_2(x_2)}{\displaystyle\int_{\mathbb{R}} f_{12}(x_1 \,|\, x_2) f_2(x_2) \,\mathrm{d}x_2}$$

▶ **Beispiel 4.4.12.** Akzeptanzkorrektur

Es sei X eine Zufallsvariable mit der Dichtefunktion $f(x)$, etwa der bei einem Streuexperiment gemessene Streuwinkel. Es kann nun sein, dass der Detektor nicht überall die gleiche Ansprechwahrscheinlichkeit hat. Die Wahrscheinlichkeit, dass die Realisierung x von X auch tatsächlich beobachtet wird, sei gleich $A(x)$. $A(x)$ wird auch *Nachweiswahrscheinlichkeit* oder *Akzeptanzfunktion* genannt. Die Bestimmung der Akzeptanzfunktion ist ein wichtiger Schritt in der Kalibration einer Messaparatur.

Es sei nun J eine Zufallsvariable, die den Wert 1 annimmt, wenn x beobachtet wird, und 0 sonst. J ist dann unter der Bedingung $X = x$ alternativverteilt:

$$W(J=1\,|\,X=x)=A(x),\quad W(J=0\,|\,X=x)=1-A(x)$$

Die tatsächlich beobachtete Verteilung von X kann nun mit dem Satz von Bayes bestimmt werden:

$$f(x\,|\,J=1)=\frac{A(x)f(x)}{\displaystyle\int_{\mathbb{R}}A(x)f(x)\,\mathrm{d}x}$$

► **Beispiel 4.4.13.** Einschränkung des Messbereichs

Eine Größe X, etwa ein Streuwinkel, wird gemessen, wobei der Messbereich auf das Intervall $[a,b]$ eingeschränkt ist. Die Dichtefunktion der Messgröße sei $f(x)$. Die Akzeptanz $A(x)$ ist dann die Indikatorfunktion $\mathbb{1}_{[a,b]}(x)$, und die Dichtefunktion der Beobachtung ist gleich:

$$f(x\,|\,x\in[a,b])=\frac{f(x)\cdot\mathbb{1}_{[a,b]}(x)}{\displaystyle\int_a^b f(x)\,\mathrm{d}x}$$

◄

► **Beispiel 4.4.14.** Lebensdauermessung

Es wird die mittlere Lebensdauer von Myonen gemessen. Die Lebensdauer ist exponentialverteilt mit Mittelwert $\tau=2.197\cdot10^{-6}\,\mathrm{s}$. Um den Untergrund bei kleinen und großen gemessenen Zeiten zu unterdrücken, ist es ratsam, die Zeitmessung auf ein Intervall $[a,b]$ mit $a>0$ einzuschränken. Die Dichtefunktion der gemessen Zeiten ist dann gleich

$$f(x\,|\,x\in[a,b])=\frac{(1/\tau)\exp(-x/\tau)\cdot\mathbb{1}_{[a,b]}(x)}{\exp(-a/\tau)-\exp(-b/\tau)}$$

◄

4.4.2 Die multivariate Normalverteilung

☞ **Definition 4.4.15.** Dichtefunktion der multivariaten Normalverteilung

Die Dichtefunktion der multivariaten Normalverteilung $\mathrm{No}(\boldsymbol{\mu}, \mathbf{V})$ *lautet:*

$$f(\boldsymbol{x}\,|\,\boldsymbol{\mu}, \mathbf{V}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

$\mathbf{V}$ *und* $\mathbf{V}^{-1}$ *sind symmetrische positiv definite* $d \times d$*-Matrizen. Ist* $\boldsymbol{\mu} = \mathbf{0}$ *und* $\mathbf{V} = \mathbf{I}$*, liegt die multivariate Standardnormalverteilung der Dimension* d *vor.*

☛ **Satz 4.4.16.** Eigenschaften der multivariaten Normalverteilung

Die multivariate Normalverteilung $\mathrm{No}(\boldsymbol{\mu}, \mathbf{V})$ *hat folgende Eigenschaften:*

1. *Der Mittelwert der Verteilung ist* $\boldsymbol{\mu}$*, die Kovarianzmatrix ist* $\mathbf{V}$*.*
2. *Ist* $\boldsymbol{X} \sim \mathrm{No}(\boldsymbol{\mu}, \mathbf{V})$ *und* $\mathbf{H}$ *eine* $m \times d$ *Matrix, so ist* $\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} \sim \mathrm{No}(\mathbf{H}\boldsymbol{\mu}, \mathbf{H}\mathbf{V}\mathbf{H}^{\mathrm{T}})$*.*
3. *Jede Randverteilung einer multivariaten Normalverteilung ist wieder eine Normalverteilung. Mittelwert und Kovarianzmatrix der Randverteilung von* m *Komponenten von* $\boldsymbol{X}$ *entstehen durch Streichen der Spalten und Zeilen der anderen* $n - m$ *Komponenten.*
4. *Jede bedingte Verteilung einer Normalverteilung ist wieder eine Normalverteilung.*
5. *Ist* $\mathbf{V}$ *eine Diagonalmatrix, so sind die Komponenten von* $\boldsymbol{X}$ *nicht nur unkorreliert, sondern sogar unabhängig.*

Ist $\boldsymbol{X} \sim \mathrm{No}(\boldsymbol{\mu}, \mathbf{V})$, so kann $\mathbf{V}$ als positiv definite symmetrische Matrix mittels einer orthogonalen Transformation auf Diagonalform gebracht werden:

$$\mathbf{U}\mathbf{V}\mathbf{U}^{\mathrm{T}} = \mathbf{D}^2$$

Alle Digonalelemente von $\mathbf{D}^2$ sind positiv. Die Zufallsvariable $\boldsymbol{Y} = \mathbf{D}^{-1}\mathbf{U}(\boldsymbol{X} - \boldsymbol{\mu})$ ist dann standardnormalverteilt. Die Drehung $\mathbf{U}$ heißt *Hauptachsentransformation*. Ist $\boldsymbol{Y}$ standardnormalverteilt, so ist $\sum_{i=1}^{d} {Y_i}^2$ χ^2_d-verteilt mit d Freiheitsgraden.

4.4.3 Die bivariate Normalverteilung

Für $d = 2$ und $\boldsymbol{\mu} = \mathbf{0}$ kann die Dichtefunktion folgendermaßen angeschrieben werden:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right]$$

$\rho = \sigma_{12}/(\sigma_1\sigma_2)$ ist der Korrelationskoeffizient. Sind X_1 und X_2 unkorreliert, also $\rho = 0$, folgt:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2}\right)\right] = f_1(x_1) \cdot f_2(x_2)$$

Das zeigt, dass zwei unkorrelierte normalverteilte Zufallsvariable mit gemeinsamer Normalverteilung unabhängig sind.

Abbildung 4.13 zeigt die Dichtefunktion und einige Höhenschichtlinien der bivariaten Normalverteilung mit $\sigma_1 = 1, \sigma_2 = 1, \rho = 0.6$.

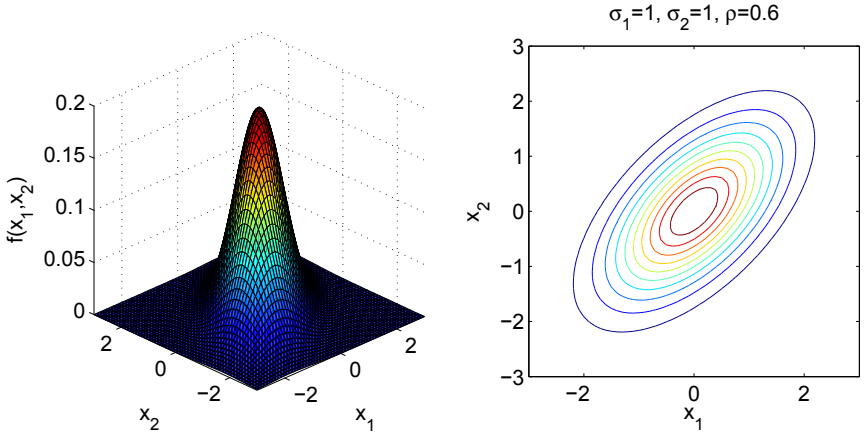

Abbildung 4.13: Dichtefunktion (links) und Höhenschichtlinien (rechts) der bivariaten Normalverteilung mit $\sigma_1 = 1, \sigma_2 = 1, \rho = 0.6$. MATLAB: `make_norm2`

☛ **Satz 4.4.17.**
Ist $\boldsymbol{X} = (X_1, X_2)$ bivariat standardnormalverteilt, so ist $Y = {X_1}^2 + {X_2}^2$ $\chi^2(2)$-verteilt mit 2 Freiheitsgraden, d.h. exponentialverteilt gemäß Ex(2).

☛ **Satz 4.4.18.** BEDINGTE DICHTEFUNKTION
Die bedingte Dichtefunktion $f_{21}(x_2 \,|\, x_1)$ ist gegeben durch

$$f_{21}(x_2 \,|\, x_1) = \frac{f(x_2, x_1)}{f(x_1)} = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\,\sigma_1^2\,(1-\rho^2)}\left(x_2 - \frac{\rho\, x_1\, \sigma_1}{\sigma_2}\right)^2\right]$$

$X_2 \,|\, X_1 = x_1$ ist also eine normalverteilte Zufallsvariable mit dem Erwartungswert:

$$\mathsf{E}[X_2 \,|\, X_1] = \rho\, x_1 \sigma_1/\sigma_2$$

☞ **Definition 4.4.19.** BEDINGTE ERWARTUNG
$\mathsf{E}[X_2 \,|\, X_1]$ *heißt die durch X_1 bedingte Erwartung von X_2.*

Je nach Vorzeichen von ρ fällt oder wächst die bedingte Erwartung von X_2, wenn X_1 wächst. Ist $\rho = 1$, sind X_1 und X_2 strikt proportional: $X_2 = X_1\, \sigma_2/\sigma_1$. Die Höhenschichtlinien der Dichtefunktion sind *Ellipsen*. Die *Hauptachsentransformation* ist jene Drehung, die die Ellipsen in achsenparallele Lage bringt. Sie hängt im Fall $d = 2$ nur von ρ ab. Ist $\rho = 0$, sind X_1 und X_2 bereits unabhängig, und der Drehwinkel ist gleich 0. Ist $\rho \neq 0$, ist die Drehmatrix U gleich

$$U = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \text{ mit } \varphi = -\frac{1}{2}\operatorname{arccot}\frac{\sigma_2^2 - \sigma_1^2}{2\rho\sigma_1\sigma_2}$$

4.4.4 Funktionen von Zufallsvektoren, Fehlerfortpflanzung

Es sei $\boldsymbol{X}$ ein Zufallsvektor der Dimension n und $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ eine (messbare) Funktion. Dann ist $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X})$ ein Zufallsvektor der Dimension m.

☛ **Satz 4.4.20.** TRANSFORMATION VON STETIGEN ZUFALLSVEKTOREN
Es sei $\boldsymbol{X}$ ein stetiger Zufallsvektor und $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X})$, wobei $\boldsymbol{h}$ eine umkehrbare und stetig differenzierbare Funktion ist. Dann gilt:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{h}^{-1}(\boldsymbol{y})) \cdot \left|\frac{\partial \boldsymbol{h}^{-1}(\boldsymbol{y})}{\partial \boldsymbol{y}}\right|$$

► **Beispiel 4.4.21.** Affine Transformation eines Zufallsvektors

Es sei $\boldsymbol{X}$ eine stetiger Zufallsvektor mit der Dichtefunktion $f_{\boldsymbol{X}}(\boldsymbol{x})$ und $\boldsymbol{Y} = \mathbf{A}\boldsymbol{X} + \boldsymbol{b}$, mit regulärer Matrix $\mathbf{A}$. Dann gilt:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\mathbf{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})) \cdot \frac{1}{|\det \mathbf{A}|}$$ ◄

► **Beispiel 4.4.22.** Transformation in Polarkoordinaten

In einem idealen Gas im thermischen Gleichgewicht sind aufgrund des zentralen Grenzwertsatzes die Komponenten des Geschwindigkeitsvektors $\boldsymbol{V} = (V_1, V_2, V_3)$ der Gasmoleküle in guter Näherung normalverteilt. Wegen der räumlichen Isotropie sind die Komponenten unabhängig, haben Mittelwert $\mu = 0$ und die gleiche Standardabweichung σ. Die Dichtefunktion von $\boldsymbol{V}$ hat daher die folgende Form:

$$f(\boldsymbol{v}) = \frac{1}{\sigma^3(2\pi)^{3/2}} \exp\left(-\frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}{2\sigma^2}\right)$$

Um die Verteilung des Betrags der Geschwindigkeit zu bestimmen, wird von kartesischen in Polarkoordinaten transformiert:

$$v = |\boldsymbol{v}|\,, \quad \varphi = \arctan(v_2/v_1), \quad \theta = \arccos(v_3/v)$$

Die Determinante der Jacobi-Matrix der Umkehrtransformation ist bekanntlich $v^2 \sin\theta$. Damit ergibt sich die gemeinsame Dichtefunktion des transformierten Zufallsvektors (V, Φ, Θ) zu:

$$g(v, \varphi, \theta) = \frac{1}{\sigma^3(2\pi)^{3/2}} v^2 \sin\theta \exp\left(-\frac{v^2}{2\sigma^2}\right) = \frac{1}{2\pi} \cdot \frac{1}{2} \sin\theta \cdot \sqrt{\frac{2}{\pi}} \frac{v^2}{\sigma^3} \exp\left(-\frac{v^2}{2\sigma^2}\right)$$

Da die gemeinsame Dichtefunktion als Produkt der drei Randdichten dargestellt werden kann, sind V, Φ, Θ unabhängig. Φ ist gleichverteilt im Intervall $[0, 2\pi]$, $\cos\Theta$ ist gleichverteilt im Intervall $[-1, 1]$ (siehe Beispiel 4.3.6), und V ist Maxwell-Boltzmann-verteilt. Die häufigste Geschwindigkeit, also der Modus der Verteilung von V, ist $\sqrt{2}\sigma$, die mittlere Geschwindigkeit ist $2\sigma\sqrt{2/\pi} \approx 1.6\,\sigma$. Es gilt übrigens $\sigma^2 = k_B T/m$, wo m die Molekülmasse, T die Temperatur und k_B die Boltzmannkonstante ist. ◀

Die lineare Fehlerfortpflanzung soll nun auf Zufallsvektoren verallgemeinert werden. Es sei also $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor, $\mathbf{H}$ eine $m \times n$-Matrix und $\boldsymbol{b} \in \mathbb{R}^m$. Dann ist $\boldsymbol{Y} = (Y_1, \ldots Y_m) = \mathbf{H}\boldsymbol{X} + \boldsymbol{b}$ ein Zufallsvektor der Dimension m. Es gilt das folgende exakte Resultat.

☛ **Satz 4.4.23.** Transformation von Mittelwert und Kovarianzmatrix
Es sei $\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} + \boldsymbol{b}$. Dann gilt:

$$\mathsf{E}[\boldsymbol{Y}] = \mathbf{H} \cdot \mathsf{E}[\boldsymbol{X}] + \boldsymbol{b} \quad \textit{und} \quad \mathsf{Cov}[\boldsymbol{Y}] = \mathbf{H} \cdot \mathsf{Cov}[\boldsymbol{X}] \cdot \mathbf{H}^{\mathrm{T}}$$

Für $\mathbf{H} = (a_1, \ldots, a_n)$ ergibt sich der folgende Spezialfall.

☛ **Satz 4.4.24.** Varianz einer Linearkombination
Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor und $\boldsymbol{Y} = \sum_{i=1}^n a_i X_i$. Dann gilt:

$$\mathsf{var}[\boldsymbol{Y}] = \sum_{i=1}^{n} a_i^2 \,\mathsf{var}[X_i] + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j \,\mathsf{cov}[X_i, X_j]$$

Sind alle X_i paarweise unkorreliert, gilt:

$$\mathsf{var}[\boldsymbol{Y}] = \sum_{i=1}^{n} a_i^2 \,\mathsf{var}[X_i]$$

Sind außerdem alle $a_i = 1$, folgt:

$$\mathsf{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathsf{var}[X_i]$$

Sind alle X_i paarweise unkorreliert, ist die Varianz der Summe gleich der Summe der Varianzen.

Es soll nun statt der linearen Abbildung $\mathbf{H}$ eine nichtlineare glatte (mindestens zweimal stetig differenzierbare Funktion) $\boldsymbol{h} = (h_1, \ldots, h_m)$ betrachtet werden. Es wird angenommen, dass $\boldsymbol{h}$ in jenem Bereich, in dem die Dichtefunktion von $\boldsymbol{X}$ signifikant von 0 verschieden ist, genügend gut durch die Taylorentwicklung 1. Ordnung angenähert werden kann. Als Entwicklungsstelle wählt man den Mittelwert $\mathsf{E}[\boldsymbol{X}]$.

☛ **Satz 4.4.25.** Lineare Fehlerfortpflanzung
Es sei $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X})$. Dann gilt in der Näherung 1. Ordnung:

$$\mathsf{E}[\boldsymbol{Y}] \approx \boldsymbol{h}(\mathsf{E}[\boldsymbol{X}]) \quad \textit{und} \quad \mathsf{Cov}[\boldsymbol{Y}] \approx \left(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right) \cdot \mathsf{Cov}[\boldsymbol{X}] \cdot \left(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}$$

Da der Mittelwert in der Regel nicht bekannt ist, werden die Ableitungen am beobachteten Wert berechnet. Der Näherungswert von $\mathsf{E}[\boldsymbol{Y}]$ kann noch durch eine Entwicklung bis zur 2. Ordnung verbessert werden, wie der folgende Satz zeigt (siehe auch Satz 4.3.11).

☛ **Satz 4.4.26.**

Es sei $\mathbf{C}$ *die Kovarianzmatrix von* $\boldsymbol{X}$ *und* $\mathbf{B}_i$ *die Hesse-Matrix (Matrix der zweiten partiellen Ableitungen) von* h_i*. Dann gilt in der Näherung 2. Ordnung:*

$$\mathsf{E}[Y_i] \approx h_i(\mathsf{E}[X_i]) + \tfrac{1}{2}\mathrm{spur}(\mathbf{C}\mathbf{B}_i)$$

▶ **Beispiel 4.4.27.** Lineare Fehlerfortpflanzung

An einem Gleichstrommotor wird eine Spannung $U = 120\,\mathrm{V}$ und ein Strom $I = 3\,\mathrm{A}$ gemessen. Die relativen Messfehler betragen 1% bzw. 2% und werden zunächst als unkorreliert angenommen. Es ist der relative Fehler der Leistung $P = U \cdot I$ zu bestimmen.

Es gilt:

$$\frac{\partial P}{\partial U} = I, \quad \frac{\partial P}{\partial I} = U$$

und damit:

$$\mathsf{var}[P] \approx I^2 \cdot \mathsf{var}[U] + U^2 \cdot \mathsf{var}[I] = 3^2 \cdot 1.2^2 + 120^2 \cdot 0.06^2 = 64.8$$

Folglich ist $\sigma[P] \approx 8.05$ und $\sigma[P]/\mathsf{E}[P] \approx 2.24\%$. Dies stimmt mit dem exakten Wert in allen Stellen überein.

Sind die Messungen positiv korreliert mit Korrelationskoeffizient $\rho = 0.3$, gilt:

$$\mathsf{cov}(U, I) = \rho \cdot \sigma[U] \cdot \sigma[I] = 0.0216$$

und

$$\begin{aligned}\mathsf{var}[P] &\approx I^2 \cdot \mathsf{var}[U] + U^2 \cdot \mathsf{var}[I] + 2 \cdot \mathsf{cov}(U, I) \cdot I \cdot U \\ &= 3^2 \cdot 1.2^2 + 120^2 \cdot 0.06^2 + 2 \cdot 0.0216 \cdot 3 \cdot 120 = 80.325\end{aligned}$$

Folglich ist $\sigma[P] \approx 8.964$ und $\sigma[P]/\mathsf{E}[P] \approx 2.49\%$. Der relative Fehler ist also größer geworden. Sind die Messungen negativ korreliert mit $\rho = -0.3$, ist der relative Fehler gleich $\sigma[P]/\mathsf{E}[P] \approx 1.95\%$, er ist also kleiner als im unkorrelierten Fall.

➠MATLAB: `make_ffp` ◀

▶ **Beispiel 4.4.28.** Lineare Fehlerfortpflanzung

Es seien X und Y unabhängig und normalverteilt gemäß $\mathrm{No}(4, 1)$. Es sind Mittelwert und Varianz von $Z = X/Y$ zu bestimmen. Mit linearer Fehlerfortpflanzung erhält man:

$$\frac{\partial z}{\partial x} = 1/y, \quad \frac{\partial z}{\partial y} = -x/y^2$$

und damit:

$$\mathsf{E}[Z] = 1, \quad \mathsf{var}[Z] = (1/4)^2 \cdot 1 + (4/16)^2 \cdot 1 = 0.125 \implies \sigma[Z] = 0.3536$$

Der tatsächliche Mittelwert einer simulierten Stichprobe von Z vom Umfang $N = 10^5$ ist jedoch 1.087, und die Standardabweichung ist 1.253. Da die Funktion $z = x/y$ bei $y = 0$ einen Pol hat, ist die lineare Näherung im Bereich der Verteilung von X und Y offenbar unzureichend. Wird die Standardabweichung halbiert, also $\sigma = 0.5$, ergibt die lineare Fehlerfortpflanzung:

$$\mathsf{E}[Z] = 1, \quad \mathsf{var}[Z] = (1/4)^2 \cdot 0.25 + (4/16)^2 \cdot 0.25 = 0.03125 \implies \sigma[Z] = 0.1768$$

Der tatsächliche Mittelwert einer simulierten Stichprobe von Z vom Umfang $N = 10^5$ ist jetzt 1.016, und die Standardabweichung ist 0.184. Die Übereinstimmung ist deutlich besser geworden. ◀

Teil II

Statistik

Kapitel 5

Deskriptive Statistik

5.1 Einleitung

Die deskriptive Statistik beschäftigt sich mit Aufbereitung und Präsentation von Datensätzen. Sie ist ein unentbehrliches Werkzeug, um große Datenmengen der Anschauung zugänglich zu machen. Dazu dienen sowohl die graphische Darstellung in verschiedenen Formen als auch die Zusammenfassung in wenige charakteristische Maßzahlen. Dieses Kapitel gibt eine kleine Auswahl aus der Vielzahl von gebräuchlichen Methoden.

Komplementär zur deskriptiven Statistik steht die *Inferenzstatistik*. Ihre Aufgabe ist es, die den Beobachtungen zugrunde liegenden Gesetzmäßigkeiten aufzudecken. Sie bedient sich dazu der Begriffe und Resultate der Wahrscheinlichkeitsrechnung. Der hauptsächliche Inhalt der Inferenzstatistik, soweit er hier behandelt wird, ist das Schätzen von unbekannten Größen, die Quantifizierung der Unsicherheit der Schätzung, sowie das Testen von Hypothesen über unbekannte Größen oder Verteilungen. Die Inferenzstatistik wird unter zwei verschiedenen Gesichtspunkten betrieben, die *frequentistisch* und *Bayesianisch* genannt werden. Die nachfolgenden Kapitel geben eine Einführung in die wichtigsten Begriffe und Methoden dieser beiden Ansätze.

5.2 Univariate Merkmale

5.2.1 Graphische Darstellung

Ein Bild sagt bekanntlich mehr als tausend Worte. Graphische Darstellungen von Datensätzen sind daher äußerst beliebt und nützlich. Je nach Merkmaltypus sind dabei verschiedene Methoden empfehlenswert. Für diskrete Beobachtungen bietet sich die Häufigkeitstabelle, das Tortendiagramm und das Stabdiagramm an; für stetige Beobachtungen kommen unter anderem die gruppierte Häufigkeitstabelle, das Histogramm, der Boxplot und die empirische Verteilungsfunktion in Frage. Die folgenden Abbildungen zeigen einige Beispiele dieser graphischen Möglichkeiten an drei künstlichen Datensätzen:

1. Datensatz 1 besteht aus 500 normalverteilten Werten (Abbildung 5.1).
2. Datensatz 2 enthält zusätzlich noch 100 Werte, die aus einer anderen Verteilung gezogen wurden und als „Kontamination“ betrachtet werden können (Abbildung 5.2).
3. Datensatz 3 enthält 50 Prüfungsnoten (Abbildung 5.3 und Tabelle 5.1).

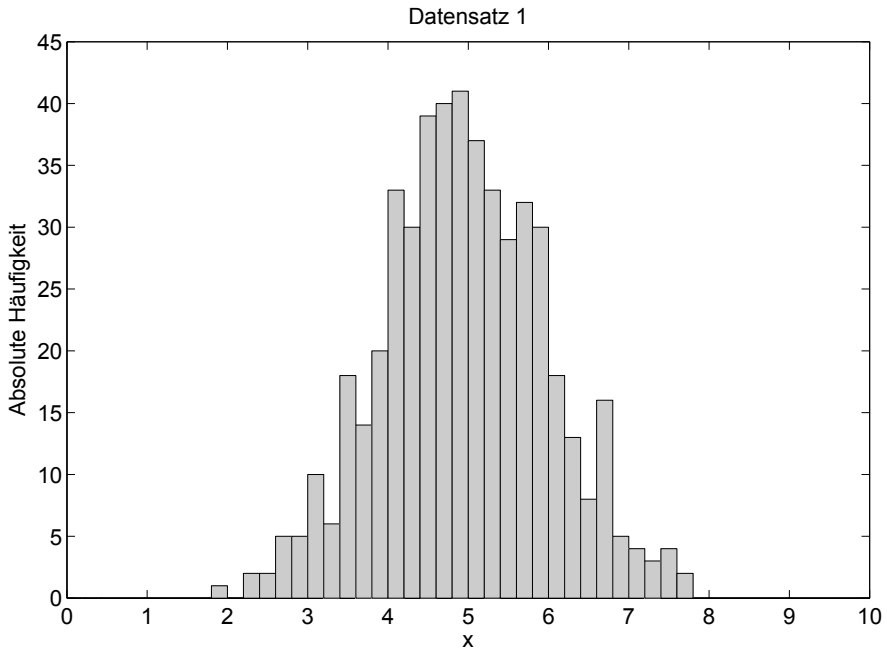

Abbildung 5.1: Histogramm von Datensatz 1. MATLAB: `make_dataset1`

5.2.2 Die Häufigkeitstabelle

Die Häufigkeitstabelle eines diskreten Merkmals gibt an, wie oft eine bestimmte Ausprägung k im Datensatz auftritt (siehe Abschnitt 1.3.3). Häufigkeiten können als *absolute* Häufigkeiten $h_n(k)$ oder als *relative* Häufigkeiten $f_n(k) = h_n(k)/n$ angegeben werden, wobei n die Größe des Datensatzes ist (siehe Tabelle 5.1).

Liegen Beobachtungen von diskreten Merkmalen mit vielen Ausprägungen oder von stetigen Merkmalen vor, ist es zielführender, die Beobachtungen in *Gruppen* einzuteilen und die *Gruppenhäufigkeiten* anzugeben. Die graphische Darstellung der absoluten Gruppenhäufigkeiten wird Histogramm genannt.

5.2.3 Empirische Verteilungsfunktion

Ab Ordinalskala ist es sinnvoll, die Daten zu ordnen. Die Häufigkeitstabelle kann dann durch absolute Summenhäufigkeiten $H_n(k)$ bzw. relative Summenhäufigkeiten $F_n(k)$

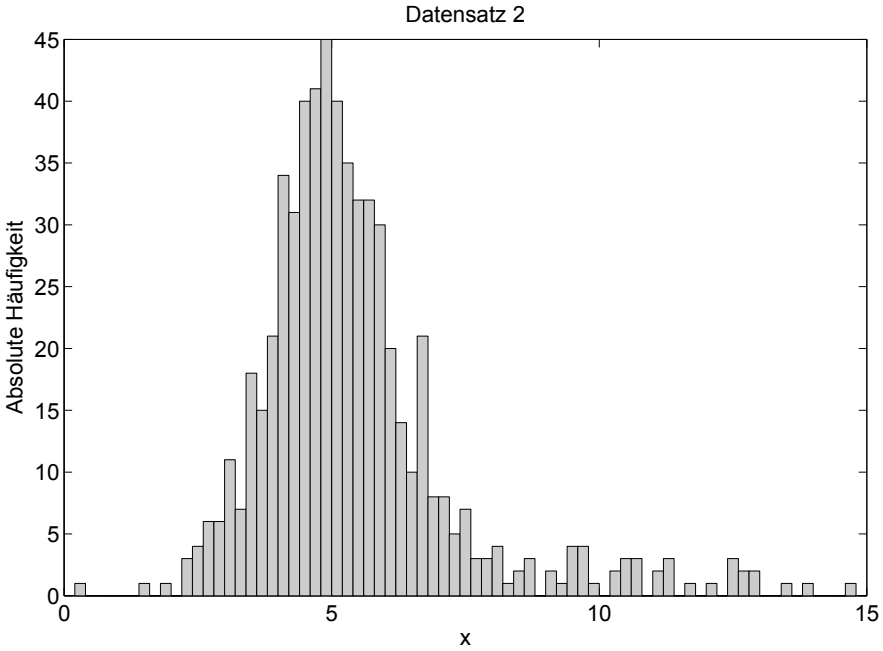

Abbildung 5.2: Histogramm von Datensatz 2. MATLAB: `make_dataset2`

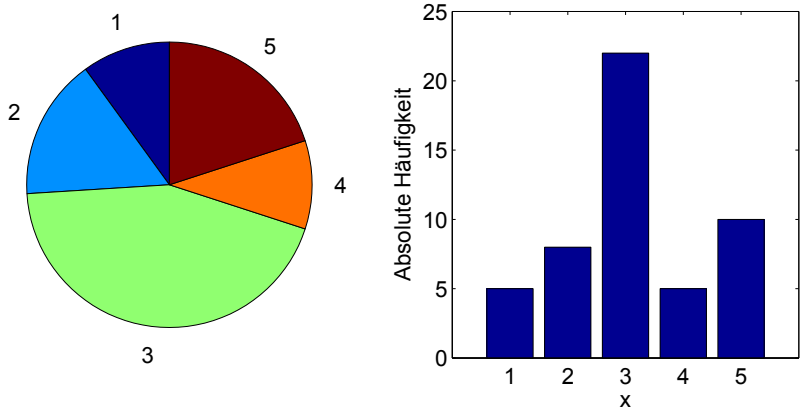

Abbildung 5.3: Tortendiagramm (links) und Stabdiagramm (rechts) von Datensatz 3. MATLAB: `make_dataset3`

Tabelle 5.1: Erweiterte Häufigkeitstabelle von Datensatz 3. $h_n(k)$: absolute Häufigkeit, $H_n(k)$: absolute Summenhäufigkeit, $f_n(k)$: relative Häufigkeit, $F_n(k)$: relative Summenhäufigkeit. MATLAB: `make_dataset3`

Note k	$h_n(k)$	$H_n(k)$	$f_n(k)$	$F_n(k)$
1	5	5	0.10	0.10
2	8	13	0.16	0.26
3	22	35	0.44	0.70
4	5	40	0.10	0.80
5	10	50	0.20	1.00

ergänzt werden (siehe Tabelle 5.1):

$$H_n(k) = \sum_{i=1}^{k} h_n(k), \quad F_n(k) = \sum_{i=1}^{k} f_n(k) = H_n(k)/n$$

Die graphische Darstellung der Summenhäufigkeiten wird die *empirische Verteilungsfunktion* der Datenliste genannt.

☞ **Definition 5.2.1.** EMPIRISCHE VERTEILUNGSFUNKTION

Der Wert der empirischen Verteilungsfunktion $F_n(x)$ der Datenliste $\boldsymbol{x} = (x_1, \ldots, x_n)$ an der Stelle x ist der Anteil der Daten, die kleiner oder gleich x sind:

$$F_n(x) = f_n(\boldsymbol{x} \leq x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(x_i)$$

Ist $x_i \leq x < x_{i+1}$, gilt:

$$F_n(x) = f_n(x_1) + \cdots + f_n(x_i)$$

F_n ist eine Sprungfunktion. Die Sprungstellen sind die Datenpunkte, die Sprunghöhen sind die relativen Häufigkeiten der Datenpunkte (Abbildungen 5.4 und 5.5).

Aus der empirischen Verteilungsfunktion können Quantile sowie Unter- und Überschreitungshäufigkeiten einfach abgelesen werden (Abbildungen 5.6 und 5.7).

5.2.4 Kernschätzer

Die Häufigkeitsverteilung in einem Histogramm kann mit einem *Kern-* oder *Dichteschätzer* geglättet werden. Dabei wird die empirische Dichtefunktion des beobachteten Merkmals durch eine Summe von Kernfunktionen $K(x)$ approximiert:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

h ist die *Bandbreite* des Kernschätzers. Der beliebteste Kern ist der Gaußkern:

$$K(x) = \frac{1}{\sqrt{2\,\pi}} \exp\left(-\frac{x^2}{2}\right)$$

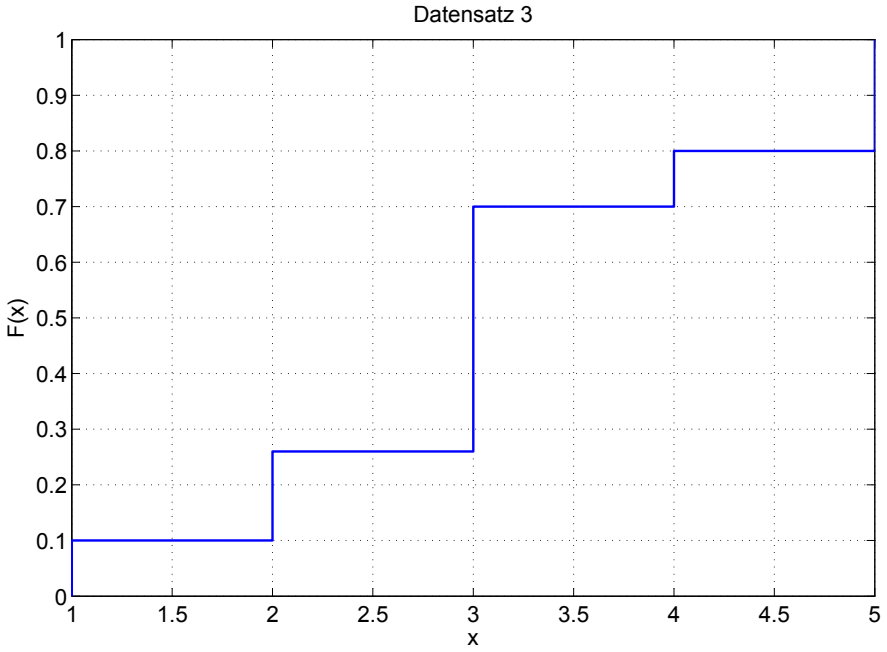

Abbildung 5.4: Empirische Verteilungsfunktion von Datensatz 3.
MATLAB: `make_dataset3`

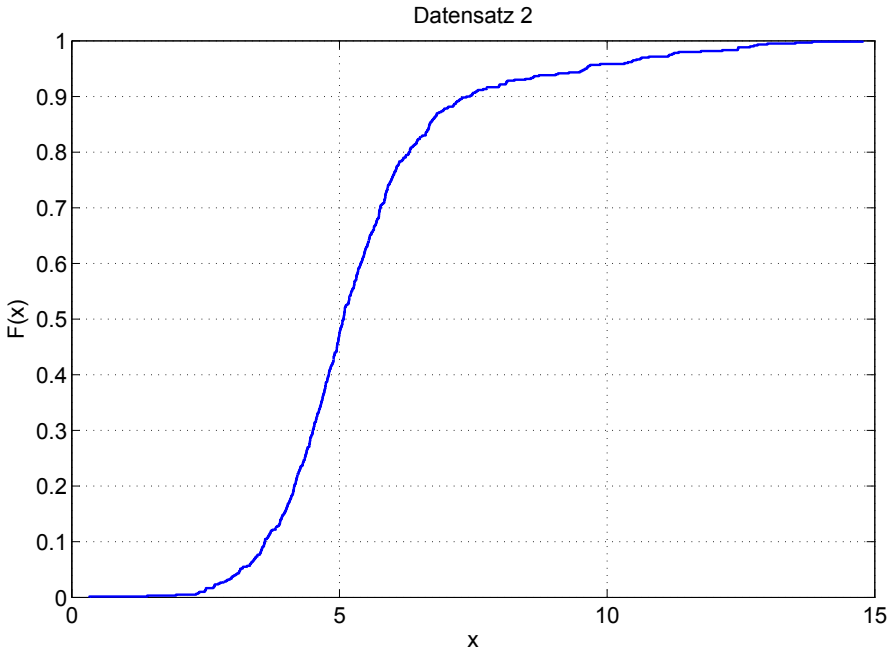

Abbildung 5.5: Empirische Verteilungsfunktion von Datensatz 2.
MATLAB: `make_dataset2`

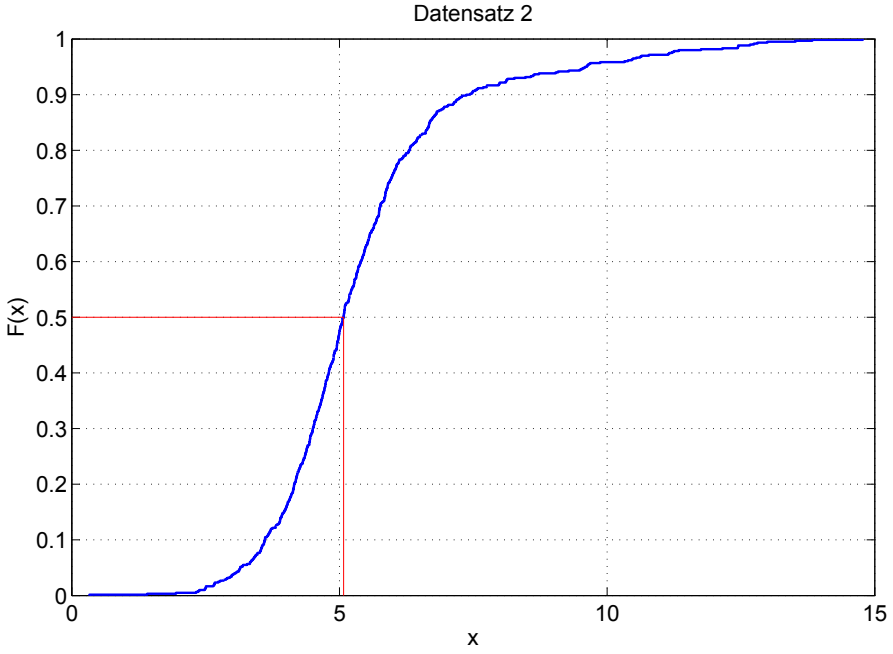

Abbildung 5.6: Empirische Verteilungsfunktion und Median von Datensatz 2.
MATLAB: `make_dataset2`

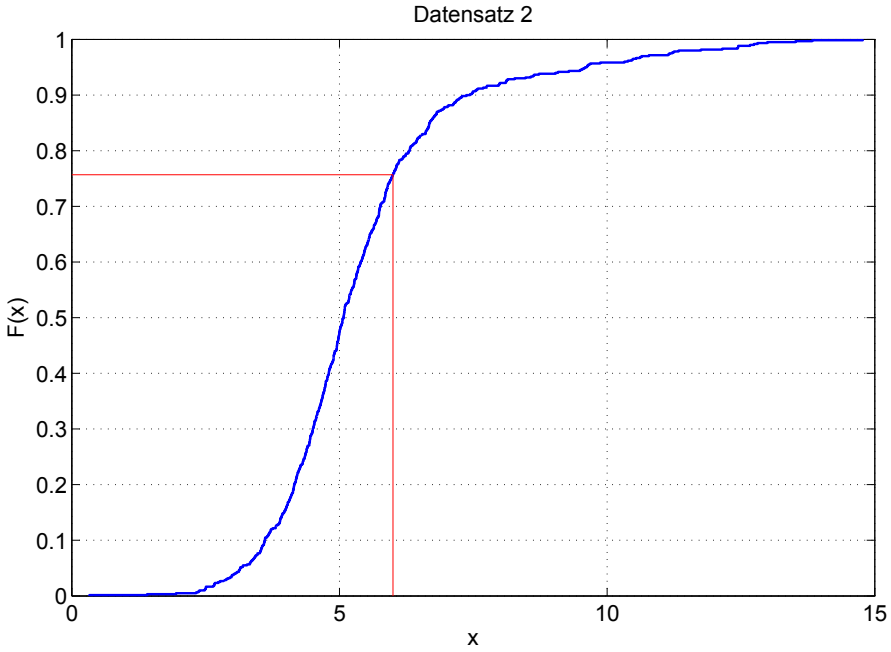

Abbildung 5.7: Anteil der Daten ≤ 6 in Datensatz 2. MATLAB: `make_dataset2`

Abbildung 5.8 zeigt die Glättung des Histogramms von Datensatz 2 durch einen Kernschätzer. Für die Wahl der Bandbreite gibt es eine Faustregel:

$$h \approx s \cdot n^{-1/5}$$

wo s die Standardabweichung des Datensatzes ist (siehe Definition 5.2.11). Die Faustregel kann auch zur Wahl der Gruppenbreite im Histogramm herangezogen werden.

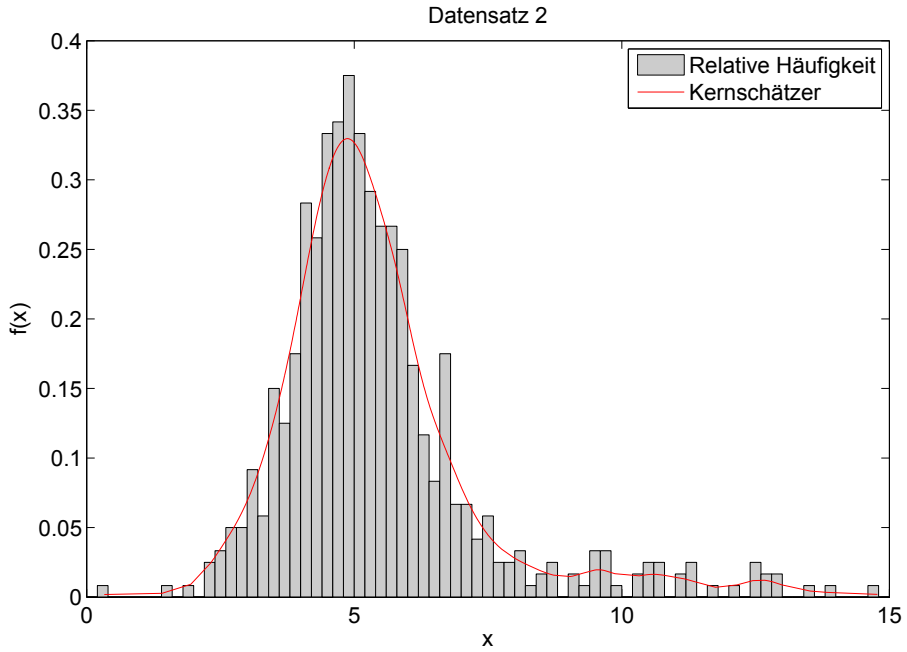

Abbildung 5.8: Glättung des Histogramms von Datensatz 2 durch einen Kernschätzer. MATLAB: `make_dataset2`

5.2.5 Das Q-Q-Diagramm

Das Q-Q-Diagramm oder Quantil-Quantil-Diagramm (engl. Q-Q plot) ist eine graphisches Werkzeug, um die Verteilung eines Datensatzes mit einer theoretischen Verteilung zu vergleichen. Dazu werden die Quantile der Daten gegen die Quantile der Verteilung aufgetragen. Stimmen die beiden Verteilungen überein, liegen die Punkte auf einer Geraden. Das Q-Q-Diagramm kann auch verwendet werden, um die Verteilungen von zwei Datensätzen zu vergleichen.

Abbildungen 5.9 und 5.10 zeigen die Q-Q-Diagramme von Datensatz 1 und 2 unter Annahme einer Normalverteilung. Die Abweichung von der Normalverteilung in Datensatz 2 ist klar zusehen.

➠MATLAB: `qqplot`

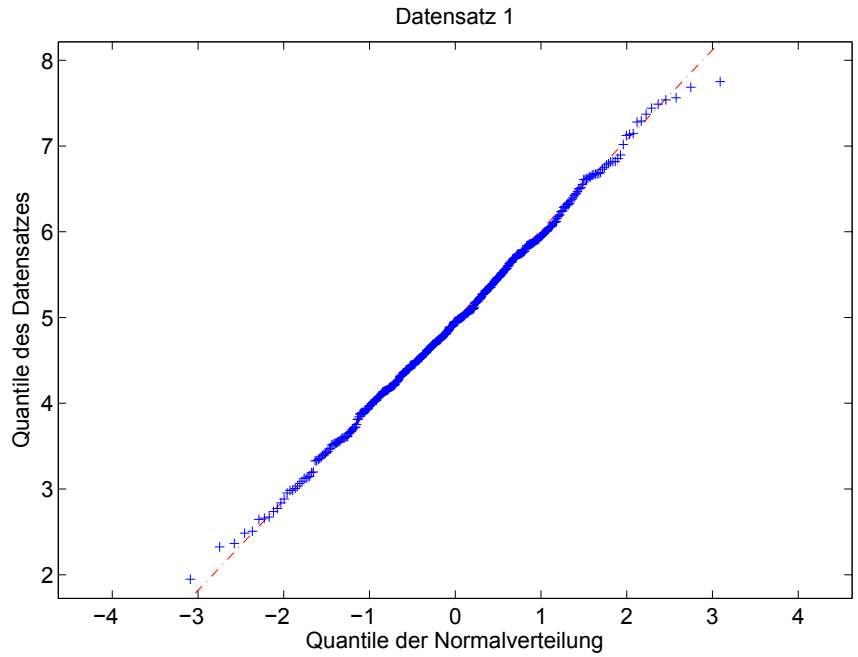

Abbildung 5.9: Q-Q-Diagramm von Datensatz 1. MATLAB: `make_dataset1`

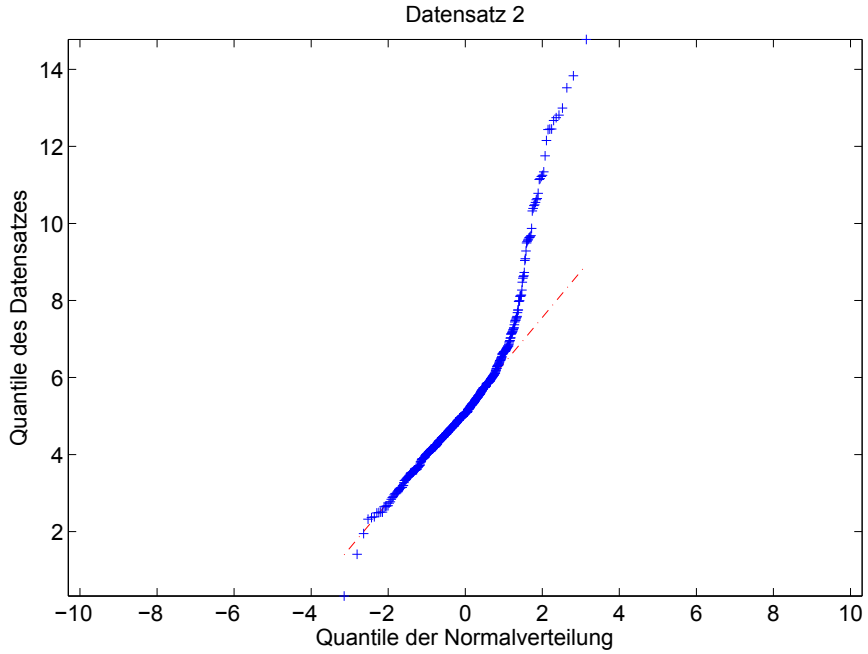

Abbildung 5.10: Q-Q-Diagramm von Datensatz 2. MATLAB: `make_dataset2`

5.2.6 Maßzahlen

Datenlisten sind oft so umfangreich, dass ihr Inhalt in einigen wenigen Maßzahlen zusammgefasst wird oder werden muss. Welche Maßzahlen dabei sinnvoll sind, hängt vom Skalentyp und von der Beschaffenheit der Daten ab. Manche Maßzahlen gehen von der geordneten Datenliste $x_{(1)}, \ldots, x_{(n)}$ aus.

Man unterscheidet Lage-, Streuungs-, und Schiefemaße. Ein *Lagemaß* gibt an, um welchen Wert die Daten konzentriert sind. Ein *Streuungsmaß* gibt an, wie groß die Schwankungen der Daten um ihren zentralen Wert sind. Ein *Schiefemaß* gibt an, wie symmetrisch die Daten um ihren zentralen Wert liegen.

5.2.6.1 Lagemaße

☞ **Definition 5.2.2.** LAGEMASS

Es sei $\boldsymbol{x} = (x_1, \ldots, x_n)$ eine Datenliste. Die Funktion $\ell(\boldsymbol{x})$ heißt ein Lagemaß für $\boldsymbol{x}$, wenn gilt:

$$\ell(a\boldsymbol{x} + b) = a\ell(\boldsymbol{x}) + b \text{ und } \min \boldsymbol{x} \leq \ell(\boldsymbol{x}) \leq \max \boldsymbol{x}$$

Sinnvolle Lagemaße geben den „typischen“ oder „zentralen“ Wert der Datenliste an. Je nach Skala sind verschiedene Lagemaße sinnvoll.

☞ **Definition 5.2.3.** MITTELWERT

Der Mittelwert $\bar{x}$ einer Datenliste ist definiert durch:

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

Der Mittelwert ist sinvoll für Merkmale auf der Intervall- und Verhältnisskala, manchmal auch Ordinalskala. Er minimiert die folgende Funktion:

$$h(x) = \sum_{i=1}^{n} (x_i - x)^2$$

➠MATLAB: `meanx=mean(x)`

☞ **Definition 5.2.4.** MEDIAN

Der Median $\tilde{x}$ einer Datenliste ist definiert durch:

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

Der Median teilt die geordnete Datenliste in zwei gleich große Teile. Ist die Länge der Liste gerade, nimmt man üblicherweise das arithmetische Mittel der beiden zentralen Werte als Median. Der Median ist sinnvoll für Merkmale auf der Ordinal-, Intervall- und Verhältnisskala. Er minimiert die folgende Funktion:

$$h(x) = \sum_{i=1}^{n} |x_i - x|$$

➠MATLAB: `medx=median(x)`.

Der Median ist ein Spezialfall eines allgemeineren Begriffs, des Quantils.

☞ **Definition 5.2.5.** QUANTIL

Das α-Quantil Q_α mit $0 \leq \alpha \leq 1$ einer Datenliste ist definiert durch:

$$Q_\alpha(\boldsymbol{x}) = x_{(\alpha n)}$$

➠MATLAB: `q=quantile(x,alpha)`

Das α-Quantil teilt die geordnete Liste im Verhältnis $\alpha : 1 - \alpha$. Quantile sind sinnvoll für Merkmale auf der Ordinal-, Intervall- und Verhältnisskala. Mit Ausnahme des Medians sind Quantile keine Lagemaße im strengen Sinn. Für $a \geq 0$ gilt zwar:

$$Q_\alpha(a\boldsymbol{x} + b) = aQ_\alpha(\boldsymbol{x}) + b$$

Für $a < 0$ gilt jedoch:

$$Q_\alpha(a\boldsymbol{x} + b) = aQ_{1-\alpha}(\boldsymbol{x}) + b$$

Q_0 ist der kleinste Wert, Q_1 ist der größte Wert der Datenliste, und $Q_{0.5}$ ist der Median. Die fünf *Quartile* $Q_0, Q_{0.25}, Q_{0.5}, Q_{0.75}, Q_1$ bilden das *five point summary* der Datenliste.

➠MATLAB: `fps=quantile(x,[0 0.25 0.5 0.75 1])`

Die graphische Darstellung des five point summary heißt *Boxplot*, siehe Abbildungen 5.11 und 5.12. Die senkrechten Striche geben die Position der fünf Quantile an. Das zentrale Rechteck (die „box“) enthält 50% der Daten; es wird durch den Median, die rote Linie, nochmals in zwei Hälften mit je 25% der Daten geteilt (siehe Abbildungen 5.11 und 5.12).

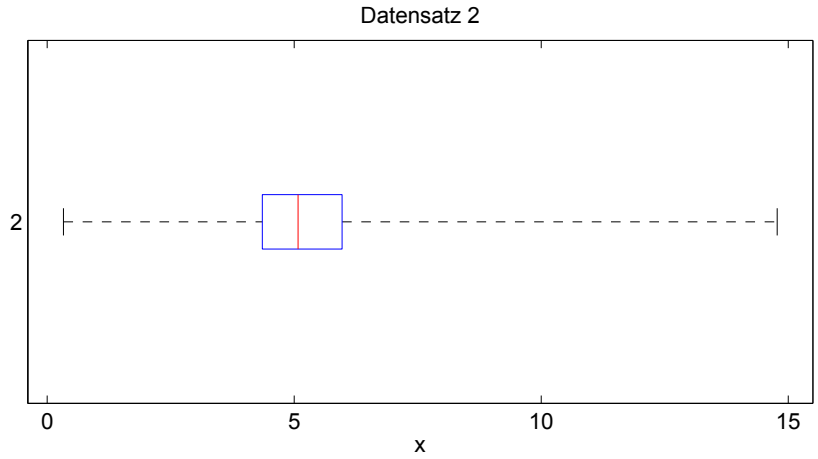

Abbildung 5.11: Boxplot von Datensatz 2. MATLAB: `make_dataset2`

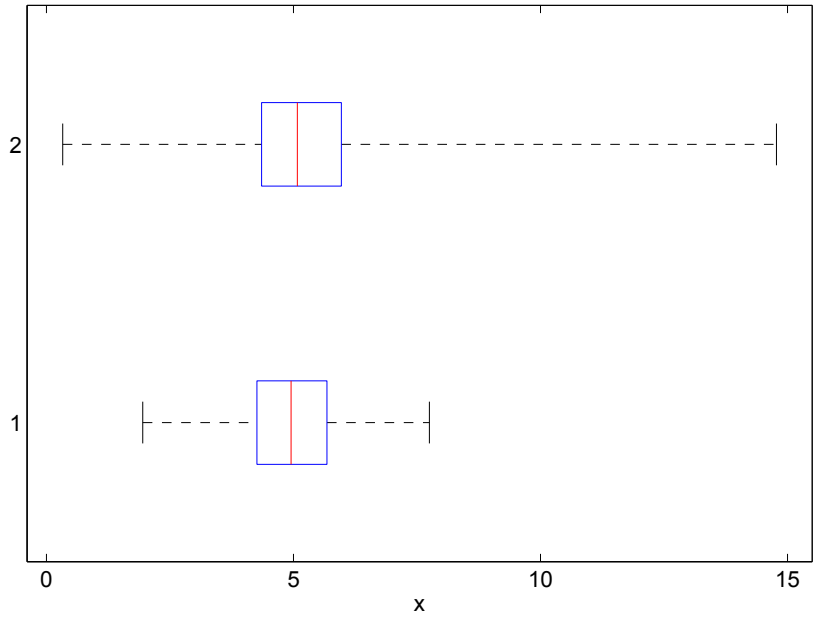

Abbildung 5.12: Vergleich von Datensatz 1 (unten) und Datensatz 2 (oben) mittels Boxplots. MATLAB: `make_boxplot`

Der Median ist *robuster* als der Mittelwert, d.h. unempfindlicher gegen extreme Werte oder Ausreißer. Es gibt noch weitere robuste Lagemaße.

