Statistik

R. Frühwirth
rudolf.fruehwirth@oeaw.ac.at

VO 142.090

5. März 2019

Übersicht über die Vorlesung

Übersicht über die Vorlesung

Teil 1

Zufallsvariable und Ereignisse

Teil 1: Zufallsvariable und Ereignisse

1.1 Einleitung

- Physiker/innen benützen Statistik zur Auswertung von Experimenten im weitesten Sinn.
- Das konkrete Ergebnis eines Experiments kann im Allgemeinen nicht genau vorausgesagt werden.
- Mehrere Gründe:
  - Mangelnde Kenntnis des Anfangszustandes: **Würfel**
  - Die beobachteten Objekte sind eine zufällige Auswahl (Stichprobe) aus einer größeren Grundgesamtheit: **Umfrage**
  - Der beobachtete Prozess ist prinzipiell indeterministisch (Quantenmechanik): **radioaktiver Zerfall**
  - Messfehler geben dem Ergebnis einen stochastischen Charakter: die meisten **Messvorgänge**
- Wir betrachten daher das konkrete Ergebnis eines Experiments als die Realisierung einer **Zufallsvariablen**.

1.1 Einleitung

- Die mathematisch exakte Definition einer Zufallsvariablen wird im Rahmen der Maßtheorie formuliert. Wir begnügen uns hier mit der intuitiven Definition.
- Ist die Realisierung eine Zahl $x \in \mathbb{R}$, wird X als **univariate** Zufallsvariable bezeichnet.
- Eine **multivariate** Zufallsvariable oder ein **Zufallsvektor** $\boldsymbol{X} = (X_1, \ldots, X_n)$ der Dimension n ist ein Vektor von n univariaten Zufallsvariablen.
- Die Realisierung einer univariaten Zufallsvariablen X wird auch als **Merkmal** bezeichnet.
- Die möglichen Werte eines Merkmals werden dessen **Ausprägungen** genannt.
- Ein Merkmal kann **diskrete** (endlich oder abzählbar unendlich viele) oder **kontinuierliche** (überabzählbar unendlich viele) Ausprägungen haben.

1.2 Merkmale

1.2.1 Merkmalstypen

1.2.1 Merkmalstypen

Qualitative Merkmale

- **binär** (ja=1/nein=0). Beispiel: EU-Bürgerschaft.
- **kategorial** (Kategorie $1, \ldots, k$). Beispiel: ledig=1/geschieden=2/verheiratet=3/verwitwet=4.
- **ordinal** (Rang). Beispiel: Noten 1–5.

Quantitative Merkmale

- **diskret** (ganzzahlig). Beispiel: Zählvorgang.
- **kontinuierlich** (reellwertig). Beispiel: Messvorgang.

1.2.1 Skalentypen

1.2.1 Skalentypen

- **Nominalskala**: Zahlenwerte sind nur Bezeichnung für sich ausschließende Kategorien.
- **Ordinalskala**: Ordnung der Zahlen ist wesentlich.
- **Intervallskala**: Ordnung und Differenzen zwischen den Werten sind sinnvoll interpretierbar, der Nullpunkt ist willkürlich festgelegt.
- **Verhältnisskala**: Ordnung, Differenzen und Größenverhältnisse sind sinnvoll interpretierbar, es gibt einen absoluten Nullpunkt.

1.2.1 Skalentypen

Beispiel

1. Der Familienstand einer Person wird durch Zahlen kodiert (1=ledig, 2=verheiratet, 3=geschieden, 4=verwitwet). Nominalskala.
2. Der Stand einer Mannschaft in der Meisterschaft wird durch den Rang in der Liga angegeben. Ordinalskala.
3. Die Jahreszahlen (2007, 2008, ...) bilden eine Intervallskala, da der Nullpunkt willkürlich festgelegt ist.
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

In der folgenden Datenmatrix D sind drei Merkmale von acht Personen erhoben:
G=Geschlecht (1=W,2=M)
A=Alter (in Jahren)
B=Ausbildung (1=Pflichtschule, 2=Höhere Schule, 3=Bachelor, 4=Master)

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

1.3.1 Der Ergebnisraum

- Eine univariate Zufallsvariable X ist das numerische Resultat eines Experiments (im weitesten Sinn).
- Die Menge $\Omega \subseteq \mathbb{R}$ aller möglichen Ausprägungen einer Zufallsvariablen heißt **Ergebnisraum** oder **Stichprobenraum**.
- Der Ergebnisraum Ω kann endlich, abzählbar unendlich oder überabzählbar unendlich sein.

Beispiel

- Beim Roulette gibt es 37 mögliche Ergebnisse. Der Ergebnisraum $\Omega = \{0, 1, \ldots, 36\}$ ist endlich.
- Wird eine radioaktive Quelle beobachtet, ist die Anzahl der Zerfälle pro Sekunde im Prinzip unbeschränkt. Der Ergebnisraum $\Omega = \{0, 1, 2, \ldots\}$ ist abzählbar unendlich.
- Die Wartezeit zwischen zwei Zerfällen kann jeden beliebigen positiven Wert annehmen. Der Ergebnisraum $\Omega = \mathbb{R}_+$ ist überabzählbar unendlich.

1.3.1 Die Ereignisalgebra

Definition (Ereignis)

Ein **Ereignis** E ist eine Teilmenge des Ergebnisraums Ω. Ein Ereignis E **tritt ein**, wenn E das Ergebnis $x \in \Omega$ des Experiments (die Realisierung der Zufallsvariablen X) enthält.

Beispiel

Der Wurf mit einem Würfel hat den Ergebnisraum $\Omega = \{1, 2, 3, 4, 5, 6\}$. Das Ereignis G (gerade Zahl) ist die Teilmenge

$$G = \{2, 4, 6\}$$

G tritt ein, wenn eine gerade Zahl geworfen wird.

Definition (Elementarereignis)

Ein **Elementarereignis** e ist eine einelementige Teilmenge des Ergebnisraums Ω.

1.3.1 Die Ereignisalgebra

Definition (Ereignisalgebra)

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

- Im endlichen oder abzählbar unendlichen Fall kann **jede** Teilmenge als Ereignis betrachtet werden. Die Ereignisalgebra heißt **diskret**.
- Im überabzählbar unendlichen Fall müssen gewisse pathologische (nicht messbare) Teilmengen ausgeschlossen werden. Die Ereignisalgebra heißt **kontinuierlich** oder **stetig**.
- Zwei Ereignisse $A \in \mathfrak{S}$ und $B \in \mathfrak{S}$ können logisch **verknüpft werden**.

1.3.1 Die Ereignisalgebra

Verknüpfung von Ereignissen

Disjunktion

Symbol	Name	Bedeutung
$A \cup B$	Disjunktion	A oder B (oder beide)

Konjunktion

Symbol	Name	Bedeutung
$A \cap B$	Konjunktion	A und B (sowohl A als auch B)

Negation

Symbol	Name	Bedeutung
A'	Negation	nicht A (das Gegenteil von A)

1.3.1 Die Ereignisalgebra

- Mit diesen Verknüpfungen ist $\mathfrak{S}$ ist eine **Boole'sche Algebra**: distributiver komplementärer Verband mit Null- und Einselement.
- Das Nullelement $\mathbf{0} = \emptyset$ ist das **unmögliche Ereignis**.
- Das Einselement $\mathbf{1} = \Omega$ ist das **sichere Ereignis**.
- Ein Ereignis, das genau ein mögliches Ergebnis enthält, ist ein Elementarereignis.

Beispiel

Es seien A, B, C drei Ereignisse. Wir können mittels Verknüpfungen die folgenden Ereignisse formulieren:

1. Alle drei Ereignisse treten ein:
$$A \cap B \cap C$$
2. A und C treten ein, B nicht:
$$A \cap B' \cap C$$
3. Genau zwei der Ereignisse treten ein:
$$(A \cap B \cap C') \cup (A \cap B' \cap C) \cup (A' \cap B \cap C)$$
4. Höchstens eines der Ereignisse tritt ein:
$$(A \cap B' \cap C') \cup (A' \cap B \cap C') \cup (A' \cap B' \cap C) \cup (A' \cap B' \cap C')$$

1.3.1 Wiederholte Experimente

- Der Wurf mit einem Würfel hat den Ergebnisraum

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

  Die Ereignisalgebra $\mathfrak{S}(\Omega)$ hat folglich sechs Elementarereignisse:

  $$e_1 = \{1\}, e_2 = \{2\}, e_3 = \{3\}, e_4 = \{4\}, e_5 = \{5\}, e_6 = \{6\}$$

  und insgesamt $2^6 = 64$ Ereignisse (Teilmengen von Ω).

- Der Ergebnisraum des zweimaligen Würfelns ist das **kartesische Produkt** $\Omega \times \Omega$:

  $$\Omega \times \Omega = \{(i, j) | i, j = 1, \ldots, 6\}$$

  Das geordnete Paar (i, j) bedeutet: i beim ersten Wurf, j beim zweiten Wurf. Die Ereignisalgebra $\mathfrak{S}(\Omega \times \Omega)$ hat folglich 36 Elementarereignisse e_{ij}:

  $$e_{1,1} = \{(1, 1)\}, \ldots, e_{6,6} = \{(6, 6)\}$$

1.3.1 Wiederholte Experimente

- Analog ist beim n-maligen Würfeln der Ergebnisraum das n-fache kartesische Produkt $\Omega \times \Omega \times \ldots \times \Omega$.

Beispiel (Ereignisalgebra des Doppelwurfs)

Beispiele für Elemente 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:	$\{(1,6),(2,5),\ldots,(6,1)\}$

Beispiel (Wiederholter Alternativversuch)

Ein Experiment, das nur zwei mögliche Ergebnisse hat, heißt ein **Alternativversuch** oder **Bernoulliexperiment**. Es gibt zwei Ergebnisse, 0 und 1. Wird ein Alternativversuch n-mal durchgeführt, ergibt sich eine Ergebnisraum mit 2^n Ergebnissen, nämlich den Folgen der Form $(i_1, \ldots, i_n)$ mit $i_j = 0$ oder 1. Jedes Ergebnis kann als Realisierung einer multivariaten Zufallsvariablen der Dimension n aufgefasst werden. Ein Beispiel ist das n-malige Werfen einer Münze.

Teil 2

Deskriptive Statistik

Teil 2: Deskriptive Statistik

2.1.1 Grundbegriffe

Definition von Statistik

1. Die Erhebung und Speicherung von Daten, z.B. durch statistische Ämter
2. Die mathematische Auswertung von Daten, z.B. die Berechnung von Maß- und Kennzahlen

Deskriptive Statistik

- Beschreibung von vorhandenen Daten durch Maßzahlen, Tabellen, Graphiken

2.1.1 Grundbegriffe

Induktive Statistik

- Untersuchung von Gesetzmäßigkeiten und Ursachen, die hinter den Daten stehen und die Daten (teilweise) erklären.
- Explorative Datenanalyse: Ziel ist, Hypothesen für die Theoriebildung zu gewinnen
- Konfirmative Datenanalyse: Ziel ist, vorhandene Theorien zu prüfen, z.B. durch Schätzen von Parametern oder Testen von Hypothesen

2.1.1 Häufigkeit

Definition (Absolute Häufigkeit)

Es sei $\boldsymbol{x} = (x_1, \ldots, x_n)$ eine Stichprobe der Zufallsvariablen X und $E \subseteq \Omega$ ein Ereignis. Die absolute Häufigkeit $h(E)$ von E in $\boldsymbol{x}$ ist die Anzahl der Beobachtungen, in denen E eintritt, d.h. $x_i \in E$.

Beispiel

E ist das Ereignis "Eine Person in Matrix D hat zumindest Bakkalaureat" $(B \in \{3, 4\})$. Dann ist $h(E) = 4$.

2.1.1 Häufigkeit

Definition (Relative Häufigkeit)

Die relative Häufigkeit $f(E) = h(E)/n$ von E ist die Anzahl der Fälle, in denen E eintritt, dividiert durch die Größe der Stichprobe.

Beispiel

E ist das Ereignis "Eine Person x in Matrix D ist älter als dreißig Jahre" ($A \geq 30$). Dann ist $f(E) = 7/8$.

Spezielle Ereignisse

- $E = \emptyset$: E trifft niemals zu, $h(E) = f(E) = 0$.
- $E = \Omega$: E trifft immer zu, $h(E) = n, f(E) = 1$.

Rechengesetze für Häufigkeiten

Additionsgesetz

$$E \cap F = \emptyset \implies \begin{cases} h(E \cup F) = h(E) + h(F) \\ f(E \cup F) = f(E) + f(F) \end{cases}$$

Siebformel

$$h(E \cup F) = h(E) + h(F) - h(E \cap F)$$
$$f(E \cup F) = f(E) + f(F) - f(E \cap F)$$

Beispiel

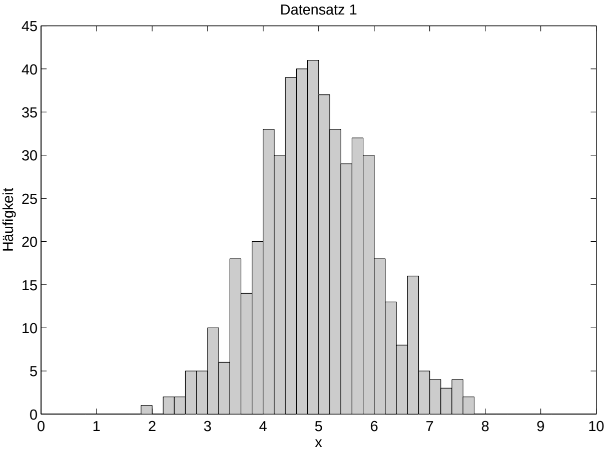
33% der Kunden einer Bank haben einen Wohnungskredit, 24% haben einen Kredit zur Finanzierung von Konsumgütern, 11% haben beides. Wie groß ist der Anteil der Kunden, die weder Wohnungs- noch Konsumgüterkredit haben?

2.2.1 Graphische Darstellung

2.2.1 Graphische Darstellung

- Ein Bild sagt mehr als tausend Worte!
- Graphische Darstellungen von Datensätzen sind daher äußerst beliebt und nützlich.
- Qualitative Variable: Häufigkeitstabelle, Tortendiagramm, Stabdiagramm
- Quantitative Variable: gruppierte Häufigkeitstabelle, Histogramm, Boxplot, empirische Verteilungsfunktion

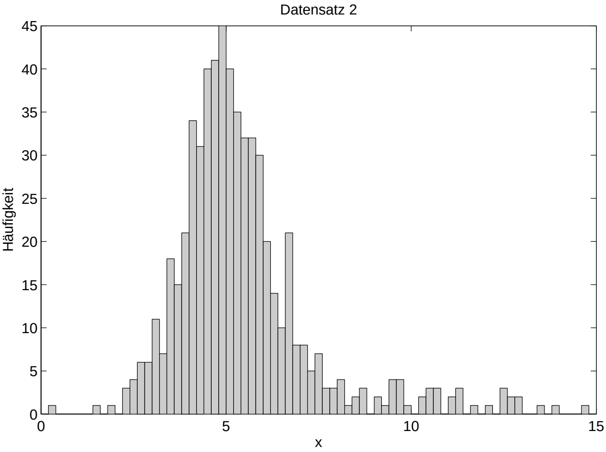
- Datensatz 1 (500 normalverteilte Werte):



Histogramm

2.2.1 Graphische Darstellung

- Datensatz 2 = Datensatz 1 + Kontamination (100 Werte):



Histogramm

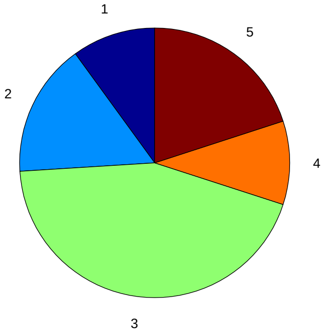
- Datensatz 3 (50 Püfungsnoten):

Note k	$h(k)$	$f(k)$
1	5	0.10
2	8	0.16
3	22	0.44
4	5	0.10
5	10	0.20
	50	1.00

Häufigkeitstabelle

- MATLAB: `make_dataset3`

- Datensatz 3 (50 Püfungsnoten):

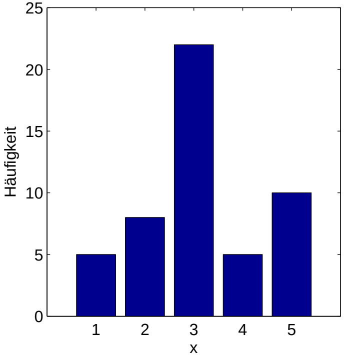


Tortendiagramm

- MATLAB: `make_dataset3`

2.2.1 Graphische Darstellung

- Datensatz 3 (50 Püfungsnoten):



Stabdiagramm

- MATLAB: `make_dataset3`

2.2.1 Empirische Verteilungsfunktion

- Ab Ordinalskala ist es sinnvoll, die Daten zu ordnen.
- Die Häufigkeitstabelle kann durch Summenhäufigkeiten ergänzt werden.
- Datensatz 3 (50 Prüfungsnoten):

Note k	$h(k)$	$H(k)$	$f(k)$	$F(k)$
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

Häufigkeitstabelle mit Summenhäufigkeiten

- MATLAB: `make_dataset3`

2.2.1 Empirische Verteilungsfunktion

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

Definition (Empirische Verteilungsfunktion)

Die empirische Verteilungsfunktion $F_n(x)$ der Datenliste $\vec{x} = (x_1, \ldots, x_n)$ ist der Anteil der Daten, die kleiner oder gleich x sind:

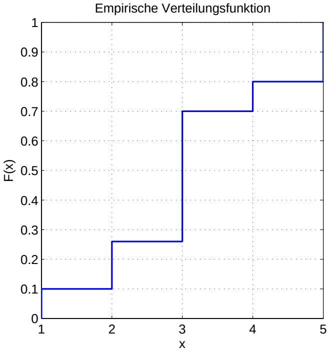
$$F_n(x) = f(\vec{x} \leq x).$$

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

$$F_n(x) = f(x_1) + \cdots + f(x_i).$$

- F_n ist eine Sprungfunktion. Die Sprungstellen sind die Datenpunkte, die Sprunghöhen sind die relativen Häufigkeiten der Datenpunkte.

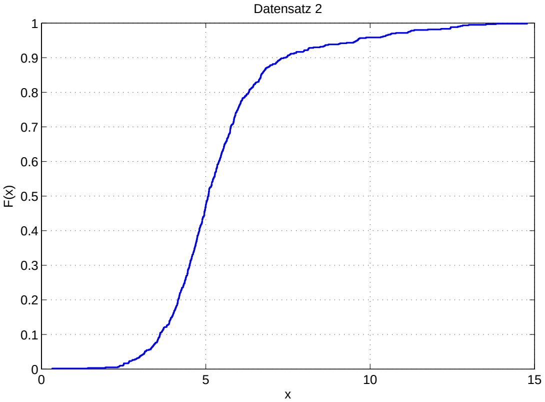
- Datensatz 3: (50 Prüfungsnoten):



Empirische Verteilungsfunktion

- MATLAB: `make_dataset3`

- Datensatz 2 (500 Werte + Kontamination):



Empirische Verteilungsfunktion

- MATLAB: `make_dataset2`

2.2.1 Kernschätzer

- Die Häufigkeitsverteilung (Histogramm) kann mit einem Kern- oder Dichteschätzer geglättet werden.
- Die Dichte des beobachteten Merkmals wird dabei durch eine Summe von Kernen $K(\cdot)$ approximiert:

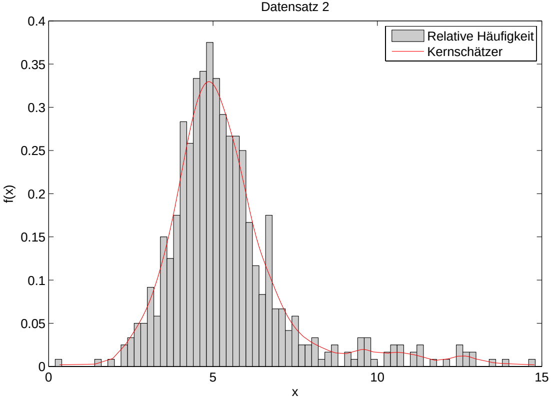
$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

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

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

Statistik
R. Frühwirth
2.1 Einleitung
2.2 Eindimensionale Merkmale
2.2.1 Graphische Darstellung
2.2.1 Empirische Verteilungsfunktion
2.2.1 Kernschätzer
2.2.1 Maßzahlen
2.2.1 Beispiele
2.3 Zweidimensionale Merkmale

- Datensatz 2:



Glättung des Histogramms durch Kernschätzer

- MATLAB: `make_dataset2`

2.2.1 Maßzahlen

2.2.1 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 ab.
- Manche Maßzahlen gehen von der geordneten Datenliste $x_{(1)}, \ldots, x_{(n)}$ aus.
- Wir unterscheiden 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.

Lagemaße

Definition (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:

- $\ell(a\boldsymbol{x} + b) = a\ell(\boldsymbol{x}) + b$ für $a > 0$
- $\min \boldsymbol{x} \leq \ell(\boldsymbol{x}) \leq \max(\boldsymbol{x})$

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

Mittelwert

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

- Sinnvoll für Intervall- und Verhältnisskala.
- Der Mittelwert minimiert die folgende Funktion:

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

- MATLAB: `xbar=mean(x)`

2.2.1 Maßzahlen

Median

$$\tilde{x} = x_{(n/2)}$$

- Der Median teilt die geordnete Liste in zwei gleich große Teile.
- Sinnvoll für Ordinal-, Intervall- und Verhältnisskala.
- Der Median minimiert die folgende Funktion:

$$\tilde{x} = \arg_x \min \sum_{i=1}^{n} |x_i - x|$$

- MATLAB: `xmed=median(x)`

2.2.1 Maßzahlen

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

α-Quantil

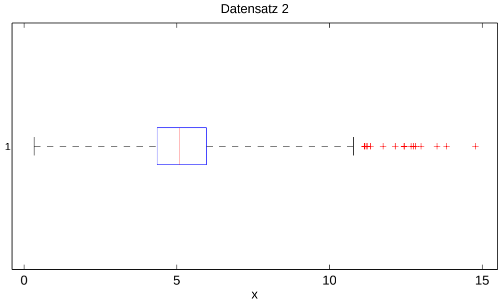
$$Q_\alpha = x_{(\alpha n)}$$

- Das α-Quantil teilt die geordnete Liste im Verhältnis $\alpha : 1 - \alpha$.
- Sinnvoll für Ordinal-, Intervall- und Verhältnisskala.
- MATLAB: `qa=quantile(x,alpha)`
- Q_0 ist der kleinste Wert, Q_1 ist der größte Wert der Datenliste.
- $Q_{0.5}$ 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])`

2.2.1 Maßzahlen

- Der **Boxplot** ist die graphische Darstellung des five point summary.
- Datensatz 2 (500 Werte + Kontamination):

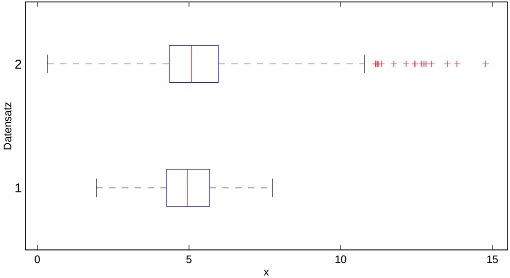


Boxplot

- MATLAB: `make_dataset2`

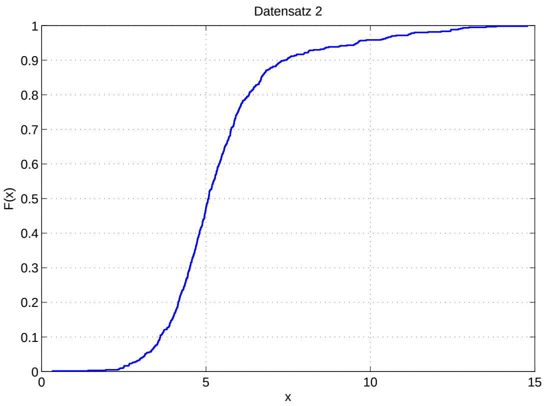
2.2.1 Maßzahlen

- Vergleich von Datensatz 1 und Datensatz 2 mittels Boxplot:



Boxplot

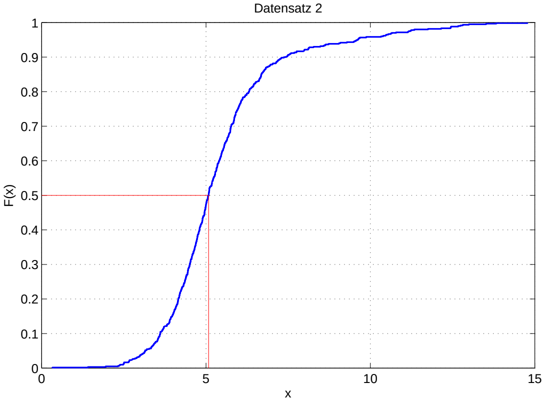
- Aus der empirischen Verteilungsfunktion können Quantile einfach abgelesen werden.
- Median von Datensatz 2:



Empirische Verteilungsfunktion

2.2.1 Maßzahlen

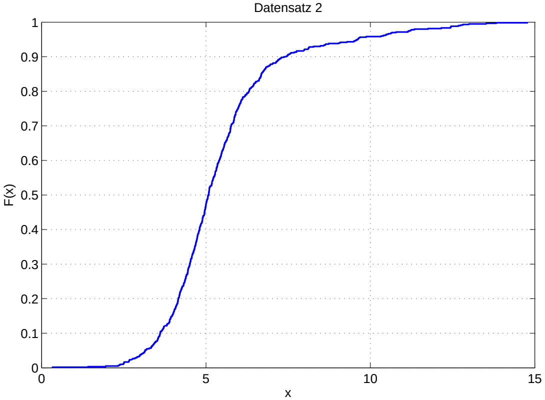
- Aus der empirischen Verteilungsfunktion können Quantile einfach abgelesen werden.
- Median von Datensatz 2:



Empirische Verteilungsfunktion

2.2.1 Maßzahlen

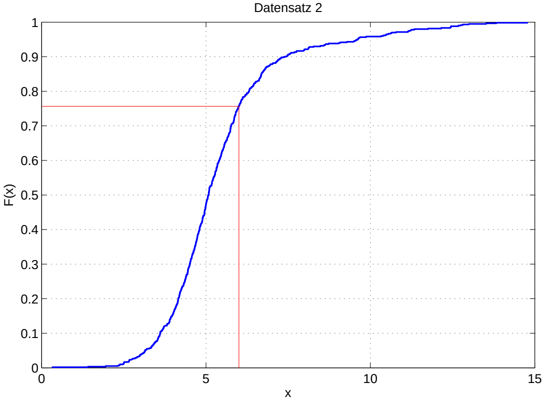
- Es können auch Unter- und Überschreitungshäufigkeiten abgelesen werden.
- Welcher Anteil der Daten ist kleiner oder gleich 6?



Empirische Verteilungsfunktion

2.2.1 Maßzahlen

- Es können auch Unter- und Überschreitungshäufigkeiten abgelesen werden.
- Welcher Anteil der Daten ist kleiner oder gleich 6?



Empirische Verteilungsfunktion

LMS (Least Median of Squares)

Der LMS-Wert ist der Mittelpunkt des kürzesten Intervalls, das $h = \lfloor n/2 \rfloor + 1$ Datenpunkte enthält.

- Der LMS-Wert ist extrem unempfindlich gegen fehlerhafte oder untypische Daten.
- Der LMS-Wert minimiert die folgende Funktion:

$$\tilde{x} = \arg_x \min \mathsf{med}_{i=1}^{n} (x_i - x)^2$$

- Ein verwandtes Lagemaß ist der “shorth”, der Mittelwert aller Daten im kürzesten Intervall, das h Datenpunkte enthält.

- MATLAB: `xlms=lms(x)`
- MATLAB: `xshorth=shorth(x)`

Modus

Der Modus ist der häufigste Wert einer Datenliste

- Sinnvoll vor allem für qualitative Merkmale.
- Für quantitative Merkmale kann der Modus aus dem Kernschätzer der Dichte bestimmt werden.
- MATLAB: `xmode=mode(x)`

HSM (Half-sample mode)

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 Mittel der beiden letzten Daten.

- MATLAB: `xhsm=hsm(x)`

Streuungsmaße

Definition (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:

- $\sigma(\boldsymbol{x}) \geq 0$
- $\sigma(a\boldsymbol{x} + b) = |a|\,\sigma(\boldsymbol{x})$

- Sinnvolle Streuungsmaße messen die Abweichung der Daten von ihrem zentralen Wert.
- Streuungsmaße sind invariant unter Verschiebung der Daten.
- Je nach Skala sind verschiedene Streuungsmaße sinnvoll.

Standardabweichung

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

- Sinnvoll für Intervall- und Verhältnisskala.
- Die Standardabweichung hat die gleiche Dimension wie die Daten.
- Das Quadrat der Standardabweichung heißt **Varianz**.

- MATLAB: `xstd=std(x,1)`
- MATLAB: `xvar=var(x,1)`

Interquartilsdistanz

$$IQR = Q_{0.75} - Q_{0.25}$$

- Die Interquartilsdistanz ist die Länge des Intervalls, das die zentralen 50% der Daten enthält.
- Sinnvoll für Ordinal-, Intervall- und Verhältnisskala.
- MATLAB: `xiqr=iqr(x)`

LoS (Length of the Shorth)

LoS ist die Länge des kürzesten Intervalls, das $h = \lfloor n/2 \rfloor + 1$ Datenpunkte enthält.

- Sinnvoll für Ordinal-, Intervall- und Verhältnisskala.

- MATLAB: `xlos=LoS(x)`

Schiefemaße

Definition (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:

- $\mathfrak{s}(a\boldsymbol{x} + b) = \text{sgn}(a)\,\mathfrak{s}(\boldsymbol{x})$
- $\mathfrak{s}(\boldsymbol{x}) = 0$, wenn $\exists b : \boldsymbol{x} - b = b - \boldsymbol{x}$

- Sinnvolle Schiefemaße messen die Asymmetrie der Daten.
- Schiefemaße sind invariant unter Verschiebung der Daten.
- Je nach Skala sind verschiedene Schiefemaße sinnvoll.

Schiefe

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

- Die Schiefe γ ist gleich 0 für symmetrische Daten.
- Ist $\gamma < 0$, heißen die Daten **linksschief**.
- Ist $\gamma > 0$, heißen die Daten **rechtsschief**.
- Sinnvoll für Intervall- und Verhältnisskala.

- MATLAB: `xgamma=skewness(x,1)`

Schiefekoeffizient

$$SK = \frac{R - L}{R + L}$$

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**.
- Sinnvoll für Ordinal-, Intervall- und Verhältnisskala.

- MATLAB: `xsk=SK(x)`

2.2.1 Beispiele

- Datensatz 1: Symmetrisch, 500 Werte

- Lagemaße:

Mittelwert:	4.9532
Median:	4.9518
LMS:	4.8080
Shorth:	4.8002
HSM:	5.0830

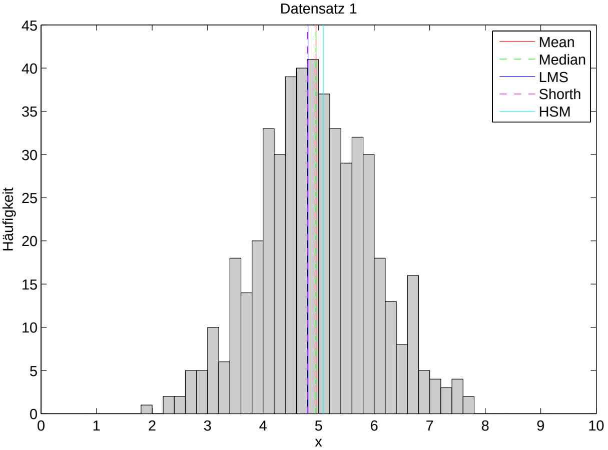
- Streuungsmaße:

Standardabweichung:	1.0255
Interquartilsdistanz:	1.4168
Length of the Shorth:	1.3520

- Schiefemaße:

Schiefe:	0.0375
Schiefekoeffizient:	0.0258

2.2.1 Beispiele



Datensatz 1: Mittelwert, Median, LMS, Shorth, HSM

2.2.1 Beispiele

- Datensatz 2: Datensatz 1 + Kontamination (100 Werte)

- Lagemaße:

Mittelwert:	5.4343
Median:	5.0777
LMS:	5.1100
Shorth:	5.0740
HSM:	4.9985

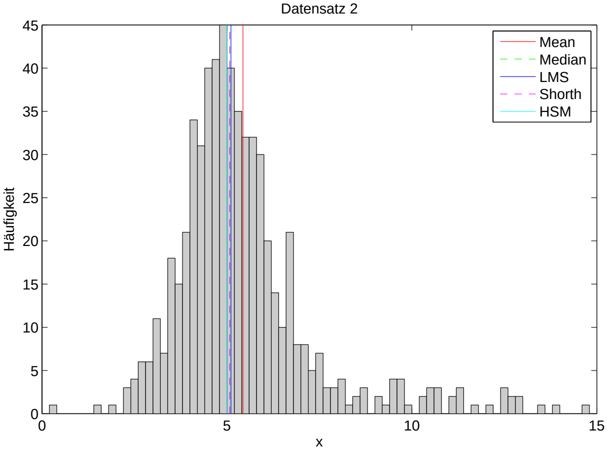
- Streuungsmaße:

Standardabweichung:	1.8959
Interquartilsdistanz:	1.6152
Length of the Shorth:	1.5918

- Schiefemaße:

Schiefe:	1.7696
Schiefekoeffizient:	0.1046

2.2.1 Beispiele



Datensatz 2: Mittelwert, Median, LMS, Shorth, HSM

2.2.1 Beispiele

- Datensatz 3: 50 Prüfungsnoten

- Lagemaße:

Mittelwert:	3.14
Median:	3
Modus:	3

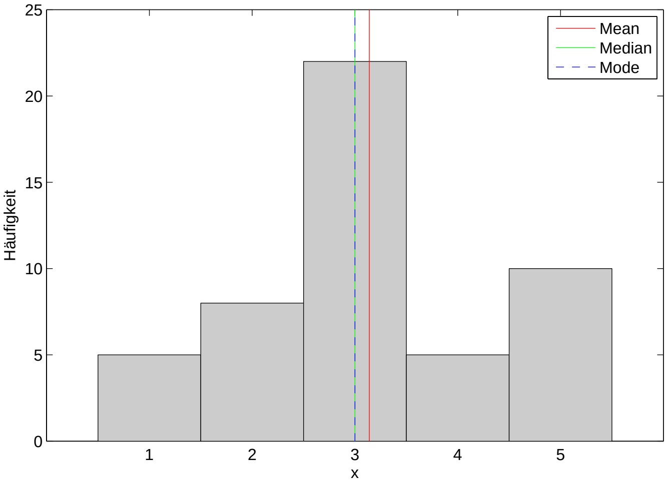
- Streuungsmaße:

Standardabweichung:	1.2
Interquartilsdistanz:	2

- Schiefemaße:

Schiefe:	0.0765
Schiefekoeffizient:	0

2.2.1 Beispiele



Datensatz 3: Mittelwert, Median, Modus

2.3 Zweidimensionale Merkmale

- Oft werden zwei oder mehr Merkmale eines Objekts **gleichzeitig** beobachtet.
- Beispiele:
  - Körpergröße und Gewicht einer Person
  - Alter und Einkommen einer Person
  - Schulbildung und Geschlecht einer Person
- Der Zusammenhang zwischen den beiden Merkmalen gibt zusätzliche Information.