Der LMS-Wert (Least Median of Squares) ist der Mittelpunkt des kürzesten Intervalls, das $h = \lfloor n/2 \rfloor + 1$ Datenpunkte enthält.

➠MATLAB: `lmsx=LMS(x)`

Der LMS-Wert ist extrem unempfindlich gegen fehlerhafte oder untypische Daten. Er minimiert nicht die Summe, sondern den Median der quadratischen Differenzen, d.h. die folgende Funktion:

$$h(x) = \mathrm{med}_{i=1}^{n}(x_i - x)^2$$

☞ **Definition 5.2.7.** SHORTH

Der Shorth ist der Mittelwert aller Daten im kürzesten Intervall, das $h = \lfloor n/2 \rfloor + 1$ *Datenpunkte enthält.*

➠MATLAB: `shorthx=shorth(x)`

☞ **Definition 5.2.8.** MODUS

Der Modus ist der häufigste Wert einer Datenliste.

Der Modus ist nur für diskrete Daten ein sinnvolles Lagemaß. Für stetige Daten kann der Modus aus einem Histogramm oder einem Kernschätzer der Dichtefunktion bestimmt werden.

➠MATLAB: `modex=mode(x)`

☞ **Definition 5.2.9.** HSM

Der HSM (Half-sample mode) ist durch den folgenden Algorithmus definiert:

1. *Bestimme das kürzeste Intervall, das* $h = \lfloor n/2 \rfloor + 1$ *Datenpunkte enthält.*
2. *Wiederhole den Vorgang auf den Daten in diesem Intervall, bis zwei Datenpunkte übrig sind.*
3. *Der HSM-Wert ist das arithmetische Mittel der beiden letzten Daten.*

➠MATLAB: `hsmx=hsm(x)`

5.2.6.2 Streuungsmaße

☞ **Definition 5.2.10.** Streuungsmass

Es sei $\boldsymbol{x} = (x_1, \ldots, x_n)$ eine Datenliste. Die Funktion $\sigma(\boldsymbol{x})$ heißt ein Streuungsmaß für $\boldsymbol{x}$, wenn gilt:

$$\sigma(a\boldsymbol{x} + b) = |a|\ \sigma(\boldsymbol{x}) \text{ und } \sigma(\boldsymbol{x}) \geq 0$$

Sinnvolle Streuungsmaße messen die Abweichung der Daten von ihrem zentralen Wert. Sie sind invariant unter Verschiebung der Daten. Je nach Skala sind verschiedene Streuungsmaße sinnvoll.

☞ **Definition 5.2.11.** Standardabweichung

Die Standardabweichung s einer Datenliste ist definiert durch:

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

Das Quadrat s^2 der Standardabweichung heißt Varianz der Daten.

Die Standardabweichung ist sinnvoll für Intervall- und Verhältnisskala, manchmal auch für Ordinalskala. Sie hat die gleiche Dimension wie die Daten, während die Varianz die Dimension der Daten hoch zwei hat.

➠Matlab: `sx=std(x,1), varx=var(x,1)`

☞ **Definition 5.2.12.** Interquartilsdistanz

Die Interquartilsdistanz IQR ist definiert durch:

$$IQR = Q_{0.75} - Q_{0.25}$$

Die Interquartilsdistanz ist die Länge des Intervalls, das die zentralen 50% der Daten enthält, also die Länge der „box“ im Boxplot. Sie ist sinnvoll für Ordinal-, Intervall- und Verhältnisskala. Sie ist robuster als die Standardabweichung.

➠Matlab: `iqrx=iqr(x)`

☞ **Definition 5.2.13.** LoS

LoS (Length of the Shorth) ist die Länge des kürzesten Intervalls, das $h = \lfloor n/2 \rfloor + 1$ Datenpunkte enthält.

LoS ist sinnvoll für Ordinal-, Intervall- und Verhältnisskala.

➠Matlab: `los=LoS(x)`

5.2.6.3 Schiefemaße

☞ **Definition 5.2.14.** Schiefemass

Es sei $\boldsymbol{x} = (x_1, \ldots, x_n)$ eine Datenliste. Die Funktion $\mathfrak{s}(\boldsymbol{x})$ heißt ein Schiefemaß für $\boldsymbol{x}$, wenn gilt:

$$\mathfrak{s}(a\boldsymbol{x} + b) = \operatorname{sgn}(a)\,\mathfrak{s}(\boldsymbol{x}) \text{ und } \mathfrak{s}(\boldsymbol{x}) = 0 \iff \text{es gibt ein } b \text{ mit } \boldsymbol{x} - b = b - \boldsymbol{x}$$

Sinnvolle Schiefemaße messen die Asymmetrie der Daten und sind 0 für symmetrische Daten. Sie sind invariant unter Verschiebung und Streckung der Daten, ändern jedoch ihr Vorzeichen unter einer Spiegelung. Je nach Skalentypus sind verschiedene Schiefemaße sinnvoll.

☞ **Definition 5.2.15.** SCHIEFE

Die Schiefe γ einer Datenliste ist definiert durch:

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

➠MATLAB: `gammax=skewness(x,1)`

Die Schiefe γ ist gleich 0 für symmetrische Daten. Ist $\gamma < 0$, heißen die Daten *linksschief*; ist $\gamma > 0$, heißen die Daten *rechtsschief*. Die Schiefe ist sinnvoll für Intervall- und Verhältnisskala, mitunter auch für Ordinalskala.

☞ **Definition 5.2.16.** SCHIEFEKOEFFIZIENT

Der Schiefekoeffizient SK ist definiert durch:

$$SK = \frac{R - L}{R + L},$$

mit $R = Q_{0.75} - Q_{0.5}, L = Q_{0.5} - Q_{0.25}$.

SK liegt zwischen -1 ($R = 0$) und $+1$ ($L = 0$). Der Schiefekoeffizient ist gleich 0 für symmetrische Daten. Ist $SK < 0$, heißen die Daten *linksschief*; ist $SK > 0$, heißen die Daten *rechtsschief*. Der Schiefekoeffizient ist sinnvoll für Ordinal-, Intervall- und Verhältnisskala.

➠MATLAB: `skx=SK(x)`

5.2.7 Beispiele

Die folgenden Beispiele zeigen die Werte der oben diskutierten Maßzahlen für die drei Datensätze.

▶ **Beispiel 5.2.17.** Datensatz 1

Datensatz 1: Symmetrisch, 500 Werte (Abbildung 5.13).

Lagemaße		Streuungsmaße		Schiefemaße	
Mittelwert:	4.9532	Standardabweichung:	1.0255	Schiefe:	0.0375
Median:	4.9518	Interquartilsdistanz:	1.4168	Schiefekoeff.:	0.0258
LMS:	4.8080	Length of the Shorth:	1.3520		
Shorth:	4.8002				
HSM:	5.0830				

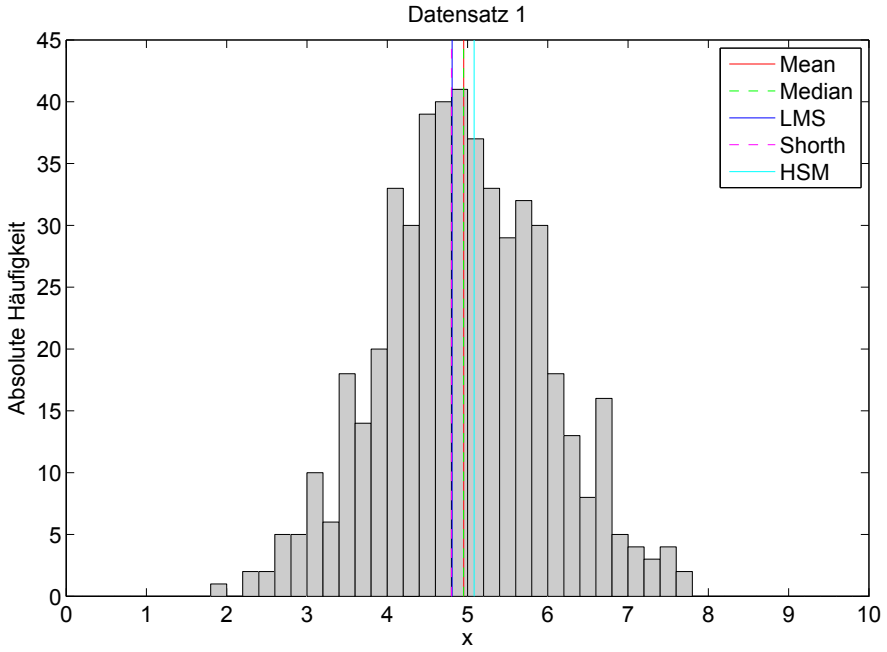

Abbildung 5.13: Datensatz 1: Mittelwert, Median, LMS, Shorth, HSM.
Matlab: `make_dataset1`

▶ **Beispiel 5.2.18.** Datensatz 2

Datensatz 2: Datensatz 1 + Kontamination (Abbildung 5.14).

Lagemaße		Streuungsmaße		Schiefemaße	
Mittelwert:	5.4343	Standardabweichung:	1.8959	Schiefe:	1.7696
Median:	5.0777	Interquartilsdistanz:	1.6152	Schiefekoeff.:	0.1046
LMS:	5.1100	Length of the Shorth:	1.5918		
Shorth:	5.0740				
HSM:	4.9985				

Hier ist zu beobachten, dass Mittelwert, Standardabweichung und Schiefe von der Kontamination wesentlich stärker beeinflusst werden als die anderen, robusten Maßzahlen.

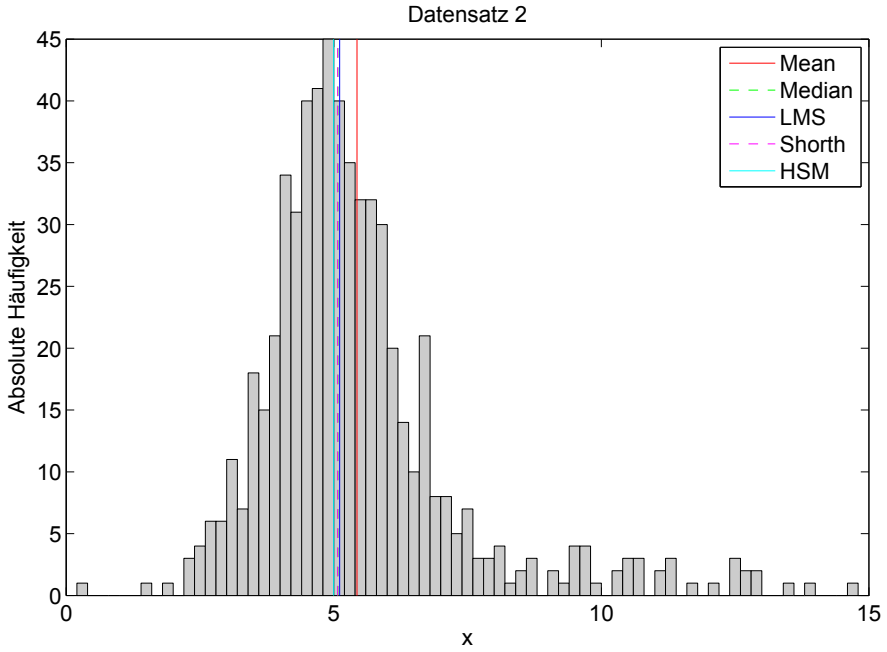

Abbildung 5.14: Datensatz 2: Mittelwert, Median, LMS, Shorth, HSM.
MATLAB: `make_dataset2`

▶ **Beispiel 5.2.19.** Datensatz 3

Datensatz 3: 50 Prüfungsnoten (Abbildung 5.15).

Lagemaße		Streuungsmaße		Schiefemaße	
Mittelwert:	3.14	Standardabweichung:	1.2	Schiefe:	0.0765
Median:	3	Interquartilsdistanz:	2	Schiefekoeffizient:	0
Modus:	3				

Wie weit ein Mittelwert von Prüfungsnoten zulässig ist, ist diskutabel; auch der Median ist bei nur fünf Ausprägungen nicht besonders aussagekräftig.

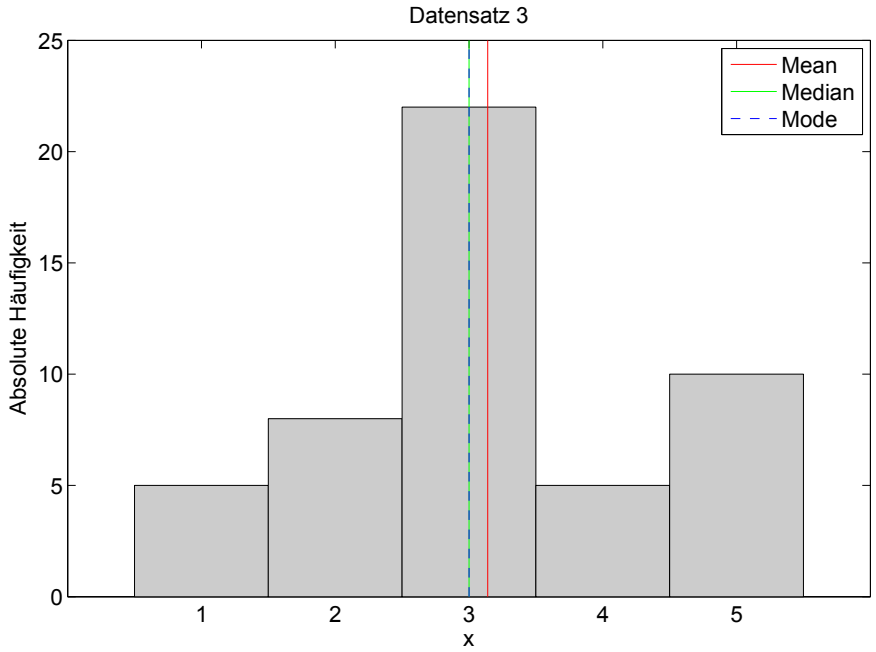

Abbildung 5.15: Datensatz 3: Mittelwert, Median, Modus.
MATLAB: `make_dataset3`

5.3 Multivariate Merkmale

5.3.1 Graphische Darstellung

Oft werden zwei oder mehr Merkmale eines Objekts *gleichzeitig* beobachtet. Einige Beispiele dafür sind Körpergröße und Gewicht einer Person, Alter und Einkommen einer Person, Schulbildung und Geschlecht einer Person, oder in der Physik Azimut- und Polarwinkel eines gestreuten Teilchens. Der Zusammenhang zwischen den beiden Merkmalen kann zusätzliche Information geben.

Es werden in diesem Abschnitt nur quantitative Merkmale betrachtet, die in der Physik von größerem Interesse sind als qualitative Merkmale. Binäre Merkmale wurden bereits behandelt, insbesondere die Darstellung in einer Vierfeldertafel (Abschnitt 2.3.1) und die Beschreibung der Kopplung mittels Vierfelderkorrelation (Abschnitt 2.3.3).

Eine übliche Darstellung von bivariaten quantitativen Merkmalen ist das *Streudiagramm* (engl. scatter plot). Jeder Punkt im Streudiagramm entspricht einer Realisierung des bivariaten Merkmals, wobei die Koordinaten des Punktes in der x-y-Ebene durch die Werte der Realisierung gegeben sind. Als Beispiel im Folgenden dient Datensatz 4, der Körpergröße und Gewicht von 100 Personen enthält (Abbildung 5.16).

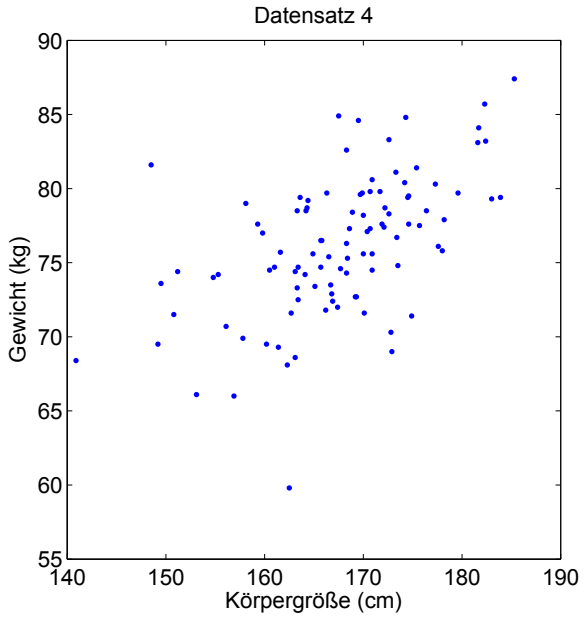

Abbildung 5.16: Streudiagramm von Datensatz 4. MATLAB: `make_dataset4`

Höherdimensionale Merkmale können durch Histogramme und paarweise Streudiagramme dargestellt werden. Dabei geht jedoch ein Teil der Information verloren. Datensatz 5 (Abbildung 5.17) enthält drei beobachtete Merkmale von 100 Personen, und zwar $x_1 =$ Körpergröße (in cm); $x_2 =$ Gewicht (in kg); $x_3 =$ Alter (in Jahren).

5.3.2 Empirische Korrelation

Das Streudiagramm hat folgende Eigenschaften:

1. $(\bar{x}, \bar{y})$ ist der Schwerpunkt der Punktwolke, wenn allen Punkte die gleiche Masse zugeteilt wird.
2. Die Projektion der Punktwolke auf die x-Achse ergibt das Punktediagramm der Datenliste $x_1, \ldots, x_n$.
3. Die Projektion der Punktwolke auf die y-Achse ergibt das Punktediagramm der Datenliste $y_1, \ldots, y_n$.

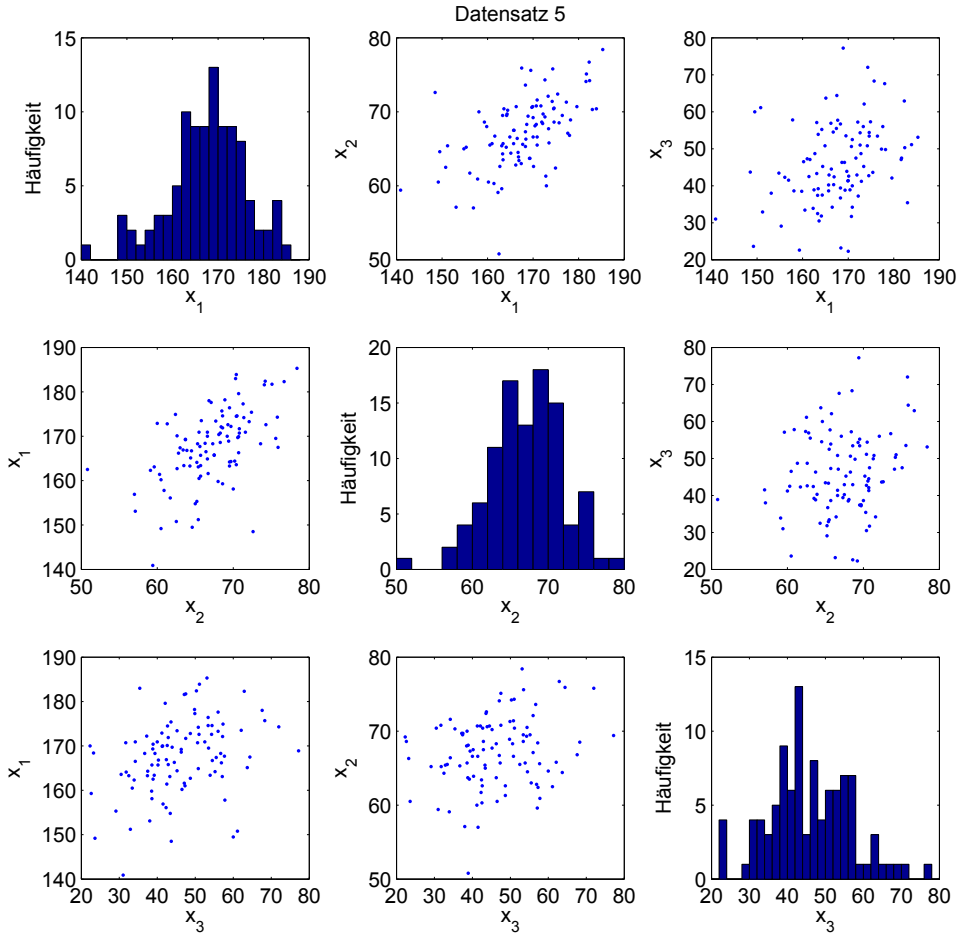

Abbildung 5.17: Histogramme und paarweise Streudiagramme von Datensatz 5. MATLAB: `make_dataset5`

Aus dem Streudiagramm von Datensatz 4 (Abbildung 5.16) ist ersichtlich, dass *tendenziell* größere Körpergröße mit größerem Gewicht einhergeht. Zwischen den beiden Merkmalen x und y besteht offensichtlich ein Zusammenhang, der auch intuitiv völlig klar ist. Eine nützliche Maßzahl zur Quantifizierung des Zusammenhangs ist der *empirische Korrelationskoeffizient*.

☞ **Definition 5.3.1.** STANDARDSCORE

Sei $(x_1, y_1), \ldots, (x_n, y_n)$ eine bivariate Datenliste. Die Standardscores sind definiert durch:

$$z_{x,i} = \frac{x_i - \bar{x}}{s_x}, \quad z_{y,i} = \frac{y_i - \bar{y}}{s_y},$$

wo s_x bzw. s_y die Standardabweichungen der x- bzw. der y-Werte sind.

☞ **Definition 5.3.2.** EMPIRISCHER KORRELATIONSKOEFFIZIENT

Der empirische Korrelationskoeffizient r_{xy} ist definiert durch:

$$r_{xy} = \frac{1}{n} \sum_{i=1}^{n} z_{x,i} z_{y,i} = \frac{1}{n} \left(z_{x,1} z_{y,1} + \cdots + z_{x,n} z_{y,n} \right)$$

Er ist also der Mittelwert der paarweisen Produkte der Standardscores.

➠MATLAB: `r=rxy(x,y)`

☛ **Satz 5.3.3.**

Es gilt immer:

$$-1 \leq r_{xy} \leq 1$$

r_{xy} ist positiv, wenn viele Produkte positiv sind, d.h. viele Paare von Standardscores das gleiche Vorzeichen haben. Das ist der Fall, wenn die Paare der Standardscores vorwiegend im 1. oder 3. Quadranten liegen. x und y heißen dann *positiv korreliert*. r_{xy} ist negativ, wenn viele Produkte negativ sind, d.h. viele Paare von Standardscores verschiedenes Vorzeichen haben. Das ist der Fall, wenn die Paare der Standardscores vorwiegend im 2. oder 4. Quadranten liegen. x und y heißen dann *negativ korreliert*.

In Abbildung 5.18 sind x und y offensichtlich positiv korreliert, da die meisten Punkte im ersten und dritten Quadranten liegen. Tatsächlich ist $r_{xy} = 0.5562$. Eine positive Korrelation muss nicht unbedingt einen kausalen Zusammenhang bedeuten. Sie kann auch durch eine gemeinsame Ursache oder einen parallel laufenden Trend verursacht sein.

Der Korrelationskoeffizient misst die *Korrelation* der Daten. Die Korrelation gibt die Bindung der Punktwolke an eine steigende oder fallende *Gerade*, die *Hauptachse* an. Die Korrelation gibt also das Ausmaß der *linearen* Kopplung an. Abbildung 5.19 zeigt Streudiagramme von Standardscores mit verschieden starker Korrelation. Besteht zwischen x und y ein starker, aber *nichtlinearer* Zusammenhang, kann die Korrelation aus Symmetriegründen trotzdem sehr klein sein (Abbildung 5.20).

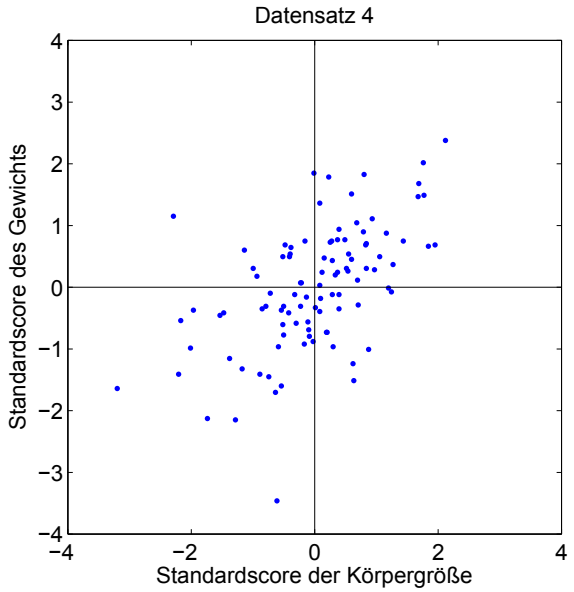

Abbildung 5.18: Streudiagramm der Standardscores von Datensatz 4.
MATLAB: `make_dataset4`

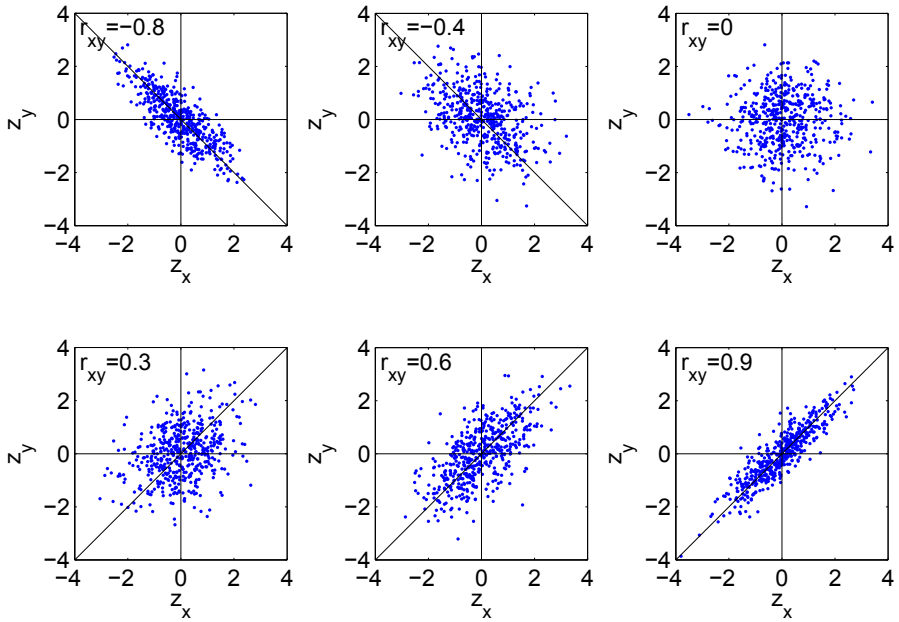

Abbildung 5.19: Standardscores mit verschiedenen Korrelationskoeffizienten.
MATLAB: `make_scatterplot`

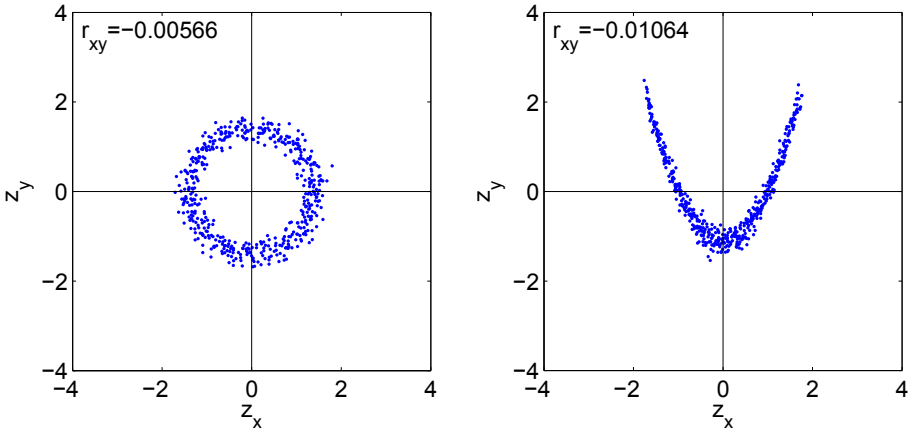

Abbildung 5.20: Nichtlinearer Zusammenhang zwischen x und y.
MATLAB: `make_scatterplot`

Kapitel 6

Schätzung

6.1 Stichprobenfunktionen

6.1.1 Grundbegriffe

$X_1, \ldots, X_n$ seien unabhängige Zufallsvariable, die alle die gleiche Verteilungsfunktion $F(x)$ haben. Sie bilden dann eine *zufällige Stichprobe* der Verteilung F.* Die gemeinsame Dichtefunktion g von $X_1, \ldots, X_n$ ist wegen der Unabhängigkeit das Produkt der Einzeldichten:

$$g(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Eine Zufallsvariable

$$Y = h(X_1, \ldots, X_n)$$

heißt eine *Stichprobenfunktion*. In vielen Fällen sind die Verteilung oder zumindest Momente von Y zu bestimmen.

6.1.2 Stichprobenmittel

☞ **Definition 6.1.1.** Stichprobenmittel
Das Stichprobenmittel $\overline{X}$ der Stichprobe $X_1, \ldots, X_n$ ist definiert durch:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

☛ **Satz 6.1.2.** Erwartung und Varianz des Stichprobenmittels
Hat die Verteilung F den Mittelwert μ und die Varianz σ^2, gilt:

1. $\mathsf{E}[\overline{X}] = \mu$
2. $\mathsf{var}[\overline{X}] = \dfrac{\sigma^2}{n}$
3. *Ist F eine Normalverteilung, so ist $\overline{X}$ normalverteilt.*

☛ **Satz 6.1.3.** Zentraler Grenzwertsatz für Stichprobenmittel
Es sei $X_1, \ldots, X_n$ eine zufällige Stichprobe aus der Verteilung F. Hat F den Mittelwert μ und die Varianz σ^2, so konvergiert die Verteilung von

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

für wachsendes gegen die Standardnormalverteilung. Ist F eine Normalverteilung, so ist Z für alle n standardnormalverteilt.

*Eine Verteilung wird hier durch ihre Verteilungsfunktion F charakterisiert.

▶ **Beispiel 6.1.4.** Stichprobenmittel der Exponentialverteilung

Es sei $X_1, \ldots, X_n$ eine zufällige Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau) = \mathrm{Ga}(1, \tau)$. Die Stichprobensumme ist dann gammaverteilt gemäß $\mathrm{Ga}(n, \tau)$ (siehe Beispiel 4.3.22), und $\overline{X}$ ist gammaverteilt gemäß $\mathrm{Ga}(n, \tau/n)$ (siehe Satz 4.2.20). ◀

▶ **Beispiel 6.1.5.** Stichprobenmittel der Cauchyverteilung

Es sei $X_1, \ldots, X_n$ eine zufällige Stichprobe aus der Cauchyverteilung $\mathrm{T}(1) = \mathrm{T}(1|0, 1)$ (siehe Abschnitt 4.2.8). Die Stichprobensumme ist dann verteilt gemäß $\mathrm{T}(1|0, n)$ (siehe Beispiel 4.3.24), und $\overline{X}$ ist verteilt gemäß $\mathrm{T}(1|0, 1)$. Das Stichprobenmittel ist also genauso verteilt wie *eine einzelne Beobachtung*. Der zentrale Grenzwertsatz gilt hier nicht, weil die Cauchyverteilung keine endlichen Momente besitzt. ◀

6.1.3 Stichprobenvarianz

☞ **Definition 6.1.6.** Stichprobenvarianz

Die Stichprobenvarianz S^2 der Stichprobe $X_1, \ldots, X_n$ ist definiert durch:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Der Faktor $1/(n-1)$ vor der Summe garantiert, dass die Erwartung der Stichprobenvarianz gleich der Verteilungsvarianz σ^2 ist.

☛ **Satz 6.1.7.** Erwartung und Varianz der Stichprobenvarianz
Hat F die Varianz σ^2, gilt:

$$\mathsf{E}[S^2] = \sigma^2$$

Hat F das vierte zentrale Moment μ_4, gilt:

$$\mathsf{var}[S^2] = \frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$$

☛ **Satz 6.1.8.** Stichprobenvarianz unter Normalverteilung
Ist F eine Normalverteilung mit Mittelwert μ und Varianz σ^2, so gilt:

1. $(n-1)S^2/\sigma^2$ *ist* χ^2*-verteilt mit* $n-1$ *Freiheitsgraden.*
2. $\overline{X}$ *und* S^2 *sind unabhängig.*
3. $\mathsf{var}[S^2] = \dfrac{2\sigma^4}{n-1}$
4. $T = \dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ *ist* $\mathrm{T}(n-1)$*-verteilt mit* $n-1$ *Freiheitsgraden.*

6.1.4 Stichprobenmedian

Der Stichprobenmedian wird aus der geordneten Stichprobe $X_{(1)}, \ldots, X_{(n)}$ berechnet.

☞ **Definition 6.1.9.** Stichprobenmedian
Der Stichprobenmedian $\widetilde{X}$ der Stichprobe $X_1, \ldots, X_n$ ist definiert durch:

$$\widetilde{X} = \begin{cases} X_{((n+1)/2)}, & n \text{ ungerade} \\ \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right), & n \text{ gerade} \end{cases}$$

☛ **Satz 6.1.10.** Momente des Stichprobenmedians
Hat F den eindeutigen Median m und die Dichtefunktion f, gilt:

1. $\lim\limits_{n\to\infty} \mathsf{E}[\widetilde{X}] = m$
2. $\lim\limits_{n\to\infty} \mathsf{var}[\widetilde{X}] = \dfrac{1}{4\,n f^2(m)}$, *sofern* $f(m) > 0$.
3. $\widetilde{X}$ *ist asymptotisch normalverteilt, sofern* $f(m) > 0$.

6.2 Punktschätzer

6.2.1 Eigenschaften von Punktschätzern

Ein Punktschätzer T von ϑ ist eine Stichprobenfunktion, die einen möglichst genauen Näherungswert für einen unbekannten Verteilungsparameter ϑ liefern soll:

$$T = g(X_1, \ldots, X_n)$$

Die Funktion $g(x_1, \ldots, x_n)$ wird die Schätzfunktion genannt. T ist als Funktion der Stichprobe eine Zufallsvariable. Die Konstruktion von sinnvollen Punktschätzern für einen Parameter ϑ ist Aufgabe der Schätztheorie. Für einen Parameter ϑ sind viele Punktschätzer möglich. Ein „guter“ Punktschätzer sollte jedoch gewisse Anforderungen erfüllen. Insbesondere sollte er keine oder nur eine kleine systematische Abweichung vom wahren Wert von ϑ haben, und seine Standardabweichung, oft auch *Standardfehler* genannt, sollte möglichst klein sein.

☞ **Definition 6.2.1.** Erwartungstreue

Ein Punktschätzer T für den Parameter ϑ heißt erwartungstreu oder unverzerrt, wenn für alle zulässigen Werte von ϑ gilt:

$$\mathsf{E}_\vartheta[T] = \vartheta$$

T heißt asymptotisch unverzerrt, wenn gilt:

$$\lim_{n\to\infty} \mathsf{E}_\vartheta[T] = \vartheta$$

Die Differenz

$$b[T] = \mathsf{E}_\vartheta[T] - \vartheta$$

wird als die Verzerrung (engl. bias) von T bezeichnet.

Ist der unbekannte Parameter gleich ϑ, dann ist die Erwartung eines erwartungstreuen Punktschätzers gleich ϑ. Ein solcher Punktschätzer hat zwar zufällige Abweichungen vom wahren Wert ϑ, aber keine systematische Abweichung, d.h. er ist „im Mittel“ korrekt.

☞ **Definition 6.2.2.** MSE

Die mittlere quadratische Abweichung (mean squared error, MSE) eines Punktschätzers T für den Parameter ϑ ist definiert durch:

$$\mathsf{MSE}[T] = \mathsf{E}_\vartheta[(T-\vartheta)^2]$$

Für erwartungstreue Punktschätzer ist die mittlere quadratische Abweichung gleich der Varianz. Für verzerrte Punktschätzer gilt:

$$\mathsf{MSE}[T] = \mathsf{var}[T] + (b[T])^2$$

☞ **Definition 6.2.3.** MSE-Konsistenz

Ein Punktschätzer T für den Parameter ϑ heißt konsistent im quadratischen Mittel (MSE-konsistent), wenn gilt:

$$\lim_{n\to\infty} \mathsf{MSE}[T] = 0$$

MSE-Konsistenz bedeutet, dass mit wachsender Stichprobengröße die Verteilung des Schätzers sich immer stärker um den wahren Wert des Parameters konzentriert. Abweichungen über einer gewissen Schranke sind zwar möglich, werden aber immer unwahrscheinlicher. Konsistenz ist eine Minimalanforderung an einen Schätzer. Nicht konsistente Schätzer sollten in den Anwendungen auf jeden Fall vermieden werden.

Abbildung 6.1 zeigt die Dichtefunktion des Stichprobenmittels einer Stichprobe aus der Exponentialverteilung Ex(3) für verschiedene Stichprobengrößen ($n = 10, 50, 250, 1000$). Der Schätzer ist offensichtlich MSE-konsistent (siehe auch Satz 6.2.9).

☞ **Definition 6.2.4.** MSE-Effizienz

Ein Punktschätzer T_1 heißt MSE-effizienter als der Punktschätzer T_2, wenn für alle zulässigen ϑ gilt:

$$\mathsf{MSE}[T_1] \leq \mathsf{MSE}[T_2]$$

Zum Beispiel ist für normalverteilte Stichproben das Stichprobenmittel MSE-effizienter als der Stichprobenmedian (siehe Beispiel 6.2.14).

☞ **Definition 6.2.5.** Effizienz

Ein unverzerrter Punktschätzer T_1 heißt effizienter als der unverzerrte Punktschätzer T_2, wenn für alle zulässigen ϑ gilt:

$$\mathsf{var}[T_1] \leq \mathsf{var}[T_2]$$

Ein unverzerrter Punktschätzer T heißt effizient, wenn seine Varianz den kleinsten möglichen Wert annimmt.

Einige Beispiele von effizienten Punktschätzern sind in Beispiel 6.2.10 zu finden.

☞ **Definition 6.2.6.** Fisher-Information

Es sei $X_1, \ldots, X_n$ eine Stichprobe mit der gemeinsamen Dichtefunktion g:

$$g(x_1, \ldots, x_n \,|\, \vartheta) = \prod_{i=1}^{n} f(x_i \,|\, \vartheta)$$

Die Erwartung

$$I_n(\vartheta) = \mathsf{E}\left[-\frac{\partial^2 \ln g(X_1, \ldots, X_n \,|\, \vartheta)}{\partial \vartheta^2}\right]$$

heißt die (erwartete) Fisher-Information der Stichprobe.

Der folgende Satz besagt, dass die Varianz eines unverzerrten Schätzers von ϑ bei gegebener Stichprobengröße n nicht kleiner als die inverse Fisher-Information $I_n(\vartheta)$ werden kann, wenn gewisse Regularitätsbedingungen erfüllt sind. Insbesondere darf der Bereich, in dem die Dichtefunktion der Beobachtungen positiv ist, nicht vom Parameter

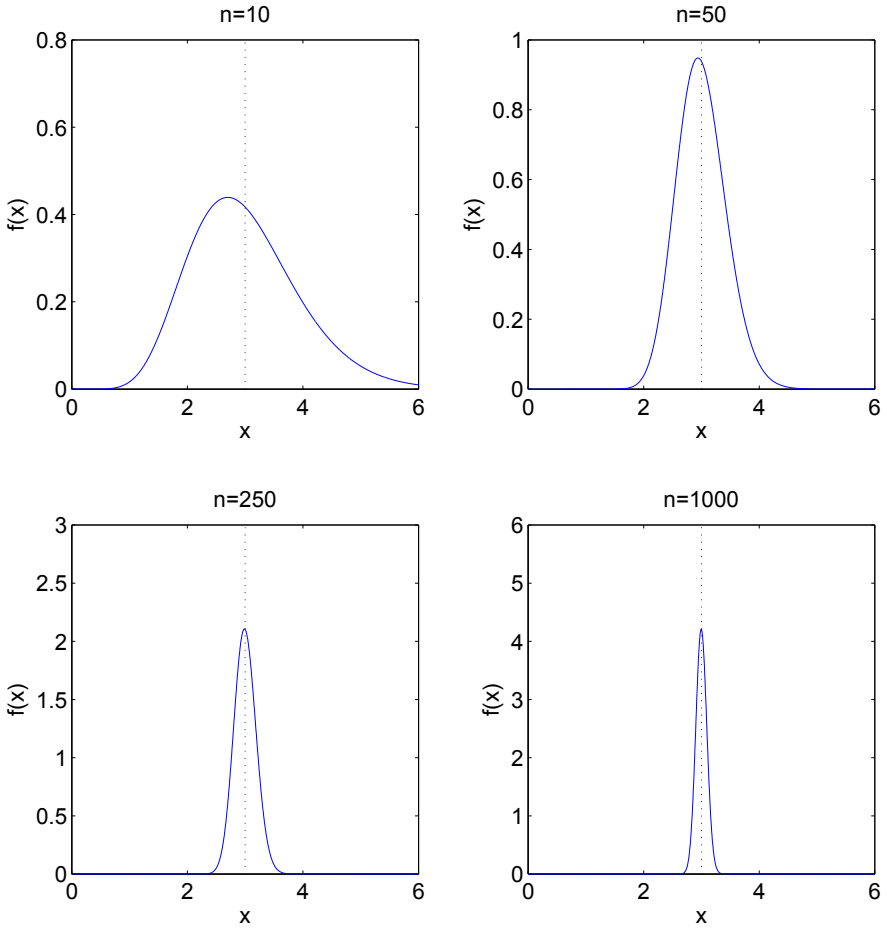

Abbildung 6.1: Dichtefunktion des Stichprobenmittels einer Stichprobe aus der Exponentialverteilung Ex(3) für die Stichprobengrößen $n = 10, 50, 250, 1000$.
MATLAB: `make_konsistenz`

ϑ abhängen. Dieser Bereich wird der *Träger* der Dichtefunktion genannt. Die Regularitätsbedingungen können in jedem Lehrbuch über mathematische Statistik nachgelesen werden (siehe z.B. das Buch von C. Czado in Anhang L). Sie werden von den meisten hier besprochenen Verteilungen erfüllt. Eine Ausnahme ist die Gleichverteilung mit unbekannter Intervallbegrenzung (siehe Beispiel 6.3.9), da in diesem Fall der Träger der Dichte vom zu schätzenden Parameter abhängt.

☛ **Satz 6.2.7.** SATZ VON RAO UND CRAMÈR

Es sei $X_1, \ldots, X_n$ eine Stichprobe mit der gemeinsamen Dichtefunktion $g(x_1, \ldots, x_n | \vartheta)$. Die Varianz eines unverzerrten Schätzers T für den Parameter ϑ ist nach unten beschränkt durch:

$$\mathsf{var}[T] \geq 1/I_n(\vartheta)$$

wenn geeignete Regularitätsbedingungen erfüllt sind.

► **Beispiel 6.2.8.**
Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$. Die gemeinsame Dichtefunktion ist dann gleich

$$g(x_1, \ldots, x_n \,|\, \tau) = \frac{1}{\tau^n} \exp\left(-\sum_{i=1}^{n} x_i/\tau \right)$$

Daraus folgt:

$$\ln g(x_1, \ldots, x_n \,|\, \tau) = -\, n \ln \tau - \sum_{i=1}^{n} x_i/\tau$$

$$\frac{\partial^2}{\partial \tau^2} \ln g(x_1, \ldots, x_n \,|\, \tau) = \frac{n}{\tau^2} - \frac{2}{\tau^3} \sum_{i=1}^{n} x_i$$

$$\mathsf{E}\left[\frac{\partial^2}{\partial \tau^2} \ln g(X_1, \ldots, X_n \,|\, \tau) \right] = \frac{n}{\tau^2} - \frac{2\,n\tau}{\tau^3} = -\frac{n}{\tau^2}$$

Die Fisher-Information ist also gleich:

$$I_n(\tau) = \frac{n}{\tau^2}$$

Für jeden unverzerrten Schätzer T von τ gilt folglich:

$$\mathsf{var}[T] \geq \frac{\tau^2}{n}$$ ◄

6.2.2 Schätzung des Mittelwerts

☛ Satz 6.2.9.

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit Mittelwert μ. Dann ist das Stichprobenmittel $\overline{X}$ ein unverzerrter Punktschätzer von μ. Hat F die endliche Varianz σ^2, so ist $\mathsf{var}[\overline{X}] = \sigma^2/n$ und $\overline{X}$ ist MSE-konsistent.

Der Standardfehler des Stichprobenmittels ist daher verkehrt proportional zur Wurzel aus der Stichprobengröße. Wie das folgende Beispiel zeigt, ist das Stichprobenmittel für gewisse einfache Verteilungen der *effiziente Schätzer* für das Verteilungsmittel.

▶ Beispiel 6.2.10. Effiziente Schätzer des Mittelwerts

A. Normalverteilung Ist F die Normalverteilung $\mathrm{No}(\mu, \sigma^2)$, so ist $\overline{X}$ normalverteilt gemäß $\mathrm{No}(\mu, \sigma^2/n)$. Da die Fisher-Information für μ gleich $I_n(\mu) = n/\sigma^2$ ist, ist $\overline{X}$ effizient für μ.

B. Exponentialverteilung Ist F die Exponentialverteilung $\mathrm{Ex}(\tau)$, so ist $\overline{X}$ gammaverteilt mit den Mittelwert τ und Varianz τ^2/n. Da die Fisher-Information für τ gleich $I_n(\tau) = n/\tau^2$ ist, ist $\overline{X}$ effizient für τ.

C. Poissonverteilung Ist F die Poissonverteilung $\mathrm{Po}(\lambda)$, hat $\overline{X}$ Mittelwert λ und Varianz λ/n. Da die Fisher-Information für λ gleich $I_n(\lambda) = n/\lambda$ ist, ist $\overline{X}$ effizient für λ.

D. Alternativverteilung Ist F die Alternativverteilung $\mathrm{Al}(p)$, hat $\overline{X}$ Mittelwert p und Varianz $p(1-p)/n$. Da die Fisher-Information für p gleich $I_n(p) = n/[p(1-p)]$ ist, ist $\overline{X}$ effizient für p. ◀

6.2.3 Schätzung der Varianz

☛ Satz 6.2.11.

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit Mittelwert μ und Varianz σ^2. Dann ist die Stichprobenvarianz S^2 ein unverzerrter Punktschätzer von σ^2. Hat F das endliche vierte zentrale Moment μ_4, so ist

$$\mathsf{var}[S^2] = \frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$$

In diesem Fall ist S^2 MSE-konsistent.

▶ Beispiel 6.2.12.

Ist F die Normalverteilung $\mathrm{No}(\mu, \sigma^2)$, so ist $(n-1)S^2/\sigma^2$ χ^2-verteilt mit $n-1$ Freiheitsgraden (siehe Satz 6.1.8). Die Varianz von S^2 ist dann gleich:

$$\mathsf{var}[S^2] = \frac{2\sigma^4}{n-1}$$

Die Fisher-Information für σ^2 ist gleich:

$$I_n(\sigma^2) = \frac{n}{2\sigma^4}$$

S^2 ist also ein asymptotisch effizienter Punktschätzer für σ^2. ◀

6.2.4 Schätzung des Medians

☛ **Satz 6.2.13.**

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der stetigen Verteilung F mit Dichtefunktion $f(x)$ und Median m. Dann ist der Stichprobenmedian $\widetilde{X}$ ein asymptotisch unverzerrter Punktschätzer von m. Für symmetrisches F ist $\widetilde{X}$ unverzerrt. Der Stichprobenmedian $\widetilde{X}$ hat asymptotisch die Varianz

$$\mathsf{var}[\widetilde{X}] \approx \frac{1}{4nf(m)^2}$$

Der Stichprobenmedian ist MSE-konsistent, sofern $f(m) > 0$.

▶ **Beispiel 6.2.14.**

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$. Die Varianz von $\overline{X}$ ist gleich:

$$\mathsf{var}[\overline{X}] = \frac{\sigma^2}{n}$$

Die Varianz von $\widetilde{X}$ ist für großes n ungefähr gleich:

$$\mathsf{var}[\widetilde{X}] \approx \frac{2\,\pi\sigma^2}{4\,n} \approx 1.57 \cdot \frac{\sigma^2}{n}$$

Sie ist also um mehr als 50% größer als die Varianz von $\overline{X}$. ◀

▶ **Beispiel 6.2.15.**

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der T-Verteilung T(3). Die Varianz von $\overline{X}$ ist gleich:

$$\mathsf{var}[\overline{X}] = \frac{3}{n}$$

Die Varianz von $\widetilde{X}$ ist für großes n ungefähr gleich:

$$\mathsf{var}[\widetilde{X}] \approx \frac{1}{4\,nf(0)^2} = \frac{1.8506}{n} \approx 0.62 \cdot \frac{3}{n}$$

Sie ist also fast um 40% kleiner als die Varianz von $\overline{X}$. Es gibt aber Schätzer mit noch kleinerer Varianz als $\widetilde{X}$. ◀

6.3 Spezielle Schätzverfahren

6.3.1 Maximum-Likelihood-Schätzer

Das Maximum-Likelihood-Verfahren ist eines der wichtigsten Verfahren der klassischen Inferenzstatistik. Die so konstruierten Schätzer (ML-Schätzer) haben unter relativ allgemeinen Bedingungen optimale Eigenschaften und sind daher in der experimentellen Praxis überaus beliebt.

☞ **Definition 6.3.1.** ML-Schätzer

Es sei $X_1, \ldots, X_n$ eine Stichprobe mit der gemeinsamen Dichtefunktion $g(x_1, \ldots, x_n \,|\, \vartheta)$. Die Funktion

$$\mathcal{L}(\vartheta \,|\, X_1, \ldots, X_n) = g(X_1, \ldots, X_n \,|\, \vartheta)$$

heißt die Likelihood-Funktion der Stichprobe. Der plausible oder Maximum-Likelihood-Schätzer $\hat{\vartheta}$ ist jener Wert von ϑ, der die Likelihood-Funktion der Stichprobe maximiert.

Der ML-Schätzer kann ohne weiteres auch auch auf einen Vektor $\boldsymbol{\vartheta}$ von Parametern verallgemeinert werden. Im folgenden wird der ML-Schätzer durch ein „Dach“ auf dem geschätzten Parameter bezeichnet.

Oft wird statt der Likelihood-Funktion ihr Logarithmus, die Log-Likelihood-Funktion $\ell(\vartheta) = \ln \mathcal{L}(\vartheta)$ maximiert. Der ML-Schätzer kann allerdings nur in einfachen Fällen in geschlossener Form angegeben werden; in der Praxis muss er in der Regel durch numerische Maximierung der Likelihood-Funktion oder der Log-Likelihood-Funktion ermittelt werden.

► **Beispiel 6.3.2.** ML-Schätzung des Bernoulli-Parameters

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Alternativverteilung $\mathrm{Al}(p)$. Die gemeinsame Dichtefunktion lautet:

$$g(x_1, \ldots, x_n \,|\, p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Die Log-Likelihood-Funktion ist daher:

$$\ell(p) = \sum_{i=1}^{n} X_i \, \ln p + \left(n - \sum_{i=1}^{n} X_i \right) \ln(1-p)$$

Ableiten nach p ergibt:

$$\frac{\partial \ell(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right)$$

Nullsetzen der Ableitung und Auflösen nach p ergibt:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Der ML-Schätzer $\hat{p}$ ist unverzerrt und effizient. ◄

► **Beispiel 6.3.3.** ML-Schätzung des Poisson-Parameters

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Poissonverteilung $\mathrm{Po}(\lambda)$. Die gemeinsame Dichtefunktion lautet:

$$g(x_1, \ldots, x_n \,|\, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} \mathrm{e}^{-\lambda}}{x_i!}$$

Die Log-Likelihood-Funktion ist daher:

$$\ell(\lambda) = \sum_{i=1}^{n} [X_i \, \ln \lambda - \lambda - \ln(X_i!)]$$

Ableiten nach λ ergibt:

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} X_i - n$$

Nullsetzen der Ableitung und Auflösen nach λ ergibt:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Der ML-Schätzer $\hat{\lambda}$ ist unverzerrt und effizient. ◄

► **Beispiel 6.3.4.** ML-Schätzung der mittleren Lebensdauer

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$. Die gemeinsame Dichtefunktion lautet:

$$g(x_1, \ldots, x_n \,|\, \tau) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-x_i/\tau}}{\tau}$$

Die Log-Likelihood-Funktion ist daher:

$$\ell(\tau) = \sum_{i=1}^{n} [-\ln \tau - X_i/\tau]$$

Ableiten nach τ ergibt:

$$\frac{\partial \ell(\tau)}{\partial \tau} = -\frac{n}{\tau} + \frac{1}{\tau^2}\sum_{i=1}^{n} X_i$$

Nullsetzen der Ableitung und Auflösen nach τ ergibt:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Der ML-Schätzer $\hat{\tau}$ ist unverzerrt und effizient. ◄

► **Beispiel 6.3.5.** ML-Schätzung der Parameter der Normalverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$. Die gemeinsame Dichtefunktion lautet:

$$g(x_1, \ldots, x_n \,|\, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\,\sigma^2}\right]$$

Die Log-Likelihood-Funktion ist daher:

$$\ell(\mu, \sigma^2) = \sum_{i=1}^{n} \left[-\ln\sqrt{2\pi} - \tfrac{1}{2}\ln\sigma^2 - \frac{(X_i - \mu)^2}{2\,\sigma^2}\right]$$

Ableiten nach μ und σ^2 ergibt:

$$\frac{\partial\ell(\mu, \sigma^2)}{\partial\mu} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma^2}, \quad \frac{\partial\ell(\mu, \sigma^2)}{\partial\sigma^2} = \sum_{i=1}^{n} \left[-\frac{1}{2\,\sigma^2} + \frac{(X_i - \mu)^2}{2\,\sigma^4}\right]$$

Nullsetzen der Ableitungen und Auflösen nach μ und σ^2 ergibt:

$$\begin{aligned}\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n-1}{n} S^2\end{aligned}$$

Der ML-Schätzer von μ ist unverzerrt und effizient. Der ML-Schätzer von σ^2 ist asymptotisch unverzerrt und asymptotisch effizient, seine Verzerrung kann jedoch leicht korrigiert werden. Ist der Mittelwert μ bekannt, ist

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)^2$$

der unverzerrte ML-Schätzer von σ^2. ◄

► **Beispiel 6.3.6.** ML-Schätzung der Parameter der Multinomialverteilung

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ eine Beobachtung aus $\mathrm{Mu}(n, p_1, \ldots, p_k)$, zum Beispiel der Vektor der Gruppeninhalte eines Histogramms. Es sollen die Gruppenwahrscheinlichkeiten $p_1, \ldots, p_k$ geschätzt werden. Die Wahrscheinlichkeitsfunktion lautet:

$$g(n_1, \ldots, n_k \,|\, n, p_1, \ldots, p_k) = \frac{n!}{n_1! \ldots n_k!} \prod_{i=1}^{k} p_i^{n_i}, \ \sum_{i=1}^{k} n_i = n, \ \sum_{i=1}^{k} p_i = 1$$

Die Log-Likelihood-Funktion ist daher:

$$\ell(p_1, \ldots, p_k) = \ln n! - \sum_{i=1}^{k} \ln X_i! + \sum_{i=1}^{k} X_i \ln p_i$$

Die Log-Likelihood-Funktion muss unter der Bedingung $\sum_{i=1}^{k} p_i = 1$ maximiert werden. Dazu wird die sie um den Term $\lambda(\sum_{i=1}^{k} p_i - 1)$ mit dem Lagrangemultiplikator λ ergänzt. Nullsetzen der partiellen Ableitungen nach allen p_i und λ ergibt das lineare Gleichungssystem:

$$\frac{X_i}{p_i} + \lambda = 0, \quad i = 1, \ldots, k \quad \text{und} \quad \sum_{i=1}^{k} p_i = 1$$

Es ist leicht zu verifizieren, dass die Lösung gegeben ist durch:

$$\hat{p}_i = \frac{X_i}{n}, \quad i = 1, \ldots, k \quad \text{und} \quad \lambda = -n$$

Die ML-Schätzer $\hat{p}_i$ sind unverzerrt und effizient. ◀

Der nächste Satz zeigt, dass ML-Schätzer tatsächlich unter sehr allgemeinen Voraussetzungen optimale Eigenschften haben.