- Wir betrachten hier nur den Fall von zwei binären Merkmalen A und B.
- Die Häufigkeit des Eintretens von A und B kann in einer **Vierfeldertafel** oder **Kontingenztafel** zusammengefasst werden.
- **Beispiel**:

  A=*"Die Person ist weiblich"*
  B=*"Die Person ist Raucher/in"*

- Vierfeldertafel für 1000 Personen:

	B	B'	
A	228	372	600
A'	136	264	400
	364	636	1000

- Allgemeiner Aufbau einer Vierfeldertafel:

	B	B'	
A	$h(A \cap B)$	$h(A \cap B')$	$h(A)$
A'	$h(A' \cap B)$	$h(A' \cap B')$	$h(A')$
	$h(B)$	$h(B')$	n

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

- Die Vierfeldertafel kann mittels Division durch n auf **relative Häufigkeiten** umgerechnet werden:

	B	B'	
A	$f(A \cap B)$	$f(A \cap B')$	$f(A)$
A'	$f(A' \cap B)$	$f(A' \cap B')$	$f(A')$
	$f(B)$	$f(B')$	1

- Zeilen- und Spaltensummen sind die relativen Häufigkeiten der Ausprägungen A, A' und B, B'.

- Der Zusammenhang der beiden Merkmale kann durch die **Vierfelderkorrelation** gemessen werden:

Vierfelderkorrelation

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

- Es gilt stets: $-1 \le \rho(A,B) \le 1$
- Ist $\rho(A,B) > 0$, heißen A und B **positiv gekoppelt**.
- Ist $\rho(A,B) < 0$, heißen A und B **negativ gekoppelt**.

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

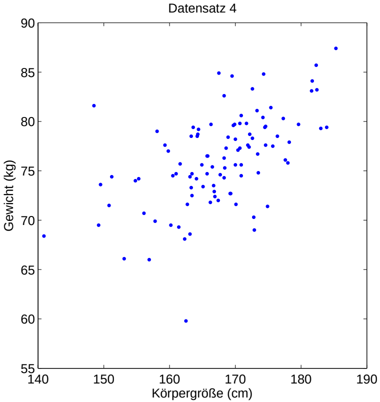
$$A = B \implies \rho(A, B) = 1$$
$$A = B' \implies \rho(A, B) = -1$$

- Eine bestehende Koppelung ist kein Beweis für einen kausalen Zusammenhang!
- Die Koppelung kann auch durch eine gemeinsame Ursache für beide Merkmale entstehen.

2.3.1 Quantitative Merkmale

- Bevorzugte Darstellung von zweidimensionalen Merkmalen: Streudiagramm (Scatter Plot)
- Jeder Punkt entspricht einem Objekt.
- Die beobachteten Merkmale bestimmen die Position des Punktes in der x-y-Ebene.
- Mehrdimensionale Merkmale können durch Histogramme und Streudiagramme dargestellt werden. Dabei geht natürlich ein Teil der Information verloren.

- Datensatz 4: Körpergröße und Gewicht von 100 Personen



Streudiagramm

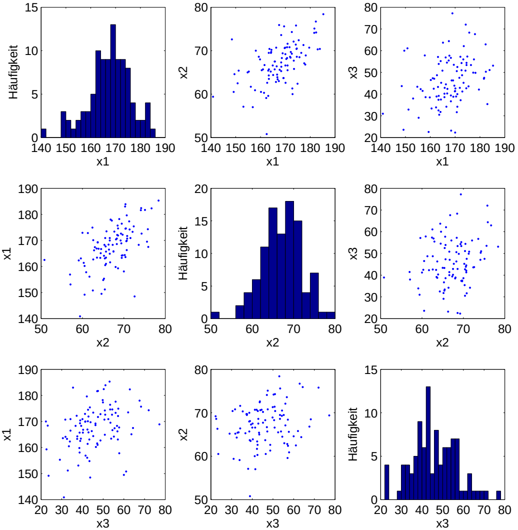
- MATLAB: `make_dataset4`

- Datensatz 5:
  Körpergröße, Gewicht und Alter von 100 Personen

Merkmal x_1:	Körpergröße (in cm)
Merkmal x_2:	Gewicht (in kg)
Merkmal x_3:	Alter (in Jahren)

- MATLAB: `make_dataset5`

2.3.1 Quantitative Merkmale



2.3.1 Empirische Korrelation

Eigenschaften des Streudiagramms

1. $(\bar{x}, \bar{y})$ ist der Mittelpunkt der Punktwolke.
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 $y_1, \ldots, y_n$.

- Aus dem Streudiagramm von Datensatz 4 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.

2.3.1 Empirische Korrelation

- Wir brauchen eine **Maßzahl** für diesen Zusammenhang.
- Eine nützliche Maßzahl ist der **empirische Korrelationskoeffizient**.
- Sei $(x_1, y_1), \ldots, (x_n, y_n)$ eine bivariate Stichprobe.
- Wir berechnen die **Standardscores**:

$$z_{x,i} = \frac{x_i - \bar{x}}{s_x}, \quad z_{y,i} = \frac{y_i - \bar{y}}{s_y}$$

- Wir erinnern uns, dass

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{und} \quad s_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- Der empirische Korrelationskoeffizient ist der **Mittelwert der Produkte** der Standardscores.

2.3.1 Empirische Korrelation

Definition (Empirischer Korrelationskoeffizient)

Der **empirische Korrelationskoeffizient** r_{xy} ist definiert als

$$r_{xy} = \frac{1}{n} \sum_{i=1}^{n} z_{x,i} z_{y,i} = \frac{1}{n} \left(z_{x,1} z_{y,1} + \cdots + z_{x,n} z_{y,n} \right)$$

- Es gilt immer:

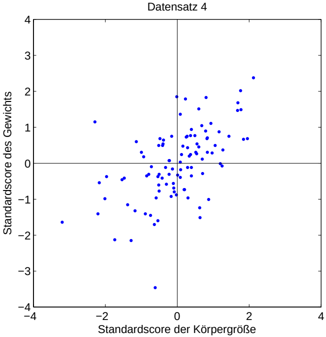
$$-1 \leq r_{xy} \leq 1$$

2.3.1 Empirische Korrelation

- 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 Standscores verschiedene 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**.

2.3.1 Empirische Korrelation

- Streudiagramm der Standardscores von Datensatz 4:



- Offensichtlich sind x und y positiv korreliert, da die meisten Punkte im 1. und 3. Quadranten liegen.
- $r_{xy} = 0.5562$

2.3.1 Empirische Korrelation

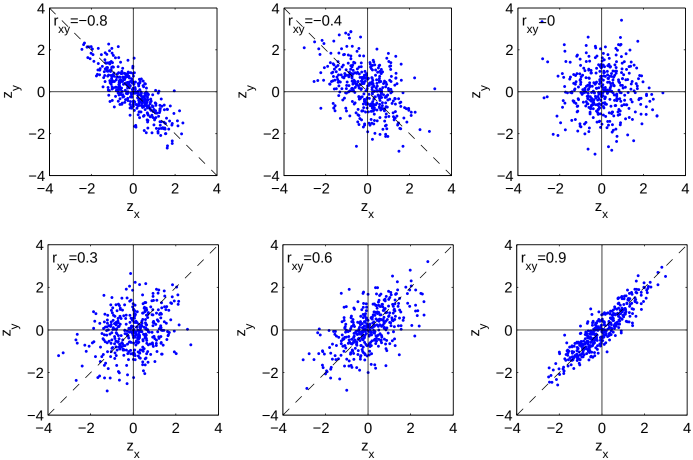
- Eine positive Korrelation muss nicht unbedingt einen kausalen Zusammenhang bedeuten.
- Die positive Korrelation kann auch durch eine gemeinsame Ursache oder einen parallel laufenden Trend verursacht sein.

Beispiel

Zwischen der Kinderzahl und der Anzahl der Störche in Österreich in den letzten 30 Jahren besteht eine positive Korrelation. Warum?

Beispiel

Zwischen dem Butterpreis und dem Brotpreis der letzten 20 Jahre besteht eine positive Korrelation. Warum?

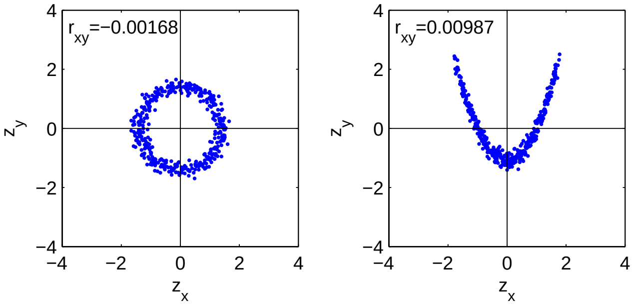


Standardscores mit verschiedenen Korrelationskoeffizienten

2.3.1 Empirische Korrelation

- 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** Koppelung an.
- Besteht zwischen x und y ein starker, aber **nichtlinearer** Zusammenhang, kann die Korrelation trotzdem sehr klein sein.

2.3.1 Empirische Korrelation



Nichtlinearer Zusammenhang zwischen x und y

Teil 3

Wahrscheinlichkeiten

Teil 3: Wahrscheinlichkeiten

3.1 Einleitung

3.1 Einleitung

- Der konkrete Ausgang eines Experiments kann im Allgemeinen nicht genau vorausgesagt werden.
- Die möglichen Ausgänge sind jedoch bekannt.
- Ziel der Wahrscheinlichkeitsrechnung ist es, den Ereignissen im Kontext des Experiments Wahrscheinlichkeiten zuzuweisen.
- Zwei Interpretationen der Wahrscheinlichkeit sind möglich.

3.1 Einleitung

Häufigkeitsinterpretation

- Die Wahrscheinlichkeit eines Ereignisses ist die relative Häufigkeit, wenn das Experiment sehr oft unter den gleichen Bedingungen wiederholt wird.
- Die darauf basierende Statistik wird „frequentistisch“ genannt.

Beispiel

Die Wahrscheinlichkeit des Ereignisses $e_1 = \{1\}$ beim Würfeln ist der Grenzwert der Häufigkeit für eine große Zahl von Würfen.

3.1 Einleitung

Subjektive Interpretation

- Die Wahrscheinlichkeit eines Ereignisses ist eine Aussage über den Glauben der Person, die die Wahrscheinlichkeit angibt.
- Die darauf basierende Statistik wird „bayesianisch“ genannt.

Beispiel

„Die Wahrscheinlichkeit, dass es morgen regnet, ist 40 Prozent“ ist eine Aussage über den Glauben der Person, die diese Aussage tätigt.

3.1 Einleitung

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

3.2.1 Wahrscheinlichkeitsmaße

3.2.1 Wahrscheinlichkeitsmaße

Definition (Wahrscheinlichkeitsmaß)

Es sei Σ eine Ereignisalgebra, A und B Ereignisse in Σ, und W eine Abbildung von Σ in $\mathbb{R}$. W heißt ein **Wahrscheinlichkeitsmaß**, wenn gilt:

1. Positivität: $W(A) \geq 0, \ \forall A \in \Sigma$
2. Additivität: $A \cap B = \mathbf{0} \implies W(A \cup B) = W(A) + W(B)$
3. Normierung: $W(\mathbf{1}) = 1$

W wird auch als **Wahrscheinlichkeitsverteilung** bezeichnet.

Definition (Wahrscheinlichkeitsraum)

Ist Σ eine Ereignisalgebra und W ein Wahrscheinlichkeitsmaß, so heißt Σ normiert, und das Paar (Σ, W) ein **Wahrscheinlichkeitsraum**.

3.2.1 Wahrscheinlichkeitsmaße

Rechengesetze für Wahrscheinlichkeit

Ist (Σ, W) ein Wahrscheinlichkeitsraum, so gilt:

1. $W(\mathbf{0}) = 0$
2. $W(A) \leq 1, \ \forall A \in \Sigma$
3. $W(A') = 1 - W(A), \ \forall A \in \Sigma$
4. $A \subseteq B \implies W(A) \leq W(B), \ \forall A, B \in \Sigma$
5. $W(A \cup B) = W(A) + W(B) - W(A \cap B), \ \forall A, B \in \Sigma$
6. Hat Σ höchstens abzählbar viele Elementarereignisse $\{e_i \mid i \in I\}$, so ist $\sum_{i \in I} W(e_i) = 1$.

3.2.1 Wahrscheinlichkeitsmaße

- In einer diskreten Ereignisalgebra ist die Wahrscheinlichkeit eines Ereignisses gleich der Summe der Wahrscheinlichkeiten der Elementarereignisse, deren $\cup$-Verbindung es ist.
- 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.
- Man kann also auf einer diskreten Ereignisalgebra Σ unendlich viele Verteilungen definieren.