☛ **Satz 6.3.7.**

Es sei ϑ_0 der wahre Wert des Parameters ϑ. Existieren die ersten beiden Ableitungen von $\mathcal{L}(\vartheta)$, existiert die Fisher-Information $I_n(\vartheta)$ für alle zulässigen ϑ und ist $E\left[(\ln \mathcal{L})'\right] = 0$, so ist die Größe

$$\sqrt{n}(\hat{\vartheta} - \vartheta_0)$$

asymptotisch normalverteilt mit Erwartungswert 0 und Varianz $1/I_1(\vartheta_0)$. Der ML-Schätzer $\hat{\vartheta}$ ist daher asymptotisch unverzerrt, MSE-konsistent und asymptotisch effizient.

▶ **Beispiel 6.3.8.** Schätzung des Lageparameters der Cauchyverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Cauchyverteilung $\mathrm{T}(1\,|\,\mu, 1)$ mit Lageparameter μ und Skalenparameter 1. Die gemeinsame Dichtefunktion lautet:

$$g(x_1, \ldots, x_n\,|\,\mu) = \prod_{i=1}^{n} \frac{1}{\pi[1 + (x_i - \mu)^2]}$$

Die Log-Likelihood-Funktion ist daher:

$$\ell(\mu) = -n \ln \pi - \sum_{i=1}^{n} \ln[1 + (X_i - \mu)^2]$$

Das Maximum $\hat{\mu}$ von $\ell(\mu)$ muss numerisch gefunden werden. Man kann zeigen, dass die Fisher-Information der Stichprobe gleich $I_n(\mu) = n/2$ ist. Für große Stichproben muss daher die Varianz des ML-Schätzers $\hat{\mu}$ ungefähr gleich $2/n$ sein. Abbildung 6.2 zeigt das Ergebnis der Simulation von $N = 10^4$ Stichproben der Größe $n = 100$. Der Standardfehler des ML-Schätzers ist 0.1435, also sehr nahe an $\sqrt{2/100} = 0.1414$.

Der Stichprobenmedian $\widetilde{X}$ ist ebenfalls ein MSE-konsistenter Schätzer für μ. Seine Varianz ist asymptotisch gleich $\pi^2/(4n) \approx 2.47/n$. Sie ist also um etwa 23% größer als die Varianz des ML-Schätzers (Abbildung 6.2). Die Korrelation zwischen $\widetilde{X}$ und $\hat{\mu}$ ist etwa 90%.

➠MATLAB: `make_ML_cauchy` ◀

Da in der Praxis meist nur eine einzige Stichprobe vorliegt, ist es wichtig, gleichzeitig mit dem ML-Schätzwert auch seinen Standardfehler zu bestimmen. Tatsächlich kann der Standardfehler des ML-Schätzers näherungsweise aus der normierten Likelihood-Funktion einer einzelnen Stichprobe abgelesen werden. Dazu werden die folgenden Schritte durchgeführt:

1. Adjustieren der Log-Likelihood-Funktion und Berechnung der Likelihood-Funktion:

$$\ell'(\mu) = \ell(\mu) - \max(\ell(\mu)), \quad \mathcal{L}(\mu) = \exp[\ell'(\mu)]$$

2. Normieren der Likelihood-Funktion:

$$C = \int \mathcal{L}(\mu)\,\mathrm{d}\mu, \quad \mathcal{L}'(\mu) = \mathcal{L}(\mu)/C$$

3. Berechnung des Standardfehlers:

$$m_1 = \int \mu \cdot \mathcal{L}'(\mu)\,\mathrm{d}\mu, \quad m_2 = \int \mu^2 \cdot \mathcal{L}'(\mu)\,\mathrm{d}\mu, \quad \sigma = \sqrt{m_2 - m_1^2}$$

Falls wie in Beispiel 6.3.8 die Likelihood-Funktion nicht in geschlossener Form vorliegt, werden die Integrale numerisch, zum Beipiel mittels Trapez- oder Simpsonregel ausgewertet. In diesem Beispiel liefert das Verfahren einen Standardfehler von 0.1314 (Abbildung 6.3).

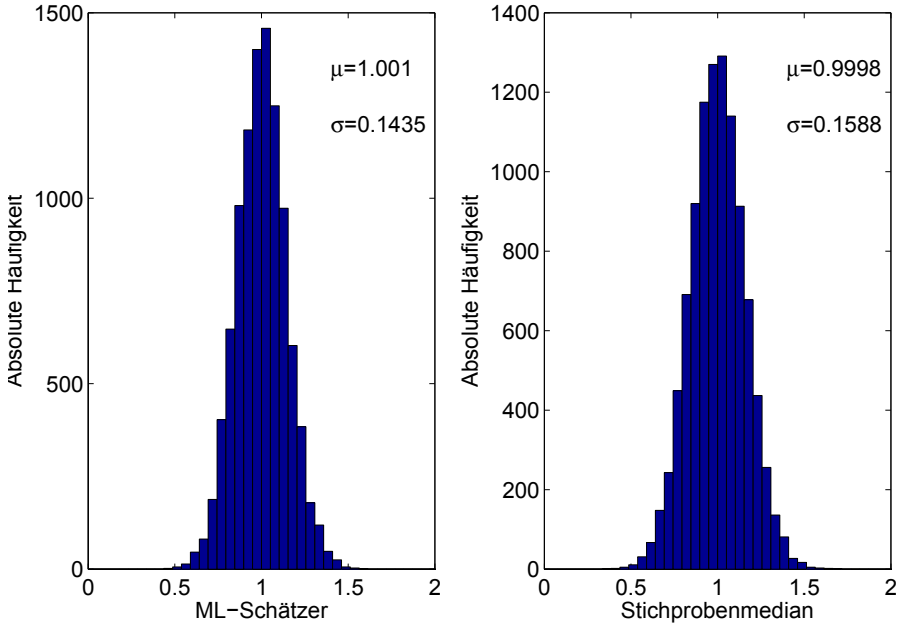

Abbildung 6.2: Verteilung von ML-Schätzer (links) und Stichprobenmedian (rechts). MATLAB: `make_ML_cauchy`

► **Beispiel 6.3.9.** ML-SCHÄTZUNG DER OBERGRENZE DER GLEICHVERTEILUNG

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Gleichverteilung $\mathrm{Un}(0, b)$ mit unbekannter Obergrenze b. Die gemeinsame Dichtefunktion lautet:

$$g(x_1, \ldots, x_n \,|\, b) = \frac{1}{b^n}, 0 \leq x_1, \ldots, x_n \leq b$$

Der größte Wert der Likelihood-Funktion ist daher bei

$$\hat{b} = \max_i X_i$$

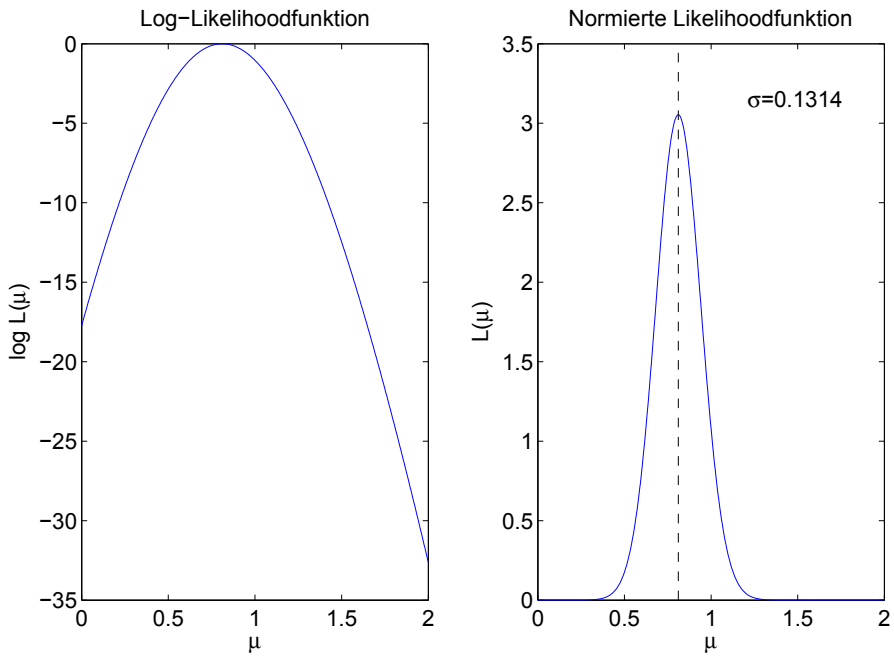

Abbildung 6.3: Die Log-Likelihood-Funktion (links) und die normierte Likelihood-Funktion (rechts) einer Stichprobe. MATLAB: `make_ML_cauchy`

Da ein Randmaximum vorliegt, gelten die üblichen asymptotischen Eigenschaften nicht. Die Verteilungsfunktion F von $\hat{b} = \max_i X_i$ kann folgendermaßen bestimmt werden:

$$F(x) = W(\max_i X_i \le x) = W(X_1 \le x \cap \ldots \cap X_n \le x) = (x/b)^n$$

Die Dichtefunktion von $\hat{b} = \max_i X_i$ lautet daher:

$$f(x) = \frac{n x^{n-1}}{b^n}$$

Daraus können Erwartungswert und Varianz berechnet werden:

$$\mathsf{E}[\hat{b}] = \frac{b\,n}{n+1}, \quad \mathsf{var}[\hat{b}] = \frac{b^2 n}{(n+2)(n+1)^2}$$

Der Schätzer ist asymptotisch unverzerrt, die Varianz konvergiert aber wesentlich rascher gegen 0 als üblich, nämlich wie $1/n^2$. Der Schätzer ist auch nicht asymptotisch normalverteilt. Abbildung 6.4 zeigt das Ergebnis der Simulation von 10000 Stichproben der Größe $n = 25$ bzw. $n = 100$ mit $b = 1$.

➠MATLAB: `make_ML_uniform` ◀

Wird statt des Parameters ϑ eine Funktion $\eta = h(\vartheta)$ des Parameters geschätzt, ist der ML-Schätzer $\hat{\eta}$ von η gleich $h(\hat{\vartheta})$. Es wird angenommen, dass h eindeutig umkehrbar ist, sodass $\mathrm{d}h^{-1}(\eta)/\,\mathrm{d}\eta$ nicht 0 ist. Ist $\mathcal{L}_\vartheta(\vartheta)$ die Likelihood-Funktion von ϑ und

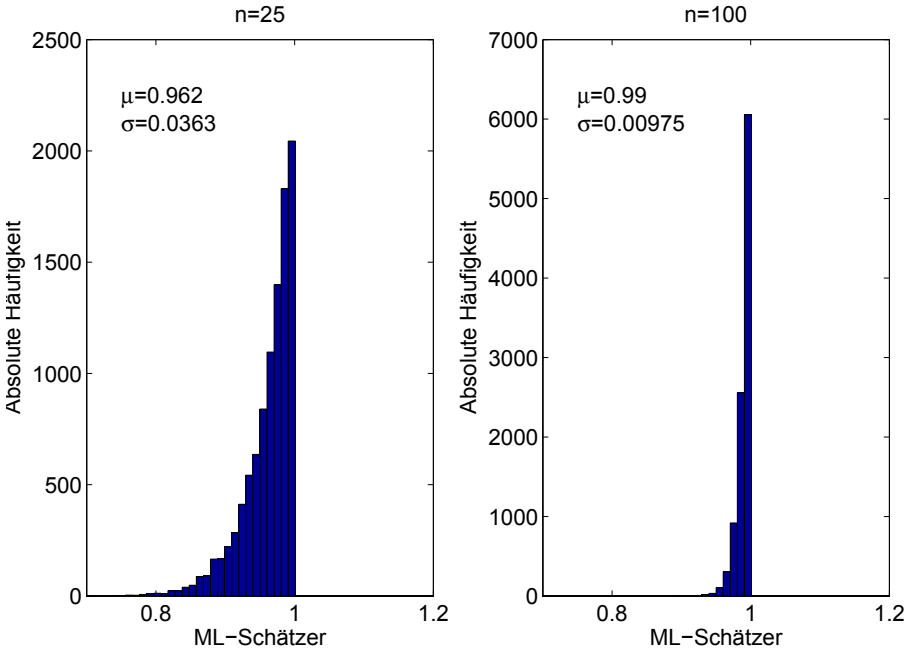

Abbildung 6.4: Verteilung des ML-Schätzers aus Stichproben der Größe $n = 25$ (links) bzw. $n = 100$ (rechts). MATLAB: `make_ML_uniform`

$\mathcal{L}_\eta(\eta) = \mathcal{L}_\vartheta(h^{-1}(\eta))$ die Likelihood-Funktion von η, gilt für die Ableitung $\mathcal{L}'_\eta(\eta)$ nach der Kettenregel:

$$\mathcal{L}'_\eta(\eta) = \mathcal{L}'_\vartheta(h^{-1}(\eta)) \cdot \frac{\mathrm{d}h^{-1}(\eta)}{\mathrm{d}\eta}$$

Ist $\hat{\eta}$ eine Nullstelle von $\mathcal{L}_\eta$, dann muss $h^{-1}(\hat{\eta})$ eine Nullstelle von $\mathcal{L}_\vartheta$ sein. Der ML-Schätzer $\hat{\vartheta}$ ist also gleich $h^{-1}(\hat{\eta})$, oder $\hat{\eta} = h(\hat{\vartheta})$.

▶ **Beispiel 6.3.10.** ML-SCHÄTZUNG DER MITTLEREN RATE

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$. Aus der Stichprobe soll die mittlere Rate $\lambda = 1/\tau$ geschätzt werden. Der ML-Schätzer von τ ist $\hat{\tau} = \overline{X}$. Folglich ist der ML-Schätzer von $\lambda = 1/\tau$ gleich:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i}$$

Der ML-Schätzer von λ ist im Gegenteil zu $\hat{\tau}$ nicht unverzerrt. Es ist mittels der Jensenschen Ungleichung[a] leicht zu zeigen, dass $\mathsf{E}[\hat{\lambda}] > \lambda$. ◀

▶ **Beispiel 6.3.11.** ML-SCHÄTZUNG DER MITTLEREN WARTEZEIT

Es seien $X_1, \ldots, X_n$ Beobachtungen aus einem Poissonprozess mit Rate λ. Es soll die mittlere Wartezeit $\tau = 1/\lambda$ geschätzt werden. Der ML-Schätzer von λ ist $\hat{\lambda} = \overline{X}$. Folglich ist der ML-Schätzer von $\tau = 1/\lambda$ gleich:

$$\hat{\tau} = \frac{n}{\sum_{i=1}^n X_i}$$

Der ML-Schätzer von τ ist im Gegenteil zu $\hat{\lambda}$ nicht unverzerrt. Es gilt analog zum vorigen Beispiel $\mathsf{E}[\hat{\tau}] > \tau$. ◀

[a] http://de.wikipedia.org/wiki/Jensensche_Ungleichung

6.3.2 Momentenschätzer

Der ML-Schätzer hat im Allgemeinen asymptotisch optimale Eigenschaften, ist aber in der Praxis häufig nur durch numerische Maximierung der Log-Likelihood-Funktion zu bestimmen. Als leicht zu berechnende Alternative bietet sich ein *Momentenschätzer* an. Dazu werden die zu schätzenden Parameter als Funktionen von Verteilungsmomenten dargestellt und dann die Verteilungsmomente durch die entsprechenden Stichprobenmomente ersetzt. Das Verfahren wird im Folgenden an einigen Beispielen illustriert.

6.3.2.1 Momentenschätzer für die hypergeometrische Verteilung

Es sei $X \sim \mathrm{Hy}(N, M, n)$ (siehe Abschnit 3.2.2). N ist die Größe der Grundgesamtheit, M ist die Anzahl der Merkmalsträger in der Grundgesamtheit, und X ist die Anzahl der Merkmalsträger in einer zufälligen Stichprobe der Größe n. Ferner gilt:

$$\mathsf{E}[X] = \frac{nM}{N} \quad \text{und} \quad \mathsf{var}[X] = \frac{nM(N-n)(N-M)}{N^2(N-1)}$$

Es soll zunächst der Fall behandelt werden, dass N bekannt und M unbekannt ist. Es sei m der beobachtete Wert von X. Der Erwartungswert $\mathsf{E}[X]$ wird durch die Beobachtung ersetzt und ergibt so den Momentenschätzer von M:

$$\check{M} = N \cdot \frac{m}{n}$$

Der Erwartungswert von $\check{M}$ ist dann gleich:

$$\mathsf{E}[\check{M}] = N \cdot \frac{\mathsf{E}[X]}{n} = M$$

Der Schätzer $\check{M}$ ist also unverzerrt. Seine Varianz ist gleich:

$$\mathsf{var}[\check{M}] = N^2 \cdot \frac{\mathsf{var}[X]}{n^2} = \frac{M(N-n)(N-M)}{n(N-1)}$$

Da der wahre Wert von M nicht bekannt ist, wird er in der Praxis durch den Schätzwert ersetzt:

$$\mathsf{var}[\check{M}] \approx N^2 \cdot \frac{\mathsf{var}[X]}{n^2} = \frac{\check{M}(N-n)(N-\check{M})}{n(N-1)}$$

Man kann zeigen, dass der Momentenschätzer $\check{M}$ auch der ML-Schätzer ist.

▶ **Beispiel 6.3.12.** Schätzung der Anzahl der Merkmalsträger

Aus einer Lieferung von $N = 1000$ Widerständen wird ohne Zurücklegen eine Stichprobe vom Umfang $n = 50$ gezogen. Die Stichprobe enthält $m = 4$ defekte Stücke. Der Schätzwert von M und sein Standardfehler ergibt sich zu:

$$\begin{aligned} \check{M} &= \frac{1000 \cdot 4}{50} = 80 \\ \mathsf{var}[\check{M}] &\approx \frac{80 \cdot 950 \cdot 920}{50 \cdot 999} = 1399.80 \\ \sigma[\check{M}] &\approx 37.4 \end{aligned}$$

Eine Faustregel besagt, dass der wahre Wert von M mit einer Sicherheit von ca. 95% zwischen $\check{M} - 2\sigma[\check{M}]$ und $\check{M} + 2\sigma[\check{M}]$ liegt (siehe auch Abschnitt 6.4). ◀

Ist M bekannt und die Größe der Grundgesamtheit N unbekannt, kann eine Beobachtung von X dazu verwendet werden, um N zu schätzen. Sei m der beobachtete Wert von X. Der Erwartungswert $\mathsf{E}[X]$ wird wieder durch die Beobachtung ersetzt und ergibt so den Momentenschätzer von N:

$$\check{N} = M \cdot \frac{n}{m}$$

Der Erwartungswert von $\check{N}$ ist dann *ungefähr* gleich:

$$\mathsf{E}[\check{N}] \approx M \cdot \frac{n}{\mathsf{E}[X]} = N$$

Der Schätzer $\check{N}$ ist aber nicht exakt erwartungstreu. Die Varianz von $\check{N}$ ist *ungefähr* gleich:

$$\mathsf{var}[\check{N}] \approx \frac{N^2(N-n)(N-M)}{nM(N-1)}$$

Da der wahre Wert von N nicht bekannt ist, wird er in der Praxis durch den Schätzwert ersetzt:

$$\mathsf{var}[\check{N}] \approx \frac{\check{N}^2(\check{N}-n)(\check{N}-M)}{nM(\check{N}-1)}$$

► **Beispiel 6.3.13.** Schätzung der Grösse der Grundgesamtheit

In einem Teich soll die Größe des Fischbestandes geschätzt werden. Zunächst werden $M = 200$ Fische gefangen, markiert, und wieder freigesetzt. Nach einiger Zeit werden $n = 100$ Fische gefangen, von denen $m = 21$ eine Markierung aufweisen. Der Schätzwert von N und sein Standardfehler ergibt sich zu:

$$\check{N} = \frac{200 \cdot 100}{21} \approx 952$$
$$\mathsf{var}[\check{N}] \approx \frac{952^2 \cdot 852 \cdot 752}{100 \cdot 200 \cdot 951} \approx 30530$$
$$\sigma[\check{N}] \approx 175$$

Die Qualität der Schätzung kann durch ein Simulationsexperiment überprüft werden. Abbildung 6.5 zeigt die Häufigkeitsverteilung von $\check{N}$ für 10000 Experimente mit $N = 952, M = 200, n = 100$. Der Mittelwert der Verteilung ist ca. 983. Die Größe des Fischbestandes wird also etwas zu groß geschätzt.

➠MATLAB: `make_hypgeom_estN` ◀

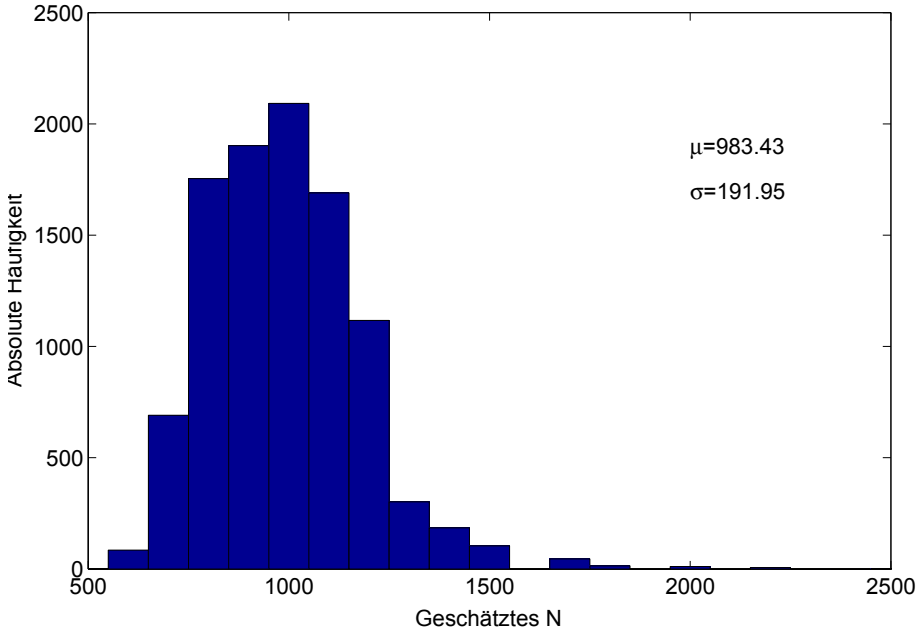

Abbildung 6.5: Häufigkeitsverteilung von $\check{N}$ für 10000 Experimente mit $N = 952, M = 200, n = 100$. MATLAB: `make_hypgeom_estN`

6.3.2.2 Momentenschätzer für die Gammaverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Gammaverteilung $\mathrm{Ga}(a, b)$ mit Formparameter a und Skalenparameter b. Es gilt:

$$\mathsf{E}[X] = ab, \quad \mathsf{var}[X] = ab^2$$

Auflösen dieser beiden Gleichungen nach a und b ergibt:

$$a = \mathsf{E}[X]^2/\mathsf{var}[X], \quad b = \mathsf{var}[X]/\mathsf{E}[X]$$

Werden der Erwartungswert und die Varianz durch das Stichprobenmittel und die Stichprobenvarianz ersetzt, erhält man die Momentenschätzer:

$$\check{a} = \overline{X}^2/S^2, \quad \check{b} = S^2/\overline{X}$$

► **Beispiel 6.3.14.** Vergleich mit ML-Schätzer

Da die ML-Schätzer asymptotisch effizient sind, ist zu erwarten, dass die Momentenschätzer eine größere Streuung als die ML-Schätzer haben. Ist zum Beispiel $a = 3$, $b = 2$ und $n = 500$, kann man die Erwartung und Standardfehler von $\check{a}$ und $\check{b}$ leicht durch Simulation von $N = 10^5$ Stichproben abschätzen und erhält:

$$\mathsf{E}[\check{a}] \approx 3.018, \quad \mathsf{E}[\check{b}] = 1.999, \quad \sigma[\check{a}] = 0.220, \quad \sigma[\check{b}] = 0.155$$

Die Momentenschätzer sind also nur geringfügig verzerrt, haben aber deutlich größere Standardfehler als die ML-Schätzer. Unter Benützung der aymptotischen Kovarianzmatrix erhält man nämlich:

$$\sigma[\hat{a}] = 0.180, \quad \sigma[\hat{b}] = 0.131$$

➠Matlab: `make_MS_gamma` ◄

6.3.2.3 Momentenschätzer für die Rayleighverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Rayleighverteilung mit Skalenparameter ϑ (siehe auch Beispiel 4.3.7). Ihre Dichtefunktion lautet:

$$f(x\,|\,\vartheta) = \frac{x}{\vartheta^2} \exp\left[-\frac{x^2}{2\vartheta^2}\right]$$

Weiters gilt:

$$\mathsf{E}[X] = \vartheta\sqrt{\pi/2}, \quad \mathsf{var}[X] = \vartheta^2\frac{4-\pi}{2}, \quad \mathsf{E}[X^2] = \mathsf{var}[X] + \mathsf{E}[X]^2 = 2\vartheta^2$$

Daraus können zwei Momentenschätzer konstruiert werden, indem $\mathsf{E}[X]$ durch den Mittelwert der Beobachtungen und $\mathsf{E}[X^2]$ durch den Mittelwert der quadrierten Beobachtungen ersetzt wird:

$$\check{\vartheta}_1 = \overline{X}\sqrt{2/\pi}, \quad \check{\vartheta}_2 = \sqrt{\frac{\sum X_i^2}{2n}}$$

► **Beispiel 6.3.15.** Vergleich mit ML-Schätzer

Der ML-Schätzer kann in geschlossener Form berechnet werden und ist identisch mit $\check{\vartheta}_2$. Die Simulation von $N = 10^5$ Stichproben der Größe $n = 100$ mit $\vartheta = 2$ ergibt:

$$\mathsf{E}[\check{\vartheta}_1] = 2.000, \quad \sigma[\check{\vartheta}_1] = 0.105$$
$$\mathsf{E}[\check{\vartheta}_2] = 1.998, \quad \sigma[\check{\vartheta}_2] = 0.100$$

$\check{\vartheta}_1$ ist unverzerrt, hat aber einen etwas größeren Standardfehler als $\check{\vartheta}_2$, das nur asymptotisch unverzerrt ist. ϑ^2 kann durch den ML-Schätzer unverzerrt und effizient geschätzt werden, wobei der ML-Schätzer auch der Momentenschätzer ist:

$$\hat{\vartheta}^2 = \frac{\sum X_i^2}{2n}$$

➠Matlab: `make_MS_rayleigh` ◄

6.4 Konfidenzintervalle

6.4.1 Grundbegriffe

Neben dem Schätzwert selbst ist auch seine Streuung um den wahren Wert von Interesse. Zu diesem Zweck soll aus einer Stichprobe ein Intervall bestimmt werden, das den wahren Wert mit einer gewissen Sicherheit enthält.

☞ **Definition 6.4.1.** KONFIDENZINTERVALL

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit dem unbekannten Parameter ϑ. Ein Intervall mit den Grenzen $G_1 = g_1(X_1, \ldots, X_n)$ und $G_2 = g_2(X_1, \ldots, X_n)$ heißt ein Konfidenzintervall mit Sicherheit $1 - \alpha$, wenn gilt:

$$W(G_1 \leq G_2) = 1$$
$$W(G_1 \leq \vartheta \leq G_2) = 1 - \alpha$$

Ein solches Intervall wird kurz als $(1 - \alpha)$-Konfidenzintervall bezeichnet.

Zu jedem Wert der Sicherheit $1 - \alpha$ gibt es viele verschiedene Konfidenzintervalle. Ist F stetig, gibt es unendlich viele Konfidenzintervalle mit Sicherheit $1 - \alpha$. Ein symmetrisches Konfidenzintervall liegt vor, wenn gilt:

$$W(\vartheta \leq G_1) = W(\vartheta \geq G_2)$$

Ein links- bzw. rechtsseitiges Konfidenzintervall liegt vor, wenn gilt:

$$W(\vartheta \leq G_2) = 1 - \alpha \qquad \text{bzw.} \qquad W(G_1 \leq \vartheta) = 1 - \alpha$$

6.4.2 Allgemeine Konstruktion nach Neyman

Es sei $Y = h(X_1, \ldots, X_n)$ eine Stichprobenfunktion. Die Verteilung von Y hängt dann ebenfalls vom unbekannten Parameter ϑ ab. Für jeden Wert von ϑ kann ein *Prognoseintervall* $[y_1(\vartheta), y_2(\vartheta)]$ mit Wahrscheinlichkeitsinhalt $1 - \alpha$ bestimmt werden:

$$W(y_1(\vartheta) \leq Y \leq y_2(\vartheta)) \geq 1 - \alpha$$

Ist die Beobachtung gleich $Y = y_0$, so ist das Konfidenzintervall $[G_1(y_0), G_2(y_0)]$ gegeben durch:

$$G_1(y_0) = \min_{\vartheta} \{\vartheta \,|\, y_1(\vartheta) \leq y_0 \leq y_2(\vartheta)\}$$
$$G_2(y_0) = \max_{\vartheta} \{\vartheta \,|\, y_1(\vartheta) \leq y_0 \leq y_2(\vartheta)\}$$

G_1 ist also der kleinste Wert des Parameters ϑ, unter dem das Prognoseintervall die Beobachtung enthält, und G_2 ist der größte derartige Wert von ϑ (siehe Abbildung 6.6).

► **Beispiel 6.4.2.**

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus $\mathrm{No}(0, \sigma^2)$ mit unbekannter Varianz σ. Dann ist $(n-1)S^2/\sigma^2$ χ^2-verteilt mit $n - 1$ Freiheitsgraden. Für Varianz σ^2 und $Y = S^2$ gilt daher:

$$W\left(\chi^2_{\alpha/2|n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2|n-1}\right) = 1 - \alpha$$

Dies kann umgeformt werden zu:

$$W\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2|n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2|n-1}}\right) = 1-\alpha$$

Daraus folgt:

$$G_1(S^2) = \frac{(n-1)S^2}{\chi^2_{1-\alpha/2|n-1}}, \quad G_2(S^2) = \frac{(n-1)S^2}{\chi^2_{\alpha/2|n-1}}$$

Eine Veranschaulichung des Verfahrens ist in Abbildung 6.6 zu sehen. ◀

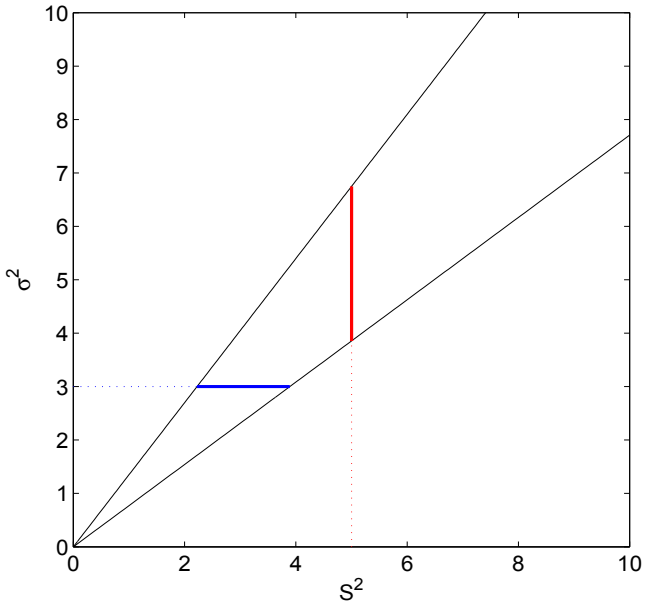

Abbildung 6.6: Prognoseintervall für die Stichprobenvarianz S^2 unter der Annahme $\sigma^2 = 3$ (blau); Konfidenzintervall für σ^2 mit der Beobachtung $S^2 = 5$ (rot).
MATLAB: `make_confidence_belt`

6.4.3 Binomialverteilung

Es sei k eine Beobachtung der Zufallsvariablen Y aus der Binomialverteilung $\mathrm{Bi}(n,p)$. Gesucht ist ein Konfidenzintervall für p. Je nach Konstruktion des Prognoseintervalls

$[y_1(p), y_2(p)]$ ergeben sich verschiedene Konfidenzintervalle.

6.4.3.1 Konfidenzintervall nach Clopper und Pearson

$y_1(p), y_2(p)$ sind die Quantile der Binomialverteilung $\mathrm{Bi}(n,p)$:

$$y_1(p) = \max_k \sum_{i=0}^{k} W(k|n,p) \leq \alpha/2, \quad y_2(p) = \min_k \sum_{i=k}^{n} W(k|n,p) \leq \alpha/2$$

Für die praktische Berechnung des Konfidenzintervalls können die Quantile der Betaverteilung benützt werden:

$$G_1(k) = \beta_{\alpha/2|k,n-k+1}, \quad G_2(k) = \beta_{1-\alpha/2|k+1,n-k}$$

Dieses Intervall ist *konservativ* in dem Sinn, dass die Sicherheit praktisch immer größer als $1-\alpha$ ist.

Soll eine obere Grenze für p angegeben werden, wird ein linksseitiges Konfidenzintervall konstruiert:

$$G_1(k) = 0, \quad G_2(k) = \beta_{1-\alpha|k+1,n-k}$$

Soll eine untere Grenze für p angegeben werden, wird ein rechtsseitiges Konfidenzintervall konstruiert:

$$G_1(k) = \beta_{\alpha|k+1,n-k}, \quad G_2(k) = 1$$

6.4.3.2 Näherung durch Normalverteilung

Für genügend großes n ist $\hat{p} = Y/n$ annähernd normalverteilt gemäß $\mathrm{No}(p, p(1-p)/n)$. Daraus folgt $\mathsf{E}[\hat{p}] = p$ und $\sigma[\hat{p}] = \sqrt{p(1-p)/n}$. Das Standardscore

$$Z = \frac{\hat{p} - p}{\sigma[\hat{p}]}$$

ist dann annähernd standardnormalverteilt. Aus

$$W(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1 - \alpha$$

folgt:

$$W(\hat{p} - z_{1-\alpha/2}\sigma[\hat{p}] \le p \le \hat{p} + z_{1-\alpha/2}\sigma[\hat{p}]) = 1 - \alpha$$

und damit:

$$G_1(k) = \hat{p} - z_{1-\alpha/2}\sigma[\hat{p}], \quad G_2(k) = \hat{p} + z_{1-\alpha/2}\sigma[\hat{p}]$$

Da p nicht bekannt ist, muss $\sigma[\hat{p}]$ näherungsweise bestimmt werden. Dazu gibt es zwei Möglichkeiten: Im sogenannten *Bootstrap-Verfahren* wird p durch $\hat{p}$ ersetzt; im sogenannten *robusten Verfahren* wird p so gewählt, dass $\sigma[\hat{p}]$ maximal ist, also $p = 0.5$.

6.4.3.3 Korrektur nach Agresti und Coull

Das Konfidenzintervall nach dem Bootstrap-Verfahren kann eine kleinere Sicherheit als $1-\alpha$ haben, besonders dann, wenn p nahe bei 0 oder 1 ist. Wenn $k = 0$ oder $k = n$, schrumpft es überhaupt auf einen Punkt zusammen. Eine Verbesserung wird durch die Definition:

$$\hat{p} = \frac{k + c^2/2}{n + c^2}$$

erzielt, wo $c = z_{1-\alpha/2}$ ein Quantil der Standardnormalverteilung ist. Für ein 95%-iges Konfidenzintervall ist $\alpha = 0.05$ und $z_{0.975} = 1.96 \approx 2$. Daraus ergibt sich die Formel:

$$\hat{p} = \frac{k+2}{n+4}$$

Die weitere Konstruktion verläuft wie im Bootstrap-Verfahren.

► **Beispiel 6.4.3.** Vergleich der Konfidenzintervalle

Bei einer Umfrage unter $n = 400$ Personen geben $k = 157$ Personen an, Produkt P zu kennen. Gesucht ist ein 95%-Konfidenzintervall für den Bekanntheitsgrad p. Das Konfidenzintervall nach Clopper und Pearson lautet:

$$G_1(k) = \beta_{0.025|157,244} = 0.3443, \quad G_2(k) = \beta_{0.975|158,243} = 0.4423$$

Die Näherung durch Normalverteilung benützt $\hat{p} = 0.3925$ und $z_{0.975} = 1.96$. Mit dem Bootstrap-Verfahren ergibt sich $\sigma[\hat{p}] = 0.0244$. Die Grenzen des Konfidenzintervalls sind daher

$$G_1 = 0.3925 - 1.96 \cdot 0.0244 = 0.3446, \quad G_2 = 0.3925 + 1.96 \cdot 0.0244 = 0.4404$$

Mit dem robusten Verfahren ergibt sich $\sigma[\hat{p}] = 0.025$ und die Grenzen

$$G_1 = 0.3925 - 1.96 \cdot 0.025 = 0.3435, \quad G_2 = 0.3925 + 1.96 \cdot 0.025 = 0.4415$$

Das robuste Intervall ist nur unwesentlich länger als das Bootstrap-Intervall.

Mit der Korrektur von Agresti-Coull ergibt sich $\hat{p} = 0.3936$. Die Grenzen des Konfidenzintervalls sind dann

$$G_1 = 0.3936 - 1.96 \cdot 0.0244 = 0.3457, \quad G_2 = 0.3936 + 1.96 \cdot 0.0244 = 0.4414$$

Abbildung 6.7 zeigt die tatsächliche Sicherheit von verschiedenen Konfidenzintervallen. Die nominelle Sicherheit ist 95%, die Stichprobengröße ist $n = 100$.

➠Matlab: `make_KI_binomial` ◄

► **Beispiel 6.4.4.** Konfidenzintervall für das Verzweigungsverhältnis

In einer Stichprobe von $n = 1000$ Zerfällen eines instabilen Teilchens wird ein gewisser Zerfallskanal nicht beobachtet. Gesucht ist ein rechtsseitiges 95%-iges Konfidenzintervall für das Verzzeigungsverhältnis dieses Kanals. Das Intervall nach Clopper und Pearson ist dann:

$$G_1(k) = 0, \quad G_2(0) = \beta_{0.95|1,n} = 0.003$$

Das Verzweigungsverhältnis ist mit 95%-iger Sicherheit nicht größer als 0.3%. ◄

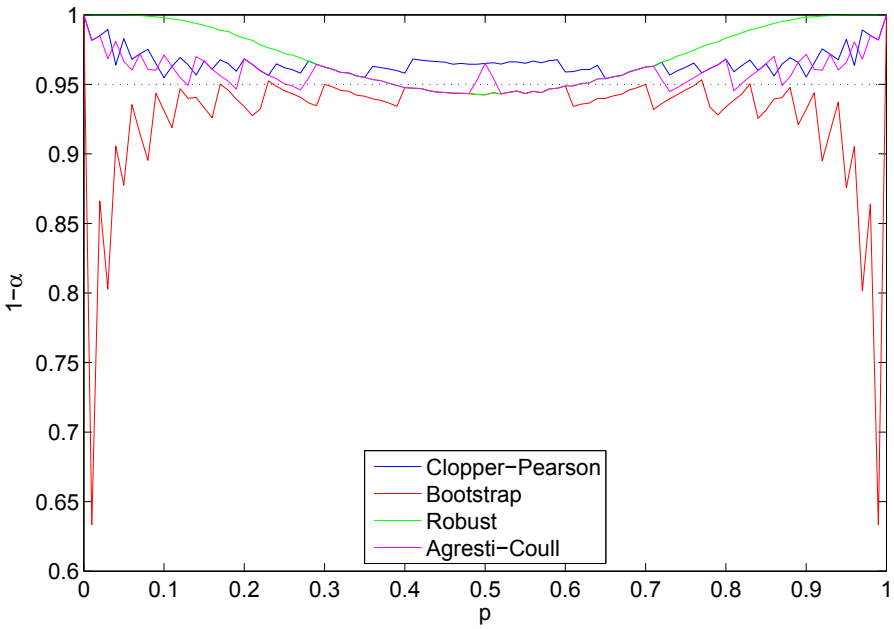

Abbildung 6.7: Sicherheit von verschiedenen Konfidenzintervallen für p. Die nominelle Sicherheit ist 95%, die Stichprobengröße ist $n = 100$. MATLAB: `make_KI_binomial`

6.4.4 Mittelwert der Poissonverteilung

Es sei k eine Beobachtung der Zufallsvariablen Y aus der Poissonverteilung $\mathrm{Po}(\lambda)$. Gesucht ist ein Konfidenzintervall für λ. Je nach Konstruktion des Prognoseintervalls $[y_1(\lambda), y_2(\lambda)]$ ergeben sich verschiedene Konfidenzintervalle.

6.4.4.1 Symmetrisches Konfidenzintervall

$y_1(\lambda), y_2(\lambda)$ sind die Quantile der Poissonverteilung $\mathrm{Po}(\lambda)$:

$$y_1(\lambda) = \max_k \sum_{i=0}^{k} W(k\,|\,\lambda) \le \alpha/2, \quad y_2(\lambda) = \min_k \sum_{i=k}^{\infty} W(k\,|\,\lambda) \le \alpha/2$$

Für die praktische Berechnung des Konfidenzintervalls können die Quantile der Gammaverteilung benützt werden (siehe die Tabelle in Anhang T.3):

$$G_1(k) = \gamma_{\alpha/2|k,1}, \quad G_2(k) = \gamma_{1-\alpha/2|k+1,1}$$

Dieses Intervall ist *konservativ* in dem Sinn, dass die Sicherheit praktisch immer größer als $1 - \alpha$ ist.

Liegen n Beobachtungen $k_1, \ldots, k_n$ vor, so ist $K = \sum k_i$ poissonverteilt mit Mittelwert $n\lambda$. Das symmetrische Konfidenzintervall für λ ist daher:

$$G_1(K) = \gamma_{\alpha/2|K,1/n}, \quad G_2(K) = \gamma_{1-\alpha/2|K+1,1/n}$$

Abbildung 6.8 zeigt die tatsächliche Sicherheit des symmetrischen Konfidenzintervalls mit nominell 95% Sicherheit für $n = 1, 5, 25$.

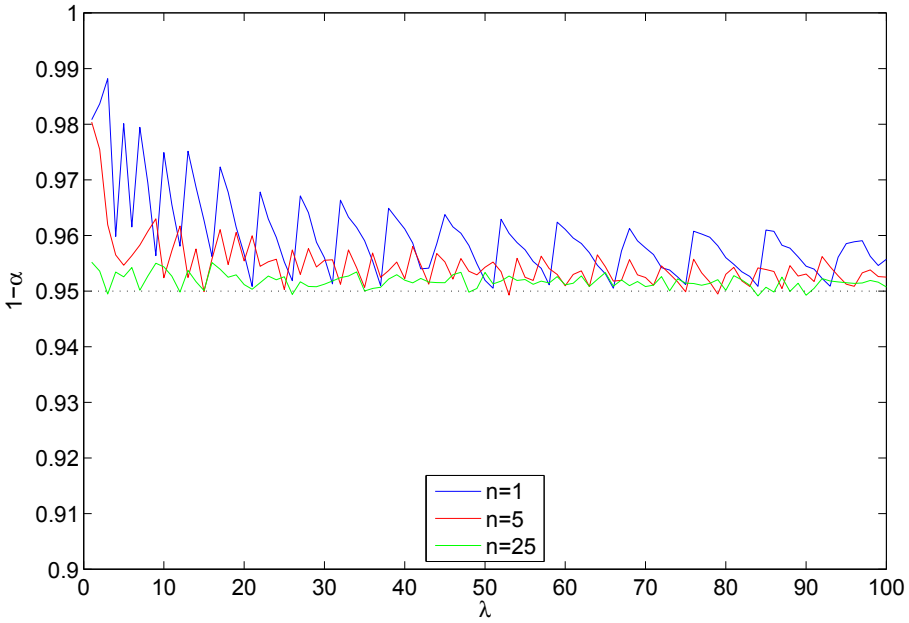

Abbildung 6.8: Sicherheit des symmetrischen Konfidenzintervalls für λ. Die nominelle Sicherheit ist 95%. MATLAB: `KI_Poisson_symmetrisch`

6.4.4.2 Einseitige Konfidenzintervalle

Soll eine obere Grenze für λ angegeben werden, wird ein linksseitiges Intervall konstruiert. Ist K wieder die Summe der n Beobachtungen, lautet das Intervall:

$$G_1(k) = 0, \quad G_2(k) = \gamma_{1-\alpha|K+1,1/n}$$

Soll eine untere Grenze für λ angegeben werden, wird ein rechtsseitiges Intervall konstruiert:

$$G_1(k) = \gamma_{\alpha|K,1/n}, \quad G_2(k) = \infty,$$

Abbildungen 6.9 bzw. 6.10 zeigen die tatsächliche Sicherheit des links- bzw. rechtsseitigen Konfidenzintervalls mit nominell 95% Sicherheit für $n = 1, 5, 25$.

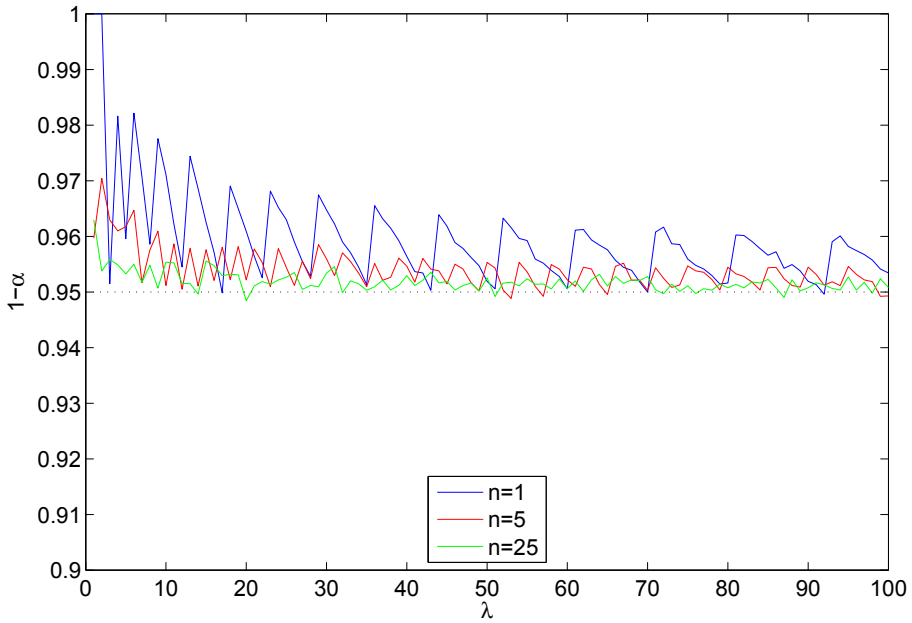

Abbildung 6.9: Sicherheit des linksseitigen Konfidenzintervalls für λ. Die nominelle Sicherheit ist 95%. MATLAB: `KI_Poisson_linksseitig`

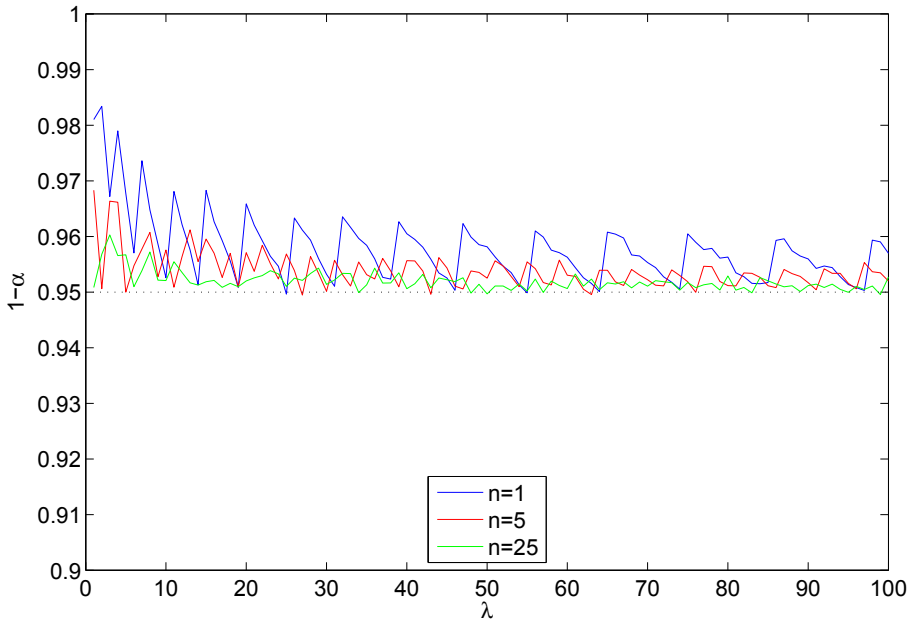

Abbildung 6.10: Sicherheit des rechtsseitigen Konfidenzintervalls für λ. Die nominelle Sicherheit ist 95%. MATLAB: `KI_Poisson_rechtsseitig`

6.4.5 Mittelwert der Exponentialverteilung

6.4.5.1 Symmetrisches Konfidenzintervall

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$. Das Stichprobenmittel $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ hat die folgende Dichtefunktion:

$$f(x) = \frac{x^{n-1}}{(\tau/n)^n \Gamma(n)} \exp\left(-\frac{x}{\tau/n}\right)$$

$\overline{X}$ ist also gammaverteilt gemäß $\mathrm{Ga}(n, \tau/n)$. Daher gilt:

$$W\left(\gamma_{\alpha/2|n,\tau/n} \le \overline{X} \le \gamma_{1-\alpha/2|n,\tau/n}\right) = 1-\alpha$$

Multiplikation der Ungleichungen mit n/τ ergibt:

$$W\left(\gamma_{\alpha/2|n,1} \le \frac{n\overline{X}}{\tau} \le \gamma_{1-\alpha/2|n,1}\right) = 1-\alpha$$

und daraus:

$$W\left(\frac{n\overline{X}}{\gamma_{1-\alpha/2|n,1}} \le \tau \le \frac{n\overline{X}}{\gamma_{\alpha/2|n,1}}\right) = 1-\alpha$$

Damit sind die Grenzen des Konfidenzintervalls gegeben durch:

$$G_1(\overline{X}) = \frac{n\overline{X}}{\gamma_{1-\alpha/2|n,1}}, \quad G_2(\overline{X}) = \frac{n\overline{X}}{\gamma_{\alpha/2|n,1}}$$

Die Quantile $\gamma_{\alpha|n,1}$ sind in Anhang T.3 tabelliert.

Das Konfidenzintervall für den unbekannten Mittelwert τ kann leicht in ein Konfidenzintervall für die mittlere Rate $\lambda = 1/\tau$ umgerechnet werden:

$$W\left(\frac{n\overline{X}}{\gamma_{1-\alpha/2|n,1}} \leq \tau \leq \frac{n\overline{X}}{\gamma_{\alpha/2|n,1}}\right) = 1-\alpha \implies W\left(\frac{\gamma_{\alpha/2|n,1}}{n\overline{X}} \leq \lambda \leq \frac{\gamma_{1-\alpha/2|n,1}}{n\overline{X}}\right) = 1-\alpha$$

6.4.5.2 Einseitige Konfidenzintervalle für den Mittelwert

Weiters gilt:

$$W\left(\gamma_{\alpha|n,\tau/n} \leq \overline{X}\right) = 1-\alpha \implies W\left(\gamma_{\alpha|n,1} \leq \frac{n\overline{X}}{\tau}\right) = 1-\alpha$$

Damit ergibt sich das linksseitige Konfidenzintervall:

$$G_1(\overline{X}) = 0, \quad G_2(\overline{X}) = \frac{n\overline{X}}{\gamma_{\alpha|n,1}}$$

und analog das rechsseitige Konfidenzintervall:

$$G_1(\overline{X}) = \frac{n\overline{X}}{\gamma_{1-\alpha|n,1}}, \quad G_2(\overline{X}) = \infty$$

► **Beispiel 6.4.5.** Konfidenzintervall für die mittlere Lebensdauer

Die Beobachtung der Lebensdauer von freien Neutronen ergibt eine Stichprobe vom Umfang $n = 1000$, mit dem Stichprobenmittel $\bar{x} = 889.15\,\mathrm{s}$. Das symmetrische 95%-ige Konfidenzintervall ist gleich:

$$G_1(\overline{X}) = 836.52, \quad G_2(\overline{X}) = 946.94$$

Das linksseitige Konfidenzintervall ist gleich:

$$G_1(\overline{X}) = 0, \quad G_2(\overline{X}) = 937.37$$

Das rechsseitige Konfidenzintervall ist gleich:

$$G_1(\overline{X}) = 844.74, \quad G_2(\overline{X}) = \infty$$

Der wahre Wert ist bekanntlich gleich $\tau = 881.5\,\mathrm{s}$. ◄

6.4.6 Normalverteilung

6.4.6.1 Mittelwert der Normalverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$. $\overline{X}$ ist normalverteilt gemäß $\mathrm{No}(\mu, \sigma^2/n)$. Ist σ^2 bekannt, ist das Standardscore

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

standardnormalverteilt. Aus

$$W(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1 - \alpha$$

folgt

$$W(\overline{X} - z_{1-\alpha/2}\sigma/\sqrt{n} \le \mu \le \overline{X} + z_{1-\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha$$

und damit:

$$G_1(\overline{X}) = \overline{X} - z_{1-\alpha/2}\sigma/\sqrt{n}\,, \quad G_2(\overline{X}) = \overline{X} + z_{1-\alpha/2}\sigma/\sqrt{n}$$

Ist σ^2 unbekannt, wird σ^2 durch die Stichprobenvarianz geschätzt, und das Standardscore

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

ist T-verteilt mit $n - 1$ Freiheitsgraden. Aus

$$W(-t_{1-\alpha/2|n-1} \le T \le t_{1-\alpha/2|n-1}) = 1 - \alpha$$

folgt

$$W(\overline{X} - t_{1-\alpha/2|n-1}S/\sqrt{n} \le \mu \le \overline{X} + t_{1-\alpha/2|n-1}S/\sqrt{n}) = 1 - \alpha$$

und damit:

$$G_1(\overline{X}, S^2) = \overline{X} - z_{1-\alpha/2}S/\sqrt{n}\,, \quad G_2(\overline{X}, S^2) = \overline{X} + z_{1-\alpha/2}S/\sqrt{n}$$

Die Quantile z_α und $t_{\alpha|n}$ können aus den Tabellen in Anhang T.2 bzw. T.5 abgelesen werden.

► **Beispiel 6.4.6.**
Eine Stichprobe vom Umfang $n = 50$ aus der Standardnormalverteilung hat das Stichprobenmittel $\overline{X} = 0.054$ und die Stichprobenvarianz $S^2 = 1.0987$. Wird die Varianz als bekannt vorausgesetzt, lautet das symmetrische 95%-Konfidenzintervall für μ:

$$G_1 = 0.054 - 1.96/\sqrt{50} = -0.2232, \quad G_2 = 0.054 + 1.96/\sqrt{50} = 0.3312$$

Wird die Varianz als unbekannt angenommen, lautet das symmetrische 95%-Konfidenzintervall für μ:

$$G_1 = 0.054 - 2.01 \cdot 1.048/\sqrt{50} = -0.2439, \quad G_2 = 0.054 + 2.01 \cdot 1.048/\sqrt{50} = 0.3519$$

Durch die zusätzliche Unsicherheit der Varianz wird das Konfidenzintervall etwas länger.

➠MATLAB: `make_KI_normal` ◄

6.4.6.2 Varianz der Normalverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$. $(n-1)S^2/\sigma^2$ ist χ^2-verteilt mit $n-1$ Freiheitsgraden. Aus

$$W\left(\chi^2_{\alpha/2|n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2|n-1}\right) = 1 - \alpha$$

folgt

$$W\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2|n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2|n-1}}\right) = 1 - \alpha$$

und damit:

$$G_1(S^2) = \frac{(n-1)S^2}{\chi^2_{1-\alpha/2|n-1}}, \quad G_2(S^2) = \frac{(n-1)S^2}{\chi^2_{\alpha/2|n-1}}$$

Die Quantile $\chi^2_{\alpha|n}$ können aus der Tabelle in Anhang T.4 abgelesen werden.

► **Beispiel 6.4.7.**
Eine Stichprobe vom Umfang $n = 50$ aus der Normalverteilung $\mathrm{No}(0, 4)$ hat die Stichprobenvarianz $S^2 = 4.3949$. Das symmetrische 95%-Konfidenzintervall für σ^2 lautet:

$$G_1 = 49 \cdot 4.3949/70.2224 = 3.0667, \quad G_2 = 49 \cdot 4.3949/31.5549 = 6.8246$$

Da die $\chi^2(n-1)$-Verteilung Mittelwert $n-1$ und Varianz $2(n-1)$ hat, können für genügend großes n ihre Quantile durch die der Normalverteilung $\mathrm{No}(n-1, 2(n-1))$ angenähert werden. Mit dieser Näherung lautet das Konfidenzintervall:

$$G_1 = 49 \cdot 4.3949/68.4027 = 3.1483, \quad G_2 = 49 \cdot 4.3949/29.5973 = 7.2760$$

Es ist also noch ein deutlicher Unterschied zu sehen.

➠MATLAB: `make_KI_normal_varianz` ◄

6.4.6.3 Mittelwert einer beliebigen Verteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit Mittelwert μ und Varianz σ^2. Auf Grund des zentralen Grenzwertsatzes ist das Standardscore Z des Stichprobenmittels:

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

für große Stichproben annähernd normalverteilt. Es gilt also *näherungsweise*:

$$W(\overline{X} - z_{1-\alpha/2}S/\sqrt{n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}S/\sqrt{n}) = 1 - \alpha$$

und damit:

$$G_1 = \overline{X} - z_{1-\alpha/2}S/\sqrt{n}, \quad G_2 = \overline{X} + z_{1-\alpha/2}S/\sqrt{n}$$

▶ **Beispiel 6.4.8.**

Für exponentialverteilte Stichproben vom Umfang n gibt die folgende Tabelle die Sicherheit des 95%-Konfidenzintervalls in Näherung durch Normalverteilung, geschätzt aus $N = 10000$ Stichproben:

n	25	50	100	400	800
$1-\alpha$	0.9129	0.9297	0.9394	0.9461	0.9458

Die Sicherheit ist also erst ab einer Stichprobengröße von etwa 400 korrekt.

➠MATLAB: `make_KI_exponential` ◀

Kapitel 7
Testen von Hypothesen

7.1 Einleitung

Es wird eine Stichprobe $X_1, \ldots, X_n$ aus einer Verteilung F beobachtet. Ein Test soll feststellen, ob die Beobachtungen mit einer gewissen Annahme über F verträglich sind. Die Annahme wird als *Nullhypothese* H_0 bezeichnet. Ist die Form von F bis auf einen oder mehrere Parameter spezifiziert, heißt der Test *parametrisch*. Ist die Form von F nicht spezifiziert, heißt der Test *nichtparametrisch* oder *parameterfrei*. Nichtparametrische Tests werden im Folgenden nicht behandelt.

7.1.1 Allgemeine Vorgangsweise

Aus der Stichprobe wird eine Testgröße (Teststatistik) T berechnet. Der Wertebereich von T wird, in Abhängigkeit von H_0, in einen *kritischen Bereich* oder *Ablehnungsbereich* C und einen *Annahmebereich* C′ unterteilt. Fällt der Wert von T in den Ablehnungsbereich, wird H_0 verworfen. Andernfalls wird H_0 vorläufig beibehalten. Das ist jedoch keine Bestätigung von H_0. Es heißt lediglich, dass die Beobachtungen mit der Hypothese vereinbar sind.

7.1.2 Signifikanz und Güte

Bei jedem Testverfahren sind zwei Arten von Fehlern möglich:

Fehler 1. Art: Die Hypothese H_0 wird abgelehnt, obwohl sie zutrifft.

Fehler 2. Art: Die Hypothese H_0 wird beibehalten, obwohl sie nicht zutrifft.

Zunächst wird die Verteilung von T mit dem Wissen über die Form der Verteilung F *unter Annahme von* H_0 bestimmt. Dann wird der Ablehnungsbereich so festgelegt, dass die Wahrscheinlichkeit eines Fehlers 1. Art maximal gleich einem Wert α ist. α heißt das *Signifikanzniveau* des Tests. Gängige Werte sind $\alpha = 0.05, 0.01, 0.005$.

Ist der Ablehnungsbereich festgelegt, kann für eine Gegenhypothese H_1 die Wahrscheinlichkeit $\beta(H_1)$ eines Fehlers 2. Art berechnet werden. $1 - \beta(H_1)$ heißt die *Güte* oder *Schärfe* (engl. power) des Tests für H_1. Die Güte sollte nie kleiner als α sein. Trifft dies zu, heißt der Test *unverzerrt*. Ein Ziel der Testtheorie ist es, unverzerrte Tests mit maximaler Güte zu konstruieren. Es gibt ein allgemeines Verfahren zur Konstruktion von solchen Tests, das sogenannte Likelihood-Quotienten-Verfahren.* Wir verzichten hier jedoch auf eine allgemeine Behandlung und geben lediglich Beispiele von Tests für einige der wichtigsten Verteilungsfamilien an.

*http://de.wikipedia.org/wiki/Likelihood-Quotienten-Test

7.2 Parametrische Tests

7.2.1 Grundlagen

Es liegt eine Stichprobe $X_1, \ldots, X_n$ aus einer Verteilung F vor, die bis auf einen oder mehrere Parameter ϑ bzw. $\boldsymbol{\vartheta}$ spezifiziert ist. Tests von Hypothesen über F heißen *parametrisch*. Eine Nullhypothese H_0 kann als eine Teilmenge des Parameterraums Θ aufgefasst werden: $\mathrm{H}_0 \subset \Theta$. Der Test entscheidet, ob die Stichprobe mit H_0 vereinbar ist. Vor der Anwendung ist zu klären, ob die angenommene parametrische Form plausibel ist.

Zunächst wird die Teststatistik T und das Signifikanzniveau α gewählt. Dann wird der Ablehnungsbereich C so festgelegt, dass

$$W(T \in \mathrm{C} \,|\, \vartheta \in \mathrm{H}_0) \leq \alpha$$

Zu einer Nullhypothese H_0 kann eine *Gegenhypothese* H_1 formuliert werden. H_1 kann ebenfalls als Teilmenge des Parameterraums Θ aufgefasst werden, mit $\mathrm{H}_1 \cap \mathrm{H}_0 = \emptyset$. Ist das Signifikanzniveau α festgelegt, kann für jedes $\vartheta_1 \in \mathrm{H}_1$ die Güte berechnet werden:

$$1 - \beta(\vartheta_1) = W(T \in \mathrm{C} \,|\, \vartheta_1 \in \mathrm{H}_1)$$

$1 - \beta(\vartheta_1)$ heißt die *Gütefunktion* des Tests. Im folgenden Beispiel werden die einzelnen Schritte im Detail gezeigt .

▶ **Beispiel 7.2.1.** Test des Mittelwerts einer Exponentialverteilung

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus $\mathrm{Ex}(\tau)$. Die Hypothese $\mathrm{H}_0 : \tau = \tau_0$ soll anhand der Stichprobe getestet werden, mit Signifikanzniveau $\alpha = 0.05$.

Als Teststatistik T dient das Stichprobenmittel: $T = \overline{X}$. Unter Annahme von H_0 hat T die folgende Dichtefunktion:

$$f_0(t) = \frac{t^{n-1}}{(\tau_0/n)^n \Gamma(n)} \exp\left(-\frac{t}{\tau_0/n}\right)$$

T ist also verteilt gemäß $\mathrm{Ga}(n, \tau_0/n)$. H_0 wird abgelehnt, wenn T von seinem Erwartungswert „weit entfernt“, also relativ klein oder relativ groß ist. Ein Ablehnungsbereich mit Signifikanzniveau α ist die Menge:

$$\mathrm{C} = [0, c_1] \cup [c_2, \infty)$$

mit

$$c_1 = \gamma_{\alpha/2|n,\tau_0/n} = \gamma_{\alpha/2|n,1} \cdot \tau_0/n, \quad c_2 = \gamma_{1-\alpha/2|n,\tau_0/n} = \gamma_{1-\alpha/2|n,1} \cdot \tau_0/n$$

Es sei nun $\tau_0 = 1$ und $n = 100$. Der Ablehnungsbereich ist dann:

$$\mathrm{C} = [0, 0.814] \cup [1.205, \infty)$$

Abbildung 7.1 zeigt die Dichtefunktion von T für diesen Fall. Der Ablehnungsbereich C ist rot, der Annahmebereich C′ ist grün eingefärbt.

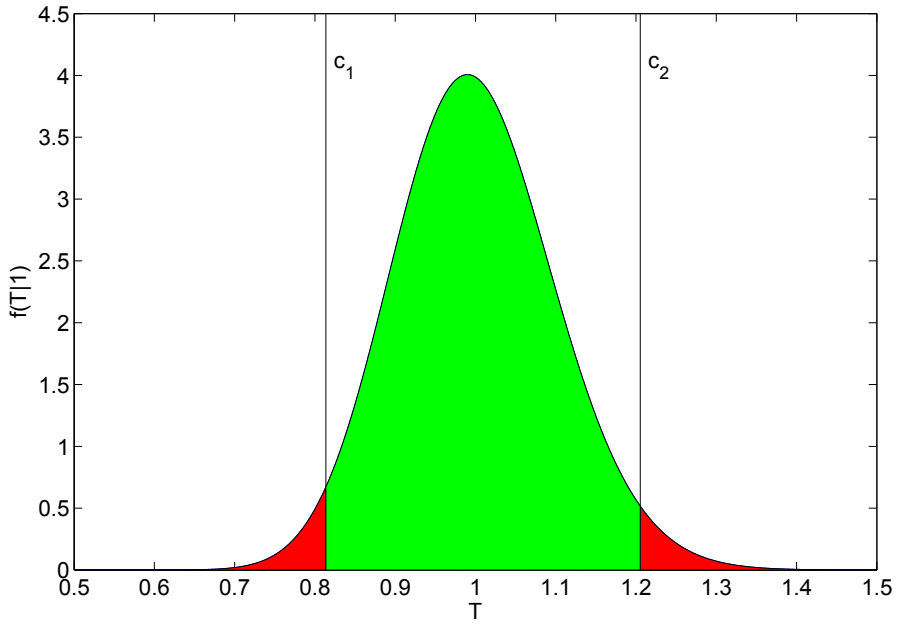

Abbildung 7.1: Dichtefunktion von T mit $\tau = 1$ und $n = 100$. Der Ablehnungsbereich mit $\alpha = 0.05$ ist rot, der Annahmebereich ist grün eingefärbt.
MATLAB: `make_test_exponential_mean`

Stammt die Stichprobe aus einer Exponentialverteilung mit Mittelwert $\tau_1 \neq \tau_0$, ist T gammaverteilt mit der Dichtefunktion:

$$f(t\,|\,\tau_1) = \frac{t^{n-1}}{(\tau_1/n)^n \Gamma(n)} \exp\left(-\frac{t}{\tau_1/n}\right)$$

Ein Fehler 2. Art tritt auf, wenn T nicht in C, d.h. in C′ liegt. Die Wahrscheinlichkeit dafür ist gleich:

$$\beta(\tau) = \int_{c_1}^{c_2} f(t\,|\,\tau_1)\,\mathrm{d}t = F(c_2\,|\,\tau_1) - F(c_1\,|\,\tau_1)$$

Hier bezeichnet $F(t\,|\,\tau_1)$ die Verteilungsfunktion von T unter der Annahme $\tau = \tau_1$. Die Güte des Test für den Wert τ_1 ist dann gleich:

$$1 - \beta(\tau_1) = 1 - F(c_2\,|\,\tau_1) + F(c_1\,|\,\tau_1)$$

Abbildung 7.2 zeigt die Dichtefunktion von T für $\tau = 1.3$. Die Güte des Tests enspricht der blau eingefärbten Fläche, die Wahrscheinlichkeit eines Fehlers 2. Art der magenta eingefärbten Fläche.

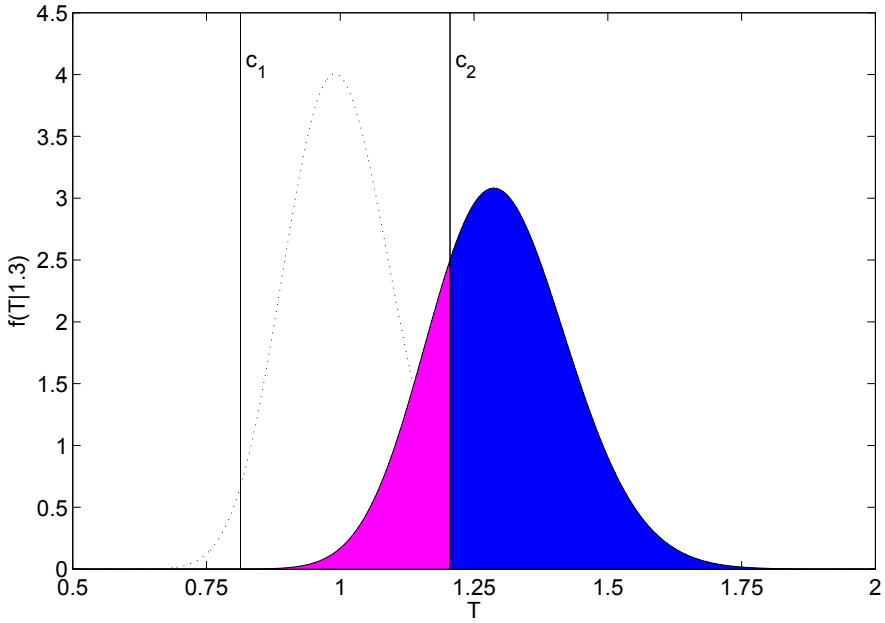

Abbildung 7.2: Dichtefunktion von T mit $\tau = 1.3$. Die Güte des Tests enspricht der blau eingefärbten Fläche, die Wahrscheinlichkeit eines Fehlers 2. Art der magenta eingefärbten Fläche. MATLAB: `make_test_exponential_mean`

Die Gütefunktion für einen beliebigen Wert τ ergibt sich durch:

$$1 - \beta(\tau) = W(T \in \mathrm{C}) = 1 - F(c_2\,|\,\tau) + F(c_1\,|\,\tau)$$

Sie ist in Abbildung 7.3 für zwei Stichprobengrößen ($n = 25, 100$) dargestellt. Klarerweise wächst die Güte mit der Stichprobengröße und mit dem Abstand zwischen τ_0 und τ. Der Test ist nicht unverzerrt, da z.B. für $\tau_0 = 1$ und $n = 25$

$$1 - \beta(0.986) = 0.0495 < 0.05$$

Für große Stichproben wird die Verzerrung vernachlässigbar, der Test ist asymptotisch unverzerrt.

➠MATLAB: `make_test_exponential_mean` ◀

7.2.2 Tests für normalverteilte Stichproben

Für normalverteilte Stichproben gibt es zahlreiche Tests, die das vorgegebene Signifikanzniveau α exakt einhalten. Das gilt allerdings nur, wenn die Beobachtungen tatsächlich aus einer Normalverteilung stammen. Andernfalls kann die Wahrscheinlichkeit

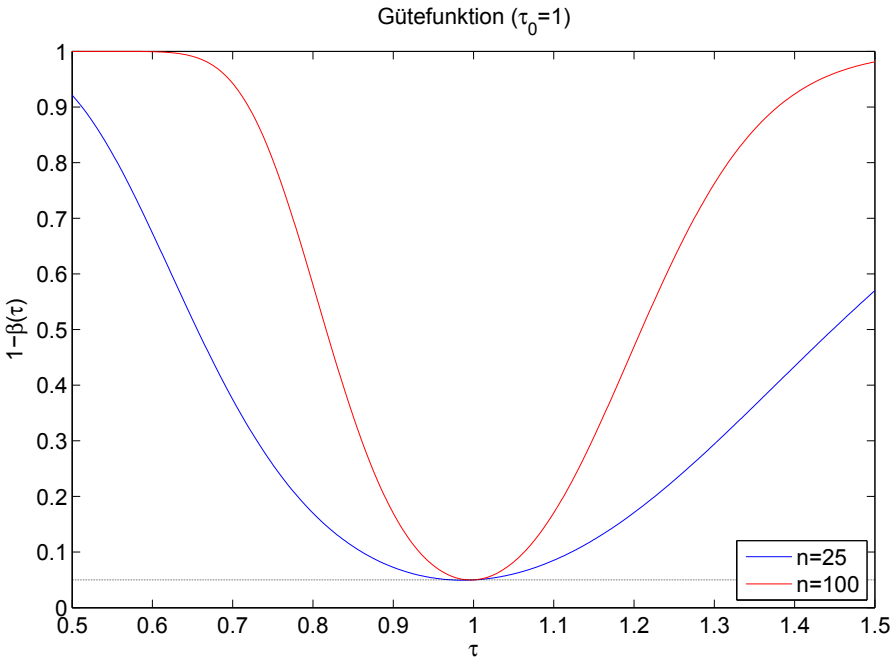

Abbildung 7.3: Gütefunktion des Tests in Beispiel 7.2.1. Es ist $\tau_0 = 1$ und $\alpha = 0.05$.
MATLAB: `make_test_exponential_mean`

eines Fehlers 1. Art erheblich vom nominellen Wert abweichen. Sind also die Beobachtungen mit Ausreißern kontaminiert, sollten die folgenden Tests mit einiger Vorsicht angewendet werden. Das gilt ganz besonders für Tests von Varianzen (Abschnitte 7.2.2.6 und 7.2.2.7).

7.2.2.1 Mittelwert bei bekannter Varianz

Es sei $X_1, \ldots, X_n$ eine normalverteilte Stichprobe aus $\mathrm{No}(\mu, \sigma^2)$ mit bekanntem σ^2. Die Hypothese $\mathrm{H}_0 : \mu = \mu_0$ soll anhand der Stichprobe gegen die Gegenhypothese $\mathrm{H}_1 : \mu \neq \mu_0$ getestet werden. Als Teststatistik T dient das Standardscore des Stichprobenmittels:

$$T = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Unter Annahme von H_0 ist $T \sim \mathrm{No}(0, 1)$. H_0 wird abgelehnt, wenn T von seinem Erwartungswert „weit entfernt“, also relativ klein oder relativ groß ist. Ein Ablehnungsbereich mit Signifikanzniveau α ist die Menge

$$\mathrm{C} = (-\infty, z_{\alpha/2}] \cup [z_{1-\alpha/2}, \infty)$$

wo z_q das Quantil der Standardnormalverteilung zum Niveau q ist. Die Hypothese H_0 wird also abgelehnt, wenn

$$|T| = \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} > z_{1-\alpha/2}$$

Stammt die Stichprobe aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$ mit $\mu \neq \mu_0$, so ist T verteilt gemäß $\mathrm{No}(\delta, 1)$ mit $\delta = \sqrt{n}(\mu - \mu_0)/\sigma$. Dann ergibt sich die Gütefunktion an der Stelle μ als:

$$1 - \beta(\mu) = W(T \in \mathrm{C}) = \Phi(z_{\alpha/2} \,|\, \delta, 1) + 1 - \Phi(z_{1-\alpha/2} \,|\, \delta, 1)$$

wo $\Phi(x \,|\, \delta, 1)$ die Verteilungsfunktion der $\mathrm{No}(\delta, 1)$-Verteilung ist. Mit Satz 4.2.11 folgt weiters:

$$1 - \beta(\mu) = \Phi(z_\alpha - \delta) + 1 - \Phi(z_{1-\alpha/2} - \delta)$$

wo $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung ist. Der Test ist unverzerrt. Abbildung 7.4 zeigt die Gütefunktion des zweiseitigen Tests.

➠MATLAB: `make_test_normal_mean`

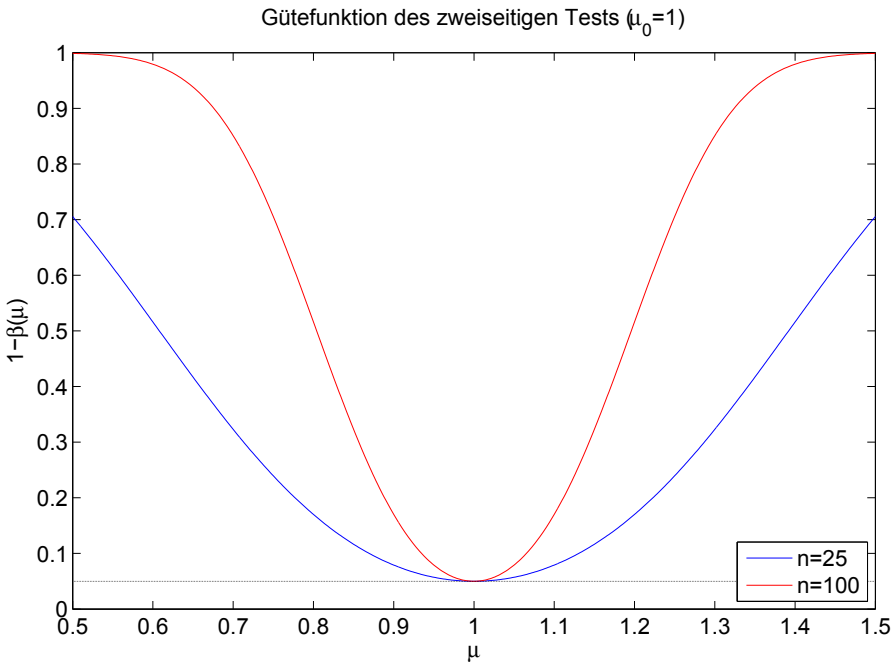

Abbildung 7.4: Gütefunktion des zweiseitigen Tests des Mittelwerts einer Normalverteilung. Es ist $\mu_0 = 1$ und $\alpha = 0.05$. MATLAB: `make_test_normal_mean`

7.2.2.2 Einseitiger Test

Die Hypothese $\mathrm{H}_0 : \mu = \mu_0$ soll mit der Teststatistik T gegen die Gegenhypothese $\mathrm{H}_1 : \mu > \mu_0$ getestet werden. H_0 wird abgelehnt, wenn T „zu groß“ ist. Ein Ablehnungsbereich mit Signifikanzniveau α ist die Menge

$$\mathrm{C} = [z_{1-\alpha}, \infty)$$

Die Hypothese H_0 wird also abgelehnt, wenn

$$T = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}$$

Die Gütefunktion für einen Wert $\mu > \mu_0$ ergibt sich durch:

$$1 - \beta(\mu) = W(T \in \mathrm{C}) = 1 - \Phi(z_{1-\alpha} \,|\, \delta, 1) = 1 - \Phi(z_{1-\alpha} - \delta)$$

mit $\delta = \sqrt{n}(\mu - \mu_0)/\sigma$. Abbildung 7.5 zeigt die Gütefunktion des einseitigenTests. Der Test mit Gegenhypothese $\mathrm{H}_1 : \mu < \mu_0$ verläuft analog.

➠MATLAB: `make_test_normal_mean`

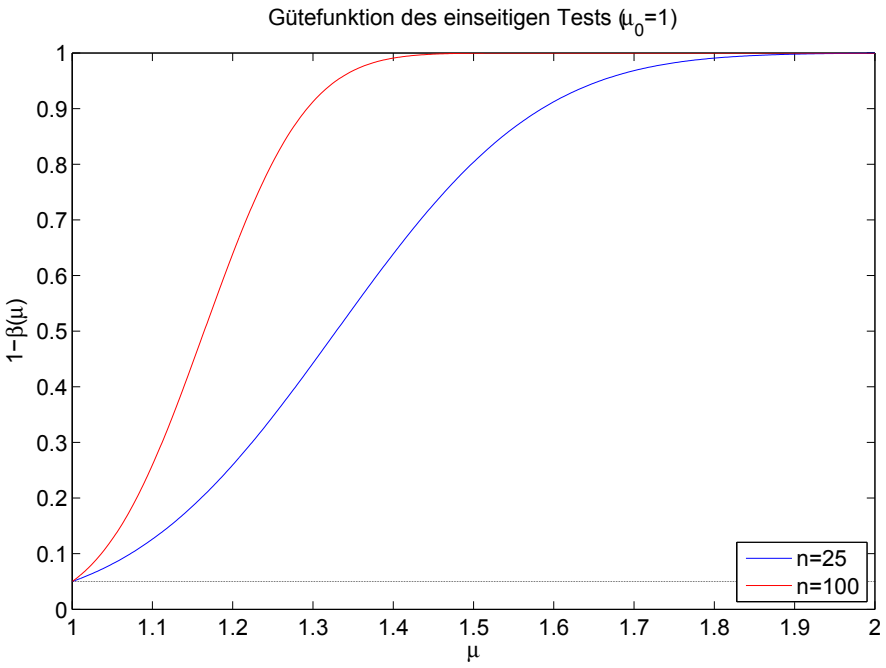

Abbildung 7.5: Gütefunktion des einseitigen Tests des Mittelwerts einer Normalverteilung. Es ist $\mu_0 = 1$ und $\alpha = 0.05$. MATLAB: `make_test_normal_mean`

► **Beispiel 7.2.2.**
Eine Stichprobe der Größe $n = 50$ stammt aus einer Normalverteilung mit $\sigma = 2.4$ mit unbekanntem Mittelwert μ. Die Nullhypothese ist $\mathrm{H}_0 : \mu = 10$. Der Ablehnungsbereich des einseitigen Tests mit der Gegenhypothese $\mathrm{H}_1 : \mu_1 > \mu_0$ und $\alpha = 0.01$ ist die Menge $\mathrm{C} = [z_{0.99}, \infty) = [2.326, \infty)$. Unter der Gegenhypothese $\mu_1 = 11$ ist T verteilt mit Erwartungswert $\delta = \sqrt{50}(11 - 10)/2.4 = 2.946$ und Standardabweichung $\sigma_T = 1$. Die Güte des Tests für $\mu_1 = 11$ ist dann gleich:

$$W(T \in \mathrm{C}) = 1 - \Phi(2.326 - 2.946) = 1 - \Phi(-0.62) = 0.732$$ ◄

7.2.2.3 Mittelwert bei unbekannter Varianz: T-Test

Es sei $X_1, \ldots, X_n$ eine normalverteilte Stichprobe aus $\mathrm{No}(\mu, \sigma^2)$ mit unbekanntem σ^2. Die Hypothese $\mathrm{H}_0 : \mu = \mu_0$ soll anhand der Stichprobe gegen die Gegenhypothese $\mathrm{H}_1 : \mu \neq \mu_0$ getestet werden. Als Teststatistik T dient das Standardscore des Stichprobenmittels, unter Benützung der Stichprobenvarianz S^2:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Unter Annahme von H_0 ist $T \sim \mathrm{T}(n-1)$. H_0 wird abgelehnt, wenn T von seinem Erwartungswert „weit entfernt“, also relativ klein oder relativ groß ist. Ein Ablehnungsbereich mit Signifikanzniveau α ist die Menge

$$\mathrm{C} = (-\infty, t_{\alpha/2|n-1}] \cup [t_{1-\alpha/2|n-1}, \infty)$$

Die Hypothese H_0 wird also abgelehnt, wenn

$$|T| = \frac{|\overline{X} - \mu_0|}{S/\sqrt{n}} > t_{1-\alpha/2|n-1}$$

Die Gütefunktion für einen Wert μ ergibt sich durch:

$$1 - \beta(\mu) = W(T \in \mathrm{C}) = G(z_{\alpha/2}) + 1 - G(z_{1-\alpha/2})$$

wo G die Verteilungsfunktion der nichtzentralen $\mathrm{T}(n-1, \delta)$-Verteilung mit

$$\delta = \sqrt{n}(\mu - \mu_0)/\sigma$$

ist.* Der Test ist unverzerrt. Abbildung 7.6 zeigt die Gütefunktion des Tests.

➠MATLAB: `make_test_normal_mean`

*http://en.wikipedia.org/wiki/Noncentral_t-distribution

7.2.2.4 Test der Gleichheit von zwei Mittelwerten

$X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ seien zwei unabhängige normalverteilte Stichproben aus $\mathrm{No}(\mu_x, \sigma_x^2)$ bzw. $\mathrm{No}(\mu_y, \sigma_y^2)$. Die Hypothese $\mathrm{H}_0 : \mu_x = \mu_y$ soll anhand der Stichproben gegen die Gegenhypothese $\mathrm{H}_1 : \mu_x \neq \mu_y$ getestet werden. Sind die Varianzen bekannt, wird die Differenz der Stichprobenmittel als Teststatistik T verwendet:

$$T = \overline{X} - \overline{Y}$$

Unter Annahme von H_0 ist $T \sim \mathrm{No}(0, \sigma_x^2/n + \sigma_y^2/m)$. Das Standardscore

$$Z = \frac{T}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}$$

ist dann standardnormalverteilt. Die Hypothese H_0 wird also abgelehnt, wenn

$$|Z| > z_{1-\alpha/2}$$

oder

$$\frac{|\overline{X} - \overline{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} > z_{1-\alpha/2}$$

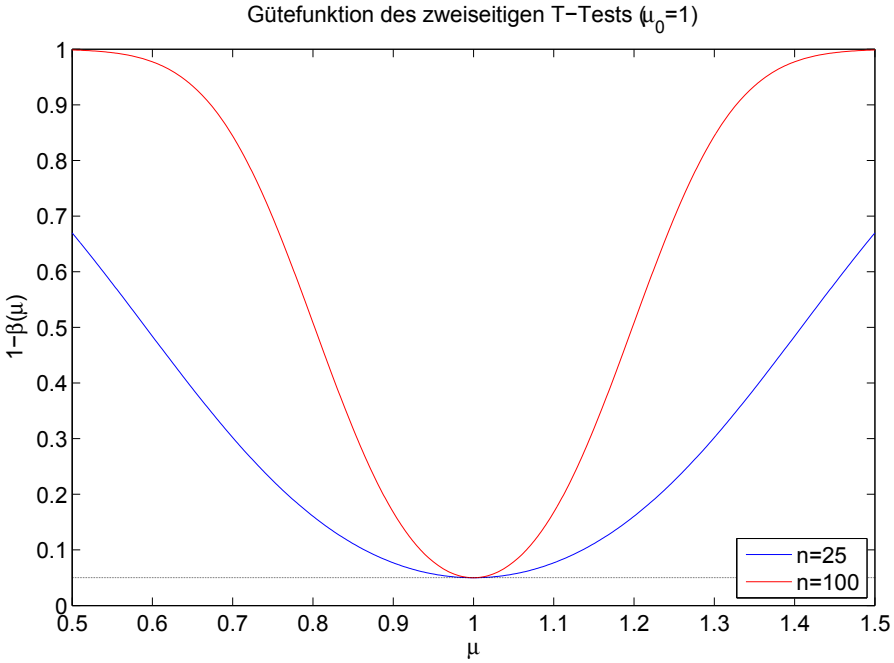

Abbildung 7.6: Gütefunktion des zweiseitigen T-Tests des Mittelwerts einer Normalverteilung. Es ist $\mu_0 = 1$ und $\alpha = 0.05$. MATLAB: `make_test_normal_mean`

Wenn die Varianzen unbekannt sind, aber als gleich groß angenommen werden können, kann unter der Nullhypothese die Varianz aus der kombinierten Stichprobe geschätzt werden:

$$S^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Unter Annahme von H_0 ist die Testgröße

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{S^2(1/n + 1/m)}}$$

T-verteilt mit $n+m-2$ Freiheitsgraden. Die Hypothese H_0 wird also abgelehnt, wenn:

$$|T| > t_{1-\alpha/2|n+m-2}$$

Können die Varianzen *nicht* als (ungefähr) gleich angenommen werden, kann der modifizierte T-Test nach Welch angewendet werden. Die Testgröße lautet:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{S_x^2/n + S_y^2/m}}$$

Die Anzahl ν der Freiheitsgrade wird näherungsweise aus den Beobachtungen bestimmt:

$$\nu = \frac{(S_x^2/n + S_y^2/m)^2}{S_x^4/(n^3 - n^2) + S_y^4/(m^3 - m^2)}$$

Es ist zu beachten, das der Welch-Test nicht das exakte Signifikanzniveau α garantiert. Alle in diesem Abschnitt besprochenen Tests können natürlich auch einseitig ausgeführt werden.

▶ **Beispiel 7.2.3.** Gleichheit der Mittelwerte

Es liegen Stichproben der Zufallsvariablen X und Y vor, vom Umfang $n = 200$ bzw. $m = 300$. X ist normalverteilt mit $\mu_x = 4,\ \sigma_x = 1$, Y ist normalverteilt mit $\mu_y = 4,\ \sigma_y = 1.2$. Es gilt:

$$\bar{x} = 3.961,\ S_x^2 = 0.827;\ \bar{y} = 3.906,\ S_y^2 = 1.361$$

Unter Annahme der Gleichheit der Varianzen ist die Schätzung der gemeinsamen Varianz gleich $S^2 = 1.147$, und die Testgröße ist gleich $T = 0.560$, mit 498 Freiheitsgraden. Ohne die Annahme der Gleichheit ist die Testgröße gleich $T = 0.588$, mit geschätzten 486 Freiheitsgraden. Die Nullhypothese $\mathrm{H}_0 : \mu_x = \mu_y$ wird in beiden Fällen nicht abgelehnt.

Ist dagegen $\mu_y = 5$, so sind die Testgrößen gleich -9.667 bzw. -10.152. Die Nullhypothese wird also in beiden Fällen abgelehnt.

➠MATLAB: `make_test_normal_2stichproben` ◀

7.2.2.5 T-Test für gepaarte Stichproben

Gepaarte Stichproben $(X_1, Y_1), \ldots, (X_n, Y_n)$ entstehen, wenn für jedes beobachtete Objekt die selbe Größe zweimal gemessen wird, vor und nach einer bestimmten Intervention. Die Wirkung der Intervention wird durch die Differenzen $W_i = Y_i - X_i, i = 1, \ldots, n$ beschrieben. Es seien $W_1, \ldots, W_n$ normalverteilt mit Mittelwert μ_w und unbekannter Varianz σ_w^2. Die Nullhypothese $\mathrm{H}_0 : \mu_w = 0$ (keine Wirkung der Intervention) soll anhand der Stichprobe gegen die Gegenhypothese $\mathrm{H}_1 : \mu_w \neq 0$ getestet werden. Dies erfolgt mit dem T-Test für einzelne Stichproben (siehe Abschnitt 7.2.2.3).

7.2.2.6 Test der Varianz

Es sei $X_1, \ldots, X_n$ eine normalverteilte Stichprobe mit unbekanntem Mittelwert μ und unbekannter Varianz σ^2. Die Hypothese $\mathrm{H}_0 : \sigma^2 = \sigma_0^2$ soll anhand der Stichprobe gegen die Gegenhypothese $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2$ getestet werden. Als Teststatistik T wählen wir:

$$T = \frac{(n-1)S^2}{\sigma_0^2}$$

Unter Annahme von H_0 ist T χ^2-verteilt mit $n-1$ Freiheitsgraden. Die Hypothese H_0 wird also abgelehnt, wenn:

$$T < \chi^2_{\alpha/2|n-1} \quad \text{oder} \quad T > \chi^2_{1-\alpha/2|n-1}$$

wo $\chi^2_{q|k}$ das Quantil der χ^2-Verteilung mit k Freiheitsgraden zum Niveau q ist. Die Gütefunktion für einen Wert σ^2 ergibt sich durch:

$$1 - \beta(\sigma^2) = G(\sigma_0^2/\sigma^2 \cdot \chi^2_{\alpha/2|n-1}) + 1 - G(\sigma_0^2/\sigma^2 \cdot \chi^2_{1-\alpha/2|n-1})$$

wo G die Verteilungsfunktion der $\chi^2(n-1)$-Verteilung ist. Der Test ist nicht unverzerrt. Abbildung 7.7 zeigt die Gütefunktion des Tests.

➠MATLAB: `make_test_normal_variance`

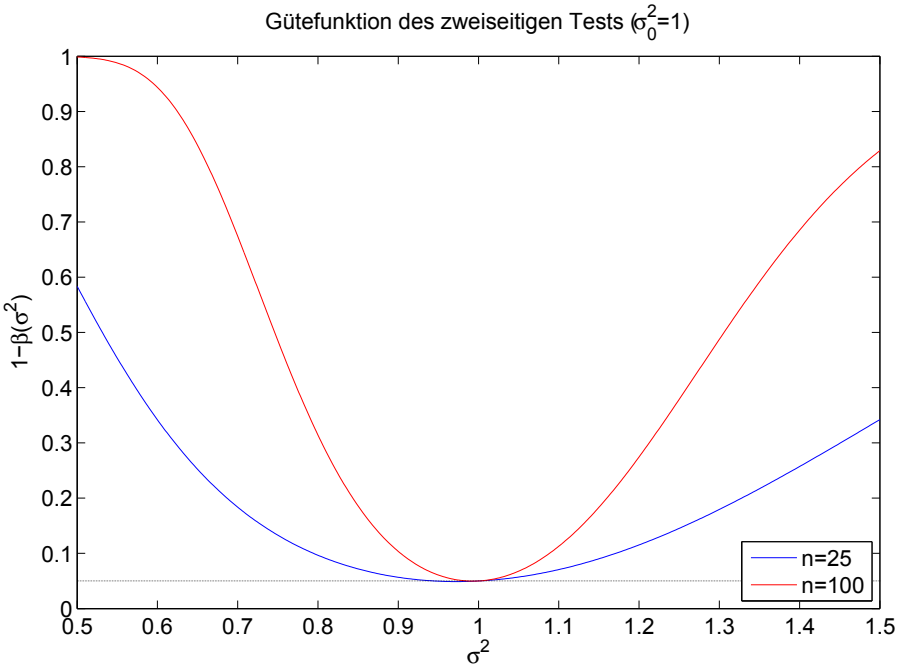

Abbildung 7.7: Gütefunktion des zweiseitigen Tests der Varianz einer Normalverteilung. Es ist $\sigma_0^2 = 1$ und $\alpha = 0.05$.
MATLAB: `make_test_normal_variance`

7.2.2.7 Test der Gleichheit von zwei Varianzen

Es seien $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ zwei unabhängige normalverteilte Stichproben aus $\mathrm{No}(\mu_x, \sigma_x^2)$ bzw. $\mathrm{No}(\mu_y, \sigma_y^2)$. Die Nullhypothese $\mathrm{H}_0 : \sigma_x^2 = \sigma_y^2$ soll anhand der Stichproben gegen die Gegenhypothese $\mathrm{H}_1 : \sigma_x^2 \neq \sigma_y^2$ getestet werden. Die Teststatistik T ist das Verhältnis der Stichprobenvarianzen:

$$T = \frac{S_x^2}{S_y^2}$$

Unter Annahme von H_0 ist T F-verteilt* gemäß $\mathrm{F}(n-1, m-1)$. Der Test wird daher als F-Test bezeichnet. Die Nullhypothese wird abgelehnt, wenn:

$$T < F_{\alpha/2} \quad \text{oder} \quad T > F_{1-\alpha/2}$$

wo F_q das Quantil der F-Verteilung mit $n-1$ bzw. $m-1$ Freiheitsgraden zum Niveau q ist. Ist $\sigma_y^2 = k\sigma_x^2$, ergibt sich die Gütefunktion für einen Wert k durch:

$$1 - \beta(k) = G(k \cdot F_{\alpha/2}) + 1 - G(k \cdot F_{1-\alpha/2})$$

wo G die Verteilungsfunktion der $\mathrm{F}(n-1, m-1)$-Verteilung ist. Der Test ist unverzerrt. Abbildung 7.8 zeigt die Gütefunktion des Tests mit $n = m = 25$ bzw. $n = m = 100$.

➠MATLAB: `make_test_normal_variance`

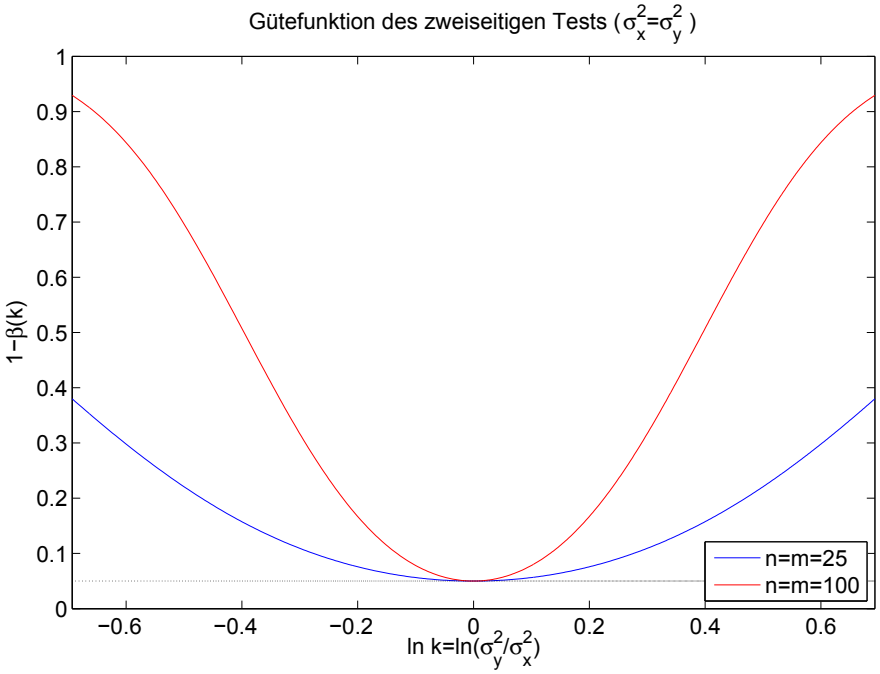

Abbildung 7.8: Gütefunktion des Tests der Gleichheit von zwei Varianzen.
MATLAB: `make_test_normal_variance`

*http://de.wikipedia.org/wiki/F-Verteilung

► **Beispiel 7.2.4.** Gleichheit der Varianzen

Die beiden Stichproben aus Beispiel 7.2.3 sollen auf Gleichheit der Varianzen getestet werden. Die Testgröße ist gleich:

$$T = \frac{S_x^2}{S_y^2} = \frac{0.827}{1.361} = 0.608$$

Unter Nullhypothese ist T F-verteilt mit 199 bzw. 299 Freiheitsgraden. Mit $\alpha = 0.05$ erhält man die Quantile:

$$F_{0.025} = 0.773, \quad F_{0.975} = 1.285$$

Die Nullhypothese wird abgelehnt.

➠Matlab: `make_test_normal_2stichproben` ◄

7.2.3 Tests für alternativverteilte Stichproben

7.2.3.1 Zweiseitiger Test der Erfolgswahrscheinlichkeit

Es sei $X_1, \ldots, X_n$ eine alternativverteilte Stichprobe aus $\mathrm{Al}(p)$. Die Nullhypothese $\mathrm{H}_0 : p = p_0$ soll anhand der Stichprobe gegen die Gegenhypothese $\mathrm{H}_1 : p \neq p_0$ getestet werden. Als Teststatistik T dient die Anzahl der Erfolge:

$$T = \sum_{i=1}^{n} X_i$$

Unter der Nullhypothese ist $T \sim \mathrm{Bi}(n, p_0)$. H_0 wird abgelehnt, wenn T „zu klein“ oder „zu groß“ ist, also wenn:

$$T < B_{\alpha/2|n,p_0} \quad \text{oder} \quad T > B_{1-\alpha/2|n,p_0}$$

wo $B_{q|n,p}$ das q-Quantil der Binomialverteilung $\mathrm{Bi}(n, p)$ ist. Ein dazu äquivalentes Ablehnungkriterium ist:

$$\sum_{i=1}^{T} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha/2 \quad \text{oder} \quad \sum_{i=T}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha/2$$

7.2.3.2 Einseitiger Test der Erfolgswahrscheinlichkeit

$X_1, \ldots, X_n$ sei wieder eine alternativverteilte Stichprobe aus $\mathrm{Al}(p)$. Die Hypothese $\mathrm{H}_0 : p \leq p_0$ soll anhand der Stichprobe gegen die Gegenhypothese $\mathrm{H}_1 : p > p_0$ getestet werden. Als Teststatistik T dient wieder die Anzahl der Erfolge:

$$T = \sum_{i=1}^{n} X_i$$

Unter der Nullhypothese ist $T \sim \mathrm{Bi}(n, p)$ mit $p \leq p_0$. H_0 wird abgelehnt, wenn T „zu groß“ ist, also wenn

$$T > B_{1-\alpha|n,p}, \quad \text{für alle } p \leq p_0$$

Da $B_{1-\alpha|n,p}$ für festes α und n eine wachsende Funktion von p ist, wird H_0 genau dann abgelehnt, wenn:

$$T > B_{1-\alpha|n,p_0} \quad \text{oder} \quad \sum_{i=T}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$$

Die analoge Nullhypothese $H_0 : p \geq p_0$ wird abgelehnt, wenn:

$$T < B_{\alpha|n,p_0} \quad \text{oder} \quad \sum_{i=1}^{T} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$$

► **Beispiel 7.2.5.**
Ein Hersteller behauptet, dass nicht mehr als 2% eines gewissen Bauteils fehlerhaft sind. In einer Stichprobe vom Umfang 300 sind 9 Stück defekt. Kann die Behauptung des Herstellers widerlegt werden? Es gilt:

$$\sum_{i=9}^{300} \binom{300}{i} 0.02^i 0.98^{300-i} = 0.1507$$

Die Behauptung des Herstellers lässt sich also auf einem Signifikanzniveau von 5% nicht widerlegen.

➠MATLAB: `make_test_alternative_mean` ◄

7.2.3.3 Näherung durch Normalverteilung

Ist n genügend groß, kann die Verteilung von T unter H_0 durch die Normalverteilung mit Mittelwert np_0 und Varianz $np_0(1-p_0)$ angenähert werden. Die Testgröße ist dann das Standardscore von T:

$$Z = \frac{T - np_0}{\sqrt{np_0(1-p_0)}}$$

Im zweiseitigen Test wird H_0 abgelehnt, wenn:

$$|Z| \geq z_{1-\alpha/2}$$

Im einseitigen Test wird $H_0 : p \geq p_0$ bzw. $H_0 : p \leq p_0$ abgelehnt, wenn:

$$Z \leq z_\alpha \quad \text{bzw.} \quad Z \geq z_{1-\alpha}$$

▶ Beispiel 7.2.6.
Mit der Angabe von Beispiel 7.2.5 ergibt die Näherung durch Normalverteilung:

$$Z = 1.2372 < z_{0.95} = 1.645$$

Die Behauptung kann also nicht widerlegt werden.

➠MATLAB: `make_test_alternative_mean` ◀

7.2.4 Tests für poissonverteilte Stichproben

7.2.4.1 Einseitiger Test des Mittelwerts

Es sei $X_1, \ldots, X_n$ eine poissonverteilte Stichprobe aus $\mathrm{Po}(\lambda)$. Die Hypothese $H_0 : \lambda \leq \lambda_0$ soll anhand der Stichprobe gegen die Gegenhypothese $H_1 : \lambda > \lambda_0$ getestet werden. Als Teststatistik T dient die Stichprobensumme:

$$T = \sum_{i=1}^{n} X_i$$

Unter der Nullhypothese ist $T \sim \mathrm{Po}(n\lambda)$ mit $\lambda \leq \lambda_0$. H_0 wird abgelehnt, wenn T „zu groß“ ist, also wenn:

$$T > P_{1-\alpha|n\lambda}, \quad \text{für alle } \lambda \leq \lambda_0$$

wo $P_{1-\alpha|n\lambda}$ das $(1-\alpha)$-Quantil der Poissonverteilung $\mathrm{Po}(n\lambda)$ ist. Da das $(1-\alpha)$-Quantil der Poissonverteilung eine wachsende Funktion des Mittelwertes ist, wird H_0 genau dann abgelehnt, wenn:

$$T > P_{1-\alpha|n\lambda_0}$$

Die analoge Nullhypothese $H_0 : \lambda \geq \lambda_0$ wird abgelehnt, wenn:

$$T < P_{\alpha|n\lambda_0}$$

Die Quantile $P_{\alpha|\lambda}$ sind für ganzzahlige Werte von λ im Anhang tabelliert.

▶ Beispiel 7.2.7.
Ein Hersteller strebt an, dass in einer Fabrik täglich im Mittel nicht mehr als 25 defekte Bauteile hergestellt werden. Eine Stichprobe von 5 Tagen ergibt 28,34,32,38 und 22 defekte Bauteile. Hat der Hersteller sein Ziel erreicht? Es gilt:

$$T = 154, \quad P_{0.99|125} = 152$$

Die Hypothese $H_0 : \lambda \leq 25$ lässt sich also auf einem Signifikanzniveau von 1% oder mehr widerlegen.

➠MATLAB: `make_test_poisson_mean` ◀

7.2.4.2 Näherung durch Normalverteilung

Ist n genügend groß, kann die Verteilung von T unter H_0 durch die Normalverteilung mit Mittelwert $n\lambda_0$ und Varianz $n\lambda_0$ angenähert werden. Die Testgröße ist dann das Standardscore von T:

$$Z = \frac{T - n\lambda_0}{\sqrt{n\lambda_0}}$$

Im zweiseitigen Test wird H_0 abgelehnt, wenn:

$$|Z| \geq z_{1-\alpha/2}$$

Im einseitigen Test wird $H_0 : \lambda \geq \lambda_0$ bzw. $H_0 : \lambda \leq \lambda_0$ abgelehnt, wenn:

$$Z \leq z_\alpha \quad \text{bzw.} \quad Z \geq z_{1-\alpha}$$

▶ **Beispiel 7.2.8.**
Mit der Angabe von Beispiel 7.2.7 ergibt die Näherung durch Normalverteilung:

$$Z = 2.5938 > z_{0.99} = 2.3263$$

Die Hypothese $H_0 : \lambda \leq 25$ lässt sich also auf einem Signifikanzniveau von 1% oder mehr widerlegen.

➠MATLAB: `make_test_poisson_mean` ◀

7.3 Anpassungstests

Ein Test, der die Hypothese überprüft, ob die Beobachtungen einer gewissen Verteilung entstammen können, heißt ein *Anpassungstest*. Die Verteilung kann völlig oder bis auf unbekannte Parameter bestimmt sein. Ein Anpassungstest kann einem parametrischen Test vorausgehen, um dessen Anwendbarkeit zu überprüfen.

7.3.1 Der Chiquadrat-Test

7.3.1.1 Der Chiquadrat-Test mit bekannter Nullhypothese

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus einer Verteilung F. Es soll getestet werden, ob die Verteilung F gleich einer hypothetischen Verteilung F_0 ist, wobei F_0 vollständig spezifiziert ist. F und F_0 sind entweder beide diskret oder beide stetig.

Zunächst wird die Ergebnismenge Ω von X in k disjunkte Gruppen $G_1, \ldots, G_k$ zerlegt (siehe Definition 2.3.6). Nun sei $p_j = W(X \in G_j \,|\, F_0)$ die Wahrscheinlichkeit unter der Nullhypothese, dass X in Gruppe G_j fällt, und Y_j die tatsächliche Anzahl der Stichprobenwerte in Gruppe G_j. Unter der Nullhypothese ist $Y_1, \ldots, Y_k$ multinomial verteilt gemäß $\mathrm{Mu}(n, p_1, \ldots, p_k)$. Dann gilt:

$$\mathsf{E}[Y_j] = np_j, \text{ mit } p_j = W(X \in G_j \,|\, F_0)$$

Die Testgröße vergleicht die beobachteten Häufigkeiten Y_j mit ihren Erwartungswerten:

$$T = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j}$$

Die Nullhypothese wird verworfen, wenn T groß ist. Der Ablehnungsbereich kann nach dem folgenden Satz bestimmt werden.

☛ **Satz 7.3.1.**

Unter Annahme der Nullhypothese ist die Zufallsvariable T asymptotisch, d.h. für $n \to \infty$, χ^2-verteilt mit $k-1$ Freiheitsgraden.

Soll der Test Signifikanzniveau α haben, wird F_0 abgelehnt, wenn:

$$T \geq \chi^2_{1-\alpha|k-1}$$

Der Grund dafür, dass T nur $k-1$ Freiheitsgrade hat, ist der lineare Zusammenhang zwischen den Y_j:

$$\sum_{j=1}^{k} Y_j = n$$

Als Faustregel gilt: n sollte so groß sein, dass $np_j > 10$ für $j = 1, \ldots, k$. Andernfalls ist die Verteilung von T noch zu weit von einer χ^2-Verteilung entfernt.

▶ **Beispiel 7.3.2.**

Es wird anhand einer Stichprobe vom Umfang 60 mit dem χ^2-Test getestet, ob ein Würfel symmetrisch ist, d.h. ob die Augenzahl X folgende Verteilung hat:

$$W(X=1) = \ldots = W(X=6) = \frac{1}{6}$$

Es soll untersucht werden, ob die Testgröße T tatsächlich in guter Näherung $\chi^2(5)$-verteilt ist. Dazu wird der Erwartungswert und die Varianz von T mittels Simulation von $N = 10^5$ Stichproben geschätzt. Das Ergebnis ist:

$$\mathsf{E}[T] = 5.001, \quad \mathsf{var}[T] = 9.873$$

Der Mittelwert stimmt fast exakt, die Varianz ist etwas zu klein. Das 0.95-Quantil der $\chi^2(5)$-Verteilung ist $\chi^2_{0.95|5} = 11.07$, die tatsächliche Überschreitungswahrscheinlichkeit ist gleich:

$$W(T \geq 11.07) = 0.0474$$

Die Verteilung von T stimmt also sehr gut mit einer $\chi^2(5)$-Verteilung überein, wie auch das Q-Q-Diagramm (siehe Abschnitt 5.2.5) in Abbildung 7.9 zeigt.