3.2.1 Wahrscheinlichkeitsmaße

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

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

- Die Wahrscheinlichkeit eines Ereignisses A wird durch **Integration** über die Dichte ermittelt:

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

- Die Dichte muss so beschaffen sein, dass das Integral für alle zugelassenen Ereignisse existiert.

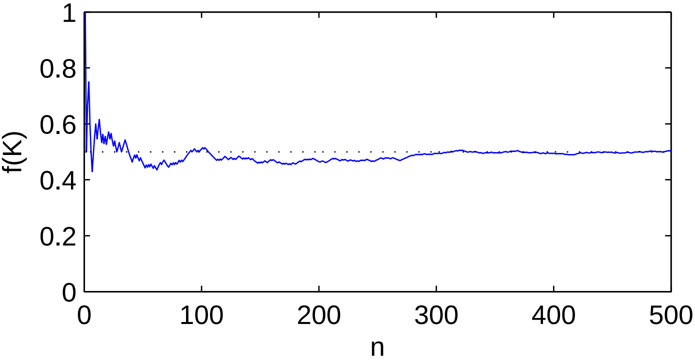
- Betrachten einfaches Zufallsexperiment: Münzwurf
- Zwei mögliche Ergebnisse: Kopf (K), Zahl (Z)
- Annahme: Münze symmetrisch, K und Z gleichwahrscheinlich
- Experiment wird n-mal wiederholt

n	$h_n(K)$	$f_n(K)$	$\lvert f_n(K) - 0.5\rvert$
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

Häufigkeitstabelle

- MATLAB: `make_coin`

3.2.1 Gesetz der großen Zahlen



Entwicklung der relativen Häufigkeit von K

- Die relative Häufigkeit des Ereignisses K scheint gegen den Grenzwert 0.5 zu streben.
- Dieser Grenzwert wird als die **Wahrscheinlichkeit** $W(K)$ bezeichnet.

Empirisches Gesetz der großen Zahlen

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

- 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 sein muss, sondern nur in einem weiteren, statistischen Sinn.

3.2.1 Kombinatorik

- Häufig ist es auf Grund von Symmetrieüberlegungen möglich, die Elementarereignisse als gleichwahrscheinlich anzusehen.
- Dies ist natürlich nur sinnvoll für endlich viele Elementarereignisse.
- Diese Annahme entspricht nur in seltenen Fällen der physikalischen Realität und muss im Zweifelsfall durch das Experiment überprüft werden.
- Sind alle m Elementarereignisse gleichwahrscheinlich, gilt:

$$W(e_1) = W(e_2) = \ldots = W(e_m) = \frac{1}{m}$$

- Für ein Ereignis A, das sich aus g Elementarereignissen zusammensetzt, gilt:

Regel von Laplace

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

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

3.2.1 Kombinatorik

Definition (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.

- Es gibt
$$V_k^n = \frac{n!}{(n-k)!} = n \cdot (n-1) \ldots (n-k+1)$$
solcher Variationen.
- Für den Sonderfall $k = n$ sieht man, dass sich die n Elemente der Menge M auf
$$n! = \prod_{i=1}^{n} i$$
verschiedene Weisen (Permutationen) anordnen lassen.

3.2.1 Kombinatorik

Definition (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.

- Es gibt $C_k^n = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ solcher Kombinationen.
- C_k^n wird auch als **Binomialkoeffizient** bezeichnet.
- Die Binomialkoeffizienten können im sogenannten Pascal'schen Dreieck angeordnedt werden:

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

3.2.1 Kombinatorik

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

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

Beispiel

Beim Roulette sind die Zahlen von 0 bis 36 als Ergebnis möglich.

1. Wie groß ist die Wahrscheinlichkeit, dass sich in einer Serie von zehn Würfen keine Zahl wiederholt?
2. Wie groß ist die Wahrscheinlichkeit, dass in einer Serie von 37 Würfen jede Zahl vorkommt?

3.3.1 Kopplung und bedingte Wahrscheinlichkeit

- Wir betrachten jetzt zwei Ereignisse A und B, die bei einem Experiment eintreten können.
- **Frage**: Besteht ein Zusammenhang zwischen den Ereignissen?
- Ein solcher Zusammenhang wird **Koppelung** genannt.
- **Positive Koppelung**: Je öfter A eintritt, desto öfter tritt tendenziell auch B ein.
- **Negative Koppelung**: Je öfter A eintritt, desto seltener tritt tendenziell auch B ein.
- Quantifizierung von „oft“ und „selten“ erfolgt durch Häufigkeitstabelle.

- Die Häufigkeit des Eintretens von A und B kann in einer **Vierfeldertafel** oder **Kontingenztafel** zusammengefaßt werden.
- **Beispiel**:

  A=*"Eine untersuchte Person ist weiblich"*
  B=*"Eine untersuchte Person hat Diabetes"*

- Vierfeldertafel für 1000 Personen:

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

- **Gewöhnliche relative Häufigkeiten** werden auf den Umfang n des gesamten Datensatzes bezogen:

$$f(A \cap B) = \frac{h(A \cap B)}{n}$$

- **Bedingte relative Häufigkeiten** werden auf das Eintreten des anderen Merkmals bezogen:

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

- $f(A|B)$ heißt die bedingte relative Häufigkeit von A unter der Bedingung B.

3.3.1 Kopplung und bedingte Wahrscheinlichkeit

- Die Vierfeldertafel U gibt folgende bedingte relative Häufigkeiten:

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

- Es ist somit zu vermuten, dass die beiden Merkmale gekoppelt sind.
- $f(A|B) > f(A)$ deutet auf eine positive Koppelung, $f(A|B) < f(A)$ auf eine negative Koppelung.

- Stammen die Daten aus einem Zufallsexperiment, dann besitzen die Ereigniskombinationen auch Wahrscheinlichkeiten.
- Wahrscheinlichkeitstabelle:

	B	B'	
A	$W(A \cap B)$	$W(A \cap B')$	$W(A)$
A'	$W(A' \cap B)$	$W(A' \cap B')$	$W(A')$
	$W(B)$	$W(B')$	1

- Nach dem empirischen Gesetz der großen Zahl sind diese Wahrscheinlichkeiten die Grenzwerte der entsprechenden relativen Häufigkeiten.

3.3.1 Kopplung und bedingte Wahrscheinlichkeit

- Die bedingten relativen Häufigkeiten konvergieren für $n \to \infty$ gegen einen Grenzwert:

$$f_n(A|B) = \frac{f_n(A \cap B)}{f_n(B)} \to W(A|B) = \frac{W(A \cap B)}{W(B)}$$

Definition (Bedingte Wahrscheinlichkeit)

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

heißt die **bedingte Wahrscheinlichkeit von A unter der Bedingung B**, sofern $W(B) \neq 0$.

Beispiel (Der symmetrische Würfel)

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

$$W(e_i) = \frac{1}{6},\ 1 \le i \le 6$$

Wir definieren die folgenden Ereignisse:

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

Dann gilt zum Beispiel

$$W(e_1|U) = \frac{W(e_1 \cap U)}{W(U)} = \frac{W(e_1)}{W(U)} = \frac{1}{3}$$

$$W(e_1|G) = \frac{W(e_1 \cap G)}{W(G)} = \frac{W(\mathbf{0})}{W(G)} = 0$$

Beispiel (Fortsetzung)

$$W(U|e_1) = \frac{W(e_1 \cap U)}{W(e_1)} = \frac{W(e_1)}{W(e_1)} = 1$$

$$W(e_1 \cup e_3|U) = \frac{W((e_1 \cup e_3) \cap U)}{W(U)} = \frac{W(e_1 \cup e_3)}{W(U)} = \frac{2}{3}$$

$$W(e_1 \cup e_2|U) = \frac{W((e_1 \cup e_2) \cap U)}{W(U)} = \frac{W(e_1)}{W(U)} = \frac{1}{3}$$

3.3.1 Kopplung und bedingte Wahrscheinlichkeit

- Aus der Definition der bedingten Wahrscheinlichkeit folgt sofort die

Produktformel

$$W(A \cap B) = W(A|B)W(B) = W(B|A)W(A)$$

und die Formel für die

Inverse Wahrscheinlichkeit

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

- Beide Formeln gelten auch für relative Häufigkeiten!

3.3.1 Satz von Bayes

Definition (Zerlegung)

Die Ereignisse $B_1, B_2, \ldots, B_m$ bilden eine **Zerlegung** der Ergebnismenge Ω, wenn gilt:

1. Unvereinbarkeit: $B_i \cap B_j = \emptyset, i \neq j$
2. Vollständigkeit: $B_1 \cup B_2 \cup \ldots \cup B_m = \Omega$

Satz

Bilden die Ereignisse $B_1, B_2, \ldots, B_m$ eine Zerlegung der Ergebnismenge Ω, dann gilt:

$$W(B_1) + W(B_2) + \ldots + W(B_m) = W(\Omega) = \mathbf{1}$$

- Es sei $B_1, \ldots, B_m$ eine Zerlegung. Dann gilt:

Totale Wahrscheinlichkeit

$$W(A) = W(A|B_1)W(B_1) + \ldots + W(A|B_m)W(B_m)$$

Beispiel

Ein Betrieb erzeugt Halogenleuchten mit 100W (35% der Produktion), 200W (45%) und 300W (20%). Nach einem Jahr sind noch 98% der 100W-Leuchten funktionsfähig, 96% der 200W-Leuchten, und 93% der 300W-Leuchten. Welcher Anteil an allen Leuchten ist nach einem Jahr noch funktionsfähig?

3.3.1 Satz von Bayes

- Es sei $B_1, \ldots, B_m$ eine Zerlegung. Dann gilt:

Satz von Bayes

$$\begin{aligned} W(B_i|A) &= \frac{W(A|B_i)W(B_i)}{W(A)} \\ &= \frac{W(A|B_i)W(B_i)}{W(A|B_1)W(B_1) + \ldots + W(A|B_m)W(B_m)} \end{aligned}$$

- $W(B_i)$ wird die **a-priori** Wahrscheinlichkeit von B genannt, $W(B_i|A)$ die **a-posteriori** Wahrscheinlichkeit.

Beispiel

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

1. Wie groß ist der totale Ausschussanteil?
2. Wie groß ist die Wahrscheinlichkeit, dass ein einwandfreier Bauteil von Anbieter 2 kommt?
3. Wie groß ist die Wahrscheinlichkeit, dass ein mangelhafter Bauteil von Anbieter 1 kommt?

3.3.1 Satz von Bayes

Beispiel

Ein Bauteil wird von vier Firmen geliefert, und zwar kommen 20% von Firma 1, 30% von Firma 2, 35% von Firma 3, und 15% von Firma 4. Die Wahrscheinlichkeit, dass der Bauteil im Testbetreib innerhalb von 24 Stunden ausfällt, ist 0.02 für Firma 1, 0.015 für Firma 2, 0.025 für Firma 3, und 0.02 für Firma 4. Ein Bauteil fällt im Testbetrieb nach 16 Stunden aus. Die Wahrscheinlichkeit, dass er von Firma i kommt, ist mittel des Satzes von Bayes zu berechnen.

3.3.1 Unabhängigkeit

- Zwei Ereignisse sind **positiv gekoppelt**, wenn

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

- Zwei Ereignisse sind **negativ gekoppelt**, wenn

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

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

Definition (Unabhängigkeit)

Zwei Ereignisse A und B heißen **stochastisch unabhängig**, wenn

$$W(A \cap B) = W(A)W(B)$$

Die Ereignisse $A_1, A_2, \ldots, A_n$ heißen unabhängig, wenn gilt:

$$W(A_1 \cap \ldots \cap A_n) = W(A_1) \cdot \ldots \cdot W(A_n)$$

Dazu genügt nicht, dass je zwei Ereignisse A_i und A_j paarweise unabhängig sind!

Beispiel

Wir betrachten den zweimaligen Wurf einer Münze (Kopf/Zahl). Die möglichen Ausgänge sind $\Omega = \{KK, KZ, ZK, ZZ\}$. Ferner definieren wir die Ereignisse:

$$E_1 = \{KK, KZ\} \ldots \text{Kopf beim ersten Wurf}$$
$$E_2 = \{KK, ZK\} \ldots \text{Kopf beim zweiten Wurf}$$
$$E_3 = \{KK, ZZ\} \ldots \text{Gerade Zahl von Köpfen}$$

Dann gilt für alle $i \neq j$

$$W(E_i \cap E_j) = \frac{1}{4} = W(E_i) \cdot W(E_j)$$

aber

$$W(E_1 \cap E_2 \cap E_3) = \frac{1}{4} \neq \frac{1}{8} = W(E_1) \cdot W(E_2) \cdot W(E_3)$$

3.3.1 Unabhängigkeit

- Sind A und B unabhängig, gilt $W(A|B) = W(A)$ und $W(B|A) = W(B)$.
- Die Vierfeldertafel für zwei unabhängige Ereignisse:

	B	B'	
A	$W(A)W(B)$	$W(A)W(B')$	$W(A)$
A'	$W(A')W(B)$	$W(A')W(B')$	$W(A')$
	$W(B)$	$W(B')$	1

3.3.1 Unabhängigkeit

- Die Koppelung kann durch die **Vierfelderkorrelation** gemessen werden:

Vierfelderkorrelation

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

Eigenschaften der Vierfelderkorrelation

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

3.3.1 Unabhängigkeit

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

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

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

3.3.1 Unabhängigkeit

- 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, da ihre $\cap$-Verbindung stets das unmögliche Ereignis ist.
- Zwei Elementarereignisse sind sogar höchst „abhängig“, weil das Eintreten des einen das Eintreten des anderen mit Sicherheit ausschließt.
- Sind E_1 und E_2 zwei unabhängige Ereignisse eines Wahrscheinlichkeitsraumes (Σ, W), so sind auch E_1 und E_2', E_1' und E_2, sowie E_1' und E_2' unabhängig.

Beispiel (Wiederholung eines Alternativversuchs)

Die Ereignisalgebra hat 2^n Elementarereignisse, nämlich die Folgen der Form $(i_1, \ldots, i_n)$, $i_j = 0$ oder 1. Sind die Wiederholungen unabhängig, und bezeichnet p die Wahrscheinlichkeit des Eintretens von 1, ist die Wahrscheinlichkeit einer Folge

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

wo k bzw. $n-k$ die Anzahl des Eintretens von 1 bzw. 0 angibt.

3.3.1 Unabhängigkeit

Die Unabhängigkeit von zwei Ereignissen A und B kann anhand einer Vierfeldertafel empirisch überprüft (getestet) werden. Es gilt die folgende Regel:

Test der Unabhängigkeit

Hat eine Vierfeldertafel für A und B n Einträge und ist die Vierfelderkorrelation gleich ρ, so wird die Annahme (Hypothese) der Unabhängigkeit von A und B abgelehnt, wenn

$$\left|\sqrt{n}\cdot\rho\right| > 2$$

- Die Wahrscheinlichkeit einer Fehlentscheidung ist etwa 5%.
- Mehr über das Testen von Hypothesen findet sich in Teil 7.

Teil 4

Diskrete Verteilungen

4.1 Grundbegriffe

4.1.1 Diskrete Verteilungen

- Diskrete Zufallsvariable sind entweder die beobachteten Ausprägungen eines **kategorialen** Merkmals oder das Resultat von **Zählvorgängen**.
- In der physikalischen Praxis kommen beide Fälle häufig vor.

Beispiel

Die meisten instabilen Teilchen können auf viele 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

Wird die Anzahl der Zerfälle einer radioaktiven Quelle pro Zeiteinheit gezählt, ist dadurch eine diskrete Zufallsvariable definiert.

4.1.1 Diskrete Verteilungen

- Im folgenden nehmen wir an, dass die Werte der diskreten Zufallsvariablen X nichtnegative ganze Zahlen $k \in \mathbb{N}_0$ sind.
- Dies ist keine Einschränkung, weil jede abzählbare Menge von reellen Zahlen bijektiv auf (eine Teilmenge von) $\mathbb{N}_0$ abgebildet werden kann.
- Jede Teilmenge von $\mathbb{N}_0$ ist ein Ereignis.
- Die Elementarereignisse sind $\{0\}, \{1\}, \{2\}, \ldots$

4.1.1 Diskrete Verteilungen

- Die Wahrscheinlichkeiten $W_X(\{k\})$ der Elementarereignisse können als Werte einer Funktion f_X angesehen werden:

$$f_X(x) = \begin{cases} W_X(\{k\}), & \text{wenn } x = k \\ 0, & \text{sonst} \end{cases}$$

Definition (Diskrete Dichtefunktion)

Die Funktion $f_X(k)$ wird als **Wahrscheinlichkeitsfunktion** oder auch Dichte(funktion) der Zufallsvariablen X bezeichnet.

4.1.1 Diskrete Verteilungen

- Die Wahrscheinlichkeit $W_X(E)$ eines Ereignisses E lässt sich dann mit Hilfe der Dichte f_X von X berechnen:

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

Definition (Diskrete Verteilungsfunktion)

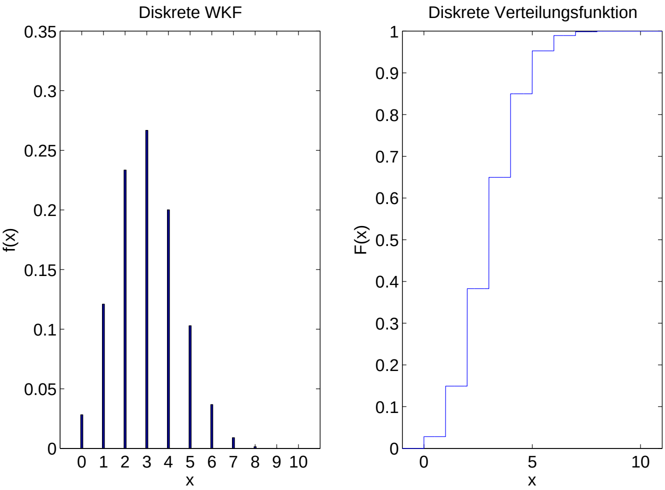
Ist X eine diskrete Zufallsvariable, so ist die **Verteilungsfunktion** F_X von X definiert durch:

$$F_X(x) = W(X \leq x)$$

Es gilt offenbar:

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

4.1.1 Diskrete Verteilungen



Eigenschaften einer diskreten Verteilungsfunktion F

1. F hat eine Sprungstelle in allen Punkten des Wertebereichs
2. Die Sprunghöhe im Punkt k ist gleich $f_X(k)$
3. $0 \le F(x) \le 1, \; \forall x \in \mathbb{R}$
4. $x \le y \implies F(x) \le F(y), \; \forall x, y \in \mathbb{R}$
5. $\lim_{x \to -\infty} F(x) = 0; \; \lim_{x \to \infty} F(x) = 1$
6. Die Wahrscheinlichkeit, dass X in das Intervall $(a, b]$ fällt, ist gleich $F(b) - F(a)$:

$$W(r \le a) + W(a < r \le b) = W(r \le b) \implies$$
$$W(a < r \le b) = F(b) - F(a)$$

4.1.1 Erwartungswert und Varianz

Definition (Erwartung)

Es sei X eine diskrete Zufallsvariable mit der Dichte $f_X(x)$. Ferner sei $g : \mathbb{N}_0 \to \mathbb{R}$ eine beliebige Funktion. Man definiert $\mathsf{E}_X[g] = \mathsf{E}[g(X)]$ durch:

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

$\mathsf{E}_X[g] = \mathsf{E}[g(X)]$ heißt die **Erwartung** von $g(X)$.

Definition (Erwartung einer Zufallsvariablen)

Ist $g(k) = k$, so heißt $\mathsf{E}[g(X)] = \mathsf{E}[X]$ die **Erwartung** oder der **Mittelwert** von X.

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

4.1.1 Erwartungswert und Varianz

- Der Mittelwert ist der **Schwerpunkt** der durch die Dichte definierten Massenverteilung.
- Der Mittelwert kann als **Lageparameter** betrachtet werden.
- Der Mittelwert ist ein **linearer Operator**:

$$\mathsf{E}[aX + b] = a\,\mathsf{E}[X] + b$$
$$\mathsf{E}[X_1 + X_2] = \mathsf{E}[X_1] + \mathsf{E}[X_2]$$

Beispiel

Ist X das Ergebnis des Wurfs mit einem symmetrischen Würfel, so gilt:

$$\mathsf{E}[X] = \sum_{k=1}^{6} k \cdot \frac{1}{6} = 3.5$$

4.1.1 Erwartungswert und Varianz

Definition (Varianz)

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

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

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.
- Die Standardabweichung hat die gleiche Dimension wie X.
- Es gilt:

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

- Varianz und Standardabweichung sind **invariant** gegen Translationen.

4.1.1 Erwartungswert und Varianz

Beispiel

Ist X das Ergebnis des Wurfs mit einem symmetrischen Würfel, so gilt:

$$\mathsf{var}[X] = \sum_{k=1}^{6} (k-3.5)^2 \cdot \frac{1}{6} = \frac{35}{12} = 2.9167$$

$$\sigma[X] = \sqrt{\mathsf{var}[X]} = 1.7078$$

Definition (Diskreter Zufallsvektor)

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

Beispiel

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

Beispiel

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

4.1.1 Zufallsvektoren

Definition (Dichtefunktion 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 Dichte(funktion) des Zufallsvektors $\boldsymbol{X}$ bezeichnet:

$$\begin{aligned} f_{\boldsymbol{X}}(k_1, \ldots, k_n) &= W_{\boldsymbol{X}}(k_1, \ldots, k_n) \\ &= W(X_1 = k_1 \cap \ldots \cap X_n = k_n) \end{aligned}$$

Definition (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 Dichte von X_i wird als **Randdichte** bezeichnet.

Berechnung der Randverteilung

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein diskreter Zufallsvektor. Die Randverteilung der Komponente X_i wird durch Summation 'über alle übrigen Komponenten berechnet:

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

Definition (Unabhängigkeit)

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

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

4.1.1 Zufallsvektoren

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

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 (Kovarianz)

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

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

- X_1 und X_2 heißen **unkorreliert**, wenn ihre Kovarianz gleich 0 ist.
- Unabhängige Zufallsvariable sind stets unkorreliert, jedoch brauchen unkorrelierte Zufallsvariable nicht unabhängig zu sein.

Definition (Kovarianzmatrix)

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

$$\mathbf{C}_{ij} = \mathsf{E}[(X_i - \mathsf{E}[X_i])(X_j - \mathsf{E}[X_j])]$$

heißt die Varianz-Kovarianz- oder kurz **Kovarianzmatrix** von $\boldsymbol{X}$.

- 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 (Korrelationsmatrix)

Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein Zufallsvektor. 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 symmetrisch und positiv (semi-)definit. Sie enthält in der Diagonale 1 und neben der Diagonale die paarweisen Korrelationskoeffizienten ρ_{ij} aller Komponenten von $\boldsymbol{X}$.

Definition (Bedingte Dichte)

Es sei $\boldsymbol{X} = (X_1, X_2)$ eine 2-dimensionale diskrete Zufallsvariable mit der Dichte $f(k_1, k_2)$ und den Randverteilungsdichten $f_1(k_1)$ bzw. $f_2(k_2)$. Die Funktion $f(k_1|k_2)$, definiert durch:

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

heißt die durch X_2 **bedingte Dichte** von X_1.

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

Beispiel

$f(k_1, k_2)$	$k_2 = 1$	$k_2 = 2$	$k_2 = 3$	$k_2 = 4$	$f_1(k_1)$
$k_1 = 1$	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“.

4.1.1 Zufallsvektoren

Beispiel (Fortsetzung)

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:

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

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

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

4.2.1 Grundbegriffe

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

4.2.1 Ziehen mit Zurücklegen

- Wir gehen von einer Grundgesamtheit mit beliebiger Größe aus.
- Der Anteil der Objekte mit einer gewissen Eigenschaft E (Merkmalsträger) ist gleich p.
- Es werden n Objekte zufällig gezogen, wobei ein gezogenes Objekte **wieder zurückgelegt** wird.
- Bei jeder Ziehung hat jedes Objekt die gleiche Wahrscheinlichkeit, gezogen zu werden.
- Die Anzahl der gezogenen Objekte mit der Eigenschaft E ist eine diskrete Zufallsvariable X.
- Die Verteilung von X wird **Binomialverteilung** $\mathrm{Bi}(n, p)$ genannt.