➠MATLAB: `make_chi2test_wuerfel` ◀

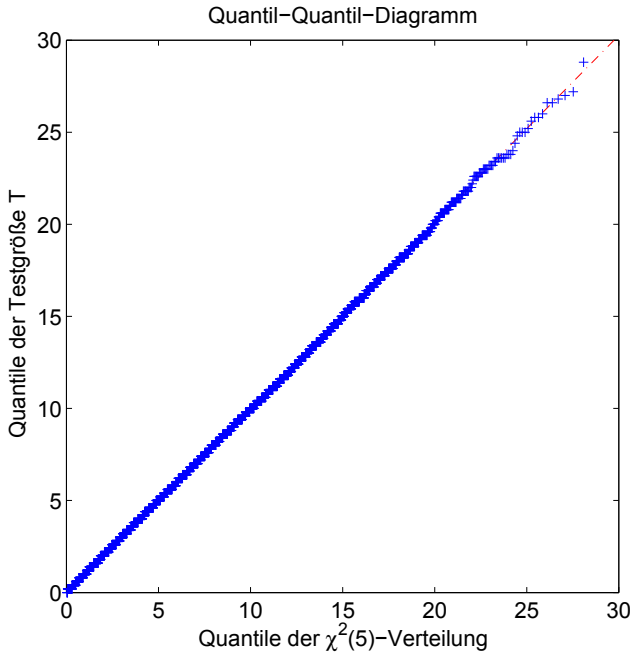

Abbildung 7.9: Q-Q-Diagramm der Testgröße T. MATLAB: `make_chi2test_wuerfel`

7.3.1.2 Nullhypothese mit unbekannten Parametern

Die hypothetische Verteilung von X muss nicht vollständig spezifiziert sein. In diesem Fall hängen die Verteilungsfunktion F_0 und damit auch die Gruppenwahrscheinlichkeiten p_j noch von unbekannten Parametern $\boldsymbol{\vartheta}$ ab:

$$W(X \in \mathrm{G}_j) = p_j(\boldsymbol{\vartheta})$$

Die Teststatistik T ist nun ebenfalls eine Funktion der unbekannten Parameter:

$$T(\boldsymbol{\vartheta}) = \sum_{j=1}^{k} \frac{(Y_j - np_j(\boldsymbol{\vartheta}))^2}{np_j(\boldsymbol{\vartheta})}$$

Zunächst werden die Parameter geschätzt, durch ML-Schätzung oder durch Minimierung von $T(\boldsymbol{\vartheta})$:

$$T(\tilde{\boldsymbol{\vartheta}}) = \min_{\boldsymbol{\vartheta}} T(\boldsymbol{\vartheta})$$

Der Ablehnungsbereich kann dann mit Hilfe des folgenden Satzes bestimmt werden.

☛ **Satz 7.3.3.**
Werden m Parameter aus der Stichprobe geschätzt, so ist $T(\tilde{\boldsymbol{\vartheta}})$ asymptotisch χ^2-verteilt mit $k-1-m$ Freiheitsgraden.

Soll der Test Signifikanzniveau α haben, wird H_0 abgelehnt, wenn:

$$T \geq \chi^2_{1-\alpha|k-1-m}$$

Wird die Nullhypothese nicht abgelehnt und wird nur ein Parameter ϑ geschätzt, kann der Standardfehler von $\tilde{\vartheta}$ näherungsweise aus der Funktion $T(\vartheta)$ abgelesen werden, vorausgesetzt, dass die Funktion in einer Umgebung des Minimums gut durch eine Parabel angenähert weden kann. Ist dies der Fall, gibt es zwei Werte $\vartheta_1 < \vartheta_2$ mit:

$$T(\vartheta_i) = T(\tilde{\vartheta}) + 1, \ i = 1, 2$$

Der Standardfehler von $\tilde{\vartheta}$ kann dann folgendermaßen abgeschätzt werden:

$$\sigma[\tilde{\vartheta}] = \frac{\vartheta_2 - \vartheta_1}{2}$$

▶ **Beispiel 7.3.4.**
Bei der Messung der Hintergrundstrahlung in einem Labor wurden 100 Wartezeiten $X_1, \ldots, X_{100}$ zwischen den Signalen gemessen. Eine Einteilung in fünf Gruppen ergab folgende Häufigkeiten Y_j:

Gruppe	1	2	3	4	5
X	$(0,1]$	$(1,2]$	$(2,4]$	$(4,6]$	>6
Y	34	15	23	12	16

Es soll getestet werden, ob die gemessenen Zeiten exponentialverteilt sind. Der unbekannte Mittelwert τ der Exponentialverteilung wird durch numerische Minimierung der Funktion $T(\tau)$ geschätzt. Abbildung 7.10 zeigt die Funktion und das Minimum bei $\tilde{\tau} = 3.0648$. Die Testgröße, also der Funktionswert am Minimum, ist gleich $T(\tilde{\tau}) = 3.1328$. Das 95%-Quantil der χ^2-Verteilung mit drei Freiheitsgraden ist gleich $\chi^2_{0.95|3} = 8.987$. Die Hypothese, dass die Beobachtungen exponentialverteilt sind, kann also mit dem Signifikanzniveau $\alpha = 0.05$ aufrecht erhalten werden.

Die Werte von τ mit $T(\tau_i) = T(\tilde{\tau}) + 1$ sind $\tau_1 = 2.776$ bzw. $\tau_2 = 3.408$. Daraus ergibt sich die Abschätzung des Standardfehlers von $\tilde{\tau}$ zu $\sigma[\tilde{\tau}] \approx 0.316$. Abbildung 7.10 zeigt jedoch, daß die Funktion $T(\tau)$ im Intervall $[\tau_1, \tau_2]$ noch eine leichte Asymmetrie aufweist.

Wird τ aus den ungruppierten Daten geschätzt, ist der ML-Schätzer von τ gleich $\hat{\tau} = \overline{X} = 3.117$, und der Standardfehler ungefähr gleich $\sigma[\hat{\tau}] \approx \tau/10 = 0.3117$.

➠MATLAB: `make_chi2test_exponential` ◀

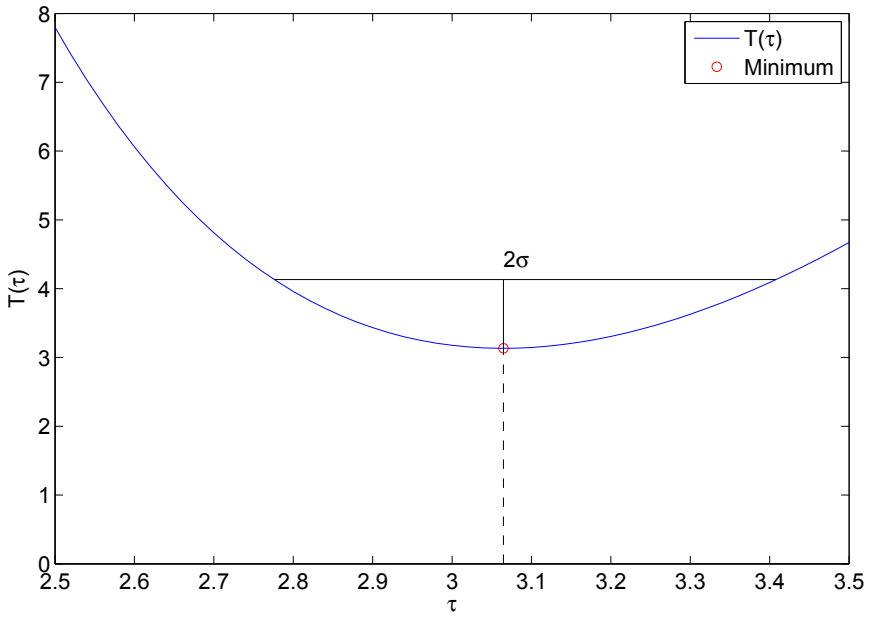

Abbildung 7.10: Die Funktion $T(\tau)$. Der Schätzwert $\tilde{\tau}$ ist die Stelle des Minimums der Funktion. MATLAB: `make_chi2test_exponential`

7.3.2 Der Kolmogorow-Smirnow-Test

7.3.2.1 Eine Stichprobe

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus einer stetigen Verteilung mit der Verteilungsfunktion F. Die Hypothese $\mathrm{H}_0 : F(x) = F_0(x)$ soll getestet werden. Die Testgröße D_n ist die maximale absolute Abweichung der empirischen Verteilungsfunktion $F_n(x)$ der Stichprobe (siehe Definition 5.2.1) von der hypothetischen Verteilungsfunktion $F_0(x)$:

$$D_n = \max_x |F_n(x) - F_0(x)|$$

Abbildung 7.11 zeigt beispielhaft die empirische Verteilungsfunktion F_n, die hypothetische Verteilungsfunktion F_0, und die Testgröße D_n.

Für Stichproben aus F_0 ist die Verteilung von D_n unabhängig von F_0, und die Verteilungsfunktion von $\sqrt{n}D_n$ strebt für $n \to \infty$ gegen:

$$K(x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}\mathrm{e}^{-2k^2x^2}$$

Aus der asymptotischen Verteilungsfunktion können Quantile $K_{1-\alpha}$ berechnet werden. Die Nullhypothese wird abgelehnt, wenn

$$\sqrt{n}D_n > K_{1-\alpha}$$

Werden vor dem Test Parameter von F_0 geschätzt, sind die Quantile nicht mehr gültig. In diesem Fall muss der Ablehnungsbereich durch Simulation ermittelt werden.

Der Test ist nach den russischen Mathematikern A. Kolmogorow* und N. Smirnow† benannt.

➠MATLAB: `kstest`

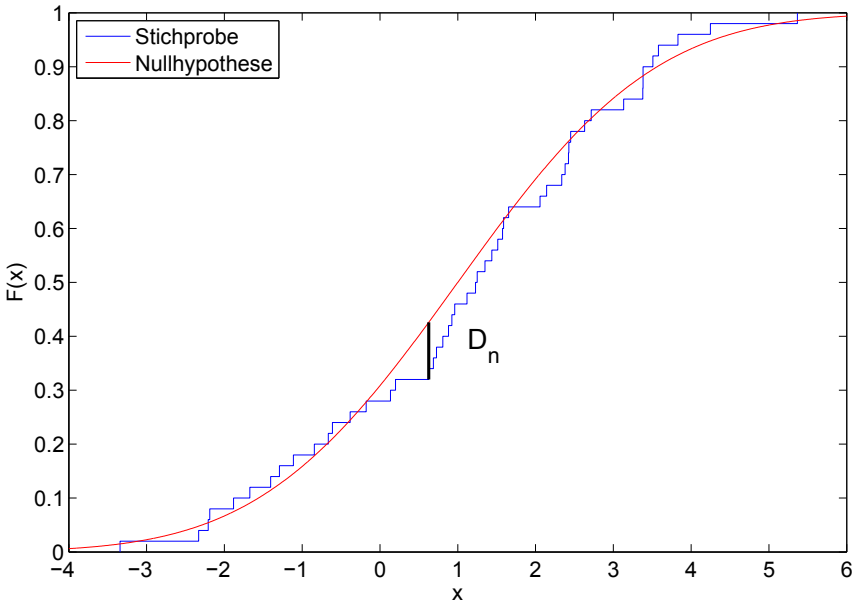

Abbildung 7.11: Stichprobe, Nullhypothese und Testgröße des Kolmogorow-Smirnow-Tests. MATLAB: `make_kstest_normal`

*http://de.wikipedia.org/wiki/Andrei_Nikolajewitsch_Kolmogorow
†http://en.wikipedia.org/wiki/Nikolai_Smirnov_(mathematician)

7.3.2.2 Zwei Stichproben

Es soll getestet werden, ob zwei Stichproben vom Umfang n bzw. m aus der gleichen unbekannten Verteilung F stammen. Die Testgröße ist die maximale absolute Differenz der beiden empirischen Verteilungsfunktionen:

$$D_{n,m} = \max_x \left| F_n^1(x) - F_m^2(x) \right|$$

Die Nullhypothese wird abgelehnt, wenn

$$\sqrt{\frac{nm}{n+m}} D_{n,m} > K_{1-\alpha}$$

➠MATLAB: `kstest2`

Kapitel 8

Elemente der Bayes-Statistik

8.1 Einleitung

In der Bayes-Statistik werden Wahrscheinlichkeiten nicht als Grenzwerte von relativen Häufigkeiten, sondern als rationale Einschätzungen von Sachverhalten interpretiert. Während im frequentistischen Ansatz die Beobachtungen als zufällige und die Parameter von Verteilungen als unbekannte, aber feste Größen betrachtet werden, ist es in der Bayes-Statistik in gewisser Weise umgekehrt: die Beobachtungen werden als gegebene feste Größen angesehen, und die aktuelle Information über die Verteilungsparameter wird durch eine geeignete Dichtefunktion beschrieben. Die Vorinformation über die Parameter wird in einer *a-priori-Dichte* zusammengefasst; die Wahrscheinlichkeit der Beobachtung bedingt durch einen bestimmten Wert der Parameter wird durch die Likelihood-Funktion ausgedrückt. Mit Hilfe des Satzes von Bayes (siehe Satz 4.4.11) wird die Information über die Parameter anhand der vorliegenden Beobachtungen aktualisiert und in der *a-posteriori-Dichte* zusammengefasst. Die a-posteriori-Dichte kann dann zum Schätzen der Parameter, zur Berechnung von Vertrauensintervallen, zum Testen, zur Prognose, zur Modellwahl etc. verwendet werden.

Durch die Wahl der a-priori-Verteilung wird eine subjektive Komponente in die Datenanalyse eingeführt. Dies kann besonders bei kleiner Anzahl von Beobachtungen einen merkbaren Einfluss auf das Resultat haben. Es gibt jedoch Verfahren zur Konstruktion von a-priori-Verteilungen, die diesen Einfluss minimieren. Liegen viele Beobachtungen vor und bleibt die Anzahl der Parameter gleich, so wird der Einfluss der a-priori-Verteilung im Allgemeinen vernachlässigbar. In diesem Fall liefert die Bayes-Analyse annähernd die gleichen Resultate wie klassische "frequentistische" Verfahren. In der a-posteriori-Dichte ist jedoch mehr Information enthalten als in klassischen Punktschätzern oder Konfidenzintervallen.

8.2 Grundbegriffe

Der *Stichprobenraum* $\mathcal{Y}$ ist die Menge aller möglichen Stichproben $\boldsymbol{y}$. Da die Beobachtungen in diesem Kapitel als feste Größen betrachtet werden, werden sie mit Kleinbuchstaben notiert. Der *Parameterraum* Θ ist die Menge aller möglichen Werte des Parameters ϑ. Die *a-priori-Dichte* $\pi(\vartheta)$ beschreibt die Einschätzung, ob ein bestimmter Wert ϑ die wahre Verteilung der Beobachtungen beschreibt. Die *Likelihood-Funktion* $p(\boldsymbol{y}\,|\,\vartheta)$ beschreibt die Wahrscheinlichkeit, dass die Stichprobe $\boldsymbol{y}$ beobachtet wird, wenn ϑ wahr ist. Wird $\boldsymbol{y}$ beobachtet, kann die Einschätzung von ϑ aktualisiert werden, indem mit dem Satz von Bayes (siehe Satz 4.4.11) die *a-posteriori-Dichte* berechnet wird:

$$p(\vartheta\,|\,\boldsymbol{y}) = \frac{p(\boldsymbol{y}\,|\,\vartheta)\pi(\vartheta)}{\int_\Theta p(\boldsymbol{y}\,|\,\vartheta)\pi(\vartheta)\,\mathrm{d}\vartheta}$$

Die a-posteriori-Verteilung beschreibt die Einschätzung von ϑ im Lichte der Beobachtungen $\boldsymbol{y}$. Kommen neue Beobachtungen dazu, kann die a-posteriori-Dichte als die neue a-priori-Dichte benützt werden. Kann das Integral im Nenner nicht analytisch berechnet werden, muss man zu numerischen oder Monte-Carlo-Methoden greifen. Als Punktschätzer von ϑ werden häufig der *Mittelwert (posterior mean)* oder der *Modus (posterior mode)* von $p(\vartheta\,|\,\boldsymbol{y})$ verwendet, gelegentlich auch der *Median (posterior median)*. Analog zu Konfidenzintervallen können aus der a-posteriori-Dichte *Vertrauensintervalle* berechnet werden. Eine weitergehende Einführung in die Prinzipien der Bayes-Statistik bietet unter vielen anderen das Buch von Hoff (siehe Anhang L).

8.3 A-priori-Verteilungen

Die Wahl der a-priori-Verteilung kann im Fall von kleinen Stichproben das Ergebnis merkbar beeinflussen. Man unterscheidet *informative* und *nicht informative* a-priori-Verteilungen. Eine informative a-priori-Verteilung beschreibt tatsächliche Information über den unbekannten Parameter ϑ. Diese kann subjektiv oder objektiv sein. Eine nicht informative a-priori-Verteilung beschreibt die Abwesenheit jeglicher solcher Information. In jedem Fall sollte die Sensitivität des Resultats bezüglich der a-priori-Verteilung untersucht werden.

Die a-priori-Dichte braucht nicht normiert zu sein, da sich im Satz von Bayes der Normierungsfaktor wegkürzt. Kann die a-priori-Dichte normiert werden, wird sie als *proper* bezeichnet, sonst als *improper*. Auch eine impropere a-priori-Dichte kann zu einer properen a-posteriori-Dichte führen. Die Wahl der a-priori-Verteilung wird auch oft von rein rechnerischen Überlegungen beeinflusst. Für einige Formen der Likelihood-Funktion gibt es a-priori-Verteilungen, die a-posteriori-Verteilungen aus der gleichen Verteilungsfamilie erzeugen. Solche a-priori-Verteilungen heißen *konjugiert*. In einigen Fällen ist eine nicht informative a-priori-Verteilung auch die konjugierte a-priori-Verteilung.

Es gibt verschiedene Vorschläge zur Wahl einer nicht informativen a-priori-Verteilung, unter anderem die folgenden:

- Prinzip der *maximalen Entropie*
- Invarianz unter Transformationen des Parameters: *Jeffrey's prior*
- Minimaler Einfluss der a-priori-Verteilung auf die a-posteriori-Verteilung: *reference prior*

Verteilungen mit maximaler Entropie in einer gegebenen Klasse von Verteilungen minimieren das Ausmaß an Information, das der Verteilung inhärent ist. Für eine diskrete Verteilung F mit Wahrscheinlichkeitsfunktion $f(k)$ ist die Entropie definiert durch:

$$\mathrm{Ent}(F) = -\sum_{k=1}^{\infty} f(k) \ln f(k)$$

Für eine stetige Verteilung F mit Dichtefunktion $f(x)$ ist die Entropie definiert durch:

$$\mathrm{Ent}(F) = -\int_{\mathrm{A}} f(x) \ln f(x)\,\mathrm{d}x$$

mit $\mathrm{A} = \{x \in \mathbf{R}\,|\,f(x) > 0\}$.

☛ **Satz 8.3.1.** Verteilungen mit maximaler Entropie

Stetige Verteilungen:

- $\mathsf{E}[X]$ *und* $\mathsf{var}[X]$ *gegeben: Normalverteilung*
- $X \geq 0$, $\mathsf{E}[\ln X]$ *und* $\mathsf{var}[\ln X]$ *gegeben: Log-Normalverteilung*
- $X \geq 0$, $\mathsf{E}[X]$ *und* $\mathsf{E}[\ln X]$ *gegeben: Gammaverteilung*
- $X \geq 0$, $\mathsf{E}[X]$ *gegeben: Exponentialverteilung*
- $X \in [a, b]$*: Gleichverteilung auf* $[a, b]$

Diskrete Verteilungen:

- $X \in \{1, \ldots, n\}$*: Gleichverteilung auf* $\{1, \ldots, n\}$
- $X \in \mathbb{N}$, $\mathsf{E}[X]$ *gegeben: Geometrische Verteilung*

Jeffrey's prior wird so konstruiert, dass die a-priori-Dichte invariant unter einer Transformation des Parameters ϑ ist. Es sei $\tau = h(\vartheta)$ der transformierte Parameter und $\pi_{\mathrm{J}}(\vartheta)$ Jeffrey's prior in ϑ. Dann ist die transformierte a-priori-Dichte in τ gleich (siehe Satz 4.3.1):

$$\pi(\tau) = \pi_{\mathrm{J}}(\vartheta) \left| \frac{\mathrm{d}\vartheta}{\mathrm{d}\tau} \right|$$

Die transformierte Fisher-Information ist gleich:

$$I(\tau) = I(\vartheta) \left(\frac{\mathrm{d}\vartheta}{\mathrm{d}\tau} \right)^2$$

Mit der Wahl von Jeffrey's prior als:

$$\pi_{\mathrm{J}}(\vartheta) \propto \sqrt{I(\vartheta)}$$

ergibt sich das gewünschte Transformationsverhalten. Jeffrey's prior und reference prior sind für eindimensionales ϑ meist identisch; im mehrdimensionalen Fall gilt dies nicht mehr. Wir gehen hier auf die Konstruktion der reference prior nicht ein.

8.4 Binomialverteilte Beobachtungen

Ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit ϑ wird n mal wiederholt. Die Anzahl Y der Erfolge ist dann binomialverteilt gemäß $\mathrm{Bi}(n, \vartheta)$. Aus einer Beobachtung y von Y soll eine Aussage über die Erfolgswahrscheinlichkeit ϑ gewonnen werden. Dazu wird eine a-priori-Verteilung von ϑ benötigt.

8.4.1 A-priori-Gleichverteilung

Das Prinzip der maximalen Entropie ergibt die Gleichverteilung $\mathrm{Un}(0, 1)$ als a-priori-Verteilung:

$$\pi(\vartheta) = \mathbb{1}_{[0,1]}(\vartheta)$$

Die Likelihood-Funktion ist gleich:

$$p(y \,|\, \vartheta) = \binom{n}{y} \vartheta^y (1 - \vartheta)^{n-y}$$

Der Satz von Bayes liefert die a-posteriori-Dichte:

$$p(\vartheta \,|\, y) \propto \vartheta^y (1 - \vartheta)^{n-y} \mathbb{1}_{[0,1]}(\vartheta)$$

Da die a-posteriori-Dichte proportional zur Dichtefunktion der Betaverteilung $\mathrm{Be}(y + 1, n - y + 1)$ ist (siehe Definition 4.2.24), muss sie mit dieser identisch sein. Der Mittelwert der a-posteriori-Verteilung ist gleich:

$$\mathsf{E}[\vartheta \,|\, y] = \frac{y + 1}{n + 2}$$

Der Modus der a-posteriori-Verteilung ist gleich:

$$\hat{\vartheta} = \frac{y}{n}$$

und damit der Maximum-Likelihood-Schätzer von ϑ. Abbildung 8.1 zeigt die a-posteriori-Dichten von ϑ für $n = 20$ und einige Werte von y.

➠MATLAB: `make_posterior_binomial`

8.4.2 Jeffrey's prior

Die Fisher-Information einer Beobachtung ist gegeben durch:

$$I(\vartheta) = \frac{n}{\vartheta(1 - \vartheta)} \mathbb{1}_{[0,1]}(\vartheta)$$

Jeffrey's prior ist also die Betaverteilung $\mathrm{Be}(0.5, 0.5)$. Der Satz von Bayes liefert wieder die a-posteriori-Dichte:

$$p(\vartheta \,|\, y) \propto \vartheta^{y-0.5}(1-\vartheta)^{n-y-0.5}\mathbb{1}_{[0,1]}(\vartheta)$$

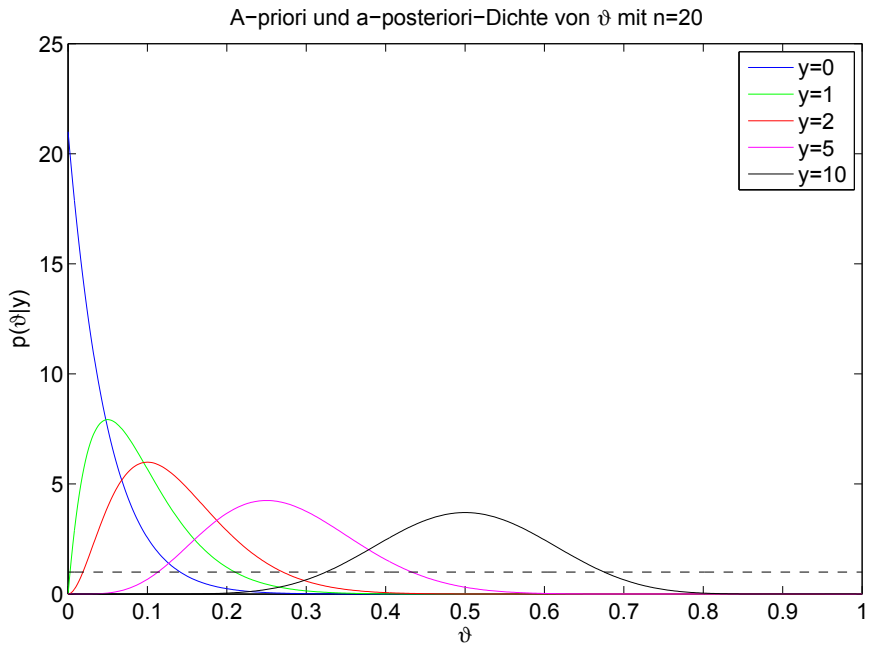

Abbildung 8.1: A-priori-Dichte (strichliert) und a-posteriori-Dichten für $n = 20$ und einige Werte von y. MATLAB: `make_posterior_binomial`

Da die a-posteriori-Dichte proportional zur Dichtefunktion der Betaverteilung $\mathrm{Be}(y + 0.5, n - y + 0.5)$ ist, muss sie mit dieser identisch sein. Der Mittelwert der a-posteriori-Verteilung ist gleich:

$$\mathsf{E}[\vartheta \,|\, y] = \frac{y + 0.5}{n + 1}$$

Der Modus der a-posteriori-Verteilung ist gleich:

$$\hat{\vartheta} = \frac{y - 0.5}{n - 1}$$

Abbildung 8.2 zeigt die a-posteriori-Dichten von ϑ für $n = 20$ und einige Werte von y.

➠MATLAB: `make_posterior_binomial`

8.4.3 Informative a-priori-Verteilungen

Liegt Vorinformation über ϑ vor, kann diese durch eine geeignete a-priori-Verteilung inkludiert werden. Die rechnerisch einfachste Behandlung ergibt sich, wenn auch die a-priori-Verteilung eine Betaverteilung ist. Sei also:

$$\pi(\vartheta) = \frac{\vartheta^{a-1}(1-\vartheta)^{b-1}}{B(a, b)}$$

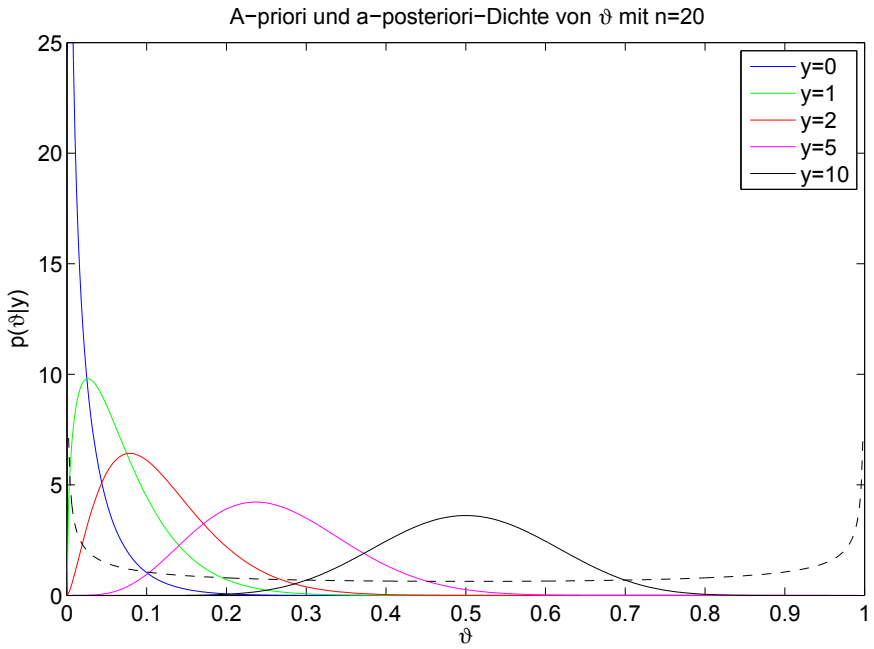

Abbildung 8.2: A-priori-Dichte (strichliert) und a-posteriori-Dichten für $n = 20$ und einige Werte von y. MATLAB: `make_posterior_binomial`

Dann ist die a-posteriori-Dichte gleich:

$$\begin{aligned} p(\vartheta \,|\, y) &\propto \vartheta^y (1-\vartheta)^{n-y} \vartheta^{a-1} (1-\vartheta)^{b-1} \\ &= \vartheta^{y+a-1} (1-\vartheta)^{n-y+b-1} \end{aligned}$$

Die a-posteriori-Verteilung ist wieder eine Betaverteilung, nämlich $\mathrm{Be}(y+a, n-y+b)$. Die Betaverteilung ist daher die konjugierte a-priori-Verteilung zur Binomialverteilung. Der Mittelwert der a-posteriori-Verteilung ist gleich:

$$\mathsf{E}[\vartheta\,|\,y] = \frac{a+y}{a+b+n}$$

Der Modus der a-posteriori-Verteilung ist gleich:

$$\hat{\vartheta} = \frac{a+y-1}{a+b+n-2}$$

Der Mittelwert der a-posteriori-Verteilung kann als gewichtetes Mittel aus a-priori-Information und Beobachtungen angeschrieben werden:

$$\begin{aligned}\mathsf{E}[\vartheta\,|\,y] =& \frac{a+y}{a+b+n} = \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{y}{n}\\ =& \frac{a+b}{a+b+n} \times \text{a-priori-Erwartung}\\ &+ \frac{n}{a+b+n} \times \text{Mittel der Beobachtungen}\end{aligned}$$

a und b können als "a-priori-Beobachtungen" interpetiert werden. a ist die Anzahl der Erfolge a-priori, b ist dieAnzahl der Misserfolge a-priori, und $a+b$ ist die Anzahl der Versuche a-priori.

Abbildungen 8.3 und 8.4 zeigen den Einfluss einer informativen a-priori-Dichte auf die a-posteriori-Dichte. Korrespondierende a-priori-Dichten und a-posteriori-Dichten sind in der gleichen Farbe dargestellt. Während für $n = 10$ noch ein Einfluss sichtbar ist, sind für $n = 100$ alle a-posteriori-Dichten praktisch identisch.

➠MATLAB: `make_posterior_binomial`

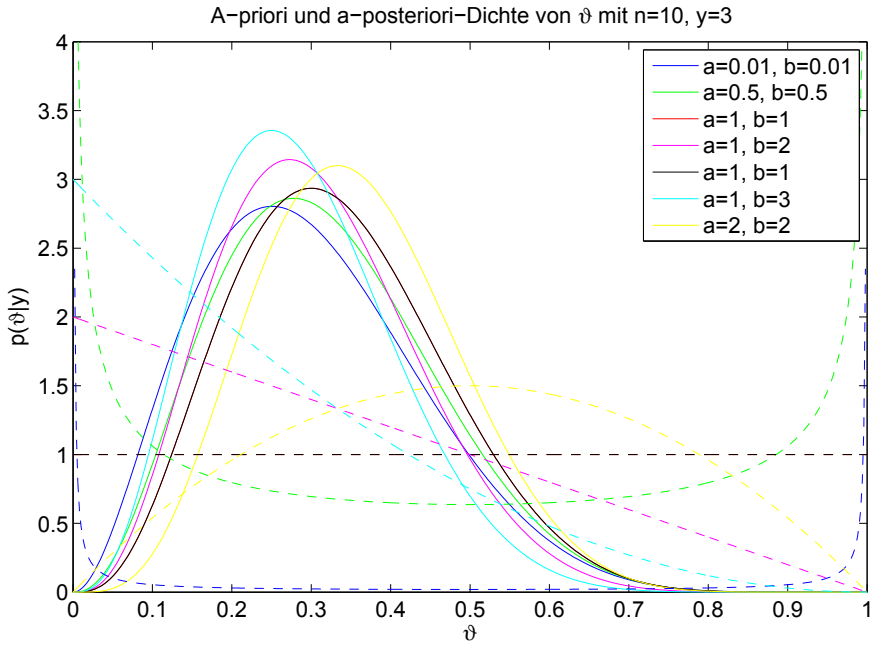

Abbildung 8.3: A-posteriori-Dichten für verschiedene informative a-priori-Dichten (strichliert) mit $n = 10, y = 3$. MATLAB: `make_posterior_binomial`

8.4.4 Haldane's prior

Wird die Anzahl der Versuche a-priori gleich 0 gesetzt, erhält man *Haldane's prior*, die wieder nicht informativ ist:

$$\pi_{\mathrm{H}}(\vartheta) = \frac{1}{\vartheta(1-\vartheta)}$$

Haldane's prior kann als $\mathrm{Be}(0,0)$ interpretiert werden, ist allerdings improper. Der a-posteriori-Mittelwert ist dann gleich dem ML-Schätzer. Paradoxerweise gibt die a-priori-Dichte ohne jegliche Vorinformation den Werten $\vartheta = 0$ und $\vartheta = 1$ die größte Wahrscheinlichkeit. Sind alle Versuche Erfolge oder Misserfolge, ist die a-posteriori-Dichte allerdings ebenfalls improper, es können dann also weder Mittelwert noch Median berechnet werden.

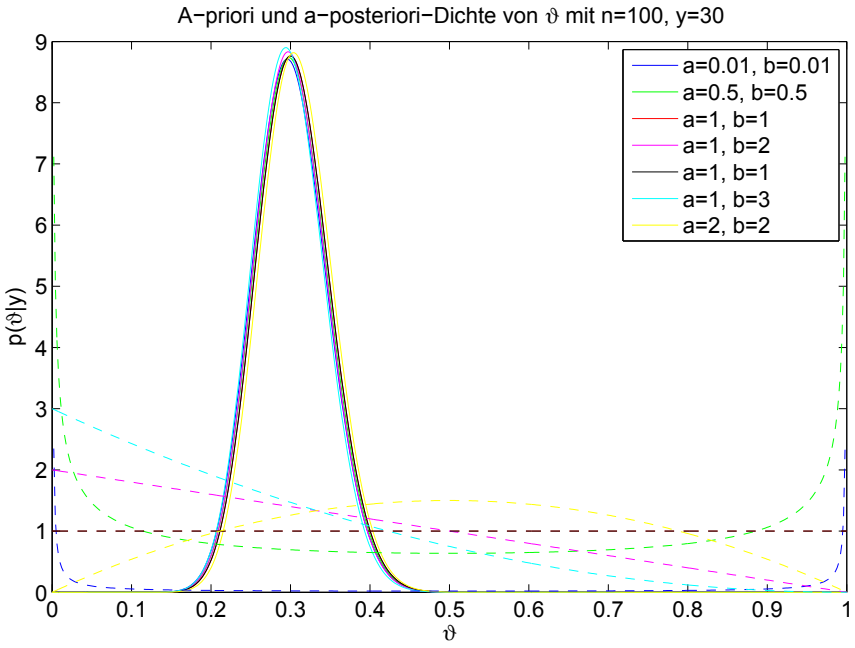

Abbildung 8.4: A-posteriori-Dichten für verschiedene informative a-priori-Dichten (strichliert) mit $n = 100, y = 30$. MATLAB: `make_posterior_binomial`

8.4.5 Vertrauensintervalle und HPD-Bereiche

Es sollen nun Teilbereiche des Parameterraums Θ konstruiert werden, in denen ϑ hohe a-posteriori-Wahrscheinlichkeit $1-\alpha$ hat. Ein solcher Bereich wird ein *Vertrauensbereich* genannt. Er ist meistens ein Intervall (Vertrauensintervall, engl. credible interval). Die einfachste Konstruktion eines Vertrauensintervalls $[\vartheta_1(y), \vartheta_2(y)]$ benützt die Quantile der a-posteriori-Dichte. Das symmetrische Vertrauensintervall lautet:

$$\vartheta_1(y) = Q_{\alpha/2}, \quad \vartheta_2(y) = Q_{1-\alpha/2}$$

wo Q_q das q-Quantil der a-posteriori-Dichte $p(\vartheta \,|\, y)$ ist. Selbstverständlich können auch asymmetrische und einseitige Vertrauensintervalle konstruiert werden.

▶ **Beispiel 8.4.1.** VERTRAUENSINTERVALL

Es sei $y = 4$ die Anzahl der Erfolge in $n = 20$ unabhängigen Alternativversuchen mit Erfolgswahrscheinlichkeit ϑ. Mit der Gleichverteilung als a-priori-Verteilung ist die a-posteriori-Verteilung von ϑ eine $\mathrm{Be}(5, 17)$-Verteilung. Das symmetrische Vertrauensintervall mit $1 - \alpha = 0.95$ hat dann die Grenzen:

$$\vartheta_1(y) = \beta_{0.025|5,17} = 0.0822$$
$$\vartheta_2(y) = \beta_{0.975|5,17} = 0.4191$$

Der Mittelwert der a-posteriori-Verteilungist gleich:

$$\mathsf{E}[\vartheta\,|\,y] = \frac{5}{22} = 0.2273$$

Der Modus der a-posteriori-Verteilung ist gleich:

$$\hat{\vartheta} = \frac{4}{20} = 0.2$$

Abbildung 8.5 zeigt die a-posteriori-Dichte und das Vertrauensintervall.

➠MATLAB: `make_posterior_binomial` ◀

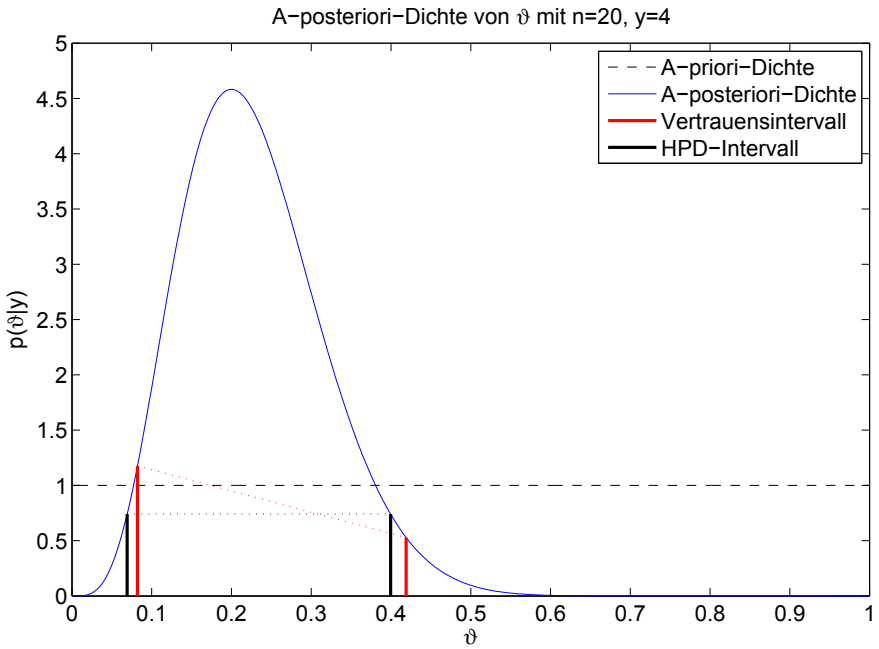

Abbildung 8.5: A-posteriori-Dichte, symmetrisches Vertrauensintervall und HPD-Intervall mit $n = 20, y = 4, \alpha = 0.05$, a-priori-Gleichverteilung.
MATLAB: `make_posterior_binomial`

► **Beispiel 8.4.2.** VERTRAUENSINTERVALL MIT JEFFREY'S PRIOR

Mit Jeffrey's prior ist die a-posteriori-Verteilung von ϑ eine $\mathrm{Be}(4.5, 16.5)$-Verteilung. Das symmetrische Vertrauensintervall mit $1 - \alpha = 0.95$ hat dann die Grenzen:

$$\vartheta_1(y) = \beta_{0.025|4.5,16.5} = 0.0715$$
$$\vartheta_2(y) = \beta_{0.975|4.5,16.5} = 0.4082$$

Der Mittelwert der a-posteriori-Verteilung ist gleich:

$$\mathsf{E}[\vartheta\,|\,y] = \frac{4.5}{21} = 0.2143$$

Der Modus der a-posteriori-Verteilung ist gleich:

$$\hat{\vartheta} = \frac{3.5}{19.5} = 0.1795$$

Abbildung 8.6 zeigt die a-posteriori-Dichte und das Vertrauensintervall.

➠MATLAB: `make_posterior_binomial` ◄

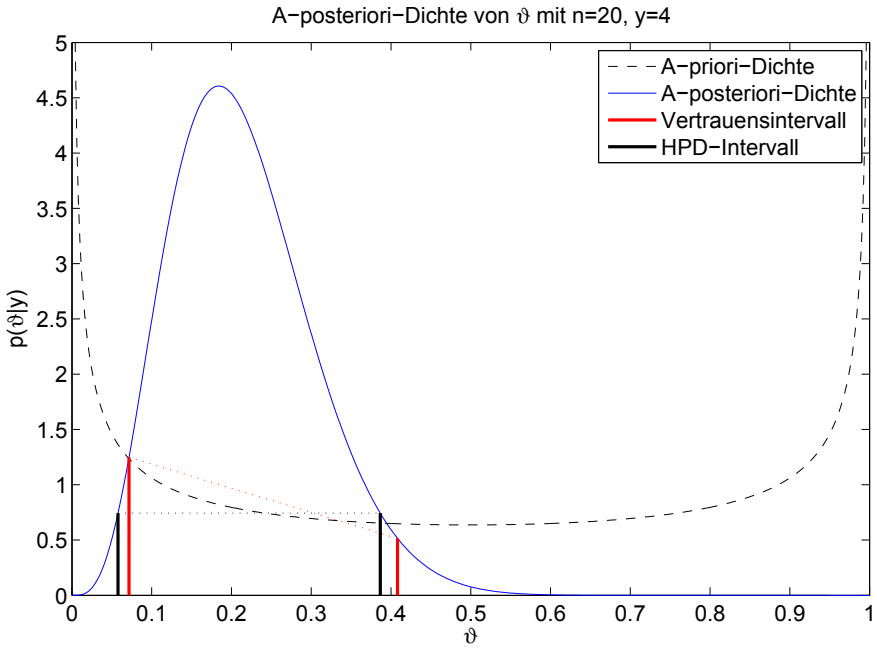

Abbildung 8.6: A-posteriori-Dichte, symmetrisches Vertrauensintervall und HPD-Intervall mit $n = 20, y = 4, \alpha = 0.05$, Jeffrey's prior.
MATLAB: `make_posterior_binomial`

Das symmetrische Vertrauensintervall enthält Werte von ϑ, die geringere a-posteriori-Wahrscheinlichkeit haben als gewisse Werte außerhalb des Intervalls. Ein Bereich, in dem alle Werte von ϑ höhere a-posteriori-Wahrscheinlichkeit haben als alle Werte außerhalb des Bereichs, heißt ein High-Posterior-Density-Bereich, kurz *HPD-Bereich*. Ist die a-posteriori-Dichte unimodal, ist der HPD-Bereich ein HPD-Intervall. In diesem Fall erhält man die Grenzen ϑ_1, ϑ_2 des HPD-Intervalls als Lösung des Gleichungssystems:

$$\begin{aligned} p(\vartheta_2 \,|\, y) - p(\vartheta_1 \,|\, y) &= 0 \\ P(\vartheta_2 \,|\, y) - P(\vartheta_1 \,|\, y) &= 1 - \alpha \end{aligned}$$

Hier ist $P(\vartheta \,|\, y)$ die a-posteriori-Verteilungsfunktion. Das Gleichungssystem muss in der Regel numerisch gelöst werden.

▶ **Beispiel 8.4.3.** HPD-Intervall

Das HPD-Intervall mit $1 - \alpha = 0.95$ für den Versuch in Beispiel 8.4.1 und a-priori-Gleichverteilung hat die Grenzen:

$$\vartheta_1(y) = 0.06921, \quad \vartheta_2(y) = 0.3995$$

Es ist mit einer Länge von 0.3303 kürzer als das symmetrische Vertrauensintervall, das eine Länge von 0.3369 hat. Abbildungen 8.5 und 8.6 zeigen das HPD-Intervall für a-priori-Gleichverteilung bzw. Jeffrey's prior.

➠MATLAB: `make_posterior_binomial` ◀

▶ **Beispiel 8.4.4.**

Ein Alternativversuch wird $n = 20$ mal wiederholt und ergibt keinen einzigen Erfolg ($y = 0$). Was kann man über die Erfolgswahrscheinlichkeit ϑ aussagen?

Mit der a-priori-Dichte $\pi(\vartheta) = 1$ ist die a-posteriori-Verteilung $\mathrm{Be}(1, 21)$. Der Modus ist gleich 0, der Mittelwert ist gleich 0.0455. Das HPD-Intervall mit $1 - \alpha = 0.95$ ist gleich $[0, 0.1329]$ und ist identisch mit dem einseitigen Vertrauensintervall. Mit Jeffrey's prior ist die a-posteriori-Verteilung $\mathrm{Be}(0.5, 20.5)$. Der Modus ist gleich 0, der Mittelwert ist gleich 0.0238. Das HPD-Intervall mit $1 - \alpha = 0.95$ ist gleich $[0, 0.0905]$. Ist bekannt, dass ϑ eher klein ist, kann z.B. $\mathrm{Be}(0.5, 5)$ als a-priori-Verteilung gewählt werden. Die a-posteriori-Verteilung ist dann $\mathrm{Be}(0.5, 25)$. Der Modus ist gleich 0, der Mittelwert gleich 0.0196, und das HPD-Intervall gleich $[0, 0.0747]$.

Zum Vergleich werden die „frequentistischen" Resultate angeführt. Der ML-Schätzer von ϑ ist gleich 0. Das einseitige Konfidenzintervall nach Clopper-Pearson (siehe Abschnitt 6.4.3) ist gleich $[0, 0.1391]$. Die Näherung durch Normalverteilung ist nur sinnvoll mit der Korrektur gemäß Agresti-Coull, das sonst das Konfidenzintervall auf den Nullpunkt schrumpft. Die Schätzung ist dann $\hat{\vartheta} = 0.0833$, das linksseitige Konfidenzintervall ist $[0, 0.1850]$. Das symmetrische Konfidenzintervall ist $[-0.0378, 0.2045]$. Es ist problematisch, da es auch negative Werte von ϑ enthält. ◀

8.5 Poissonverteilte Beobachtungen

Die Beobachtung eines Poissonprozesses ergibt die Werte $\boldsymbol{y} = (y_1, \ldots, y_n)$ mit der Summe $s = \sum y_i$. Aus den Beobachtungen soll eine Einschätzung der Intensität $\lambda > 0$ des Prozesses gewonnen werden. Die Likelihood-Funktion lautet:

$$p(\boldsymbol{y} \,|\, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} \mathrm{e}^{-\lambda}}{y_i!} = \prod_{i=1}^{n} \frac{1}{y_i!} \cdot \lambda^s \mathrm{e}^{-n\lambda}$$

Ihre Form hängt von den Beobachtungen nur über $s = \sum y_i$ ab, da der Vorfaktor nur in die Normierung eingeht. Als nicht informative a-priori-Dichten kommen in Frage: die impropere a-priori-Dichte $\pi(\lambda) = \mathbb{1}_{(0,\infty)}(\lambda)$, oder Jeffrey's prior $\pi_J(\lambda) = \lambda^{-1/2},\ \lambda > 0$, die ebenfalls improper ist. Im folgenden wird stets $\lambda > 0$ angenommen.

8.5.1 A-priori-Gleichverteilung

Mit $\pi(\lambda) = \mathbb{1}_{(0,\infty)}$ ist die a-posteriori-Dichte proportional zur Likelihood-Funktion:

$$p(\lambda\,|\,s) \propto \lambda^s \mathrm{e}^{-n\lambda}$$

Da die a-posteriori-Dichte proportional zur Dichtefunktion der Gammaverteilung $\mathrm{Ga}(s+1, 1/n)$ ist, muss sie mit dieser identisch sein:

$$p(\lambda\,|\,s) = \frac{\lambda^s \mathrm{e}^{-n\lambda}}{n^{-(s+1)}\Gamma(s+1)}$$

Der Mittelwert der a-posteriori-Verteilung ist gleich

$$\mathsf{E}[\lambda\,|\,s] = \frac{s+1}{n}$$

Der Modus der a-posteriori-Verteilung ist gleich

$$\hat{\lambda} = \frac{s}{n}$$

und damit der Maximum-Likelihood-Schätzer von λ. Abbildung 8.7 zeigt a-posteriori-Dichten für $n = 1$ und einige Werte von y.

➠MATLAB: `make_posterior_poisson`

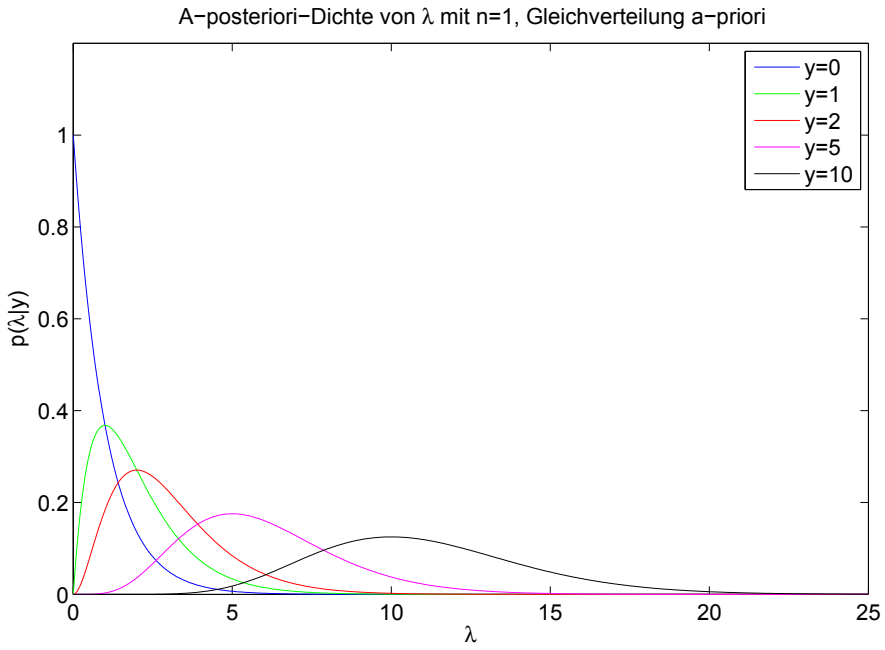

Abbildung 8.7: A-posteriori-Dichten für $n = 1$ und einige Werte von y, mit a-priori-Gleichverteilung. MATLAB: `make_posterior_poisson`

8.5.2 Jeffrey's prior

Mit $\pi(\lambda) = \lambda^{-1/2}$ lautet die a-posteriori-Dichte:

$$p(\lambda|s) \propto \lambda^{s-0.5}\mathrm{e}^{-n\lambda}$$

Da die a-posteriori-Dichte proportional zur Dichtefunktion der Gammaverteilung $\mathrm{Ga}(s+0.5, 1/n)$ ist, muss sie mit dieser identisch sein:

$$p(\lambda|s) = \frac{\lambda^{s-0.5}\mathrm{e}^{-n\lambda}}{n^{-(s+0.5)}\Gamma(s+0.5)}$$

Der Mittelwert der a-posteriori-Verteilung ist gleich

$$\mathsf{E}[\lambda|s] = \frac{s+0.5}{n}$$

Der Modus der a-posteriori-Verteilung ist gleich

$$\hat{\lambda} = \frac{s-0.5}{n}$$

Abbildung 8.8 zeigt a-posteriori-Dichten für $n = 1$ und einige Werte von y.

➠MATLAB: `make_posterior_poisson`

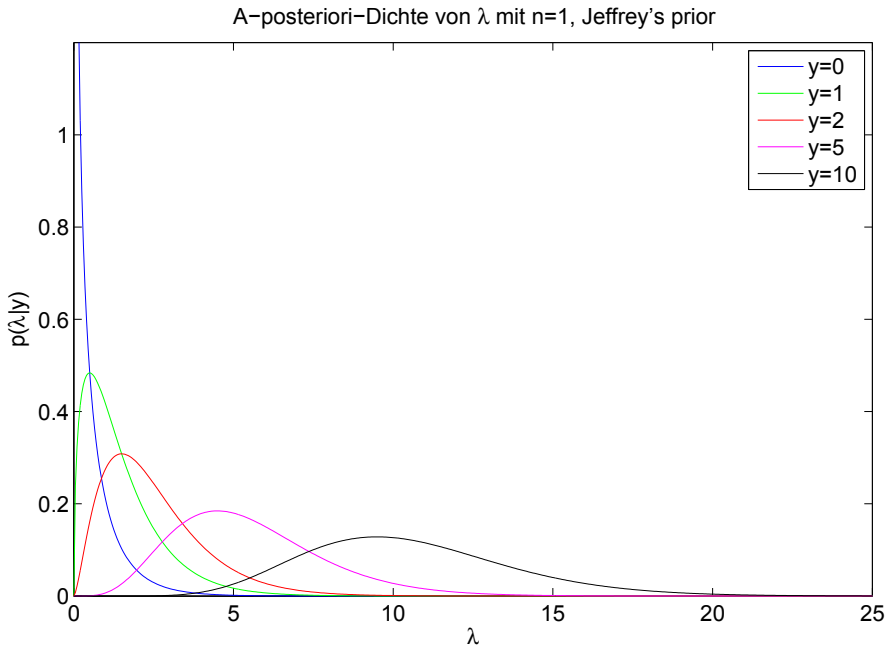

Abbildung 8.8: A-posteriori-Dichten für $n = 1$ und einige Werte von y, mit Jeffrey's prior. MATLAB: `make_posterior_poisson`

8.5.3 Informative a-priori-Verteilungen

Liegt Vorinformation über λ vor, kann diese durch eine geeignete a-priori-Verteilung inkludiert werden. Die rechnerisch einfachste Behandlung ergibt sich, wenn die a-priori-Verteilung eine Gammaverteilung ist. Sei also:

$$\pi(\lambda) = \frac{b^a \lambda^{a-1} \mathrm{e}^{-b\lambda}}{\Gamma(a)}$$

Dies ist die Dichtefunktion der Gammaverteilung $\mathrm{Ga}(a, 1/b)$. Dann ist die a-posteriori-Dichte gleich:

$$\begin{aligned} p(\lambda \,|\, s) &\propto \lambda^s \mathrm{e}^{-n\lambda} \lambda^{a-1} \mathrm{e}^{-b\lambda} \\ &= \lambda^{s+a-1} \mathrm{e}^{-(b+n)\lambda} \end{aligned}$$

Die a-posteriori-Verteilung ist $\mathrm{Ga}(a+s, 1/(b+n))$. Die Gammaverteilung ist also konjugiert zur Poissonverteilung.

Der Mittelwert der a-posteriori-Verteilung ist gleich

$$\mathsf{E}[\lambda \,|\, s] = \frac{a+s}{b+n}$$

Der Modus der a-posteriori-Verteilung ist gleich

$$\hat{\lambda} = \frac{a+s-1}{b+n}$$

Der Mittelwert der a-posteriori-Verteilung kann als gewichtetes Mittel aus a-priori-Information und Beobachtungen angeschrieben werden:

$$\begin{aligned}\mathsf{E}[\lambda\,|\,s] &= \frac{a+s}{b+n} = \frac{b}{b+n}\frac{a}{b} + \frac{n}{b+n}\frac{s}{n} \\ &= \frac{b}{b+n} \times \text{a-priori-Erwartung} \\ &+ \frac{n}{b+n} \times \text{Mittel der Beobachtungen}\end{aligned}$$

Die a-priori-Parameter a und b können folgendermaßen interpetiert werden: a ist die Summe der Beobachtungen a-priori, b ist die Anzahl der Beobachtungen a-priori. Ist $n \gg b$, dominieren die Beobachtungen:

$$\mathsf{E}[\lambda\,|\,s] \approx \frac{s}{n}$$

Abbildungen 8.9 und 8.10 zeigen den Einfluss einer informativen a-priori-Dichte auf die a-posteriori-Dichte. Korrespondierende a-priori-Dichten und a-posteriori-Dichten sind wieder in der gleichen Farbe dargestellt. Während für $n = 10$ noch ein Einfluss sichtbar ist, sind für $n = 100$ alle a-posteriori-Dichten praktisch identisch.

➠MATLAB: `make_posterior_poisson`

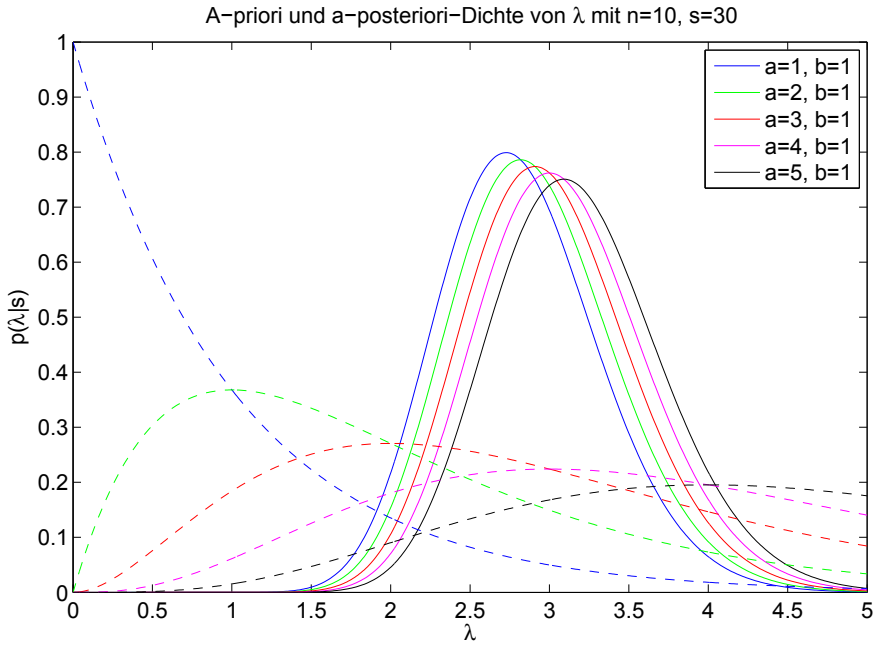

Abbildung 8.9: A-posteriori-Dichten für verschiedene informative a-priori-Dichten (strichliert) mit $n = 10, s = 30$. MATLAB: `make_posterior_poisson`

8.5.4 Vertrauensintervalle und HPD-Bereiche

Vertrauensintervalle $[\lambda_1(s), \lambda_2(s)]$ können leicht mittels der Quantile der a-posteriori-Gammaverteilung konstruiert werden. Das symmetrische Vertrauensintervall ist gleich:

$$[\gamma_{\alpha/2|a+s,1/(b+n)}, \gamma_{1-\alpha/2|a+s,1/(b+n)}]$$

Die einseitigen Vertrauensintervalle sind

$$[0, \gamma_{\alpha|a+s,1/(b+n)}] \quad \text{bzw.} \quad [\gamma_{\alpha|a+s,1/(b+n)}, \infty)$$

HPD-Bereiche können mit numerischen Methoden bestimmt werden.

▶ **Beispiel 8.5.1.**

Die Hintergrundstrahlung in einem Labor wird über eine Periode von 20 Sekunden gemessen. Die Zählwerte sind

$$6, 2, 6, 1, 6, 8, 5, 3, 8, 4, 2, 5, 7, 8, 5, 4, 7, 9, 4, 4$$

Ihre Summe ist gleich $s = 104$. Mit Jeffrey's prior ist die a-posteriori-Verteilung die Gammaverteilung $\mathrm{Ga}(104.5, 0.05)$. Ihr Mittelwert ist 5.225, ihr Modus ist 5.1750. Das symmetrische Vertrauensintervall mit $1 - \alpha = 0.95$ ist $[4.2714, 6.2733]$, das HPD-Intervall ist $[4.2403, 6.2379]$. Da die a-posteriori-Dichte fast symmetrisch ist, sind die beiden Intervalle praktisch gleich lang (siehe Abbildung 8.11).

➠MATLAB: `make_posterior_poisson` ◀

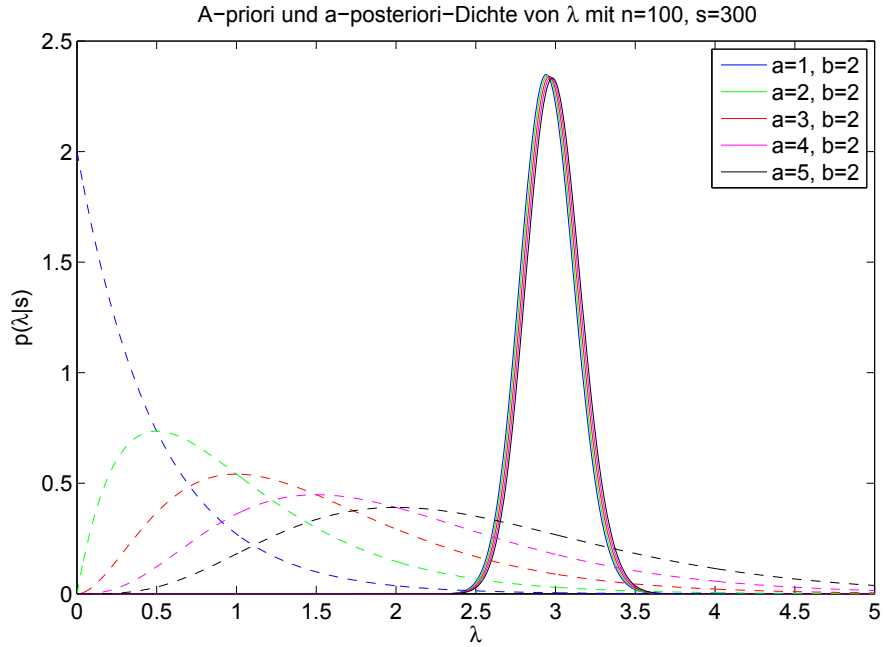

Abbildung 8.10: A-posteriori-Dichten für verschiedene informative a-priori-Dichten (strichliert) mit $n = 100, s = 300$. MATLAB: `make_posterior_poisson`

► **Beispiel 8.5.2.**

Eine Beobachtung aus einer Poissonverteilung hat den Wert $k = 0$. Was kann über den Mittelwert λ ausgesagt werden?

Mit der a-priori-Dichte $\pi(\lambda) = 1$ ist die a-posteriori-Verteilung $\mathrm{Ga}(1, 1)$. Der Modus ist 0, der Mittelwert ist 1. Das HPD-Intervall mit $1-\alpha = 0.95$ ist $[0, 2.9957]$. Mit Jeffrey's prior ist die a-posteriori-Verteilung $\mathrm{Ga}(0.5, 1)$. Der Modus ist 0, der Mittelwert ist 0.5. Das HPD-Intervall mit $1 - \alpha = 0.95$ ist $[0, 1.9207]$.

Ist bekannt, dass λ deutlich kleiner als 1 ist, kann z.B. $\mathrm{Ga}(0.5, 0.5)$ als a-priori-Verteilung gewählt werden. Die a-posteriori-Verteilung ist dann $\mathrm{Ga}(0.5, 0.6667)$. Der Modus ist 0, der Mittelwert ist 0.3333, das HPD-Intervall ist $[0, 1.2805]$.

Der Likelihoodschätzer von λ ist 0. Das linksseitige Konfidenzintervall (siehe Abschnitt 6.4.4) ist $[0, 2.9957]$, ist also identisch mit dem HPD-Intervall mit a-priori-Gleichverteilung auf $(0, \infty)$.

➠MATLAB: `make_posterior_poisson` ◄

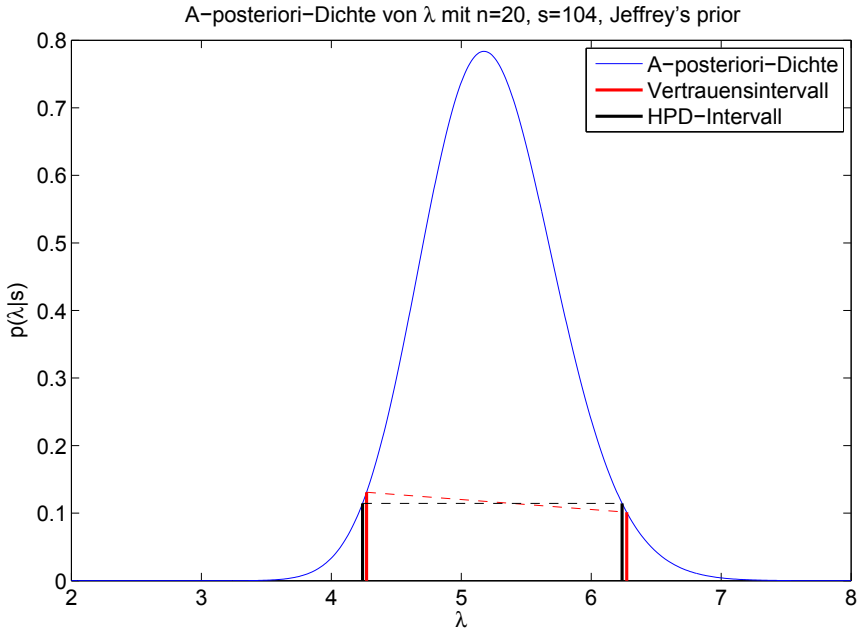

Abbildung 8.11: A-posteriori-Dichte, symmetrisches Vertrauensintervall und HPD-Intervall mit Jeffrey's prior und $n = 20, s = 104, \alpha = 0.05$.
MATLAB: `make_posterior_poisson`

8.6 Normalverteilte Beobachtungen

Es liegt eine Stichprobe $\boldsymbol{y} = (y_1, \ldots, y_n)$ aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$ vor. Ist die Varianz σ^2 bekannt, ist die Berechnung der a-posteriori-Dichte von μ ziemlich einfach. Ist σ^2 unbekannt, wird die Rechnung etwas aufwendiger. Die hier ausgelassenen Zwischenschritte in den folgenden Ableitungen können in Lehrbüchern der Bayes-Statistik nachgelesen werden (siehe Anhang L).

8.6.1 Bekannte Varianz

Zunächst wird angenommen, dass die Varianz σ^2 bekannt ist. Dann lautet die Likelihood-Funktion:

$$p(\boldsymbol{y} \,|\, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_i - \mu)^2}{2\,\sigma^2}\right]$$

In Abwesenheit jeglicher Vorinformation über den tatsächlichen Wert von μ wird die impropere a-priori-Dichte $\pi(\mu) = 1$ gewählt, die gleichzeitig auch Jeffrey's prior ist. Dann ist die a-posteriori-Dichte gleich:

$$p(\mu \,|\, \boldsymbol{y}, \sigma^2) \propto \prod_{i=1}^{n} \exp\left[-\frac{(y_i - \mu)^2}{2\,\sigma^2}\right] \propto \exp\left[-\frac{n(\mu - \bar{y})^2}{2\,\sigma^2}\right]$$

Da die a-posteriori-Dichte proportional zur Dichtefunktion der Normalverteilung $\mathrm{No}(\bar{y}, \sigma^2/n)$ ist, muss sie mit dieser identisch sein. Der Mittelwert der a-posteriori-Verteilung ist daher gleich:

$$\mathsf{E}[\mu \,|\, \boldsymbol{y}] = \bar{y}$$

und damit der Maximum-Likelihood-Schätzer von μ. Der Modus ist ebenfalls gleich $\bar{y}$.

Liegt Vorinformation über μ vor, kann diese durch eine geeignete a-priori-Verteilung inkludiert werden. Die rechnerisch einfachste Behandlung ergibt sich, wenn die a-priori-Verteilung ebenfalls eine Normalverteilung ist. Sei also:

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}\tau_0} \exp\left[-\frac{(\mu-\mu_0)^2}{2\,\tau_0^2}\right]$$

Dann ist die a-posteriori-Dichte gleich:

$$p(\mu\,|\,\boldsymbol{y}) \propto \exp\left[-\frac{n(\mu-\bar{y})^2}{2\,\sigma^2}\right] \exp\left[-\frac{(\mu-\mu_0)^2}{2\,\tau_0^2}\right] \propto \exp\left[-\frac{a(\mu-b/a)^2}{2}\right]$$

mit:

$$a = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \quad \text{und} \quad b = \frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}$$

Die a-posteriori-Verteilung ist daher die Normalverteilung $\mathrm{No}(\mu_n, \tau_n^2)$ mit

$$\mu_n = \frac{b}{a} = \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2}, \quad \tau_n^2 = \frac{1}{a} = \frac{1}{1/\tau_0^2 + n/\sigma^2}$$

Der Mittelwert μ_n ist das gewichtete Mittel aus der Vorinformation μ_0 und dem Mittelwert der Beobachtungen $\bar{y}$, wobei die Gewichte durch die *Präzision*, d.h. die inverse Varianz gegeben sind. Die Präzision $1/\tau_n^2$ ist die Summe der Präzision der Vorinformation μ_0 und der Präzision des Mittelwerts der Beobachtungen $\bar{y}$. Je kleiner die Präzision der Vorinformation ist, d.h. je größer τ_0^2 ist, desto kleiner ist der Einfluss der Vorinformation auf die a-posteriori-Verteilung.

8.6.2 Unbekannte Varianz

Ist die Varianz σ^2 unbekannt, wird eine gemeinsame a-priori-Verteilung für μ und σ^2 benötigt. Die Berechnung der a-posteriori-Dichte wird einfacher, wenn statt σ^2 die Präzision $\zeta = 1/\sigma^2$ verwendet wird. Eine mögliche Wahl ist die impropere a-priori-Dichte $\pi(\mu, \zeta) = 1/\zeta$. Dann gilt:

$$\begin{aligned} p(\mu, \zeta \,|\, \boldsymbol{y}) &\propto \frac{1}{\zeta} \prod_{i=1}^{n} \sqrt{\zeta} \exp\left[-\frac{\zeta}{2} (y_i - \mu)^2 \right] \\ &\propto \zeta^{n/2-1} \exp\left[-\frac{\zeta}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right] \end{aligned}$$

Integration über μ gibt die a-posteriori-Dichte von ζ:

$$\begin{aligned} p(\zeta \,|\, \boldsymbol{y}) &\propto \int \zeta^{n/2-1} \exp\left[-\frac{\zeta}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right] \mathrm{d}\mu \\ &\propto \zeta^{(n-3)/2} \exp\left[-\frac{\zeta}{2} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right] \end{aligned}$$

ζ ist daher a-posteriori gammaverteilt gemäß $\mathrm{Ga}((n-1)/2, 2/\sum (y_i - \bar{y})^2)$. Wegen

$$\sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$$

hängt die Verteilung nur von $s_1 = \sum y_i$ und $s_2 = \sum y_i^2$ bzw. nur vom Mittel $\bar{y}$ und der Varianz s^2 der Beobachtungen ab (siehe Definition 5.2.11).

Die a-posteriori-Dichte von σ^2 erhält man durch die Transformation $\sigma^2 = 1/\zeta$:

$$g(\sigma^2) = \frac{p(1/\sigma^2 \,|\, \boldsymbol{y})}{\sigma^4}$$

Die Verteilung von σ^2 ist die inverse Gammaverteilung $\mathrm{IG}(\alpha, \beta)$ mit:

$$\alpha = \frac{n-1}{2} \quad \text{und} \quad \beta = \frac{\sum (y_i - \bar{y})^2}{2}$$

☞ **Definition 8.6.1.** Die Inverse Gammaverteilung

Die Dichtefunktion der inversen Gammaverteilung $\mathrm{IG}(\alpha, \beta)$, $\alpha, \beta > 0$ *lautet:*

$$f_{\mathrm{IG}}(x|\alpha, \beta) = \frac{\beta^{\alpha} \mathrm{e}^{-\beta/x}}{x^{\alpha+1} \Gamma(\alpha)} . \mathrm{I}_{(0,\infty)}(x)$$

☛ **Satz 8.6.2.**

Ist $X \sim \mathrm{Ga}(a, b)$, *so ist* $Y = 1/X$ *invers gammaverteilt gemäß* $\mathrm{IG}(\alpha, \beta)$ *mit* $\alpha = a$ *und* $\beta = 1/b$.

☛ **Satz 8.6.3.** Mittelwert und Modus der inversen Gammaverteilung

Es sei X *invers gammaverteilt gemäß* $\mathrm{IG}(\alpha, \beta)$. *Dann ist der Modus der Verteilung bei* $\beta/(\alpha+1)$. *Der Mittelwert exisitiert für* $\alpha > 1$ *und ist dann gleich:*

$$\mathsf{E}[X] = \frac{\beta}{\alpha - 1}$$

☛ **Satz 8.6.4.** Verteilungsfunktion und Quantile der inversen Gammaverteilung

Die Verteilungsfunktion F_{IG} der inversen Gammaverteilung $\mathrm{IG}(\alpha,\beta)$ kann durch die Verteilungsfunktion F_{Ga} der Gammverteilung ausgedrückt werden und lautet:

$$F_{\mathrm{IG}}(x\,|\,\alpha,\beta) = 1 - F_{\mathrm{Ga}}(1/x\,|\,\alpha,1/\beta)$$

Das q-Quantil von $\mathrm{IG}(\alpha,\beta)$ ist gleich $\gamma_{1-q|\alpha,1/\beta}$

Die durch ζ bedingte a-posteriori-Dichte von μ erhält man aus:

$$\begin{aligned} p(\mu\,|\,\zeta,\boldsymbol{y}) &= \frac{p(\mu,\zeta\,|\,\boldsymbol{y})}{p(\zeta\,|\,\boldsymbol{y})} \\ &\propto \zeta^{0.5}\exp\left\{-\frac{\zeta}{2}\left[\sum(y_i-\mu)^2 - \sum(y_i-\bar{y})^2\right]\right\} \\ &\propto \zeta^{0.5}\exp\left[-\frac{n\zeta}{2}(\mu-\bar{y})^2\right] \\ &= \frac{1}{\sigma}\exp\left[-\frac{n}{2\sigma^2}(\mu-\bar{y})^2\right] \end{aligned}$$

Das ist die Normalverteilung $\mathrm{No}(\bar{y},\sigma^2/n)$. Bedingt durch einen festen Wert von ζ oder σ^2 ist μ a-posteriori also normalverteilt mit Mittel $\bar{y}$ und Varianz σ^2/n.

Schließlich ergibt sich die a-posteriori-Dichte von μ durch Integration über ζ:

$$\begin{aligned} p(\mu\,|\,\boldsymbol{y}) &\propto \int \zeta^{n/2-1}\exp\left[-\frac{\zeta}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\,\mathrm{d}\zeta \\ &\propto \frac{1}{\left[\sum(\mu-y_i)^2\right]^{n/2}} \end{aligned}$$

Daraus folgt durch eine längere Rechnung, dass das Standardscore:

$$t = \frac{\mu-\bar{y}}{S/\sqrt{n}}$$

a-posteriori T-verteilt ist mit $n-1$ Freiheitsgraden. Die a-posteriori-Verteilung von μ ist dann die mit $S/\sqrt{n}$ skalierte und um $\bar{y}$ verschobene $\mathrm{T}(n-1)$-Verteilung.

Will man eine propere a-priori-Dichte für ζ, bietet sich die Gammaverteilung an, da diese die konjugierte Verteilung ist. Die a-posteriori-Verteilung $p(\zeta\,|\,\boldsymbol{y})$ ist dann wieder eine Gammaverteilung.

▶ **Beispiel 8.6.5.**

Von einem normalverteilten Datensatz des Umfangs $n = 100$ aus $\mathrm{No}(10,2)$ sind $\bar{y} = 9.906$ und $S^2 = 1.822$ gegeben. Die a-posteriori-Verteilung von μ mit der a-priori-Dichte $\pi(\mu,\zeta) = 1/\zeta$ ist dann die $\mathrm{T}(99)$-Verteilung mit Lageparameter $\bar{y} = 9.906$ und Skalenparameter $S/\sqrt{n} = 0.135$. Sie ist in Abbildung 8.12 zu sehen. Wegen der groäsen Anzahl der Freiheitsgrade ist sie praktisch identisch mit der Normalverteilung $\mathrm{No}(\bar{y},S^2/n)$. Das HPD-Intervall mit $1-\alpha = 0.95$ ist gleich $[9.6382, 10.1739]$.

Die a-posteriori-Dichte $p(\sigma^2\,|\,\boldsymbol{y})$ der Varianz σ^2 ist die inverse Gammaverteilung

$$\mathrm{IG}((n-1)/2, S^2(n-1)/2) = \mathrm{IG}(49.5, 90.188)$$

Sie ist in Abbildung 8.13 zu sehen. Der Mittelwert ist 1.860, der Modus ist 1.786. Das HPD-Intervall mit $1-\alpha = 0.95$ ist gleich $[1.3645, 2.4002]$.

➠ MATLAB: `make_posterior_normal` ◀

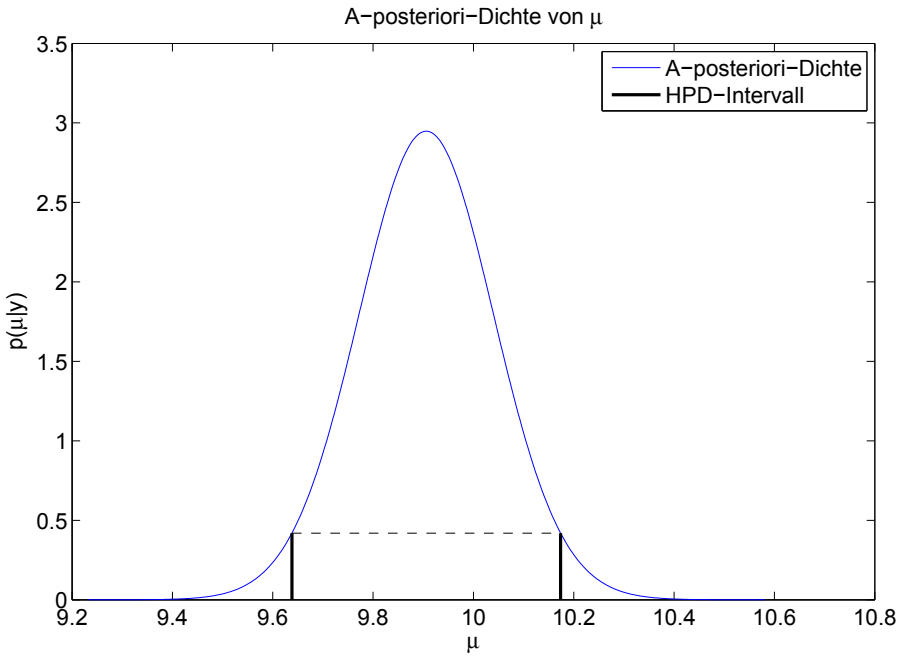

Abbildung 8.12: A-posteriori-Dichte von μ und HPD-Intervall mit $\alpha = 0.05$.
MATLAB: `make_posterior_normal`

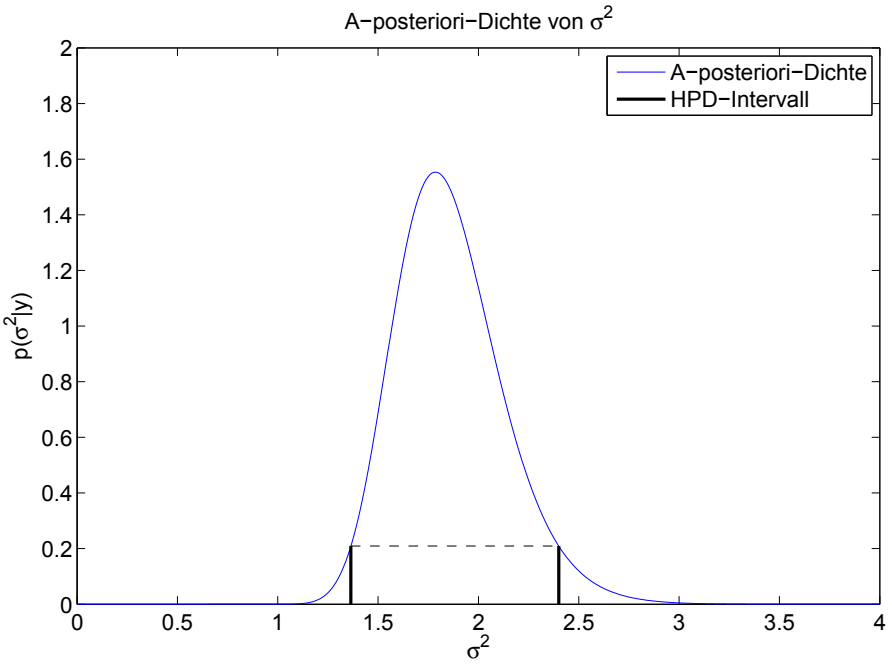

Abbildung 8.13: A-posteriori-Dichte von σ^2 und HPD-Intervall mit $\alpha = 0.05$.
MATLAB: `make_posterior_normal`

8.7 Exponentialverteilte Beobachtungen

Es liegt eine Stichprobe $\boldsymbol{y} = (y_1, \ldots, y_n)$ aus der Exponentialverteilung $\mathrm{Ex}(\tau)$ vor. Aus ihr soll eine Einschätzung des Mittelwerts τ gewonnen werden. Die Likelihood-Funktion lautet:

$$p(\boldsymbol{y}\,|\,\tau) = \prod_{i=1}^{n} \frac{1}{\tau} \exp\left(-\frac{y_i}{\tau}\right) = \frac{1}{\tau^n} \exp\left(-\frac{\sum y_i}{\tau}\right)$$

Die Fisher-Information ist gleich $I_\tau = n/\tau^2$; Jeffrey's prior ist daher gleich $\pi_J(\tau) \propto \tau^{-1}$ und damit improper. Die a-posteriori-Dichte ist dann proportional zu:

$$p(\tau\,|\,\boldsymbol{y}) \propto \frac{1}{\tau^{n+1}} \exp\left(-\frac{\sum y_i}{\tau}\right) = \frac{1}{\tau^{n+1}} \exp\left(-\frac{s}{\tau}\right)$$

Die a-posteriori-Verteilung ist daher die inverse Gammaverteilung $\mathrm{IG}(n, s)$. Der Mittelwert ist $s/(n-1)$, der Modus ist $s/(n+1)$. Vertrauens- und HPD-Intervalle können wieder mit Hilfe von Quantilen der Gammaverteilung berechnet werden.

▶ Beispiel 8.7.1.

Es liegt eine exponentialverteilte Stichprobe vom Umfang $n = 50$ vor. Die Summe der Beobachtungen ist gleich $s = 102.58$. Mit Jeffrey's prior ist der a-posteriori-Mittelwert geich 2.0935 und der a-posteriori-Modus gleich 2.0114. Das HPD-Intervall mit $1 - \alpha = 0.95$ ist $[1.5389, 2.6990]$. Abbildung 8.14 zeigt die a-posteriori-Dichte von τ.

➠MATLAB: `make_posterior_exponential` ◀

Die Likelihood-Funktion kann auch mit der mittleren Rate $\lambda = 1/\tau$ parametrisiert werden (siehe Abschnitt 4.2.4.2):

$$p(\boldsymbol{y}\,|\,\lambda) = \prod_{i=1}^{n} \lambda \exp\left(-\lambda y_i\right) = \lambda^n \exp\left(-\lambda \sum y_i\right) = \lambda^n \exp\left(-\lambda s\right)$$

Jeffrey's prior ist $\pi_{\mathrm{J}}(\lambda) = \lambda^{-1},\ \lambda > 0$. Die a-posteriori-Dichte ist dann proportional zu:

$$p(\lambda\,|\,\boldsymbol{y}) \propto \lambda^{n-1} \exp\left(-\lambda s\right)$$

Die a-posteriori-Verteilung ist daher die Gammaverteilung $\mathrm{Ga}(n, 1/s)$ mit Mittelwert n/s und Modus $(n-1)/s$. Der Mittelwert ist gleich dem Maximum-Likelihood-Schätzer von λ (siehe Beispiel 6.3.10). Die Gammaverteilung ist auch die konjugierte a-priori-Verteilung.

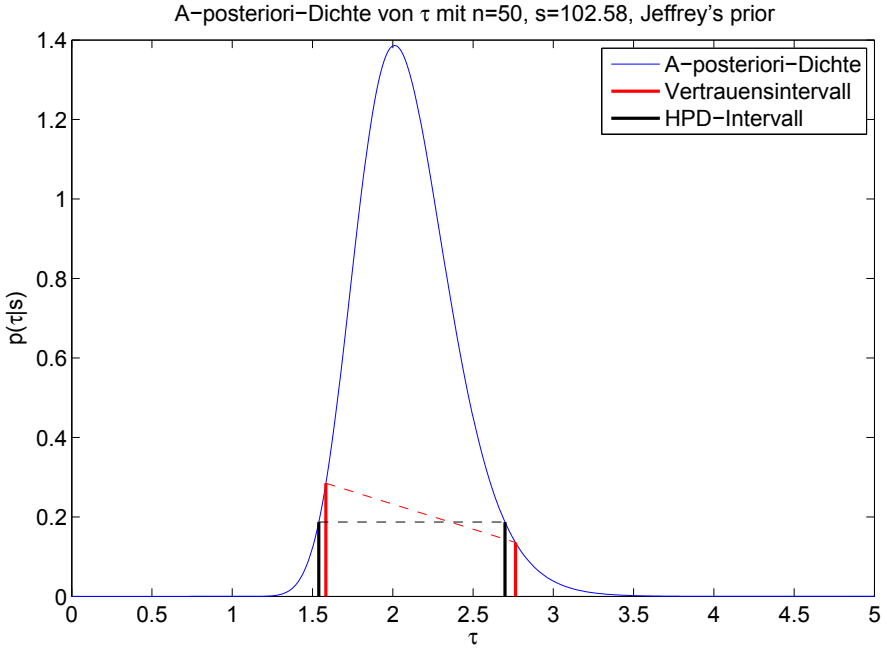

Abbildung 8.14: A-posteriori-Dichte von τ, symmetrisches Vertrauensintervall und HPD-Intervall mit $n = 50, s = 102/58, \alpha = 0.05$, Jeffrey's prior.
MATLAB: `make_posterior_exponential`

Kapitel 9

Regression

9.1 Einleitung

In einem Experiment wird oft eine physikalisch interessante Größe in Abhängigkeit von anderen Größen beobachtet, die von der Experimentatorin gewählt oder eingestellt werden. Die beobachtete Größe, also die abhängige Variable oder *Ergebnisvariable* wird mit Y bezeichnet, die unabhängigen Variablen oder *Einflussvariablen* mit $\boldsymbol{x} = (x_1, \ldots, x_r)$. Im Folgenden wird angenommen, dass die Werte der Einflussvariablen *exakt* bekannt sind, die Beobachtungen hingegen fehlerbehaftet sind.

Die Form der Abhängigkeit der Ergebnisvariablen von den Einflussvariablen wird durch ein *Regressionsmodell* beschrieben, das im einfachsten Fall die folgende Form hat:

$$Y = f(\boldsymbol{\beta}, \boldsymbol{x}) + \epsilon$$

mit den *Regressionskoeffizienten* $\boldsymbol{\beta}$ und dem *Fehlerterm* ϵ. Der Fehlerterm kann einen Messfehler oder sonstige zufällige Einflüsse auf die Beobachtung beschreiben.

Das Ziel der Regressionsanalyse ist die Schätzung von $\boldsymbol{\beta}$ anhand von Beobachtungen $Y_1, \ldots, Y_n$ für verschiedene Werte $\boldsymbol{x}_i$ der Einflussvariablen. Hängen die Beobachtungen nur von einer Einflussvariablen x ab, spricht man von einfacher Regression, ansonsten von mehrfacher (multipler) Regression.

Ein Beispiel einfacher Regression ist die Messung des Drucks p eines Gases in einem geschlossenen Behälter in Abhängigkeit von der Temperatur T. Der Druck ist die Ergebnisvariable p, die Einflussvariable ist die vom Experimentator eingestellte Temperatur. Wird die Abhängigkeit als eine lineare Funktion angesetzt, ist der Regressionskoeffizient gleich dem Proportionalitätsfaktor. Der Fehlerterm beschreibt dann sowohl den Messfehler als auch eventuelle Abweichungen vom linearen Verhalten. In der Praxis mag auch die Einflussvariable T mit einem Fehler behaftet sein; Fehler in den Einflussvariablen werden im Folgenden jedoch vernachlässigt. Es gibt aber Methoden der Regression, die Fehler in den Einflussvariablen berücksichtigen.

9.2 Einfache Regression

9.2.1 Lineare Regression

Das einfachste Regressionsmodell ist eine Gerade:

$$Y = \alpha + \beta x + \epsilon, \quad \mathsf{E}[\epsilon] = 0,\ \mathsf{var}[\epsilon] = \sigma^2$$

Es seien nun $Y_1, \ldots, Y_n$ die Ergebnisse für die Werte $x_1, \ldots, x_n$ der Einflussvariablen x, wobei angenommen wird, dass nicht alle x_i gleich sind, da sonst die Gerade nicht eindeutig bestimmbar ist. Die Schätzung von Abschnitt α und Steigung β kann nach dem *Prinzip der kleinsten Fehlerquadrate* (engl. least squares, abgekürzt LS) erfolgen. Dazu wird die folgende Zielfunktion minimiert:

$$S(\alpha, \beta) = \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2$$

Das Minimum von S kann durch Auffinden der Nullstelle des Gradienten von S bestimmt werden. Der Gradient von S ist gleich:

$$\frac{\partial S}{\partial \alpha} = -2 \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i),\ \frac{\partial S}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (Y_i - \alpha - \beta x_i)$$

Nullsetzen des Gradienten gibt die *Normalgleichungen*:

$$\sum_{i=1}^{n} Y_i = n\alpha + \beta \sum_{i=1}^{n} x_i$$
$$\sum_{i=1}^{n} x_i Y_i = \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2$$

Werden die Normalgleichungen nach α und β aufgelöst, ergeben sich die geschätzten Regressionskoeffizienten:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\overline{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$
$$\hat{\alpha} = \overline{Y} - \hat{\beta}\bar{x}$$

Es gilt $\mathsf{E}[\hat{\alpha}] = \alpha$ und $\mathsf{E}[\hat{\beta}] = \beta$. Die Varianz des Fehlerterms wird erwartungstreu geschätzt durch:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} R_i^2$$

mit den Residuen:

$$R_i = Y_i - \hat{Y}_i, \quad \hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$$

Die Kovarianzmatrix der geschätzten Regressionkoeffizienten kann gemäß Satz 4.4.23 berechnet werden:

$$\mathsf{Cov}[\hat{\alpha}, \hat{\beta}] = \sigma^2 \begin{pmatrix} \dfrac{\sum x_i^2}{n\,(\sum x_i^2 - n\bar{x}^2)} & -\dfrac{\sum x_i}{n\,(\sum x_i^2 - n\bar{x}^2)} \\ -\dfrac{\sum x_i}{n\,(\sum x_i^2 - n\bar{x}^2)} & \dfrac{1}{\sum x_i^2 - n\bar{x}^2} \end{pmatrix}$$

9.2.2 Tests, Konfidenz- und Prognoseintervalle

Ist $\beta = 0$, hängt das Ergebnis überhaupt nicht von den Einflussvariablen ab. Ein Test der Nullhypothese $\mathrm{H}_0 : \beta = 0$ gegen $\mathrm{H}_1 : \beta \neq 0$ beruht auf dem folgenden Satz.

☛ **Satz 9.2.1.**

Ist ϵ normalverteilt, so sind

$$\frac{\hat{\alpha} - \alpha}{\hat{\sigma}_{\hat{\alpha}}}, \quad \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}}$$

T-*verteilt mit* $n - 2$ *Freiheitsgraden, wobei*

$$\hat{\sigma}^2_{\hat{\alpha}} = \frac{\hat{\sigma}^2 \sum x_i^2}{n\,(\sum x_i^2 - n\bar{x}^2)}, \quad \hat{\sigma}^2_{\hat{\beta}} = \frac{\hat{\sigma}^2}{\sum x_i^2 - n\bar{x}^2}$$

Die Nullhypothese $\mathrm{H}_0 : \beta = 0$ wird abgelehnt, wenn die Testgröße

$$T = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$$

relativ klein oder relativ groß ist, also wenn

$$\frac{|\hat{\beta}|}{\hat{\sigma}_{\hat{\beta}}} > t_{1-\alpha/2|n-2}$$

Ein analoger Test kann für die Nullhypothese $\mathrm{H}_0 : \alpha = 0$ durchgeführt werden.

Die symmetrischen Konfidenzintervalle mit 95% Sicherheit lauten:

$$\hat{\alpha} \pm \hat{\sigma}_{\hat{\alpha}} \cdot t_{1-\alpha/2|n-2}, \quad \hat{\beta} \pm \hat{\sigma}_{\hat{\beta}} \cdot t_{1-\alpha/2|n-2}$$

Für $n > 30$ können die Quantile der T-Verteilung durch Quantile der Standardnormalverteilung ersetzt werden.

Es soll nun das Ergebnis $Y_0 = Y(x_0)$ für einen bestimmten Wert x_0 der Einflussvariablen x prognostiziert werden. Der Erwartungswert von Y_0 ist

$$\mathsf{E}[Y_0] = \hat{\alpha} + \hat{\beta} x_0$$

Die Varianz von $\mathsf{E}[Y_0]$ ergibt sich gemäß Satz 4.4.23 zu:

$$\mathsf{var}[\mathsf{E}[Y_0]] = \sigma^2 \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n\bar{x}^2} \right]$$

Da Y_0 um seinen Erwartungswert mit Varianz σ^2 streut, ergibt sich:

$$\mathrm{var}[Y_0] = \sigma^2 \left[\frac{n+1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n\bar{x}^2} \right]$$

Das symmetrische Prognoseintervall für Y_0 mit Sicherheit α ist daher gleich:

$$\hat{\alpha} + \hat{\beta} x_0 \pm t_{1-\alpha/2|n-2} \hat{\sigma} \sqrt{\frac{n+1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n\bar{x}^2}}$$

Die Angemessenheit des Modells kann durch Untersuchung der studentisierten Residuen (Restfehler) überprüft werden. Das Residuum R_k hat die Varianz

$$\mathrm{var}[R_k] = \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_k - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} \right]$$

Das studentisierte Residuum ist dann

$$R'_k = \frac{R_k}{\hat{\sigma} \sqrt{1 - \frac{1}{n} - \frac{(x_k - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}}}$$

Es hat Erwartungswert 0 und Varianz 1. Zwei Beispiele von studentisierten Residuen sind in den Abbildungen 9.1 und 9.2 gezeigt, wobei in Abbildung 9.1 die Daten mit einem linearen Modell erzeugt wurden, in Abbildung 9.2 mit einem parabolischen Modell. In diesem Fall sind die studentisierten Residuen offensichtlich nicht zufällig um 0 verteilt, das lineare Regressionsmodell daher nicht angemessen (siehe auch Abschnitt 9.2.4).

➠MATLAB: `make_regression_diagnostics`

9.2.3 Robuste Regression

Als LS-Schätzer ist die Regressionsgerade nicht robust, d.h. sie ist empfindlich gegen Ausreißer. In Abbildung 9.3 sind zwei Typen von Ausreißern zu sehen. In der linken Abbildung ist ein Wert der Ergebnisvariablen offensichtlich falsch. Die dadurch verursachte Verzerrung der Regressiongeraden ist zwar deutlich sichtbar, aber nicht katastrophal. In der rechten Abbildung ist ein Wert der Einflussvariablen falsch. Der entsprechende Punkt bewirkt durch seine Hebelposition, dass die Regressionsgerade den Hauptteil der Beobachtungen überhaupt nicht mehr korrekt beschreibt. Er wird deshalb als Hebelpunkt (engl. leverage point) bezeichnet.

➠MATLAB: `make_regression_outliers`

Als Abhilfe kommen unter anderem zwei robuste Regressionsverfahren in Betracht, die beide auf P. Rousseeuw* zurückgehen. Im LMS-Verfahren (Least Median of Squares) wird anstatt der Summe der Fehlerquadrate der *Median* der Fehlerquadrate minimiert, sodass bis zu 50% der Residuen beliebig groß sein können. Die LMS-Gerade

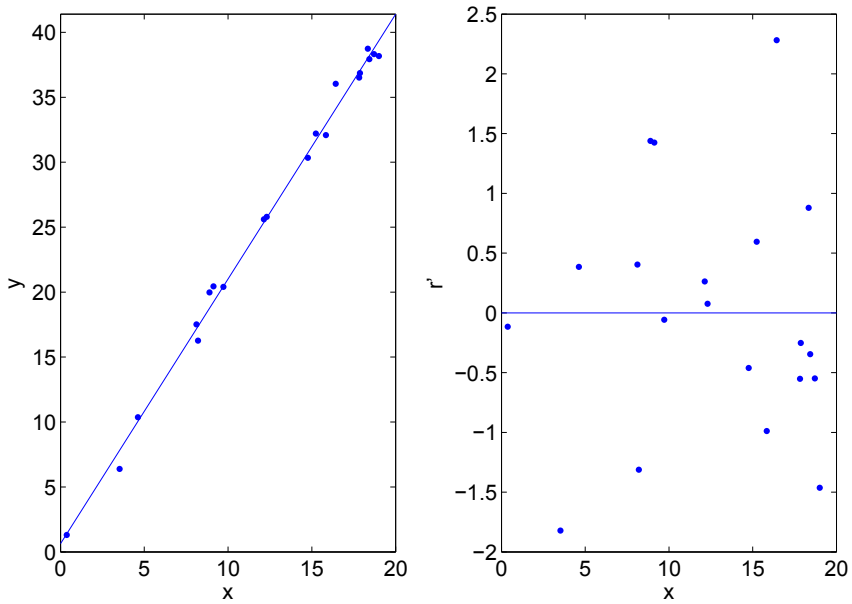

Abbildung 9.1: Regressionsgerade (links) und studentisierte Residuen (rechts). Die Datenpunkte wurden mit einem linearen Modell erzeugt.
MATLAB: `make_regression_diagnostics`

*http://en.wikipedia.org/wiki/Peter_Rousseeuw

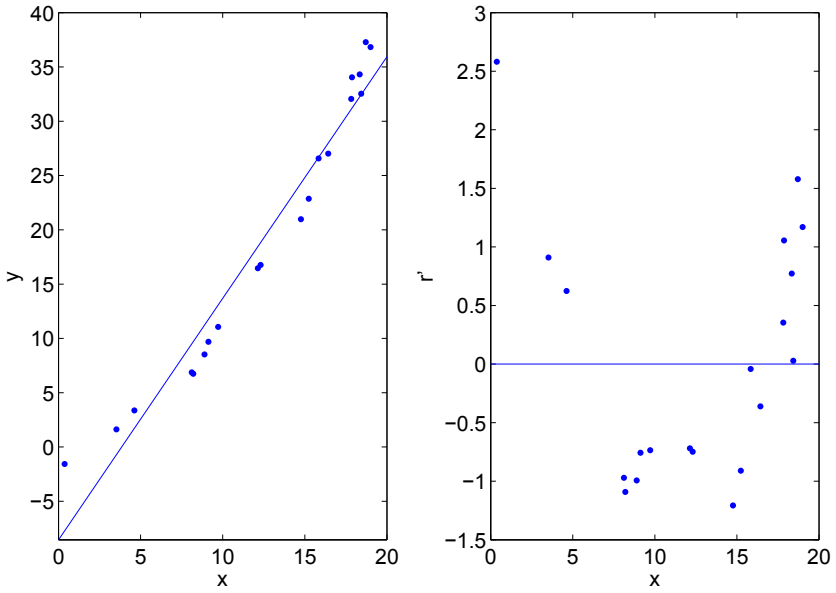

Abbildung 9.2: Regressionsgerade (links) und studentisierte Residuen (rechts). Die Datenpunkte wurden mit einem parabolischen Modell erzeugt. MATLAB: `make_regression_diagnostics`

geht durch zwei Datenpunkte, kann aber nur durch kombinatorische Optimierung ermittelt werden. Die Berechnung ist daher wesentlich aufwendiger als die der LS-Regressionsgeraden.

Im LTS-Verfahren (Least Trimmed Squares) wird die Summe einer festen Anzahl $h \leq n$ von Fehlerquadraten minimiert. Die Berechnung der Geraden erfolgt iterativ. Dazu steht ein relativ rasch konvergierendes Verfahren (FAST-LTS) zur Verfügung.

Abbildung 9.4 zeigt die mittels LMS und LTS berechneten Regressiongeraden. Die Robustheit der beiden Methoden ist klar ersichtlich. Viel mehr über robuste Regression findet sich im Buch von Rousseeuw und Leroy (siehe Anhang L).

➠MATLAB: `robust_regression`

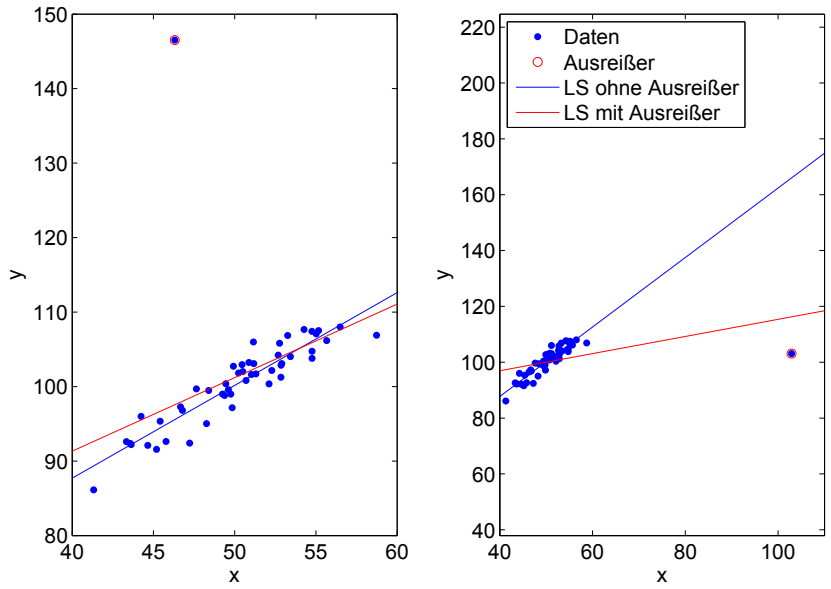

Abbildung 9.3: Lineare Regression mit Ausreißern. Links: Ausreißer in der Ergebnisvariablen, rechts: Ausreißer in der Einflussvariablen (Hebelpunkt). MATLAB: `make_regression_outliers`

9.2.4 Polynomiale Regression

Ist der Zusammenhang zwischen x und Y nicht annähernd linear, kann man versuchen, ein Polynom anzupassen. Das Modell lautet dann:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_r x^r + \epsilon, \quad \mathsf{E}[\epsilon] = 0, \ \mathsf{var}[\epsilon] = \sigma^2$$

Es seien wieder $Y_1, \ldots, Y_n$ die Ergebnisse für die Werte $x_1, \ldots, x_n$ der Einflussvariablen x. In der Regel wird der Grad des Polynoms r deutlich kleiner als die Anzahl der Beobachtungen sein. Das Modell kann folgendermaßen in Matrix-Vektor-Schreibweise

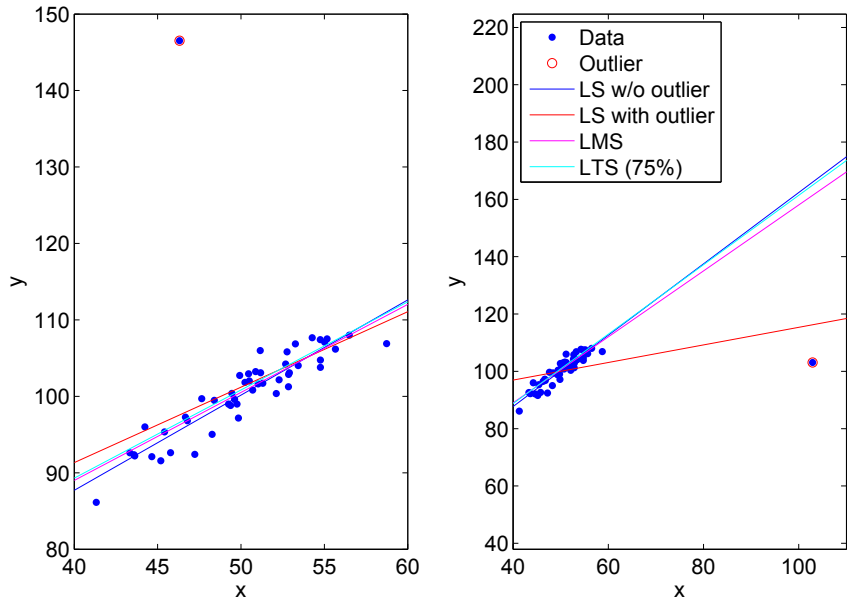

Abbildung 9.4: Robuste Regression mit Ausreißern. Links: Ausreißer in der Ergebnisvariablen, rechts: Ausreißer in der Einflussvariablen (Hebelpunkt).
MATLAB: `robust_regression`

formuliert werden:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

mit:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^r \\ 1 & x_2 & x_2^2 & \cdots & x_2^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^r \end{pmatrix}$$

Die folgende Zielfunktion wird minimiert:

$$S(\boldsymbol{\beta}) = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

Es ist zu beachten, dass die Zielfunktion nach wie vor linear in den Regressionsparametern ist. Der Gradient von S lautet:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\mathrm{T}} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

Nullsetzen des Gradienten gibt die *Normalgleichungen*:

$$\mathbf{X}^{\mathrm{T}} \boldsymbol{Y} = \mathbf{X}^{\mathrm{T}} \mathbf{X} \boldsymbol{\beta}$$

Hat $\mathbf{X}$ vollen Rang, lautet die Lösung:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}$$

Die Varianz des Fehlerterms wird erwartungstreu geschätzt durch:

$$\hat{\sigma}^2 = \frac{1}{n-r-1}\sum_{i=1}^{n} R_i^2$$

mit den Residuen:

$$\boldsymbol{R} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}, \quad \hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

Die Kovarianzmatrix der geschätzten Regressionkoeffizienten lautet:

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}] = \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

Daraus wird mit Satz 4.4.23 die Kovarianzmatrix der Residuen $\boldsymbol{R}$ berechnet:

$$\mathrm{Cov}[\boldsymbol{R}] = \sigma^2\left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]$$

Abbildung 9.5 zeigt die Regressionsparabel, die an die Beobachtungen in Abbildung 9.4 angepasst wurde. Die studentisierten Residuen streuen wieder zufällig um 0.

➠MATLAB: `make_polynomial_regression`

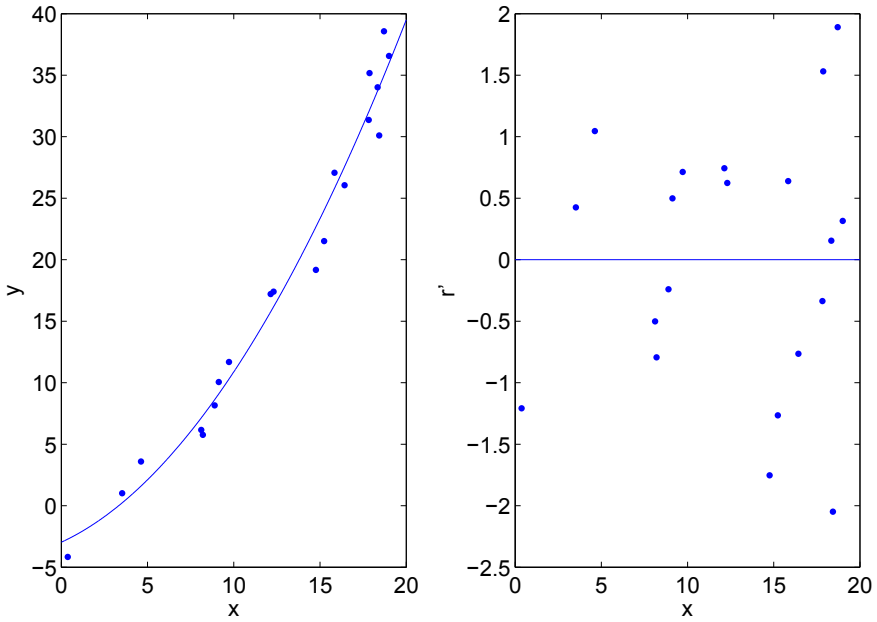

Abbildung 9.5: Regressionsparabel (links) und studentisierte Residuen (rechts).
MATLAB: `make_polynomial_regression`

Neben Polynomen können mit dem gleichen Formalismus auch andere nichtlineare Funktionen an die Beobachtungen angepasst werden, solange nur die Zielfunktion linear in den Regressionsparametern ist.

► **Beispiel 9.2.2.**
Die Daten in der Tabelle zeigen den gemessenen Energieverlust ΔE von Myonen in 500 μm Silizium, in Abhängigkeit vom Impuls p:

p [GeV/c]	2	4	6	8	10	12	14	16
ΔE [MeV]	0.2478	0.2682	0.2779	0.2967	0.2982	0.3047	0.3068	0.3191

Zunächst werden die Regressionsparameter des Modells

$$\Delta E = \beta_0 + \beta_1 \ln p$$

bestimmt. Die Einflussgrößen sind also $x_0 = 1$ und $x_1 = \ln p$. Dann gilt:

$$\mathbf{X}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.6931 & 1.3863 & 1.7918 & 2.0794 & 2.3026 & 2.4849 & 2.6391 & 2.7726 \end{pmatrix}$$

und damit:

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{pmatrix} 1.3017 & -0.5829 \\ -0.5829 & 0.2887 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 0.2232 \\ 0.0330 \end{pmatrix}$$

Die Schätzung der Varianz des Fehlerterms ist gleich $\hat{\sigma}^2 = 1.3274 \cdot 10^{-5}$; dies entspricht einer Standardabweichung $\hat{\sigma} = 0.0036$. Daraus ergibt sich die Kovarianzmatrix der geschätzten Regressionsparameter zu:

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}] = 10^{-4} \cdot \begin{pmatrix} 0.1728 & -0.0774 \\ -0.0774 & 0.0383 \end{pmatrix}$$

Der prognostizierte erwartete Energieverlust für $p = 13\,\mathrm{GeV}/c$ ist gleich 0.3080 MeV, sein Standardfehler ist gleich 0.0017 MeV. Abbildung 9.6 zeigt die gemessenen Werte (blau), die angepasste Funktion (rot) und den prognosizierten Erwartungswert mit Fehlerbalken (grün).