4.2.1 Ziehen mit Zurücklegen

Dichte der Binomialverteilung

Die Dichte der Binomialverteilung $\mathrm{Bi}(n,p)$ lautet:

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

Erwartung und Varianz

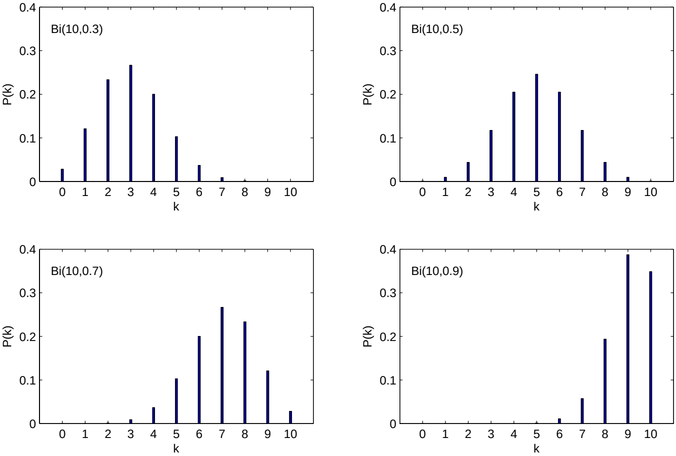
Es sei $X \sim \mathrm{Bi}(n,p)$. Dann gilt

$$\mathsf{E}[X] = np$$

und

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

4.2.1 Ziehen mit Zurücklegen



Statistik

R. Frühwirth

4.1 Grundbegriffe

4.2 Ziehungsexperimente

4.2.1 Grundbegriffe

4.2.1 Ziehen mit Zurücklegen

4.2.1 Ziehen ohne Zurücklegen

4.3 Zählexperimente

Beispiel

Sie ziehen aus einer Lieferung von Bauteilen eine Stichprobe vom Umfang $n = 50$ (mit Zurücklegen). Sie wissen aus Erfahrung, dass 1.5% der Bauteile fehlerhaft sind. Berechnen Sie die Wahrscheinlichkeit, in der Stichprobe höchstens einen fehlerhaften Bauteil zu finden.

Lösung:

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

4.2.1 Ziehen mit Zurücklegen

- Ist der Anteil der Merkmalsträger p unbekannt, kann eine Beobachtung von X dazu verwendet werden, um p zu **schätzen**.
- Sei m der beobachtete Wert von X. Wir schätzen dann p durch

$$\tilde{p} = \frac{m}{n}$$

- Die Erwartung von $\tilde{p}$ ist dann gleich

$$\mathsf{E}[\tilde{p}] = \frac{\mathsf{E}[X]}{n} = p$$

- Der Schätzer $\tilde{p}$ ist also **unverzerrt**.

4.2.1 Ziehen mit Zurücklegen

- Die Varianz von $\tilde{p}$ ist gleich

$$\mathsf{var}[\tilde{p}] = \frac{\mathsf{var}[X]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

- Die Varianz ist also **verkehrt proportional** zur Stichprobengröße.
- Da der wahre Wert von p nicht bekannt ist, wird er durch den Schätzwert ersetzt:

$$\mathsf{var}[\tilde{p}] \approx \frac{\tilde{p}(1-\tilde{p})}{n}$$

- Mehr über Schätzung in Kapitel 5!

4.2.1 Ziehen ohne Zurücklegen

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

Dichte der hypergeometrischen Verteilung

Die Dichte der hypergeometrischen Verteilung lautet:

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

Beispiel

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

Lösung:

$$W(X > 2) = 1 - W(X \le 2) = 1 - (f(0) + f(1) + f(2)) =$$

$$= 1 - \frac{\binom{20}{0}\binom{980}{50}}{\binom{1000}{50}} - \frac{\binom{20}{1}\binom{980}{49}}{\binom{1000}{50}} - \frac{\binom{20}{2}\binom{980}{48}}{\binom{1000}{50}} =$$

$$= 0.0736$$

Beispiel

Beim Lottospiel werden $n = 6$ Zahlen ohne Zurücklegen aus einer Grundgesamtheit von $N = 45$ Zahlen gezogen. Es gibt $M = 6$ Merkmalsträger, nämlich die Zahlen, auf die Sie gesetzt haben. Wie groß ist die Wahrscheinlichkeit eines Viterers, d.h. $X = 4$?

Lösung:

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

4.2.1 Ziehen ohne Zurücklegen

Erwartung und Varianz

Es sei $X \sim \mathrm{Hy}(N, M, n)$. Dann gilt

$$\mathsf{E}[X] = \frac{nM}{N}$$

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

Schreibt man $p = M/N$ für den Anteil der Merkmalsträger in der Grundgesamtheit, wird die Beziehung zur Binomialverteilung sichtbar:

$$\mathsf{E}[X] = np$$

$$\mathsf{var}[X] = np(1-p)\frac{N-n}{N-1}$$

Der Faktor $\dfrac{N-n}{N-1}$ wird **Endlichkeitskorrektur** genannt.

4.2.1 Ziehen ohne Zurücklegen

- Ist die Anzahl der Merkmalsträger M unbekannt, kann eine Beobachtung von X dazu verwendet werden, um M zu **schätzen**.
- Sei m der beobachtete Wert von X. Wir schätzen dann M durch

$$\tilde{M} = N \cdot \frac{m}{n}$$

- Die Erwartung von $\tilde{M}$ ist dann gleich

$$\mathsf{E}[\tilde{M}] = N \cdot \frac{\mathsf{E}[X]}{n} = M$$

- Der Schätzer $\tilde{M}$ ist also **unverzerrt**.

4.2.1 Ziehen ohne Zurücklegen

- Die Varianz von $\tilde{M}$ ist gleich

$$\mathsf{var}[\tilde{M}] = N^2 \cdot \frac{\mathsf{var}[X]}{n^2} = \frac{M(N-n)(N-M)}{n(N-1)}$$

- Die Varianz ist also **verkehrt proportional** zur Stichprobengröße.
- Da der wahre Wert von M nicht bekannt ist, wird er durch den Schätzwert ersetzt:

$$\mathsf{var}[\tilde{M}] \approx N^2 \cdot \frac{\mathsf{var}[X]}{n^2} = \frac{\tilde{M}(N-n)(N-\tilde{M})}{n(N-1)}$$

Beispiel

Sie ziehen aus einer Lieferung von $N = 1000$ Widerständen ohne Zurücklegen eine Stichprobe vom Umfang $n = 50$. Sie finden $m = 4$ defekte Stücke. Schätzen sie M und berechnen Sie die Standardabweichung der Schätzung.

Lösung:

$$\begin{aligned}\tilde{M} &= \frac{1000 \cdot 4}{50} = 80 \\ \mathsf{var}[\tilde{M}] &\approx \frac{80 \cdot 950 \cdot 920}{50 \cdot 999} = 1399.80 \\ \sigma[\tilde{M}] &\approx 37.4\end{aligned}$$

Eine Faustregel besagt, dass der wahre Wert von M mit einer Sicherheit von ca. 95% zwischen $\tilde{M} - 2\sigma[\tilde{M}]$ und $\tilde{M} + 2\sigma[\tilde{M}]$ liegt.

4.2.1 Ziehen ohne Zurücklegen

- Ist 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. Wir schätzen dann N durch

$$\tilde{N} = M \cdot \frac{n}{m}$$

- Die Erwartung von $\tilde{N}$ ist dann **ungefähr** gleich

$$\mathsf{E}[\tilde{N}] \approx M \cdot \frac{n}{\mathsf{E}[X]} = N$$

- Der Schätzer $\tilde{M}$ ist also nicht exakt erwartungstreu.

4.2.1 Ziehen ohne Zurücklegen

- Die Varianz von $\tilde{N}$ ist **ungefähr** gleich

$$\mathsf{var}[\tilde{N}] \approx \frac{N^2(N-n)(N-M)}{nM(N-1)}$$

- Die Varianz ist also wieder **verkehrt proportional** zur Stichprobengröße.
- Da der wahre Wert von N nicht bekannt ist, wird er durch den Schätzwert ersetzt:

$$\mathsf{var}[\tilde{N}] \approx \frac{\tilde{N}^2(\tilde{N}-n)(\tilde{N}-M)}{nM(\tilde{N}-1)}$$

Beispiel

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. Schätzen sie N und berechnen Sie die Standardabweichung der Schätzung.

Lösung:

$$\tilde{N} = \frac{200 \cdot 100}{21} \approx 952$$

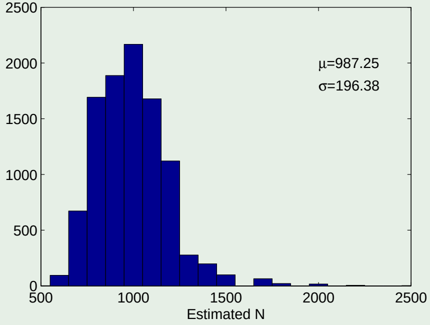
$$\text{var}[\tilde{N}] \approx \frac{952^2 \cdot 852 \cdot 752}{100 \cdot 200 \cdot 951} \approx 30530$$

$$\sigma[\tilde{N}] \approx 175$$

4.2.1 Ziehen ohne Zurücklegen

Beispiel

Wir können die Qualität der Schätzung durch ein Simulationsexperiment nachprüfen. Die Abbildung zeigt die Häufigkeitsverteilung von $\tilde{N}$ für 10000 Experimente mit $N = 952, M = 200, n = 100$.



4.3.1 Der Alternativversuch

- Ein Versuch mit nur zwei möglichen Ausgängen heißt ein **Alternativversuch** oder **Bernoulliexperiment**.
- Die beiden Ausgänge werden oft als **Erfolg** bzw. **Misserfolg** bezeichnet.
- Es sei die Wahrscheinlichkeit eines Erfolgs gleich p.
- Wird der Versuch n-mal unabhängig wiederholt, und bezeichnet X die **Anzahl der Erfolge in n Versuchen**, dann ist X **binomialverteilt** gemäß $\mathrm{Bi}(n, p)$.

4.3.1 Die Poissonverteilung

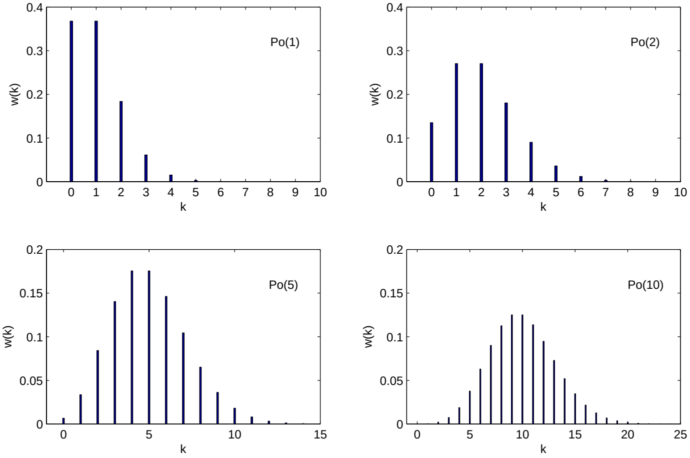
- Die **Poissonverteilung** $\mathrm{Po}(\lambda)$ entsteht aus der Binomialverteilung durch den Grenzübergang $n \to \infty$ unter der Bedingung $n \cdot p = \lambda$.
- Das klassische Beispiel einer Poisson-verteilten Zufallsvariablen ist die Anzahl X der Zerfälle pro Zeiteinheit in einer radioaktiven Quelle.
- Allgemein gilt: Ist die Wartezeit zwischen zwei Ereignissen eines Zufallsprozesses exponentialverteilt, so ist die Anzahl der Ereignisse pro Zeiteinheit Poisson-verteilt .