➠MATLAB: `make_eloss` ◀

9.3 Mehrfache Regression

9.3.1 Das lineare Modell

Hängt das Ergebnis Y von mehreren Einflussvariablen ab, lautet das einfachste lineare Regressionmodell:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r + \epsilon, \quad \mathsf{E}[\epsilon] = 0,\ \mathsf{var}[\epsilon] = \sigma^2$$

Es seien wieder $Y_1, \ldots, Y_n$ die Ergebnisse für n Werte $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ der Einflussvariablen $\boldsymbol{x} = (x_1, \ldots, x_r)$. In Matrix-Vektor-Schreibweise lautet das Modell:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

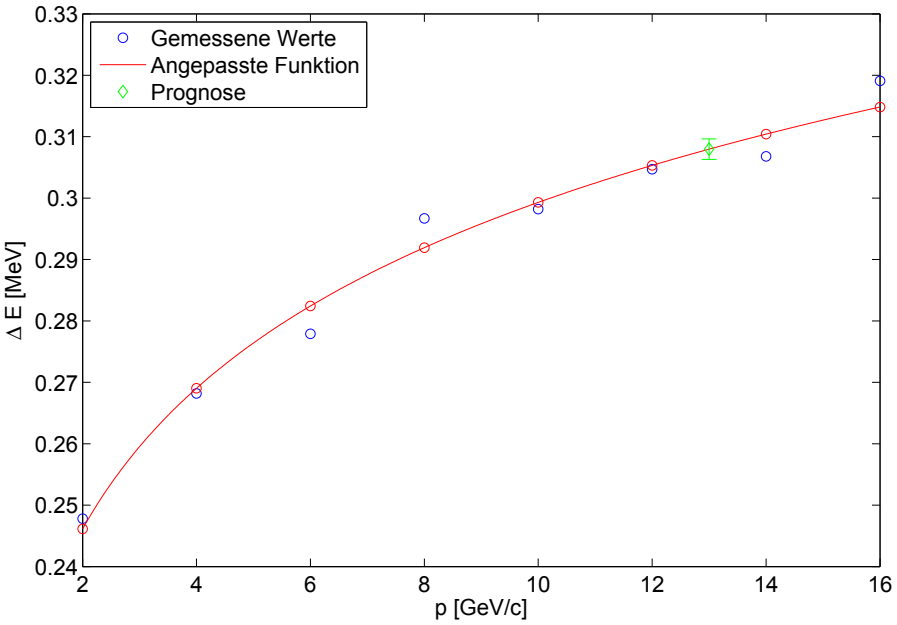

Abbildung 9.6: Gemessene Werte (blau), die angepasste Funktion (rot) und eine Prognose (grün). MATLAB: `make_eloss`

mit:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,r} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,r} \end{pmatrix}$$

Die Matrix $\mathbf{X}$ wird als die *Modellmatrix* bezeichnet. Es wird angenommen, dass sie vollen Rang hat. Auf jeden Fall muss $r \leq n - 1$ sein.

9.3.2 Schätzung, Tests und Prognoseintervalle

Die folgende Zielfunktion wird minimiert:

$$S(\boldsymbol{\beta}) = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

Gradient von S:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

Nullsetzen des Gradienten gibt die *Normalgleichungen*:

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}$$

mit der Lösung:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}$$

Die Varianz des Fehlerterms wird erwartungstreu geschätzt durch:

$$\hat{\sigma}^2 = \frac{1}{n-r-1}\sum_{i=1}^{n} R_i^2$$

mit den Residuen:

$$\boldsymbol{R} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}, \quad \hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

Die Kovarianzmatrix der geschätzten Regressionkoeffizienten ist gleich:

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}] = \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

Daraus ergibt sich mit Satz 4.4.23 die Kovarianzmatrix der Residuen $\boldsymbol{R}$ zu:

$$\mathsf{Cov}[\boldsymbol{R}] = \sigma^2 \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]$$

Ist $\beta_k = 0$, hängt das Ergebnis überhaupt nicht von der Einflussvariablen x_k ab. Ein Test der Nullhypothese $\mathrm{H}_0 : \beta_k = 0$ gegen $\mathrm{H}_1 : \beta_k \neq 0$ beruht auf dem folgenden Satz.

☛ **Satz 9.3.1.**
Ist ϵ normalverteilt, so ist

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\hat{\beta}_k}}$$

T-*verteilt mit* $n - r - 1$ *Freiheitsgraden, wobei* $\hat{\sigma}^2_{\hat{\beta}_k}$ *das* k*-te Diagonalelement der geschätzten Kovarianzmatrix*

$$\hat{\sigma}^2 \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

ist.

Die Nullhypothese $\mathrm{H}_0 : \beta_k = 0$ wird abgelehnt, wenn die Testgröße

$$T = \frac{\hat{\beta}_k}{\hat{\sigma}_{\hat{\beta}_k}}$$

relativ klein oder relativ groß ist, also wenn:

$$\frac{|\hat{\beta}_k|}{\hat{\sigma}_{\hat{\beta}_k}} > t_{1-\alpha/2|n-r-1}$$

Das symmetrische Konfidenzintervall für β_k mit 95% Sicherheit lautet:

$$\hat{\beta}_k \pm \hat{\sigma}_{\hat{\beta}_k} \cdot t_{1-\alpha/2|n-r-1}$$

Es soll nun das Ergebnis $Y_0 = Y(\boldsymbol{x}_0)$ für einen bestimmten Wert $\boldsymbol{x}_0 = (x_{01}, \ldots, x_{0r})$ der Einflussvariablen prognostiziert werden. Dazu wird $\boldsymbol{x}_0$ um den Wert 1 erweitert: $\boldsymbol{x}^+ = (1, x_{01}, \ldots, x_{0r})$. Der Erwartungswert von Y_0 ist dann:

$$\mathsf{E}[Y_0] = \boldsymbol{x}^+ \cdot \hat{\boldsymbol{\beta}}$$

Die Varianz von $\mathsf{E}[Y_0]$ ergibt sich mittels Satz 4.4.23:

$$\mathsf{var}[\mathsf{E}[Y_0]] = \sigma^2 \boldsymbol{x}^+ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \boldsymbol{x}^{+\mathrm{T}}$$

Da Y_0 um seinen Erwartungswert mit Varianz σ^2 streut, ergibt sich:

$$\mathsf{var}[Y_0] = \sigma^2 \left[1 + \boldsymbol{x}^+ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \boldsymbol{x}^{+\mathrm{T}}\right]$$

Das symmetrische Prognoseintervall für Y_0 mit Sicherheit α ist daher gleich:

$$\boldsymbol{x}^+ \cdot \hat{\boldsymbol{\beta}} \pm t_{1-\alpha/2|n-r-1} \hat{\sigma} \sqrt{1 + \boldsymbol{x}^+ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \boldsymbol{x}^{+\mathrm{T}}}$$

9.3.3 Gewichtete Regression

Im allgemeinen Fall können die Fehlerterme eine beliebige Kovarianzmatrix haben:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V}$$

Ist $\mathbf{V}$ bekannt, lautet die Zielfunktion:

$$S(\boldsymbol{\beta}) = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{G} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}), \quad \mathbf{G} = \mathbf{V}^{-1}$$

mit dem Gradienten:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

Die Matrix $\mathbf{G}$ wird als *Gewichts-* oder *Informationsmatrix* bezeichnet. Nullsetzen des Gradienten gibt wieder die *Normalgleichungen*:

$$\mathbf{X}^{\mathrm{T}}\mathbf{G}\boldsymbol{Y} = \mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\boldsymbol{\beta}$$

mit der Lösung:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}\boldsymbol{Y}$$

Die Kovarianzmatrix der geschätzten Regressionkoeffizienten ist gleich:

$$\mathsf{Cov}[\hat{\boldsymbol{\beta}}] = \left(\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right)^{-1}$$

Daraus ergibt sich mit Satz 4.4.23 die Kovarianzmatrix $\mathbf{R}$ der Residuen $\boldsymbol{R}$ zu:

$$\mathbf{R} = \mathsf{Cov}[\boldsymbol{R}] = \mathbf{V} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}$$

Tests und Prognoseintervalle können entsprechend modifizert werden.

9.3.4 Regressionsdiagnostik

Ist die Kovarianzmatrix $\mathbf{V}$ des Fehlerterms bekannt, können verschiedene Größen berechnet werden, die zur Überprüfung des Modells, d.h. von $\mathbf{X}$ und $\mathbf{V}$ dienen. Die wichtigsten sind die χ^2-Statistik und die standardisierten Residuen. Die χ^2-Statistik der Regression ist definiert durch:

$$\chi^2 = \boldsymbol{R}^{\mathrm{T}}\mathbf{G}\boldsymbol{R}$$

Sind die Modellmatrix $\mathbf{X}$ und die Kovarianzmatrix $\mathbf{V}$ korrekt und ist der Fehlerterm normalverteilt, so ist χ^2 χ^2-verteilt mit $d = \dim[Y] - \dim[\boldsymbol{\beta}]$ Freiheitsgraden. Oft wird statt χ^2 der P-Wert betrachtet. Dieser ist gleich der Wahrscheinlichkeit, einen χ^2-Wert zu beobachten, der mindestens so groß wie der tatsächlich beobachtete ist:

$$P(\chi^2) = \int_{\chi^2}^{\infty} g(x\,|\,d)\,\mathrm{d}x$$

wo $g(x\,|\,d)$ die Dichtefunktion der χ^2-Verteilung mit d Freiheitsgraden ist. Der P-Wert ist im korrekten Modell gleichverteilt im Intervall $[0, 1]$.

Die standardisierten Residuen werden folgendermaßen berechnet:

$$\dot{R}_i = \frac{R_i}{\sqrt{\mathbf{R}_{ii}}}$$

Sie sind im korrekten Modell standardnormalverteilt.

▶ **Beispiel 9.3.2.** Vermessung einer Teilchenspur

Die Spur eines geladenen Teilchens wird durch vier ebene Detektoren an den Positionen $(x = x_i, i = 1, \ldots, 4)$ vermessen, siehe Abbildung 9.7.

Jede Ebene liefert zwei gemessene Koordinaten y_i und z_i. Im feldfreien Raum sind die Messungen $\boldsymbol{Y} = (y_1, \ldots, y_4, z_1, \ldots, z_4)^{\mathrm{T}}$ lineare Funktionen von vier Spurparametern (Anfangswerten) $\boldsymbol{\beta} = (y, z, dy/dx, dz/dx)^{\mathrm{T}}$ an der Stelle x=0, plus ein Fehlerterm $\boldsymbol{\epsilon} = (\varepsilon_1, \ldots, \varepsilon_8)^{\mathrm{T}}$. $\boldsymbol{\epsilon}$ wird als normalverteilt mit Mittelwert 0 und Kovarianzmatrix $\mathbf{V} = \mathbf{G}^{-1}$ angenommen. Der Fehlerterm setzt sich aus zwei Bestandteilen zusammen: erstens die Fehler der Positionsmessung, zweitens die zufällige Ablenkung des Teilchens durch vielfache Coulombstreuung. Die Fehler der Positionsmessung können als unabhängig angenommen werden, ihre Kovarianzmatrix $\mathbf{V}_1$ ist daher diagonal. Die Bestimmung von $\mathbf{V}_1$ erfordert eine sorgtältige Kalibration der Detektoren. Die Ablenkungen durch vielfache Coulombstreuung sind dagegen korreliert, ihre Kovarianzmatrix $\mathbf{V}_2$ ist daher nichtdiagonal. Sie kann mit Hilfe theoretischer Überlegungen berechnet werden.

Das lineare Modell lautet dann:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V}_1 + \mathbf{V}_2 = \mathbf{V}$$

mit

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & x_1 & 0 \\ 1 & 0 & x_2 & 0 \\ 1 & 0 & x_3 & 0 \\ 1 & 0 & x_4 & 0 \\ 0 & 1 & 0 & x_1 \\ 0 & 1 & 0 & x_2 \\ 0 & 1 & 0 & x_3 \\ 0 & 1 & 0 & x_4 \end{pmatrix}$$

Die χ^2-Statistik der Regression hat $8 - 4 = 4$ Freiheitsgrade. Der P-Wert kann zur Unterdrückung von falschen oder kontaminierten Spuren herangezogen werden. Es ist jedoch zu beachten, dass beide Komponenten des Fehlerterms nur annähernd normalverteilt sind, sodass der P-Wert in der Praxis nicht perfekt gleichverteilt ist. ◀

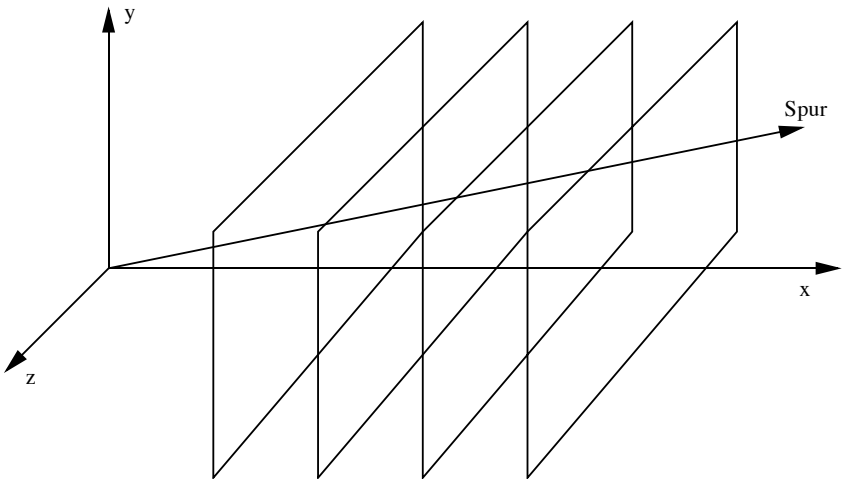

Abbildung 9.7: Vermessung einer Teilchenspur im feldfreien Raum.

9.3.5 Nichtlineare Regression

In der Praxis ist die Abhängigkeit der Ergebnisse von den Regressionskoeffizienten oft nichtlinear:

$$\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{\beta}) + \boldsymbol{\epsilon}, \quad \mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V}$$

Um die Notation zu vereinfachen, wird die Abhängigkeit von den Einflussvariablen $\boldsymbol{x}$ nicht explizit gezeigt, sie ist aber implizit in der Funktion $\boldsymbol{h}$ enthalten. Ist $\mathbf{V}$ bekannt, lautet die Zielfunktion:

$$S(\boldsymbol{\beta}) = [\boldsymbol{Y} - \boldsymbol{h}(\boldsymbol{\beta})]^{\mathrm{T}} \mathbf{G} [\boldsymbol{Y} - \boldsymbol{h}(\boldsymbol{\beta})], \quad \mathbf{G} = \mathbf{V}^{-1}$$

Die Funktion S kann zum Beispiel mit dem Gauß-Newton-Verfahren* minimiert werden. Dazu wird $\boldsymbol{h}$ an einer Stelle $\boldsymbol{\beta}_0$ in eine lineare Funktion entwickelt:

$$\boldsymbol{h}(\boldsymbol{\beta}) \approx \boldsymbol{h}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \boldsymbol{c} + \mathbf{H}\boldsymbol{\beta}, \quad \mathbf{H} = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0}, \quad \boldsymbol{c} = \boldsymbol{h}(\boldsymbol{\beta}_0) - \mathbf{H}\boldsymbol{\beta}_0$$

Die Schätzung von $\boldsymbol{\beta}$ lautet dann:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^{\mathrm{T}} \mathbf{G} \mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{G} (\boldsymbol{Y} - \boldsymbol{c})$$

$\boldsymbol{h}$ wird nun neuerlich an der Stelle $\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}$ entwickelt, und das Verfahren wird iteriert, bis die Schätzung sich nicht mehr wesentlich ändert. Neben dem Gauß-Newton-Verfahren sind noch zahlreiche andere Methoden zur Minimierung von S verfügbar. Sie können in der einschlägigen Literatur nachgelesen werden (siehe Anhang L).

*http://en.wikipedia.org/wiki/Gauss-Newton_algorithm

► **Beispiel 9.3.3.** Vermessung einer Teilchenspur im Magnetfeld.
Wenn ein geladenes Teilchen ein stationäres Magnetfeld $\vec{B}$ durchläuft, erfährt es durch die Lorentzkraft eine Ablenkung, sodass die Spur gekrümmt wird. Dies wird im Experiment zur Impulsmessung verwendet. Die Spurparameter werden zu diesem Zweck um einen vom Impuls p abhängigen Wert vermehrt, der oft als q/p gewählt wird, wo q die Ladung des Teilchens ist. Die Abhängigkeit der gemessenen Koordinaten von den Spurparametern ist dann eine nichtlineare Funktion $\boldsymbol{f}$, die durch die Lösung der Bewegungsgleichung des Teilchens im Feld ermittelt wird:

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{\beta}, \vec{B}) + \boldsymbol{\epsilon}, \quad \mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V}$$

Die Funktion $\boldsymbol{f}$ hängt nicht nur vom Magnetfeld, sondern auch von der Anordnung und der Orientierung der Sensoren ab. Oft sind auch Materialeffekte wie Energieverlust durch Ionisation in $\boldsymbol{f}$ inkludiert. Die Funktion $\boldsymbol{f}$ liegt nur in ganz einfachen Fällen in geschlossener Form vor; meistens muss sie durch numerische Integration der Bewegungsgleichung berechnet werden. Die Schätzung von $\boldsymbol{\beta}$ erfolgt mit Methoden der nichtlinearen Regression. ◄

9.4 Ausgleichsrechnung

In der Ausgleichsrechnung werden unbekannte Größen aus direkten, fehlerbehafteten Beobachtungen so geschätzt, dass sie vorgegebene *Zwangsbedingungen* erfüllen. Die Zwangsbedingungen können linear oder nichtlinear sein, und können auch weitere unbekannte Parameter enthalten. Es wird zunächst der Fall von linearen Zwangsbedingungen behandelt.

9.4.1 Lineare Zwangsbedingungen

Es sei $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ ein Vektor von Beobachtungen der n unbekannten Größen $\boldsymbol{\beta} = (\beta, \ldots, \beta_n)$:

$$\boldsymbol{Y} = \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V}$$

Die Kovarianzmatrix $\mathbf{V}$ des Fehlerterms $\boldsymbol{\epsilon}$ wird als bekannt vorausgesetzt. Außerdem sei bekannt, dass die Größen $\boldsymbol{\beta}$ m unabhängige lineare Zwangsbedingungen erfüllen müssen, die r unbekannte Parameter $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_r)^{\mathrm{T}}$ enthalten. Es existiert also eine Matrix $\mathbf{A}$ der Dimension $m \times n$, eine Matrix der $\mathbf{B}$ der Dimension $m \times r$ und ein Spaltenvektor $\boldsymbol{c}$ der Dimension $m \times 1$, sodass gilt:

$$\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\vartheta} = \boldsymbol{c}$$

$\boldsymbol{\beta}$ und $\boldsymbol{\vartheta}$ werden nach dem Prinzip der kleinsten Fehlerquadrate geschätzt. Dies erfordert die Minimierung der Zielfunktion:

$$S(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \boldsymbol{\beta}), \quad \mathbf{G} = \mathbf{V}^{-1}$$

unter der Bedingung $\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\vartheta} = \boldsymbol{c}$. Das Problem ist bestimmt, wenn $r \leq m$ und widerspruchsfrei, wenn $m \leq n + r$.

Die Minimierung von S kann mit Hilfe eines Vektors $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^{\mathrm{T}}$ von m Lagrangefaktoren erfolgen. Die erweiterte Zielfunktion ist gleich:

$$L(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \boldsymbol{\lambda}) = (\boldsymbol{Y} - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \boldsymbol{\beta}) + 2\boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\vartheta} - \boldsymbol{c})$$

mit dem Gradienten:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = -2\mathbf{G}(\boldsymbol{Y} - \boldsymbol{\beta}) + 2\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda}, \quad \frac{\partial L}{\partial \boldsymbol{\vartheta}} = 2\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda}, \quad \frac{\partial L}{\partial \boldsymbol{\lambda}} = 2(\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\vartheta} - \boldsymbol{c})$$

Nullsetzen des Gradienten führt auf das folgende lineare Gleichungssystem:

$$\begin{aligned} \mathbf{G}\boldsymbol{\beta} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} &= \mathbf{G}\boldsymbol{Y} \\ \mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{A}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\vartheta} &= \boldsymbol{c} \end{aligned}$$

Mit der zusätzlichen Annahme $m \leq n$ und den Bezeichnungen $\mathbf{G}_{\mathrm{A}} = (\mathbf{A}\mathbf{V}\mathbf{A}^{\mathrm{T}})^{-1}$, $\mathbf{V}_{\mathrm{B}} = (\mathbf{B}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{B})^{-1}$ erhält man die folgende geschlossene Lösung:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \boldsymbol{Y} + \mathbf{V}\mathbf{A}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\left[\mathbf{I} - \mathbf{B}\mathbf{V}_{\mathrm{B}}\mathbf{B}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\right](\boldsymbol{c} - \mathbf{A}\boldsymbol{Y}) \\ \hat{\boldsymbol{\vartheta}} &= \mathbf{V}_{\mathrm{B}}\mathbf{B}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}(\boldsymbol{c} - \mathbf{A}\boldsymbol{Y}) \end{aligned}$$

Die gemeinsame Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$ und $\hat{\boldsymbol{\vartheta}}$ lautet:

$$\mathrm{Cov}\left[\begin{pmatrix}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\vartheta}}\end{pmatrix}\right] = \begin{pmatrix}\mathbf{V} - \mathbf{V}\mathbf{A}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{A}\mathbf{V} + \mathbf{V}\mathbf{A}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{B}\mathbf{V}_{\mathrm{B}}\mathbf{B}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{V} & -\mathbf{V}\mathbf{A}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{B}\mathbf{V}_{\mathrm{B}} \\ -\mathbf{V}_{\mathrm{B}}\mathbf{B}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{A}\mathbf{V} & \mathbf{V}_{\mathrm{B}}\end{pmatrix}$$

Enthalten die Zwangsbedingungen keine unbekannten Parameter, gilt:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{Y} + \mathbf{V}\mathbf{A}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}(\boldsymbol{c} - \mathbf{A}\boldsymbol{Y}), \quad \mathrm{Cov}[\hat{\boldsymbol{\beta}}] = \mathbf{V} - \mathbf{V}\mathbf{A}^{\mathrm{T}}\mathbf{G}_{\mathrm{A}}\mathbf{A}\mathbf{V}$$

Ist $n < m \leq n+r$, so ist $\mathbf{AVA}^{\mathrm{T}}$ singulär, und es muss das komplette Gleichungssystem gelöst werden:

$$\begin{pmatrix} \mathbf{G} & \mathbf{O} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{O} & \mathbf{O} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{A} & \mathbf{B} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\vartheta} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{G}\boldsymbol{Y} \\ \mathbf{0} \\ \boldsymbol{c} \end{pmatrix}$$

Hier ist $\mathbf{O}$ eine Nullmatrix mit der erforderlichen Dimension. Auch dieses System hat eine geschlossene Lösung, sie ist jedoch so kompliziert, dass es in der Praxis einfacher ist, die Gleichungsmatrix numerisch zu invertieren. Sei also:

$$\mathbf{H} = \begin{pmatrix} \mathbf{G} & \mathbf{O} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{O} & \mathbf{O} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{A} & \mathbf{B} & \mathbf{O} \end{pmatrix} \quad \text{und} \quad \mathbf{H}^{-1} = \begin{pmatrix} \mathbf{X} & \mathbf{P}^{\mathrm{T}} & \mathbf{Q}^{\mathrm{T}} \\ \mathbf{P} & \mathbf{Y} & \mathbf{R}^{\mathrm{T}} \\ \mathbf{Q} & \mathbf{R} & \mathbf{Z} \end{pmatrix}$$

wobei $\mathbf{H}^{-1}$ die gleiche Blockstruktur hat wie $\mathbf{H}$. Dann gilt:

$$\hat{\boldsymbol{\beta}} = \mathbf{XG}\boldsymbol{Y} + \mathbf{Q}^{\mathrm{T}}\boldsymbol{c}$$
$$\hat{\boldsymbol{\vartheta}} = \mathbf{PG}\boldsymbol{Y} + \mathbf{R}^{\mathrm{T}}\boldsymbol{c}$$

Die gemeinsame Kovarianzmatrix der Lösung kann wieder mit Satz 4.4.23 berechnet werden:

$$\mathsf{Cov}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\vartheta}} \end{pmatrix}\right] = \begin{pmatrix} \mathbf{XGX}^{\mathrm{T}} & \mathbf{XGP}^{\mathrm{T}} \\ \mathbf{PGX}^{\mathrm{T}} & \mathbf{PGP}^{\mathrm{T}} \end{pmatrix}$$

Die χ^2-Statistik des Ausgleichs ist definiert durch:

$$\chi^2 = (\boldsymbol{Y} - \hat{\boldsymbol{\beta}})^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \hat{\boldsymbol{\beta}})$$

Ist $\boldsymbol{Y}$ normalverteilt mit Kovarianzmatrix $\mathbf{V} = \mathbf{G}^{-1}$, so ist χ^2 χ^2-verteilt mit $m - r$ Freiheitsgraden: jeder unbekannte Parameter vermindert die Anzahl der Freiheitsgrade um 1, jede Zwangsbedingung erhöht sie um 1. Die Bedingungen $r \leq m$ und $m \leq n + r$ sind äquivalent dazu, dass die Anzahl der Freiheitsgrade nicht kleiner als 0 und nicht größer als n ist. Ein kleiner P-Wert der χ^2-Statistik deutet darauf hin, dass die ausgeglichenen Werte weiter von den Beobachtungen entfernt sind, als mit ihren Fehlern vereinbar ist.

▶ **Beispiel 9.4.1.** Ausgleich der Winkel im Dreieck

In einem Dreieck werden die Winkel $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ in Einheiten von Grad gemessen. Die Messfehler sind unabhängig mit $\sigma_1 = 0.1, \sigma_2 = 0.08, \sigma_3 = 0.12$. Die gemessenen Werte sind $\boldsymbol{Y} = (34.26, 86.07, 59.52)$. Die gemessenen Winkel sollen mit der Bedingung $\beta_1 + \beta_2 + \beta_3 = 180$ ausgeglichen werden.

Es ist $\mathbf{A} = (1, 1, 1)$ und $c = 180$. $\mathbf{V}$ ist diagonal mit $\mathrm{diag}(\mathbf{V}) = (10^{-2}, 0.64 \cdot 10^{-2}, 1.44 \cdot 10^{-2})$. Weiters ist $\mathbf{G}_{\mathrm{A}} = 32.468$ und $\hat{\boldsymbol{\beta}} = (34.31, 86.10, 59.59)$, mit der Kovarianzmatrix:

$$\mathsf{Cov}[\hat{\boldsymbol{\beta}}] = 10^{-2} \cdot \begin{pmatrix} 0.675 & -0.208 & -0.468 \\ -0.208 & 0.507 & -0.299 \\ -0.468 & -0.299 & 0.767 \end{pmatrix}$$

Da die Summe der ausgeglichenen Winkel eine Konstante ist, sind die Schätzwerte negativ korreliert. Die χ^2-Statistik ist gleich 0.846, der P-Wert gleich 0.36. Dies spricht nicht gegen eine korrekte Einschätzung der Varianz der Fehlerterme.

➠MATLAB: `make_ausgleich_linear` ◀

9.4.2 Nichtlineare Zwangsbedingungen

Sind die Zwangsbedingungen nichtlinear, können sie in der folgenden allgemeinen Form angeschrieben werden:

$$\boldsymbol{h}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = \mathbf{0}$$

Zunächst wird $\boldsymbol{h}$ an einer geeigneten Stelle $(\boldsymbol{\beta}_0, \boldsymbol{\vartheta}_0)$ in eine lineare Funktion entwickelt:

$$\boldsymbol{h}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) \approx \boldsymbol{h}(\boldsymbol{\beta}_0, \boldsymbol{\vartheta}_0) + \mathbf{A}_0(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{B}_0(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) = \mathbf{A}_0\boldsymbol{\beta} + \mathbf{B}_0\boldsymbol{\vartheta} - \boldsymbol{c}_0$$

mit

$$\mathbf{A}_0 = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\beta}}, \quad \mathbf{B}_0 = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\vartheta}}, \quad \boldsymbol{c}_0 = \mathbf{A}_0\boldsymbol{\beta}_0 + \mathbf{B}_0\boldsymbol{\vartheta}_0 - \boldsymbol{h}(\boldsymbol{\beta}_0, \boldsymbol{\vartheta}_0)$$

Die Schätzung von $\boldsymbol{\beta}$ und $\boldsymbol{\vartheta}$ erfolgt nun wie im linearen Fall. Anschließend wird $\boldsymbol{h}$ an den Stellen $\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}$ und $\boldsymbol{\vartheta}_1 = \hat{\boldsymbol{\vartheta}}$ neu entwickelt. Der Vorgang wird so lange iteriert, bis die Zwangsbedingungen genügend genau erfüllt sind.

▶ **Beispiel 9.4.2.** Ausgleich von Strom, Spannung und Widerstand

An einem Widerstand wird der Spannungsabfall $U = 125\,\mathrm{V}$ und der Strom $I = 0.65\,\mathrm{A}$ gemessen, mit einer relativen Standardabweichung von 1% bzw. 0.8%. Der bekannte Wert des Widerstands ist gleich $R = 200\,\Omega$, mit einer Standardbweichung von $0.1\,\Omega$. Die nichtlineare Zwangsbedingung lautet $h(\boldsymbol{\beta}) = h(U, R, I) = U - R \cdot I = 0$. Die Taylorentwicklung an der Stelle $U_0 = 125, R_0 = 200, I_0 = 0.65$ lautet:

$$h(U, R, I) \approx U - RI_0 - IR_0 + R_0I_0$$

Daraus folgt:

$$\mathbf{A}_0 = (1, -0.65, -200), \quad c_0 = -130$$

Der erste Schätzwert ist gleich $\boldsymbol{\beta}_1 = (127.9500, 199.9877, 0.6398)$, der zweite Schätzwert ist gleich $\boldsymbol{\beta}_2 = (127.9502, 199.9879, 0.6398)$. Da $|h(\boldsymbol{\beta}_2)| < 10^{-10}$, kann nach der zweiten Iteration abgebrochen werden. Die χ^2-Statistik ist gleich 9.44, der P-Wert gleich 0.0021, also sehr klein. Das deutet darauf hin, dass die Unsicherheit der Beobachtungen unterschätzt wurde.

➠MATLAB: `make_ausgleich_nichtlinear` ◀

Teil III

Anhänge

Anhang L

Literaturverzeichnis

Wir geben von den zahllosen Büchern über Wahrscheinlichkeitsrechnung und Statistik nur diejenigen an, die während der Niederschrift des vorliegenden Werkes zu Rate gezogen wurden.

1. M. Ahsanullah, *Handbook On Univariate Statistical Distributions*, Trafford Publishing, 2011.

   Kompakte Information über die wichtigsten diskreten und stetigen univariaten Verteilungen.

2. W. Alt, *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen*, Vieweg Verlag, 2002.

   Eine gute Übersicht über Verfahren, die unter anderem in der nichtlinearen Regression eingesetzt werden können.

3. C. Czado und T. Schmidt, *Mathematische Statistik*, Springer Verlag, 2011.

   Leicht lesbare Einführung in die Theorie des Schätzens und Testens.

4. L. Fahrmeir, R. Künstler, I. Pigeot und G. Tutz, *Statistik*, 6. Auflage, Springer Verlag, 2007.

   Ein ausgezeichnetes Lehrbuch, das sich allerdings vorwiegend an Studierende der Wirtschafts- und Sozialwissenschaften wendet.

5. P. D. Hoff, *A First Course in Bayesian Statistical Methods*, Springer Verlag, 2009.

   Eine gut lesbare Einführung in Bayes-Statistik, geht über das hier gebotene deutlich hinaus.

6. J. Hartung, *Statistik*, 15. Auflage, Oldenbourg Wissenschaftsverlag, 2009.

   Ein umfangreiches Lehrbuch über Statistik, das weit über den hier gebrachten Stoff hinausgeht. Viele Themen sind für Studierende der Physik weniger interessant.

7. S. Kotz, N. Balakrishnan und N. L. Johnson, *Continuous Multivariate Distributions, Models and Applications (Volume 1)*, Wiley, 2000.

   Alles, was man über stetige multivariate Verteilungen wissen kann, ist hier zusammengetragen. Es gibt auch ein Gegenstück über diskrete multivariate Verteilungen.

8. L. Lyons, *Statistics for nuclear and particle physicists*, Cambridge University Press, 1986.

   Nett geschriebene Einführung für Kern- und Teilchenphysiker, allerdings mit recht beschränkter Thematik.

9. S. M. Ross, *Statistik für Ingenieure und Naturwissenschafter*, 3. Auflage, Spektrum Akademischer Verlag, 2006.

   Ein Lehrbuch, das sich an den hier angesprochenen Leserkreis wendet und thematisch etwas weiter gespannt ist.

10. P. Rousseeuw und A. Leroy, *Robust Regression and Outlier Detection*, Wiley, 2003.

    Das Standardwerk über robuste Regression.

11. D. S. Sivia, *Data Analysis, a Bayesian Tutorial*, Oxford Univerist Press, 1996.

    Eine Einführung in Datenanalyse mit Bayes-Methoden, setzt einige Kenntnisse in der Wahrscheinlichkeitstheorie voraus.

12. W. Stahel, *Statistische Datenanalyse*, 5. Auflage, Vieweg Verlag, 2008.

    Sehr gut zu lesendes Lehrbuch, das etwas über den hier gebrachten Stoff hinausgeht.

Anhang T

Tabellen

T.1 Verteilungsfunktion der Standardnormalverteilung

T.1.1 $\Phi(x),\ x = 0.005 : 0.005 : 1; \quad \Phi(-x) = 1 - \Phi(x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.005	0.5020	0.205	0.5812	0.405	0.6573	0.605	0.7274	0.805	0.7896
0.010	0.5040	0.210	0.5832	0.410	0.6591	0.610	0.7291	0.810	0.7910
0.015	0.5060	0.215	0.5851	0.415	0.6609	0.615	0.7307	0.815	0.7925
0.020	0.5080	0.220	0.5871	0.420	0.6628	0.620	0.7324	0.820	0.7939
0.025	0.5100	0.225	0.5890	0.425	0.6646	0.625	0.7340	0.825	0.7953
0.030	0.5120	0.230	0.5910	0.430	0.6664	0.630	0.7357	0.830	0.7967
0.035	0.5140	0.235	0.5929	0.435	0.6682	0.635	0.7373	0.835	0.7981
0.040	0.5160	0.240	0.5948	0.440	0.6700	0.640	0.7389	0.840	0.7995
0.045	0.5179	0.245	0.5968	0.445	0.6718	0.645	0.7405	0.845	0.8009
0.050	0.5199	0.250	0.5987	0.450	0.6736	0.650	0.7422	0.850	0.8023
0.055	0.5219	0.255	0.6006	0.455	0.6754	0.655	0.7438	0.855	0.8037
0.060	0.5239	0.260	0.6026	0.460	0.6772	0.660	0.7454	0.860	0.8051
0.065	0.5259	0.265	0.6045	0.465	0.6790	0.665	0.7470	0.865	0.8065
0.070	0.5279	0.270	0.6064	0.470	0.6808	0.670	0.7486	0.870	0.8078
0.075	0.5299	0.275	0.6083	0.475	0.6826	0.675	0.7502	0.875	0.8092
0.080	0.5319	0.280	0.6103	0.480	0.6844	0.680	0.7517	0.880	0.8106
0.085	0.5339	0.285	0.6122	0.485	0.6862	0.685	0.7533	0.885	0.8119
0.090	0.5359	0.290	0.6141	0.490	0.6879	0.690	0.7549	0.890	0.8133
0.095	0.5378	0.295	0.6160	0.495	0.6897	0.695	0.7565	0.895	0.8146
0.100	0.5398	0.300	0.6179	0.500	0.6915	0.700	0.7580	0.900	0.8159
0.105	0.5418	0.305	0.6198	0.505	0.6932	0.705	0.7596	0.905	0.8173
0.110	0.5438	0.310	0.6217	0.510	0.6950	0.710	0.7611	0.910	0.8186
0.115	0.5458	0.315	0.6236	0.515	0.6967	0.715	0.7627	0.915	0.8199
0.120	0.5478	0.320	0.6255	0.520	0.6985	0.720	0.7642	0.920	0.8212
0.125	0.5497	0.325	0.6274	0.525	0.7002	0.725	0.7658	0.925	0.8225
0.130	0.5517	0.330	0.6293	0.530	0.7019	0.730	0.7673	0.930	0.8238
0.135	0.5537	0.335	0.6312	0.535	0.7037	0.735	0.7688	0.935	0.8251
0.140	0.5557	0.340	0.6331	0.540	0.7054	0.740	0.7704	0.940	0.8264
0.145	0.5576	0.345	0.6350	0.545	0.7071	0.745	0.7719	0.945	0.8277
0.150	0.5596	0.350	0.6368	0.550	0.7088	0.750	0.7734	0.950	0.8289
0.155	0.5616	0.355	0.6387	0.555	0.7106	0.755	0.7749	0.955	0.8302
0.160	0.5636	0.360	0.6406	0.560	0.7123	0.760	0.7764	0.960	0.8315
0.165	0.5655	0.365	0.6424	0.565	0.7140	0.765	0.7779	0.965	0.8327
0.170	0.5675	0.370	0.6443	0.570	0.7157	0.770	0.7794	0.970	0.8340
0.175	0.5695	0.375	0.6462	0.575	0.7174	0.775	0.7808	0.975	0.8352
0.180	0.5714	0.380	0.6480	0.580	0.7190	0.780	0.7823	0.980	0.8365
0.185	0.5734	0.385	0.6499	0.585	0.7207	0.785	0.7838	0.985	0.8377
0.190	0.5753	0.390	0.6517	0.590	0.7224	0.790	0.7852	0.990	0.8389
0.195	0.5773	0.395	0.6536	0.595	0.7241	0.795	0.7867	0.995	0.8401
0.200	0.5793	0.400	0.6554	0.600	0.7257	0.800	0.7881	1.000	0.8413

T.1 Verteilungsfunktion der Standardnormalverteilung

T.1.2 $\Phi(x),\ x = 1.005 : 0.005 : 2; \quad \Phi(-x) = 1 - \Phi(x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1.005	0.8426	1.205	0.8859	1.405	0.9200	1.605	0.9458	1.805	0.9645
1.010	0.8438	1.210	0.8869	1.410	0.9207	1.610	0.9463	1.810	0.9649
1.015	0.8449	1.215	0.8878	1.415	0.9215	1.615	0.9468	1.815	0.9652
1.020	0.8461	1.220	0.8888	1.420	0.9222	1.620	0.9474	1.820	0.9656
1.025	0.8473	1.225	0.8897	1.425	0.9229	1.625	0.9479	1.825	0.9660
1.030	0.8485	1.230	0.8907	1.430	0.9236	1.630	0.9484	1.830	0.9664
1.035	0.8497	1.235	0.8916	1.435	0.9244	1.635	0.9490	1.835	0.9667
1.040	0.8508	1.240	0.8925	1.440	0.9251	1.640	0.9495	1.840	0.9671
1.045	0.8520	1.245	0.8934	1.445	0.9258	1.645	0.9500	1.845	0.9675
1.050	0.8531	1.250	0.8944	1.450	0.9265	1.650	0.9505	1.850	0.9678
1.055	0.8543	1.255	0.8953	1.455	0.9272	1.655	0.9510	1.855	0.9682
1.060	0.8554	1.260	0.8962	1.460	0.9279	1.660	0.9515	1.860	0.9686
1.065	0.8566	1.265	0.8971	1.465	0.9285	1.665	0.9520	1.865	0.9689
1.070	0.8577	1.270	0.8980	1.470	0.9292	1.670	0.9525	1.870	0.9693
1.075	0.8588	1.275	0.8988	1.475	0.9299	1.675	0.9530	1.875	0.9696
1.080	0.8599	1.280	0.8997	1.480	0.9306	1.680	0.9535	1.880	0.9699
1.085	0.8610	1.285	0.9006	1.485	0.9312	1.685	0.9540	1.885	0.9703
1.090	0.8621	1.290	0.9015	1.490	0.9319	1.690	0.9545	1.890	0.9706
1.095	0.8632	1.295	0.9023	1.495	0.9325	1.695	0.9550	1.895	0.9710
1.100	0.8643	1.300	0.9032	1.500	0.9332	1.700	0.9554	1.900	0.9713
1.105	0.8654	1.305	0.9041	1.505	0.9338	1.705	0.9559	1.905	0.9716
1.110	0.8665	1.310	0.9049	1.510	0.9345	1.710	0.9564	1.910	0.9719
1.115	0.8676	1.315	0.9057	1.515	0.9351	1.715	0.9568	1.915	0.9723
1.120	0.8686	1.320	0.9066	1.520	0.9357	1.720	0.9573	1.920	0.9726
1.125	0.8697	1.325	0.9074	1.525	0.9364	1.725	0.9577	1.925	0.9729
1.130	0.8708	1.330	0.9082	1.530	0.9370	1.730	0.9582	1.930	0.9732
1.135	0.8718	1.335	0.9091	1.535	0.9376	1.735	0.9586	1.935	0.9735
1.140	0.8729	1.340	0.9099	1.540	0.9382	1.740	0.9591	1.940	0.9738
1.145	0.8739	1.345	0.9107	1.545	0.9388	1.745	0.9595	1.945	0.9741
1.150	0.8749	1.350	0.9115	1.550	0.9394	1.750	0.9599	1.950	0.9744
1.155	0.8760	1.355	0.9123	1.555	0.9400	1.755	0.9604	1.955	0.9747
1.160	0.8770	1.360	0.9131	1.560	0.9406	1.760	0.9608	1.960	0.9750
1.165	0.8780	1.365	0.9139	1.565	0.9412	1.765	0.9612	1.965	0.9753
1.170	0.8790	1.370	0.9147	1.570	0.9418	1.770	0.9616	1.970	0.9756
1.175	0.8800	1.375	0.9154	1.575	0.9424	1.775	0.9621	1.975	0.9759
1.180	0.8810	1.380	0.9162	1.580	0.9429	1.780	0.9625	1.980	0.9761
1.185	0.8820	1.385	0.9170	1.585	0.9435	1.785	0.9629	1.985	0.9764
1.190	0.8830	1.390	0.9177	1.590	0.9441	1.790	0.9633	1.990	0.9767
1.195	0.8840	1.395	0.9185	1.595	0.9446	1.795	0.9637	1.995	0.9770
1.200	0.8849	1.400	0.9192	1.600	0.9452	1.800	0.9641	2.000	0.9772

T.1 Verteilungsfunktion der Standardnormalverteilung

T.1.3 $\Phi(x),\ x = 2.005 : 0.005 : 3; \quad \Phi(-x) = 1 - \Phi(x)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2.005	0.9775	2.205	0.9863	2.405	0.9919	2.605	0.9954	2.805	0.9975
2.010	0.9778	2.210	0.9864	2.410	0.9920	2.610	0.9955	2.810	0.9975
2.015	0.9780	2.215	0.9866	2.415	0.9921	2.615	0.9955	2.815	0.9976
2.020	0.9783	2.220	0.9868	2.420	0.9922	2.620	0.9956	2.820	0.9976
2.025	0.9786	2.225	0.9870	2.425	0.9923	2.625	0.9957	2.825	0.9976
2.030	0.9788	2.230	0.9871	2.430	0.9925	2.630	0.9957	2.830	0.9977
2.035	0.9791	2.235	0.9873	2.435	0.9926	2.635	0.9958	2.835	0.9977
2.040	0.9793	2.240	0.9875	2.440	0.9927	2.640	0.9959	2.840	0.9977
2.045	0.9796	2.245	0.9876	2.445	0.9928	2.645	0.9959	2.845	0.9978
2.050	0.9798	2.250	0.9878	2.450	0.9929	2.650	0.9960	2.850	0.9978
2.055	0.9801	2.255	0.9879	2.455	0.9930	2.655	0.9960	2.855	0.9978
2.060	0.9803	2.260	0.9881	2.460	0.9931	2.660	0.9961	2.860	0.9979
2.065	0.9805	2.265	0.9882	2.465	0.9931	2.665	0.9962	2.865	0.9979
2.070	0.9808	2.270	0.9884	2.470	0.9932	2.670	0.9962	2.870	0.9979
2.075	0.9810	2.275	0.9885	2.475	0.9933	2.675	0.9963	2.875	0.9980
2.080	0.9812	2.280	0.9887	2.480	0.9934	2.680	0.9963	2.880	0.9980
2.085	0.9815	2.285	0.9888	2.485	0.9935	2.685	0.9964	2.885	0.9980
2.090	0.9817	2.290	0.9890	2.490	0.9936	2.690	0.9964	2.890	0.9981
2.095	0.9819	2.295	0.9891	2.495	0.9937	2.695	0.9965	2.895	0.9981
2.100	0.9821	2.300	0.9893	2.500	0.9938	2.700	0.9965	2.900	0.9981
2.105	0.9824	2.305	0.9894	2.505	0.9939	2.705	0.9966	2.905	0.9982
2.110	0.9826	2.310	0.9896	2.510	0.9940	2.710	0.9966	2.910	0.9982
2.115	0.9828	2.315	0.9897	2.515	0.9940	2.715	0.9967	2.915	0.9982
2.120	0.9830	2.320	0.9898	2.520	0.9941	2.720	0.9967	2.920	0.9982
2.125	0.9832	2.325	0.9900	2.525	0.9942	2.725	0.9968	2.925	0.9983
2.130	0.9834	2.330	0.9901	2.530	0.9943	2.730	0.9968	2.930	0.9983
2.135	0.9836	2.335	0.9902	2.535	0.9944	2.735	0.9969	2.935	0.9983
2.140	0.9838	2.340	0.9904	2.540	0.9945	2.740	0.9969	2.940	0.9984
2.145	0.9840	2.345	0.9905	2.545	0.9945	2.745	0.9970	2.945	0.9984
2.150	0.9842	2.350	0.9906	2.550	0.9946	2.750	0.9970	2.950	0.9984
2.155	0.9844	2.355	0.9907	2.555	0.9947	2.755	0.9971	2.955	0.9984
2.160	0.9846	2.360	0.9909	2.560	0.9948	2.760	0.9971	2.960	0.9985
2.165	0.9848	2.365	0.9910	2.565	0.9948	2.765	0.9972	2.965	0.9985
2.170	0.9850	2.370	0.9911	2.570	0.9949	2.770	0.9972	2.970	0.9985
2.175	0.9852	2.375	0.9912	2.575	0.9950	2.775	0.9972	2.975	0.9985
2.180	0.9854	2.380	0.9913	2.580	0.9951	2.780	0.9973	2.980	0.9986
2.185	0.9856	2.385	0.9915	2.585	0.9951	2.785	0.9973	2.985	0.9986
2.190	0.9857	2.390	0.9916	2.590	0.9952	2.790	0.9974	2.990	0.9986
2.195	0.9859	2.395	0.9917	2.595	0.9953	2.795	0.9974	2.995	0.9986
2.200	0.9861	2.400	0.9918	2.600	0.9953	2.800	0.9974	3.000	0.9987

T.2 Quantile der Standardnormalverteilung

T.2.1 $z_\alpha,\ \alpha = 0.501 : 0.001 : 0.7;\quad z_{1-\alpha} = -z_\alpha$

α	z_α	α	z_α	α	z_α	α	z_α	α	z_α
0.501	0.003	0.541	0.103	0.581	0.204	0.621	0.308	0.661	0.415
0.502	0.005	0.542	0.105	0.582	0.207	0.622	0.311	0.662	0.418
0.503	0.008	0.543	0.108	0.583	0.210	0.623	0.313	0.663	0.421
0.504	0.010	0.544	0.111	0.584	0.212	0.624	0.316	0.664	0.423
0.505	0.013	0.545	0.113	0.585	0.215	0.625	0.319	0.665	0.426
0.506	0.015	0.546	0.116	0.586	0.217	0.626	0.321	0.666	0.429
0.507	0.018	0.547	0.118	0.587	0.220	0.627	0.324	0.667	0.432
0.508	0.020	0.548	0.121	0.588	0.222	0.628	0.327	0.668	0.434
0.509	0.023	0.549	0.123	0.589	0.225	0.629	0.329	0.669	0.437
0.510	0.025	0.550	0.126	0.590	0.228	0.630	0.332	0.670	0.440
0.511	0.028	0.551	0.128	0.591	0.230	0.631	0.335	0.671	0.443
0.512	0.030	0.552	0.131	0.592	0.233	0.632	0.337	0.672	0.445
0.513	0.033	0.553	0.133	0.593	0.235	0.633	0.340	0.673	0.448
0.514	0.035	0.554	0.136	0.594	0.238	0.634	0.342	0.674	0.451
0.515	0.038	0.555	0.138	0.595	0.240	0.635	0.345	0.675	0.454
0.516	0.040	0.556	0.141	0.596	0.243	0.636	0.348	0.676	0.457
0.517	0.043	0.557	0.143	0.597	0.246	0.637	0.350	0.677	0.459
0.518	0.045	0.558	0.146	0.598	0.248	0.638	0.353	0.678	0.462
0.519	0.048	0.559	0.148	0.599	0.251	0.639	0.356	0.679	0.465
0.520	0.050	0.560	0.151	0.600	0.253	0.640	0.358	0.680	0.468
0.521	0.053	0.561	0.154	0.601	0.256	0.641	0.361	0.681	0.470
0.522	0.055	0.562	0.156	0.602	0.259	0.642	0.364	0.682	0.473
0.523	0.058	0.563	0.159	0.603	0.261	0.643	0.366	0.683	0.476
0.524	0.060	0.564	0.161	0.604	0.264	0.644	0.369	0.684	0.479
0.525	0.063	0.565	0.164	0.605	0.266	0.645	0.372	0.685	0.482
0.526	0.065	0.566	0.166	0.606	0.269	0.646	0.375	0.686	0.485
0.527	0.068	0.567	0.169	0.607	0.272	0.647	0.377	0.687	0.487
0.528	0.070	0.568	0.171	0.608	0.274	0.648	0.380	0.688	0.490
0.529	0.073	0.569	0.174	0.609	0.277	0.649	0.383	0.689	0.493
0.530	0.075	0.570	0.176	0.610	0.279	0.650	0.385	0.690	0.496
0.531	0.078	0.571	0.179	0.611	0.282	0.651	0.388	0.691	0.499
0.532	0.080	0.572	0.181	0.612	0.285	0.652	0.391	0.692	0.502
0.533	0.083	0.573	0.184	0.613	0.287	0.653	0.393	0.693	0.504
0.534	0.085	0.574	0.187	0.614	0.290	0.654	0.396	0.694	0.507
0.535	0.088	0.575	0.189	0.615	0.292	0.655	0.399	0.695	0.510
0.536	0.090	0.576	0.192	0.616	0.295	0.656	0.402	0.696	0.513
0.537	0.093	0.577	0.194	0.617	0.298	0.657	0.404	0.697	0.516
0.538	0.095	0.578	0.197	0.618	0.300	0.658	0.407	0.698	0.519
0.539	0.098	0.579	0.199	0.619	0.303	0.659	0.410	0.699	0.522
0.540	0.100	0.580	0.202	0.620	0.305	0.660	0.412	0.700	0.524

T.2 Quantile der Standardnormalverteilung

T.2.2 $z_\alpha,\ \alpha = 0.701 : 0.001 : 0.9; \quad z_{1-\alpha} = -z_\alpha$

α	z_α	α	z_α	α	z_α	α	z_α	α	z_α
0.701	0.527	0.741	0.646	0.781	0.776	0.821	0.919	0.861	1.085
0.702	0.530	0.742	0.650	0.782	0.779	0.822	0.923	0.862	1.089
0.703	0.533	0.743	0.653	0.783	0.782	0.823	0.927	0.863	1.094
0.704	0.536	0.744	0.656	0.784	0.786	0.824	0.931	0.864	1.098
0.705	0.539	0.745	0.659	0.785	0.789	0.825	0.935	0.865	1.103
0.706	0.542	0.746	0.662	0.786	0.793	0.826	0.938	0.866	1.108
0.707	0.545	0.747	0.665	0.787	0.796	0.827	0.942	0.867	1.112
0.708	0.548	0.748	0.668	0.788	0.800	0.828	0.946	0.868	1.117
0.709	0.550	0.749	0.671	0.789	0.803	0.829	0.950	0.869	1.122
0.710	0.553	0.750	0.674	0.790	0.806	0.830	0.954	0.870	1.126
0.711	0.556	0.751	0.678	0.791	0.810	0.831	0.958	0.871	1.131
0.712	0.559	0.752	0.681	0.792	0.813	0.832	0.962	0.872	1.136
0.713	0.562	0.753	0.684	0.793	0.817	0.833	0.966	0.873	1.141
0.714	0.565	0.754	0.687	0.794	0.820	0.834	0.970	0.874	1.146
0.715	0.568	0.755	0.690	0.795	0.824	0.835	0.974	0.875	1.150
0.716	0.571	0.756	0.693	0.796	0.827	0.836	0.978	0.876	1.155
0.717	0.574	0.757	0.697	0.797	0.831	0.837	0.982	0.877	1.160
0.718	0.577	0.758	0.700	0.798	0.834	0.838	0.986	0.878	1.165
0.719	0.580	0.759	0.703	0.799	0.838	0.839	0.990	0.879	1.170
0.720	0.583	0.760	0.706	0.800	0.842	0.840	0.994	0.880	1.175
0.721	0.586	0.761	0.710	0.801	0.845	0.841	0.999	0.881	1.180
0.722	0.589	0.762	0.713	0.802	0.849	0.842	1.003	0.882	1.185
0.723	0.592	0.763	0.716	0.803	0.852	0.843	1.007	0.883	1.190
0.724	0.595	0.764	0.719	0.804	0.856	0.844	1.011	0.884	1.195
0.725	0.598	0.765	0.722	0.805	0.860	0.845	1.015	0.885	1.200
0.726	0.601	0.766	0.726	0.806	0.863	0.846	1.019	0.886	1.206
0.727	0.604	0.767	0.729	0.807	0.867	0.847	1.024	0.887	1.211
0.728	0.607	0.768	0.732	0.808	0.871	0.848	1.028	0.888	1.216
0.729	0.610	0.769	0.736	0.809	0.874	0.849	1.032	0.889	1.221
0.730	0.613	0.770	0.739	0.810	0.878	0.850	1.036	0.890	1.227
0.731	0.616	0.771	0.742	0.811	0.882	0.851	1.041	0.891	1.232
0.732	0.619	0.772	0.745	0.812	0.885	0.852	1.045	0.892	1.237
0.733	0.622	0.773	0.749	0.813	0.889	0.853	1.049	0.893	1.243
0.734	0.625	0.774	0.752	0.814	0.893	0.854	1.054	0.894	1.248
0.735	0.628	0.775	0.755	0.815	0.896	0.855	1.058	0.895	1.254
0.736	0.631	0.776	0.759	0.816	0.900	0.856	1.063	0.896	1.259
0.737	0.634	0.777	0.762	0.817	0.904	0.857	1.067	0.897	1.265
0.738	0.637	0.778	0.765	0.818	0.908	0.858	1.071	0.898	1.270
0.739	0.640	0.779	0.769	0.819	0.912	0.859	1.076	0.899	1.276
0.740	0.643	0.780	0.772	0.820	0.915	0.860	1.080	0.900	1.282

T.2 Quantile der Standardnormalverteilung

T.2.3 $z_\alpha,\ \alpha = 0.901 : 0.001 : 1;\quad z_{1-\alpha} = -z_\alpha$

α	z_α	α	z_α	α	z_α	α	z_α	α	z_α
0.901	1.287	0.921	1.412	0.941	1.563	0.961	1.762	0.981	2.075
0.902	1.293	0.922	1.419	0.942	1.572	0.962	1.774	0.982	2.097
0.903	1.299	0.923	1.426	0.943	1.580	0.963	1.787	0.983	2.120
0.904	1.305	0.924	1.433	0.944	1.589	0.964	1.799	0.984	2.144
0.905	1.311	0.925	1.440	0.945	1.598	0.965	1.812	0.985	2.170
0.906	1.317	0.926	1.447	0.946	1.607	0.966	1.825	0.986	2.197
0.907	1.323	0.927	1.454	0.947	1.616	0.967	1.838	0.987	2.226
0.908	1.329	0.928	1.461	0.948	1.626	0.968	1.852	0.988	2.257
0.909	1.335	0.929	1.468	0.949	1.635	0.969	1.866	0.989	2.290
0.910	1.341	0.930	1.476	0.950	1.645	0.970	1.881	0.990	2.326
0.911	1.347	0.931	1.483	0.951	1.655	0.971	1.896	0.991	2.366
0.912	1.353	0.932	1.491	0.952	1.665	0.972	1.911	0.992	2.409
0.913	1.359	0.933	1.499	0.953	1.675	0.973	1.927	0.993	2.457
0.914	1.366	0.934	1.506	0.954	1.685	0.974	1.943	0.994	2.512
0.915	1.372	0.935	1.514	0.955	1.695	0.975	1.960	0.995	2.576
0.916	1.379	0.936	1.522	0.956	1.706	0.976	1.977	0.996	2.652
0.917	1.385	0.937	1.530	0.957	1.717	0.977	1.995	0.997	2.748
0.918	1.392	0.938	1.538	0.958	1.728	0.978	2.014	0.998	2.878
0.919	1.398	0.939	1.546	0.959	1.739	0.979	2.034	0.999	3.090
0.920	1.405	0.940	1.555	0.960	1.751	0.980	2.054	1.000	∞

T.3 Quantile der Gammaverteilung

T.3.1 $\gamma_{\alpha|n,1},\ n = 1:1:40$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
1	0.005	0.010	0.025	0.051	0.693	2.996	3.689	4.605	5.298
2	0.103	0.149	0.242	0.355	1.678	4.744	5.572	6.638	7.430
3	0.338	0.436	0.619	0.818	2.674	6.296	7.225	8.406	9.274
4	0.672	0.823	1.090	1.366	3.672	7.754	8.767	10.045	10.977
5	1.078	1.279	1.623	1.970	4.671	9.154	10.242	11.605	12.594
6	1.537	1.785	2.202	2.613	5.670	10.513	11.668	13.108	14.150
7	2.037	2.330	2.814	3.285	6.670	11.842	13.059	14.571	15.660
8	2.571	2.906	3.454	3.981	7.669	13.148	14.423	16.000	17.134
9	3.132	3.507	4.115	4.695	8.669	14.435	15.763	17.403	18.578
10	3.717	4.130	4.795	5.425	9.669	15.705	17.085	18.783	19.998
11	4.321	4.771	5.491	6.169	10.669	16.962	18.390	20.145	21.398
12	4.943	5.428	6.201	6.924	11.668	18.208	19.682	21.490	22.779
13	5.580	6.099	6.922	7.690	12.668	19.443	20.962	22.821	24.145
14	6.231	6.782	7.654	8.464	13.668	20.669	22.230	24.139	25.497
15	6.893	7.477	8.395	9.246	14.668	21.886	23.490	25.446	26.836
16	7.567	8.181	9.145	10.036	15.668	23.097	24.740	26.743	28.164
17	8.251	8.895	9.903	10.832	16.668	24.301	25.983	28.030	29.482
18	8.943	9.616	10.668	11.634	17.668	25.499	27.219	29.310	30.791
19	9.644	10.346	11.439	12.442	18.668	26.692	28.448	30.581	32.091
20	10.353	11.082	12.217	13.255	19.668	27.879	29.671	31.845	33.383
21	11.069	11.825	12.999	14.072	20.668	29.062	30.888	33.103	34.668
22	11.792	12.574	13.787	14.894	21.668	30.240	32.101	34.355	35.946
23	12.521	13.329	14.580	15.719	22.668	31.415	33.308	35.601	37.218
24	13.255	14.089	15.377	16.549	23.668	32.585	34.511	36.841	38.484
25	13.995	14.853	16.179	17.382	24.667	33.752	35.710	38.077	39.745
26	14.741	15.623	16.984	18.219	25.667	34.916	36.905	39.308	41.000
27	15.491	16.397	17.793	19.058	26.667	36.077	38.096	40.534	42.251
28	16.245	17.175	18.606	19.901	27.667	37.234	39.284	41.757	43.497
29	17.004	17.957	19.422	20.746	28.667	38.389	40.468	42.975	44.738
30	17.767	18.742	20.241	21.594	29.667	39.541	41.649	44.190	45.976
31	18.534	19.532	21.063	22.445	30.667	40.691	42.827	45.401	47.209
32	19.305	20.324	21.888	23.297	31.667	41.838	44.002	46.608	48.439
33	20.079	21.120	22.716	24.153	32.667	42.982	45.174	47.813	49.665
34	20.857	21.919	23.546	25.010	33.667	44.125	46.344	49.014	50.888
35	21.638	22.721	24.379	25.870	34.667	45.266	47.512	50.213	52.107
36	22.422	23.526	25.214	26.731	35.667	46.404	48.677	51.408	53.324
37	23.208	24.333	26.051	27.595	36.667	47.541	49.839	52.601	54.537
38	23.998	25.143	26.891	28.460	37.667	48.675	51.000	53.791	55.748
39	24.791	25.955	27.733	29.327	38.667	49.808	52.158	54.979	56.955
40	25.586	26.770	28.577	30.196	39.667	50.940	53.314	56.164	58.161

T.3 Quantile der Gammaverteilung

T.3.2 $\gamma_{\alpha|n,1},\ n = 41:1:80$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
41	26.384	27.587	29.422	31.066	40.667	52.069	54.469	57.347	59.363
42	27.184	28.406	30.270	31.938	41.667	53.197	55.621	58.528	60.563
43	27.986	29.228	31.119	32.812	42.667	54.324	56.772	59.707	61.761
44	28.791	30.052	31.970	33.687	43.667	55.449	57.921	60.884	62.956
45	29.598	30.877	32.823	34.563	44.667	56.573	59.068	62.058	64.149
46	30.407	31.705	33.678	35.441	45.667	57.695	60.214	63.231	65.341
47	31.218	32.534	34.534	36.320	46.667	58.816	61.358	64.402	66.530
48	32.032	33.365	35.391	37.200	47.667	59.935	62.500	65.571	67.717
49	32.847	34.198	36.250	38.082	48.667	61.054	63.641	66.738	68.902
50	33.664	35.032	37.111	38.965	49.667	62.171	64.781	67.903	70.085
51	34.483	35.869	37.973	39.849	50.667	63.287	65.919	69.067	71.266
52	35.303	36.707	38.836	40.734	51.667	64.402	67.056	70.230	72.446
53	36.125	37.546	39.701	41.620	52.667	65.516	68.191	71.390	73.624
54	36.949	38.387	40.566	42.507	53.667	66.628	69.325	72.549	74.800
55	37.775	39.229	41.434	43.396	54.667	67.740	70.458	73.707	75.974
56	38.602	40.073	42.302	44.285	55.667	68.851	71.590	74.863	77.147
57	39.431	40.918	43.171	45.176	56.667	69.960	72.721	76.018	78.319
58	40.261	41.765	44.042	46.067	57.667	71.069	73.850	77.172	79.489
59	41.093	42.613	44.914	46.959	58.667	72.177	74.978	78.324	80.657
60	41.926	43.462	45.786	47.852	59.667	73.284	76.106	79.475	81.824
61	42.760	44.312	46.660	48.746	60.667	74.390	77.232	80.625	82.990
62	43.596	45.164	47.535	49.641	61.667	75.495	78.357	81.773	84.154
63	44.433	46.016	48.411	50.537	62.667	76.599	79.481	82.921	85.317
64	45.272	46.870	49.288	51.434	63.667	77.702	80.604	84.067	86.479
65	46.111	47.726	50.166	52.331	64.667	78.805	81.727	85.212	87.639
66	46.952	48.582	51.044	53.229	65.667	79.907	82.848	86.355	88.798
67	47.794	49.439	51.924	54.128	66.667	81.008	83.968	87.498	89.956
68	48.637	50.297	52.805	55.028	67.667	82.108	85.088	88.640	91.113
69	49.482	51.157	53.686	55.928	68.667	83.208	86.206	89.781	92.269
70	50.327	52.017	54.568	56.830	69.667	84.306	87.324	90.920	93.423
71	51.174	52.879	55.452	57.732	70.667	85.405	88.441	92.059	94.577
72	52.022	53.741	56.336	58.634	71.667	86.502	89.557	93.196	95.729
73	52.871	54.604	57.220	59.537	72.667	87.599	90.672	94.333	96.881
74	53.720	55.469	58.106	60.441	73.667	88.695	91.787	95.469	98.031
75	54.571	56.334	58.992	61.346	74.667	89.790	92.900	96.604	99.180
76	55.423	57.200	59.879	62.251	75.667	90.885	94.013	97.738	100.328
77	56.276	58.067	60.767	63.157	76.667	91.979	95.125	98.871	101.476
78	57.129	58.935	61.656	64.063	77.667	93.073	96.237	100.003	102.622
79	57.984	59.803	62.545	64.970	78.667	94.166	97.348	101.134	103.767
80	58.840	60.673	63.435	65.878	79.667	95.258	98.458	102.265	104.912

T.3 Quantile der Gammaverteilung

T.3.3 $\gamma_{\alpha|n,1},\ n = 82:2:160$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
82	60.55	62.41	65.22	67.70	81.67	97.44	100.68	104.52	107.20
84	62.27	64.16	67.00	69.51	83.67	99.62	102.89	106.78	109.48
86	63.99	65.91	68.79	71.34	85.67	101.80	105.10	109.03	111.76
88	65.72	67.66	70.58	73.16	87.67	103.98	107.31	111.28	114.04
90	67.44	69.41	72.37	74.98	89.67	106.15	109.52	113.53	116.31
92	69.17	71.17	74.16	76.81	91.67	108.32	111.73	115.77	118.58
94	70.91	72.92	75.96	78.64	93.67	110.50	113.93	118.01	120.85
96	72.64	74.69	77.76	80.47	95.67	112.66	116.13	120.25	123.11
98	74.38	76.45	79.56	82.30	97.67	114.83	118.33	122.49	125.37
100	76.12	78.22	81.36	84.14	99.67	117.00	120.53	124.72	127.63
102	77.86	79.98	83.17	85.98	101.67	119.16	122.72	126.95	129.89
104	79.61	81.76	84.98	87.81	103.67	121.32	124.92	129.18	132.14
106	81.36	83.53	86.78	89.65	105.67	123.48	127.11	131.41	134.39
108	83.11	85.30	88.59	91.49	107.67	125.64	129.30	133.64	136.64
110	84.86	87.08	90.41	93.34	109.67	127.80	131.49	135.86	138.89
112	86.62	88.86	92.22	95.18	111.67	129.96	133.67	138.08	141.13
114	88.38	90.64	94.04	97.02	113.67	132.11	135.86	140.30	143.38
116	90.14	92.42	95.85	98.87	115.67	134.27	138.04	142.52	145.62
118	91.90	94.21	97.67	100.72	117.67	136.42	140.22	144.73	147.85
120	93.66	95.99	99.49	102.57	119.67	138.57	142.40	146.94	150.09
122	95.43	97.78	101.31	104.42	121.67	140.72	144.58	149.16	152.33
124	97.20	99.57	103.14	106.27	123.67	142.87	146.76	151.37	154.56
126	98.97	101.37	104.96	108.12	125.67	145.01	148.93	153.57	156.79
128	100.74	103.16	106.79	109.98	127.67	147.16	151.11	155.78	159.02
130	102.51	104.95	108.61	111.83	129.67	149.31	153.28	157.99	161.24
132	104.28	106.75	110.44	113.69	131.67	151.45	155.45	160.19	163.47
134	106.06	108.55	112.27	115.54	133.67	153.59	157.62	162.39	165.69
136	107.84	110.35	114.10	117.40	135.67	155.73	159.79	164.59	167.91
138	109.62	112.15	115.94	119.26	137.67	157.87	161.96	166.79	170.13
140	111.40	113.95	117.77	121.12	139.67	160.01	164.12	168.99	172.35
142	113.18	115.76	119.61	122.98	141.67	162.15	166.29	171.18	174.57
144	114.97	117.56	121.44	124.85	143.67	164.29	168.45	173.38	176.78
146	116.76	119.37	123.28	126.71	145.67	166.43	170.61	175.57	179.00
148	118.54	121.18	125.12	128.57	147.67	168.56	172.78	177.76	181.21
150	120.33	122.99	126.96	130.44	149.67	170.70	174.94	179.95	183.42
152	122.12	124.80	128.80	132.31	151.67	172.83	177.10	182.14	185.63
154	123.91	126.61	130.64	134.17	153.67	174.96	179.26	184.33	187.84
156	125.71	128.42	132.48	136.04	155.67	177.10	181.41	186.52	190.05
158	127.50	130.24	134.32	137.91	157.67	179.23	183.57	188.70	192.25
160	129.30	132.05	136.17	139.78	159.67	181.36	185.72	190.89	194.46

T.3 Quantile der Gammaverteilung

T.3.4 $\gamma_{\alpha|n,1},\ n = 162:2:240$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
162	131.09	133.87	138.01	141.65	161.67	183.49	187.88	193.07	196.66
164	132.89	135.69	139.86	143.52	163.67	185.62	190.03	195.25	198.86
166	134.69	137.50	141.71	145.39	165.67	187.75	192.19	197.43	201.06
168	136.49	139.32	143.56	147.26	167.67	189.87	194.34	199.61	203.26
170	138.29	141.15	145.41	149.14	169.67	192.00	196.49	201.79	205.46
172	140.10	142.97	147.26	151.01	171.67	194.13	198.64	203.97	207.66
174	141.90	144.79	149.11	152.89	173.67	196.25	200.79	206.15	209.85
176	143.71	146.62	150.96	154.76	175.67	198.38	202.94	208.32	212.05
178	145.51	148.44	152.81	156.64	177.67	200.50	205.08	210.50	214.24
180	147.32	150.27	154.66	158.51	179.67	202.62	207.23	212.67	216.43
182	149.13	152.09	156.52	160.39	181.67	204.74	209.38	214.85	218.63
184	150.94	153.92	158.37	162.27	183.67	206.87	211.52	217.02	220.82
186	152.75	155.75	160.23	164.15	185.67	208.99	213.66	219.19	223.00
188	154.56	157.58	162.09	166.03	187.67	211.11	215.81	221.36	225.19
190	156.37	159.41	163.94	167.91	189.67	213.23	217.95	223.53	227.38
192	158.19	161.24	165.80	169.79	191.67	215.35	220.09	225.70	229.57
194	160.00	163.07	167.66	171.67	193.67	217.46	222.23	227.86	231.75
196	161.82	164.91	169.52	173.55	195.67	219.58	224.37	230.03	233.94
198	163.63	166.74	171.38	175.44	197.67	221.70	226.51	232.20	236.12
200	165.45	168.58	173.24	177.32	199.67	223.82	228.65	234.36	238.30
202	167.27	170.41	175.10	179.20	201.67	225.93	230.79	236.53	240.48
204	169.09	172.25	176.96	181.09	203.67	228.05	232.93	238.69	242.67
206	170.91	174.09	178.83	182.97	205.67	230.16	235.07	240.85	244.85
208	172.73	175.93	180.69	184.86	207.67	232.28	237.20	243.01	247.02
210	174.55	177.77	182.56	186.75	209.67	234.39	239.34	245.17	249.20
212	176.37	179.60	184.42	188.63	211.67	236.50	241.47	247.34	251.38
214	178.20	181.45	186.29	190.52	213.67	238.62	243.61	249.49	253.56
216	180.02	183.29	188.15	192.41	215.67	240.73	245.74	251.65	255.73
218	181.85	185.13	190.02	194.30	217.67	242.84	247.87	253.81	257.91
220	183.67	186.97	191.89	196.18	219.67	244.95	250.01	255.97	260.08
222	185.50	188.82	193.76	198.07	221.67	247.06	252.14	258.12	262.25
224	187.33	190.66	195.62	199.96	223.67	249.17	254.27	260.28	264.43
226	189.16	192.50	197.49	201.85	225.67	251.28	256.40	262.44	266.60
228	190.98	194.35	199.36	203.74	227.67	253.39	258.53	264.59	268.77
230	192.81	196.20	201.23	205.64	229.67	255.50	260.66	266.74	270.94
232	194.65	198.04	203.10	207.53	231.67	257.61	262.79	268.90	273.11
234	196.48	199.89	204.98	209.42	233.67	259.72	264.92	271.05	275.28
236	198.31	201.74	206.85	211.31	235.67	261.82	267.05	273.20	277.45
238	200.14	203.59	208.72	213.21	237.67	263.93	269.17	275.35	279.61
240	201.97	205.44	210.59	215.10	239.67	266.04	271.30	277.50	281.78

T.4 Quantile der Chiquadrat-Verteilung

T.4.1 $\chi^2_{\alpha|n},\ n = 1:1:40$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
1	0.000	0.000	0.001	0.004	0.455	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	1.386	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	2.366	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	3.357	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	4.351	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	5.348	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	6.346	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	7.344	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	8.343	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	9.342	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	10.341	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	11.340	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	12.340	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	13.339	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	14.339	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	15.338	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	16.338	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	17.338	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	18.338	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	19.337	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	20.337	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	21.337	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	22.337	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	23.337	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	24.337	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	25.336	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	26.336	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	27.336	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	28.336	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	29.336	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	30.336	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	31.336	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	32.336	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	33.336	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	34.336	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	35.336	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	36.336	52.192	55.668	59.893	62.883
38	19.289	20.691	22.878	24.884	37.335	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	38.335	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	39.335	55.758	59.342	63.691	66.766

T.4 Quantile der Chiquadrat-Verteilung

T.4.2 $\chi^2_{\alpha|n},\ n = 41:1:80$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
41	21.421	22.906	25.215	27.326	40.335	56.942	60.561	64.950	68.053
42	22.138	23.650	25.999	28.144	41.335	58.124	61.777	66.206	69.336
43	22.859	24.398	26.785	28.965	42.335	59.304	62.990	67.459	70.616
44	23.584	25.148	27.575	29.787	43.335	60.481	64.201	68.710	71.893
45	24.311	25.901	28.366	30.612	44.335	61.656	65.410	69.957	73.166
46	25.041	26.657	29.160	31.439	45.335	62.830	66.617	71.201	74.437
47	25.775	27.416	29.956	32.268	46.335	64.001	67.821	72.443	75.704
48	26.511	28.177	30.755	33.098	47.335	65.171	69.023	73.683	76.969
49	27.249	28.941	31.555	33.930	48.335	66.339	70.222	74.919	78.231
50	27.991	29.707	32.357	34.764	49.335	67.505	71.420	76.154	79.490
51	28.735	30.475	33.162	35.600	50.335	68.669	72.616	77.386	80.747
52	29.481	31.246	33.968	36.437	51.335	69.832	73.810	78.616	82.001
53	30.230	32.018	34.776	37.276	52.335	70.993	75.002	79.843	83.253
54	30.981	32.793	35.586	38.116	53.335	72.153	76.192	81.069	84.502
55	31.735	33.570	36.398	38.958	54.335	73.311	77.380	82.292	85.749
56	32.490	34.350	37.212	39.801	55.335	74.468	78.567	83.513	86.994
57	33.248	35.131	38.027	40.646	56.335	75.624	79.752	84.733	88.236
58	34.008	35.913	38.844	41.492	57.335	76.778	80.936	85.950	89.477
59	34.770	36.698	39.662	42.339	58.335	77.931	82.117	87.166	90.715
60	35.534	37.485	40.482	43.188	59.335	79.082	83.298	88.379	91.952
61	36.301	38.273	41.303	44.038	60.335	80.232	84.476	89.591	93.186
62	37.068	39.063	42.126	44.889	61.335	81.381	85.654	90.802	94.419
63	37.838	39.855	42.950	45.741	62.335	82.529	86.830	92.010	95.649
64	38.610	40.649	43.776	46.595	63.335	83.675	88.004	93.217	96.878
65	39.383	41.444	44.603	47.450	64.335	84.821	89.177	94.422	98.105
66	40.158	42.240	45.431	48.305	65.335	85.965	90.349	95.626	99.330
67	40.935	43.038	46.261	49.162	66.335	87.108	91.519	96.828	100.554
68	41.713	43.838	47.092	50.020	67.335	88.250	92.689	98.028	101.776
69	42.494	44.639	47.924	50.879	68.334	89.391	93.856	99.228	102.996
70	43.275	45.442	48.758	51.739	69.334	90.531	95.023	100.425	104.215
71	44.058	46.246	49.592	52.600	70.334	91.670	96.189	101.621	105.432
72	44.843	47.051	50.428	53.462	71.334	92.808	97.353	102.816	106.648
73	45.629	47.858	51.265	54.325	72.334	93.945	98.516	104.010	107.862
74	46.417	48.666	52.103	55.189	73.334	95.081	99.678	105.202	109.074
75	47.206	49.475	52.942	56.054	74.334	96.217	100.839	106.393	110.286
76	47.997	50.286	53.782	56.920	75.334	97.351	101.999	107.583	111.495
77	48.788	51.097	54.623	57.786	76.334	98.484	103.158	108.771	112.704
78	49.582	51.910	55.466	58.654	77.334	99.617	104.316	109.958	113.911
79	50.376	52.725	56.309	59.522	78.334	100.749	105.473	111.144	115.117
80	51.172	53.540	57.153	60.391	79.334	101.879	106.629	112.329	116.321

T.4 Quantile der Chiquadrat-Verteilung

T.4.3 $\chi^2_{\alpha|n}$, $n = 82:2:160$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
82	52.77	55.17	58.84	62.13	81.33	104.14	108.94	114.69	118.73
84	54.37	56.81	60.54	63.88	83.33	106.39	111.24	117.06	121.13
86	55.97	58.46	62.24	65.62	85.33	108.65	113.54	119.41	123.52
88	57.58	60.10	63.94	67.37	87.33	110.90	115.84	121.77	125.91
90	59.20	61.75	65.65	69.13	89.33	113.15	118.14	124.12	128.30
92	60.81	63.41	67.36	70.88	91.33	115.39	120.43	126.46	130.68
94	62.44	65.07	69.07	72.64	93.33	117.63	122.72	128.80	133.06
96	64.06	66.73	70.78	74.40	95.33	119.87	125.00	131.14	135.43
98	65.69	68.40	72.50	76.16	97.33	122.11	127.28	133.48	137.80
100	67.33	70.06	74.22	77.93	99.33	124.34	129.56	135.81	140.17
102	68.97	71.74	75.95	79.70	101.33	126.57	131.84	138.13	142.53
104	70.61	73.41	77.67	81.47	103.33	128.80	134.11	140.46	144.89
106	72.25	75.09	79.40	83.24	105.33	131.03	136.38	142.78	147.25
108	73.90	76.77	81.13	85.01	107.33	133.26	138.65	145.10	149.60
110	75.55	78.46	82.87	86.79	109.33	135.48	140.92	147.41	151.95
112	77.20	80.15	84.60	88.57	111.33	137.70	143.18	149.73	154.29
114	78.86	81.84	86.34	90.35	113.33	139.92	145.44	152.04	156.64
116	80.52	83.53	88.08	92.13	115.33	142.14	147.70	154.34	158.98
118	82.19	85.23	89.83	93.92	117.33	144.35	149.96	156.65	161.31
120	83.85	86.92	91.57	95.70	119.33	146.57	152.21	158.95	163.65
122	85.52	88.62	93.32	97.49	121.33	148.78	154.46	161.25	165.98
124	87.19	90.33	95.07	99.28	123.33	150.99	156.71	163.55	168.31
126	88.87	92.03	96.82	101.07	125.33	153.20	158.96	165.84	170.63
128	90.54	93.74	98.58	102.87	127.33	155.40	161.21	168.13	172.96
130	92.22	95.45	100.33	104.66	129.33	157.61	163.45	170.42	175.28
132	93.90	97.16	102.09	106.46	131.33	159.81	165.70	172.71	177.60
134	95.59	98.88	103.85	108.26	133.33	162.02	167.94	175.00	179.91
136	97.27	100.59	105.61	110.06	135.33	164.22	170.18	177.28	182.23
138	98.96	102.31	107.37	111.86	137.33	166.42	172.41	179.56	184.54
140	100.65	104.03	109.14	113.66	139.33	168.61	174.65	181.84	186.85
142	102.35	105.76	110.90	115.46	141.33	170.81	176.88	184.12	189.15
144	104.04	107.48	112.67	117.27	143.33	173.00	179.11	186.39	191.46
146	105.74	109.21	114.44	119.07	145.33	175.20	181.34	188.67	193.76
148	107.44	110.94	116.21	120.88	147.33	177.39	183.57	190.94	196.06
150	109.14	112.67	117.98	122.69	149.33	179.58	185.80	193.21	198.36
152	110.85	114.40	119.76	124.50	151.33	181.77	188.03	195.48	200.66
154	112.55	116.13	121.53	126.31	153.33	183.96	190.25	197.74	202.95
156	114.26	117.87	123.31	128.13	155.33	186.15	192.47	200.01	205.24
158	115.97	119.61	125.09	129.94	157.33	188.33	194.70	202.27	207.53
160	117.68	121.35	126.87	131.76	159.33	190.52	196.92	204.53	209.82

T.4 Quantile der Chiquadrat-Verteilung

T.4.4 $\chi^2_{\alpha|n},\ n = 162:2:240$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
162	119.39	123.09	128.65	133.57	161.33	192.70	199.13	206.79	212.11
164	121.11	124.83	130.43	135.39	163.33	194.88	201.35	209.05	214.40
166	122.82	126.57	132.22	137.21	165.33	197.06	203.57	211.30	216.68
168	124.54	128.32	134.00	139.03	167.33	199.24	205.78	213.56	218.96
170	126.26	130.06	135.79	140.85	169.33	201.42	208.00	215.81	221.24
172	127.98	131.81	137.58	142.67	171.33	203.60	210.21	218.06	223.52
174	129.71	133.56	139.37	144.49	173.33	205.78	212.42	220.31	225.80
176	131.43	135.31	141.16	146.32	175.33	207.95	214.63	222.56	228.07
178	133.16	137.07	142.95	148.14	177.33	210.13	216.84	224.81	230.35
180	134.88	138.82	144.74	149.97	179.33	212.30	219.04	227.06	232.62
182	136.61	140.58	146.54	151.80	181.33	214.48	221.25	229.30	234.89
184	138.34	142.33	148.33	153.62	183.33	216.65	223.46	231.54	237.16
186	140.08	144.09	150.13	155.45	185.33	218.82	225.66	233.79	239.43
188	141.81	145.85	151.92	157.28	187.33	220.99	227.86	236.03	241.69
190	143.55	147.61	153.72	159.11	189.33	223.16	230.06	238.27	243.96
192	145.28	149.37	155.52	160.94	191.33	225.33	232.27	240.50	246.22
194	147.02	151.14	157.32	162.78	193.33	227.50	234.46	242.74	248.49
196	148.76	152.90	159.12	164.61	195.33	229.66	236.66	244.98	250.75
198	150.50	154.67	160.92	166.44	197.33	231.83	238.86	247.21	253.01
200	152.24	156.43	162.73	168.28	199.33	233.99	241.06	249.45	255.26
202	153.98	158.20	164.53	170.11	201.33	236.16	243.25	251.68	257.52
204	155.73	159.97	166.34	171.95	203.33	238.32	245.45	253.91	259.78
206	157.47	161.74	168.14	173.79	205.33	240.48	247.64	256.14	262.03
208	159.22	163.51	169.95	175.63	207.33	242.65	249.83	258.37	264.28
210	160.97	165.28	171.76	177.47	209.33	244.81	252.03	260.59	266.54
212	162.72	167.06	173.57	179.30	211.33	246.97	254.22	262.82	268.79
214	164.47	168.83	175.38	181.15	213.33	249.13	256.41	265.05	271.04
216	166.22	170.61	177.19	182.99	215.33	251.29	258.60	267.27	273.29
218	167.97	172.38	179.00	184.83	217.33	253.44	260.79	269.49	275.53
220	169.73	174.16	180.81	186.67	219.33	255.60	262.97	271.72	277.78
222	171.48	175.94	182.63	188.51	221.33	257.76	265.16	273.94	280.02
224	173.24	177.72	184.44	190.36	223.33	259.91	267.35	276.16	282.27
226	174.99	179.50	186.26	192.20	225.33	262.07	269.53	278.38	284.51
228	176.75	181.28	188.07	194.05	227.33	264.22	271.71	280.60	286.75
230	178.51	183.06	189.89	195.89	229.33	266.38	273.90	282.81	288.99
232	180.27	184.85	191.71	197.74	231.33	268.53	276.08	285.03	291.23
234	182.03	186.63	193.52	199.59	233.33	270.68	278.26	287.25	293.47
236	183.80	188.42	195.34	201.44	235.33	272.84	280.44	289.46	295.71
238	185.56	190.20	197.16	203.29	237.33	274.99	282.62	291.68	297.95
240	187.32	191.99	198.98	205.14	239.33	277.14	284.80	293.89	300.18

T.5 Quantile der T-Verteilung

T.5.1 $t_{\alpha|n},\ n = 1:1:40; \quad t_{1-\alpha|n} = -t_{\alpha|n}$

	α									
n	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.998	0.999
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657	159.153	318.309
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	15.764	22.327
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	8.053	10.215
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	5.951	7.173
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.030	5.893
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	4.524	5.208
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.207	4.785
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	3.991	4.501
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	3.835	4.297
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	3.716	4.144
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	3.624	4.025
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.550	3.930
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012	3.489	3.852
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977	3.438	3.787
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947	3.395	3.733
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921	3.358	3.686
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898	3.326	3.646
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878	3.298	3.610
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861	3.273	3.579
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845	3.251	3.552
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831	3.231	3.527
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819	3.214	3.505
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807	3.198	3.485
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797	3.183	3.467
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787	3.170	3.450
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779	3.158	3.435
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771	3.147	3.421
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763	3.136	3.408
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756	3.127	3.396
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750	3.118	3.385
31	0.256	0.530	0.853	1.309	1.696	2.040	2.453	2.744	3.109	3.375
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738	3.102	3.365
33	0.255	0.530	0.853	1.308	1.692	2.035	2.445	2.733	3.094	3.356
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728	3.088	3.348
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724	3.081	3.340
36	0.255	0.529	0.852	1.306	1.688	2.028	2.434	2.719	3.075	3.333
37	0.255	0.529	0.851	1.305	1.687	2.026	2.431	2.715	3.070	3.326
38	0.255	0.529	0.851	1.304	1.686	2.024	2.429	2.712	3.064	3.319
39	0.255	0.529	0.851	1.304	1.685	2.023	2.426	2.708	3.059	3.313
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704	3.055	3.307

T.5 Quantile der T-Verteilung

T.5.2 $t_{\alpha|n},\ n = 42:2:120;\quad t_{1-\alpha|n} = -t_{\alpha|n}$

	α									
n	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.998	0.999
42	0.255	0.528	0.850	1.302	1.682	2.018	2.418	2.698	3.046	3.296
44	0.255	0.528	0.850	1.301	1.680	2.015	2.414	2.692	3.038	3.286
46	0.255	0.528	0.850	1.300	1.679	2.013	2.410	2.687	3.030	3.277
48	0.255	0.528	0.849	1.299	1.677	2.011	2.407	2.682	3.024	3.269
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678	3.018	3.261
52	0.255	0.528	0.849	1.298	1.675	2.007	2.400	2.674	3.012	3.255
54	0.255	0.528	0.848	1.297	1.674	2.005	2.397	2.670	3.007	3.248
56	0.255	0.527	0.848	1.297	1.673	2.003	2.395	2.667	3.002	3.242
58	0.255	0.527	0.848	1.296	1.672	2.002	2.392	2.663	2.998	3.237
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660	2.994	3.232
62	0.254	0.527	0.847	1.295	1.670	1.999	2.388	2.657	2.990	3.227
64	0.254	0.527	0.847	1.295	1.669	1.998	2.386	2.655	2.986	3.223
66	0.254	0.527	0.847	1.295	1.668	1.997	2.384	2.652	2.983	3.218
68	0.254	0.527	0.847	1.294	1.668	1.995	2.382	2.650	2.980	3.214
70	0.254	0.527	0.847	1.294	1.667	1.994	2.381	2.648	2.977	3.211
72	0.254	0.527	0.847	1.293	1.666	1.993	2.379	2.646	2.974	3.207
74	0.254	0.527	0.847	1.293	1.666	1.993	2.378	2.644	2.971	3.204
76	0.254	0.527	0.846	1.293	1.665	1.992	2.376	2.642	2.969	3.201
78	0.254	0.527	0.846	1.292	1.665	1.991	2.375	2.640	2.966	3.198
80	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639	2.964	3.195
82	0.254	0.526	0.846	1.292	1.664	1.989	2.373	2.637	2.962	3.193
84	0.254	0.526	0.846	1.292	1.663	1.989	2.372	2.636	2.960	3.190
86	0.254	0.526	0.846	1.291	1.663	1.988	2.370	2.634	2.958	3.188
88	0.254	0.526	0.846	1.291	1.662	1.987	2.369	2.633	2.956	3.185
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632	2.954	3.183
92	0.254	0.526	0.846	1.291	1.662	1.986	2.368	2.630	2.952	3.181
94	0.254	0.526	0.845	1.291	1.661	1.986	2.367	2.629	2.951	3.179
96	0.254	0.526	0.845	1.290	1.661	1.985	2.366	2.628	2.949	3.177
98	0.254	0.526	0.845	1.290	1.661	1.984	2.365	2.627	2.948	3.175
100	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626	2.946	3.174
102	0.254	0.526	0.845	1.290	1.660	1.983	2.363	2.625	2.945	3.172
104	0.254	0.526	0.845	1.290	1.660	1.983	2.363	2.624	2.944	3.170
106	0.254	0.526	0.845	1.290	1.659	1.983	2.362	2.623	2.942	3.169
108	0.254	0.526	0.845	1.289	1.659	1.982	2.361	2.622	2.941	3.167
110	0.254	0.526	0.845	1.289	1.659	1.982	2.361	2.621	2.940	3.166
112	0.254	0.526	0.845	1.289	1.659	1.981	2.360	2.620	2.939	3.165
114	0.254	0.526	0.845	1.289	1.658	1.981	2.360	2.620	2.938	3.163
116	0.254	0.526	0.845	1.289	1.658	1.981	2.359	2.619	2.937	3.162
118	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.618	2.936	3.161
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617	2.935	3.160

T.5 Quantile der T-Verteilung

T.5.3 $t_{\alpha|n},\ n = 125:5:320;\quad t_{1-\alpha|n} = -t_{\alpha|n}$

	α									
n	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.998	0.999
125	0.254	0.526	0.845	1.288	1.657	1.979	2.357	2.616	2.933	3.157
130	0.254	0.526	0.844	1.288	1.657	1.978	2.355	2.614	2.930	3.154
135	0.254	0.526	0.844	1.288	1.656	1.978	2.354	2.613	2.928	3.152
140	0.254	0.526	0.844	1.288	1.656	1.977	2.353	2.611	2.927	3.149
145	0.254	0.526	0.844	1.287	1.655	1.976	2.352	2.610	2.925	3.147
150	0.254	0.526	0.844	1.287	1.655	1.976	2.351	2.609	2.923	3.145
155	0.254	0.525	0.844	1.287	1.655	1.975	2.351	2.608	2.922	3.144
160	0.254	0.525	0.844	1.287	1.654	1.975	2.350	2.607	2.920	3.142
165	0.254	0.525	0.844	1.287	1.654	1.974	2.349	2.606	2.919	3.140
170	0.254	0.525	0.844	1.287	1.654	1.974	2.348	2.605	2.918	3.139
175	0.254	0.525	0.844	1.286	1.654	1.974	2.348	2.604	2.917	3.137
180	0.254	0.525	0.844	1.286	1.653	1.973	2.347	2.603	2.916	3.136
185	0.254	0.525	0.844	1.286	1.653	1.973	2.347	2.603	2.915	3.135
190	0.254	0.525	0.844	1.286	1.653	1.973	2.346	2.602	2.914	3.134
195	0.254	0.525	0.843	1.286	1.653	1.972	2.346	2.601	2.913	3.133
200	0.254	0.525	0.843	1.286	1.653	1.972	2.345	2.601	2.912	3.131
205	0.254	0.525	0.843	1.286	1.652	1.972	2.345	2.600	2.911	3.130
210	0.254	0.525	0.843	1.286	1.652	1.971	2.344	2.599	2.910	3.129
215	0.254	0.525	0.843	1.286	1.652	1.971	2.344	2.599	2.910	3.129
220	0.254	0.525	0.843	1.285	1.652	1.971	2.343	2.598	2.909	3.128
225	0.254	0.525	0.843	1.285	1.652	1.971	2.343	2.598	2.908	3.127
230	0.254	0.525	0.843	1.285	1.652	1.970	2.343	2.597	2.907	3.126
235	0.254	0.525	0.843	1.285	1.651	1.970	2.342	2.597	2.907	3.125
240	0.254	0.525	0.843	1.285	1.651	1.970	2.342	2.596	2.906	3.125
245	0.254	0.525	0.843	1.285	1.651	1.970	2.342	2.596	2.906	3.124
250	0.254	0.525	0.843	1.285	1.651	1.969	2.341	2.596	2.905	3.123
255	0.254	0.525	0.843	1.285	1.651	1.969	2.341	2.595	2.905	3.122
260	0.254	0.525	0.843	1.285	1.651	1.969	2.341	2.595	2.904	3.122
265	0.254	0.525	0.843	1.285	1.651	1.969	2.341	2.595	2.904	3.121
270	0.254	0.525	0.843	1.285	1.651	1.969	2.340	2.594	2.903	3.121
275	0.254	0.525	0.843	1.285	1.650	1.969	2.340	2.594	2.903	3.120
280	0.254	0.525	0.843	1.285	1.650	1.968	2.340	2.594	2.902	3.120
285	0.254	0.525	0.843	1.285	1.650	1.968	2.340	2.593	2.902	3.119
290	0.254	0.525	0.843	1.284	1.650	1.968	2.339	2.593	2.901	3.119
295	0.254	0.525	0.843	1.284	1.650	1.968	2.339	2.593	2.901	3.118
300	0.254	0.525	0.843	1.284	1.650	1.968	2.339	2.592	2.901	3.118
305	0.254	0.525	0.843	1.284	1.650	1.968	2.339	2.592	2.900	3.117
310	0.254	0.525	0.843	1.284	1.650	1.968	2.338	2.592	2.900	3.117
315	0.254	0.525	0.843	1.284	1.650	1.968	2.338	2.592	2.900	3.116
320	0.254	0.525	0.843	1.284	1.650	1.967	2.338	2.591	2.899	3.116

T.6 Quantile der Poissonverteilung

T.6.1 $P_{\alpha|n},\ n = 1{:}1{:}40$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
1	0.0	0.0	0.0	0.0	1.0	3.0	3.0	4.0	4.0
2	0.0	0.0	0.0	0.0	2.0	5.0	5.0	6.0	6.0
3	0.0	0.0	0.0	1.0	3.0	6.0	7.0	8.0	8.0
4	0.0	0.0	1.0	1.0	4.0	8.0	8.0	9.0	10.0
5	0.0	1.0	1.0	2.0	5.0	9.0	10.0	11.0	12.0
6	1.0	1.0	2.0	2.0	6.0	10.0	11.0	12.0	13.0
7	1.0	2.0	2.0	3.0	7.0	12.0	13.0	14.0	15.0
8	2.0	2.0	3.0	4.0	8.0	13.0	14.0	15.0	16.0
9	2.0	3.0	4.0	4.0	9.0	14.0	15.0	17.0	18.0
10	3.0	3.0	4.0	5.0	10.0	15.0	17.0	18.0	19.0
11	4.0	4.0	5.0	6.0	11.0	17.0	18.0	19.0	20.0
12	4.0	5.0	6.0	7.0	12.0	18.0	19.0	21.0	22.0
13	5.0	5.0	6.0	7.0	13.0	19.0	21.0	22.0	23.0
14	5.0	6.0	7.0	8.0	14.0	20.0	22.0	23.0	25.0
15	6.0	7.0	8.0	9.0	15.0	22.0	23.0	25.0	26.0
16	7.0	8.0	9.0	10.0	16.0	23.0	24.0	26.0	27.0
17	7.0	8.0	9.0	11.0	17.0	24.0	26.0	27.0	28.0
18	8.0	9.0	10.0	11.0	18.0	25.0	27.0	29.0	30.0
19	9.0	10.0	11.0	12.0	19.0	26.0	28.0	30.0	31.0
20	10.0	10.0	12.0	13.0	20.0	28.0	29.0	31.0	32.0
21	10.0	11.0	13.0	14.0	21.0	29.0	30.0	32.0	34.0
22	11.0	12.0	13.0	15.0	22.0	30.0	32.0	34.0	35.0
23	12.0	13.0	14.0	15.0	23.0	31.0	33.0	35.0	36.0
24	12.0	13.0	15.0	16.0	24.0	32.0	34.0	36.0	38.0
25	13.0	14.0	16.0	17.0	25.0	33.0	35.0	37.0	39.0
26	14.0	15.0	17.0	18.0	26.0	35.0	36.0	39.0	40.0
27	15.0	16.0	17.0	19.0	27.0	36.0	38.0	40.0	41.0
28	15.0	16.0	18.0	20.0	28.0	37.0	39.0	41.0	43.0
29	16.0	17.0	19.0	20.0	29.0	38.0	40.0	42.0	44.0
30	17.0	18.0	20.0	21.0	30.0	39.0	41.0	43.0	45.0
31	18.0	19.0	21.0	22.0	31.0	40.0	42.0	45.0	46.0
32	18.0	20.0	21.0	23.0	32.0	42.0	44.0	46.0	47.0
33	19.0	20.0	22.0	24.0	33.0	43.0	45.0	47.0	49.0
34	20.0	21.0	23.0	25.0	34.0	44.0	46.0	48.0	50.0
35	21.0	22.0	24.0	26.0	35.0	45.0	47.0	49.0	51.0
36	22.0	23.0	25.0	26.0	36.0	46.0	48.0	51.0	52.0
37	22.0	24.0	26.0	27.0	37.0	47.0	49.0	52.0	54.0
38	23.0	24.0	26.0	28.0	38.0	48.0	51.0	53.0	55.0
39	24.0	25.0	27.0	29.0	39.0	50.0	52.0	54.0	56.0
40	25.0	26.0	28.0	30.0	40.0	51.0	53.0	55.0	57.0

T.6 Quantile der Poissonverteilung

T.6.2 $P_{\alpha|n},\ n = 41:1:80$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
41	25.0	27.0	29.0	31.0	41.0	52.0	54.0	57.0	58.0
42	26.0	28.0	30.0	32.0	42.0	53.0	55.0	58.0	60.0
43	27.0	29.0	31.0	33.0	43.0	54.0	56.0	59.0	61.0
44	28.0	29.0	31.0	33.0	44.0	55.0	57.0	60.0	62.0
45	29.0	30.0	32.0	34.0	45.0	56.0	59.0	61.0	63.0
46	30.0	31.0	33.0	35.0	46.0	57.0	60.0	62.0	64.0
47	30.0	32.0	34.0	36.0	47.0	59.0	61.0	64.0	66.0
48	31.0	33.0	35.0	37.0	48.0	60.0	62.0	65.0	67.0
49	32.0	33.0	36.0	38.0	49.0	61.0	63.0	66.0	68.0
50	33.0	34.0	37.0	39.0	50.0	62.0	64.0	67.0	69.0
51	34.0	35.0	38.0	40.0	51.0	63.0	65.0	68.0	70.0
52	34.0	36.0	38.0	40.0	52.0	64.0	67.0	69.0	71.0
53	35.0	37.0	39.0	41.0	53.0	65.0	68.0	71.0	73.0
54	36.0	38.0	40.0	42.0	54.0	66.0	69.0	72.0	74.0
55	37.0	39.0	41.0	43.0	55.0	67.0	70.0	73.0	75.0
56	38.0	39.0	42.0	44.0	56.0	69.0	71.0	74.0	76.0
57	39.0	40.0	43.0	45.0	57.0	70.0	72.0	75.0	77.0
58	39.0	41.0	44.0	46.0	58.0	71.0	73.0	76.0	79.0
59	40.0	42.0	44.0	47.0	59.0	72.0	75.0	78.0	80.0
60	41.0	43.0	45.0	48.0	60.0	73.0	76.0	79.0	81.0
61	42.0	44.0	46.0	48.0	61.0	74.0	77.0	80.0	82.0
62	43.0	44.0	47.0	49.0	62.0	75.0	78.0	81.0	83.0
63	44.0	45.0	48.0	50.0	63.0	76.0	79.0	82.0	84.0
64	44.0	46.0	49.0	51.0	64.0	77.0	80.0	83.0	86.0
65	45.0	47.0	50.0	52.0	65.0	79.0	81.0	84.0	87.0
66	46.0	48.0	51.0	53.0	66.0	80.0	82.0	86.0	88.0
67	47.0	49.0	51.0	54.0	67.0	81.0	83.0	87.0	89.0
68	48.0	50.0	52.0	55.0	68.0	82.0	85.0	88.0	90.0
69	49.0	50.0	53.0	56.0	69.0	83.0	86.0	89.0	91.0
70	49.0	51.0	54.0	57.0	70.0	84.0	87.0	90.0	92.0
71	50.0	52.0	55.0	57.0	71.0	85.0	88.0	91.0	94.0
72	51.0	53.0	56.0	58.0	72.0	86.0	89.0	92.0	95.0
73	52.0	54.0	57.0	59.0	73.0	87.0	90.0	94.0	96.0
74	53.0	55.0	58.0	60.0	74.0	88.0	91.0	95.0	97.0
75	54.0	56.0	59.0	61.0	75.0	90.0	92.0	96.0	98.0
76	55.0	56.0	59.0	62.0	76.0	91.0	94.0	97.0	99.0
77	55.0	57.0	60.0	63.0	77.0	92.0	95.0	98.0	101.0
78	56.0	58.0	61.0	64.0	78.0	93.0	96.0	99.0	102.0
79	57.0	59.0	62.0	65.0	79.0	94.0	97.0	100.0	103.0
80	58.0	60.0	63.0	66.0	80.0	95.0	98.0	102.0	104.0

T.6 Quantile der Poissonverteilung

T.6.3 $P_{\alpha|n},\ n = 82:2:160$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
82	60.0	62.0	65.0	67.0	82.0	97.0	100.0	104.0	106.0
84	61.0	63.0	67.0	69.0	84.0	99.0	102.0	106.0	109.0
86	63.0	65.0	68.0	71.0	86.0	102.0	105.0	108.0	111.0
88	65.0	67.0	70.0	73.0	88.0	104.0	107.0	111.0	113.0
90	67.0	69.0	72.0	75.0	90.0	106.0	109.0	113.0	115.0
92	68.0	70.0	74.0	77.0	92.0	108.0	111.0	115.0	118.0
94	70.0	72.0	75.0	78.0	94.0	110.0	113.0	117.0	120.0
96	72.0	74.0	77.0	80.0	96.0	112.0	116.0	120.0	122.0
98	73.0	76.0	79.0	82.0	98.0	115.0	118.0	122.0	124.0
100	75.0	77.0	81.0	84.0	100.0	117.0	120.0	124.0	127.0
102	77.0	79.0	83.0	86.0	102.0	119.0	122.0	126.0	129.0
104	79.0	81.0	85.0	88.0	104.0	121.0	124.0	128.0	131.0
106	80.0	83.0	86.0	89.0	106.0	123.0	127.0	131.0	133.0
108	82.0	85.0	88.0	91.0	108.0	125.0	129.0	133.0	136.0
110	84.0	86.0	90.0	93.0	110.0	128.0	131.0	135.0	138.0
112	86.0	88.0	92.0	95.0	112.0	130.0	133.0	137.0	140.0
114	87.0	90.0	94.0	97.0	114.0	132.0	135.0	140.0	142.0
116	89.0	92.0	95.0	99.0	116.0	134.0	138.0	142.0	145.0
118	91.0	93.0	97.0	100.0	118.0	136.0	140.0	144.0	147.0
120	93.0	95.0	99.0	102.0	120.0	138.0	142.0	146.0	149.0
122	95.0	97.0	101.0	104.0	122.0	140.0	144.0	148.0	151.0
124	96.0	99.0	103.0	106.0	124.0	143.0	146.0	151.0	154.0
126	98.0	101.0	104.0	108.0	126.0	145.0	148.0	153.0	156.0
128	100.0	102.0	106.0	110.0	128.0	147.0	151.0	155.0	158.0
130	102.0	104.0	108.0	112.0	130.0	149.0	153.0	157.0	160.0
132	103.0	106.0	110.0	113.0	132.0	151.0	155.0	159.0	163.0
134	105.0	108.0	112.0	115.0	134.0	153.0	157.0	162.0	165.0
136	107.0	110.0	114.0	117.0	136.0	155.0	159.0	164.0	167.0
138	109.0	111.0	115.0	119.0	138.0	158.0	161.0	166.0	169.0
140	110.0	113.0	117.0	121.0	140.0	160.0	164.0	168.0	171.0
142	112.0	115.0	119.0	123.0	142.0	162.0	166.0	170.0	174.0
144	114.0	117.0	121.0	125.0	144.0	164.0	168.0	173.0	176.0
146	116.0	119.0	123.0	126.0	146.0	166.0	170.0	175.0	178.0
148	118.0	120.0	125.0	128.0	148.0	168.0	172.0	177.0	180.0
150	119.0	122.0	126.0	130.0	150.0	170.0	174.0	179.0	182.0
152	121.0	124.0	128.0	132.0	152.0	173.0	177.0	181.0	185.0
154	123.0	126.0	130.0	134.0	154.0	175.0	179.0	184.0	187.0
156	125.0	128.0	132.0	136.0	156.0	177.0	181.0	186.0	189.0
158	127.0	130.0	134.0	138.0	158.0	179.0	183.0	188.0	191.0
160	128.0	131.0	136.0	139.0	160.0	181.0	185.0	190.0	193.0

T.6 Quantile der Poissonverteilung

T.6.4 $P_{\alpha|n},\ n = 162:2:240$

	α								
n	0.005	0.010	0.025	0.050	0.500	0.950	0.975	0.990	0.995
162	130.0	133.0	138.0	141.0	162.0	183.0	187.0	192.0	196.0
164	132.0	135.0	139.0	143.0	164.0	185.0	190.0	195.0	198.0
166	134.0	137.0	141.0	145.0	166.0	187.0	192.0	197.0	200.0
168	136.0	139.0	143.0	147.0	168.0	190.0	194.0	199.0	202.0
170	137.0	140.0	145.0	149.0	170.0	192.0	196.0	201.0	205.0
172	139.0	142.0	147.0	151.0	172.0	194.0	198.0	203.0	207.0
174	141.0	144.0	149.0	153.0	174.0	196.0	200.0	205.0	209.0
176	143.0	146.0	150.0	154.0	176.0	198.0	202.0	208.0	211.0
178	145.0	148.0	152.0	156.0	178.0	200.0	205.0	210.0	213.0
180	146.0	150.0	154.0	158.0	180.0	202.0	207.0	212.0	215.0
182	148.0	151.0	156.0	160.0	182.0	204.0	209.0	214.0	218.0
184	150.0	153.0	158.0	162.0	184.0	207.0	211.0	216.0	220.0
186	152.0	155.0	160.0	164.0	186.0	209.0	213.0	218.0	222.0
188	154.0	157.0	162.0	166.0	188.0	211.0	215.0	221.0	224.0
190	155.0	159.0	163.0	168.0	190.0	213.0	217.0	223.0	226.0
192	157.0	161.0	165.0	170.0	192.0	215.0	220.0	225.0	229.0
194	159.0	162.0	167.0	171.0	194.0	217.0	222.0	227.0	231.0
196	161.0	164.0	169.0	173.0	196.0	219.0	224.0	229.0	233.0
198	163.0	166.0	171.0	175.0	198.0	221.0	226.0	231.0	235.0
200	165.0	168.0	173.0	177.0	200.0	224.0	228.0	234.0	237.0
202	166.0	170.0	175.0	179.0	202.0	226.0	230.0	236.0	240.0
204	168.0	172.0	176.0	181.0	204.0	228.0	232.0	238.0	242.0
206	170.0	173.0	178.0	183.0	206.0	230.0	235.0	240.0	244.0
208	172.0	175.0	180.0	185.0	208.0	232.0	237.0	242.0	246.0
210	174.0	177.0	182.0	186.0	210.0	234.0	239.0	244.0	248.0
212	175.0	179.0	184.0	188.0	212.0	236.0	241.0	247.0	250.0
214	177.0	181.0	186.0	190.0	214.0	238.0	243.0	249.0	253.0
216	179.0	183.0	188.0	192.0	216.0	240.0	245.0	251.0	255.0
218	181.0	184.0	190.0	194.0	218.0	243.0	247.0	253.0	257.0
220	183.0	186.0	191.0	196.0	220.0	245.0	250.0	255.0	259.0
222	185.0	188.0	193.0	198.0	222.0	247.0	252.0	257.0	261.0
224	186.0	190.0	195.0	200.0	224.0	249.0	254.0	260.0	263.0
226	188.0	192.0	197.0	202.0	226.0	251.0	256.0	262.0	266.0
228	190.0	194.0	199.0	203.0	228.0	253.0	258.0	264.0	268.0
230	192.0	195.0	201.0	205.0	230.0	255.0	260.0	266.0	270.0
232	194.0	197.0	203.0	207.0	232.0	257.0	262.0	268.0	272.0
234	196.0	199.0	205.0	209.0	234.0	259.0	264.0	270.0	274.0
236	197.0	201.0	206.0	211.0	236.0	262.0	267.0	272.0	276.0
238	199.0	203.0	208.0	213.0	238.0	264.0	269.0	275.0	279.0
240	201.0	205.0	210.0	215.0	240.0	266.0	271.0	277.0	281.0