4.3.1 Die Poissonverteilung

- Die Dichte der Poissonverteilung folgt aus der Berechnung des Grenzwertes:

$$
\begin{aligned}
f(k) &= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \\
&= \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} = \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left[\prod_{i=1}^{k} \frac{\left(1-\frac{i-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)}\right] \cdot \left(1-\frac{\lambda}{n}\right)^n = \\
&= \frac{\lambda^k}{k!} \cdot e^{-\lambda}
\end{aligned}
$$

4.3.1 Die Poissonverteilung



Erwartung und Varianz

Es sei $X \sim \mathrm{Po}(\lambda)$. Dann gilt

$$\begin{aligned} \mathsf{E}[X] &= \lambda \\ \mathsf{var}[X] &= \lambda \end{aligned}$$

- Ist der Mittelwert λ unbekannt, können Beobachtungen von X dazu verwendet werden, um λ zu **schätzen**.
- Seien $k_1, \ldots, k_n$ die beobachtete Werte von X. Wir schätzen dann λ durch

$$\tilde{\lambda} = \frac{\sum_{i=1}^{n} k_i}{n}$$

- Die Erwartung von $\tilde{\lambda}$ ist dann gleich

$$\mathsf{E}[\tilde{\lambda}] = \frac{\sum_{i=1}^{n} \mathsf{E}[X_i]}{n} = \lambda$$

- Der Schätzer $\tilde{\lambda}$ ist also **unverzerrt**.

4.3.1 Die Poissonverteilung

- Die Varianz von $\tilde{\lambda}$ ist gleich

$$\mathsf{var}[\tilde{\lambda}] = \frac{\sum_{i=1}^{n} \mathsf{var}[X_i]}{n^2} = \frac{\lambda}{n}$$

- Die Varianz ist also **verkehrt proportional** zur Stichprobengröße.

4.3.1 Die Multinomialverteilung

- Der Alternativversuch kann dahingehend verallgemeinert werden, dass man nicht nur zwei, sondern d Ausgänge $e_1, \ldots, e_d$ zulässt, denen die Wahrscheinlichkeiten $p_1, \ldots, p_d$ zugeordnet werden, die nur

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

erfüllen müssen.

- Führt man den **verallgemeinerten Alternativversuch** n-mal durch, so sind die Elementarereignisse die Folgen der Form:

$$(e_{i_1}, \ldots, e_{i_n}),\ 1 \leq i_j \leq d$$

4.3.1 Die Multinomialverteilung

- Sind die n Teilversuche unabhängig, gilt:

$$W(e_{i_1},\ldots,e_{i_n}) = \prod_{j=1}^{n} W(e_{i_j}) = \prod_{j=1}^{n} p_{i_j} = \prod_{i=1}^{d} p_i^{n_i}$$

  wobei n_i die Anzahl des Eintretens von e_i ist. Die Summe der n_i ist daher n.

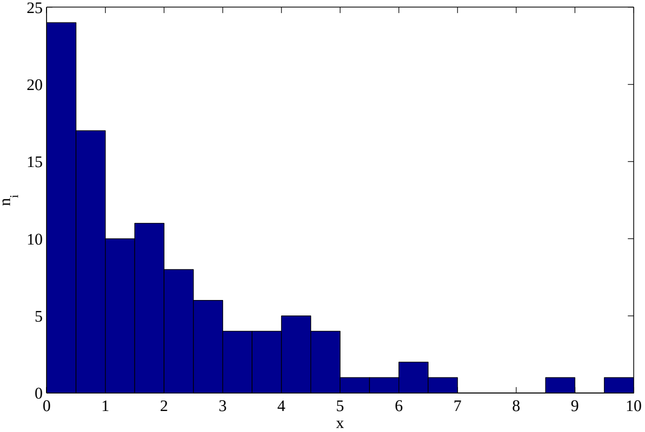
- Der d-dimensionale Zufallsvektor $\boldsymbol{X} = (X_1,\ldots,X_d)$ bildet die Folge $(e_{i_1},\ldots,e_{i_n})$ auf den Vektor $(n_1,\ldots,n_d)$ ab. Dabei werden $n!/(n_1!\cdots n_d!)$ Folgen auf den gleichen Vektor abgebildet.
- Die Dichte von $\boldsymbol{X}$ lautet daher:

$$f(n_1,\ldots,n_d) = \frac{n!}{n_1!\ldots n_d!}\prod_{i=1}^{d} p_i^{n_i}, \ \sum_{i=1}^{d} n_i = n, \ \sum_{i=1}^{d} p_i = 1$$

4.3.1 Die Multinomialverteilung

- Die Verteilung von $\boldsymbol{X}$ wird als **Multinomialverteilung** mit den Parametern n und $p_1, \ldots, p_d$ bezeichnet: $W_{\boldsymbol{X}} = \mathrm{Mu}(n, p_1, \ldots, p_d)$
- Das klassische Beispiel eines multinomialverteilten Zufallsvektors ist das Histogramm (gruppierte Häufigkeitsverteilung), das zur graphischen Darstellung der (absoluten) **experimentellen Häufigkeit** verwendet wird.
- 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 $(X_1, \ldots, X_d)$ multinomial nach $\mathrm{Mu}(n, p_1, \ldots, p_d)$ verteilt.

- Ein Histogramm



4.3.1 Die Multinomialverteilung

- Für jede Gruppe i ist die Komponente X_i **binomialverteilt** mit den Parametern n und p_i.
- Diese Verteilung ist also die **Randverteilung** von X bzgl. X_i.
- Die Komponenten X_i sind **nicht unabhängig** voneinander, da ihre Summe gleich n sein muss.
- Die Dichte von $\boldsymbol{X}$ ist daher **nicht** das Produkt ihrer Randdichten.
- Die Kovarianz von X_i und X_j ist gleich $\text{cov}[X_i, X_j] = -np_ip_j$.
- Die Kovarianzmatrix $\mathbf{C}$ von $\boldsymbol{X}$ hat daher die folgende Form:

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

Teil 5

Stetige Verteilungen

Teil 5: Stetige Verteilungen

5.1.1 Stetige Verteilungen

- Stetige Zufallsvariable sind meist das Resultat von **Messungen**.
- Der zufällige Charakter der Messungen kann mehrere Quellen haben:
  - Die fundamentale Zufälligkeit von quantenmechanischen Prozessen
  - Unkenntnis des genauen Anfangszustands eines Systems
  - Messfehler der Messapparatur
- Oft wirken mehrere Quellen zusammen.

Beispiel

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

5.1.1 Stetige Verteilungen

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

5.1.1 Stetige Verteilungen

- Es ist in diesem Fall nicht möglich, die Verteilung von X durch die Wahrscheinlichkeiten der Elementarereignisse zu spezifizieren.
- Die Verteilung muss durch die Angabe einer (stetigen) Funktion erfolgen. Drei Möglichkeiten sind im Gebrauch:
  - Dichtefunktion
  - Verteilungsfunktion (Stammfunktion der Dichtefunktion)
  - Charakteristische Funktion (Fouriertransformierte der Dichtefunktion)

Definition (Stetige Verteilungsfunktion)

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

$$F_X(x) = W(X \leq x)$$

heißt die **Verteilungsfunktion** von X. Die Wahrscheinlichkeit, dass X in ein Intervall $(x, x + \Delta x]$ fällt, ist dann:

$$W(x < X \leq x + \Delta x) = F_X(x + \Delta x) - F_X(x) = \Delta F_X.$$

5.1.1 Stetige Verteilungen

Eigenschaften einer stetigen Verteilungsfunktion

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

Definition (Quantil)

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

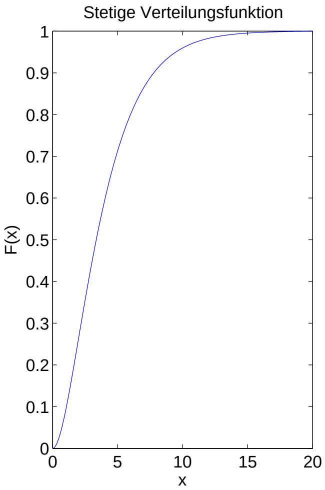
$$F_X(x_\alpha) = \alpha, \quad 0 < \alpha < 1$$

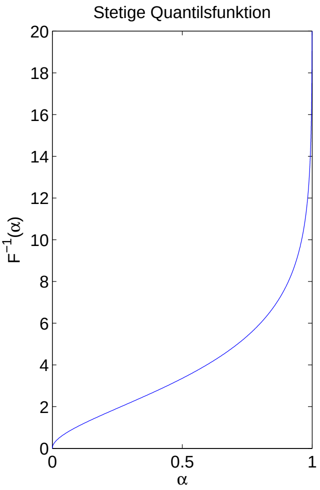
gilt, heißt das α-**Quantil** der Verteilung von X. Die Funktion

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

heißt die **Quantilsfunktion** der Verteilung von X.

5.1.1 Stetige Verteilungen





Definition (Quartil)

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

- Quantile können auch für diskrete Verteilungen definiert werden, jedoch sind sie dann meist nicht eindeutig.

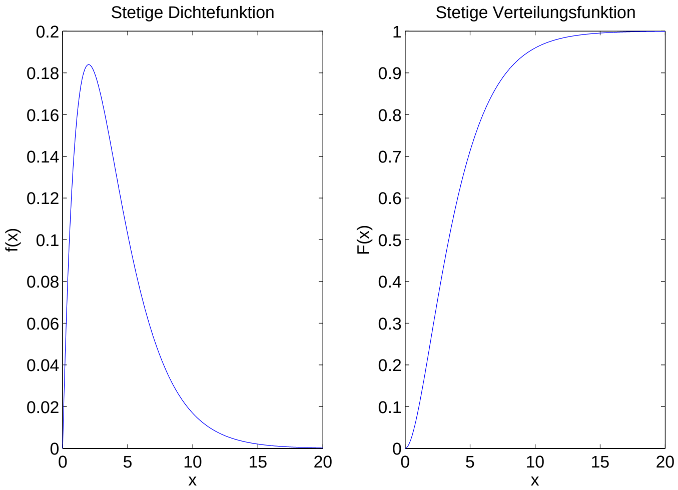
Definition (Stetige Dichtefunktion)

Ist F_X differenzierbar, gilt nach dem Hauptsatz der Integralrechnung:

$$W_X(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,\mathrm{d}x$$

wobei $f_X(x) = F'_X(x)$ ist. Die Ableitung der Verteilungsfunktion, die Funktion f_X, wird als **Wahrscheinlichkeitsdichtefunktion** oder wieder kurz Dichte von X bezeichnet.

5.1.1 Stetige Verteilungen



- Die Wahrscheinlichkeit W_X eines beliebigen Ereignisses $A \in \mathfrak{S}$ lässt sich leicht mit Hilfe der Dichte angeben:

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

- Die Wahrscheinlichkeit eines einzelnen Punktes ist immer gleich 0:

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

- Daher ist auch

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

Definition (Erwartung)

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

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

$\mathsf{E}_X[g] = \mathsf{E}[g(X)]$ heißt die **Erwartung** von $g(X)$.

Definition (Erwartung einer Zufallsvariablen)

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

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

- Die Erwartung existiert nicht immer. Ein Beispiel ist die Cauchy-Verteilung (T-Verteilung mit einem Freiheitsgrad) mit der Dichte
$$f(x) = \frac{1}{\pi(1+x^2)},\ x \in \mathbb{R}$$
- Verschwindet die Dichte außerhalb eines endlichen Intervalls, existiert die Erwartung immer.
- Der Mittelwert kann als **Lageparameter** verwendet werden.
- Der Mittelwert ist der **Schwerpunkt** der durch die Dichte definierten Massenverteilung.
- Der Mittelwert ist ein **linearer Operator**:
$$\mathsf{E}[aX+b] = a\,\mathsf{E}[X] + b$$
$$\mathsf{E}[X_1+X_2] = \mathsf{E}[X_1] + \mathsf{E}[X_2]$$
- Sind X_1 und X_2 **unabhängig**, so gilt
$$\mathsf{E}[X_1X_2] = \mathsf{E}[X_1]\cdot\mathsf{E}[X_2]$$

Definition (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.
- Die Standardabweichung hat die gleiche Dimension wie X.
- Es gilt:

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

- Varianz und Standardabweichung sind **invariant** gegen Translationen.

5.1.1 Erwartungswert und Varianz

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

Definition (Momente)

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

- Die Erwartung ist das erste Moment um 0, die Varianz ist das zweite zentrale Moment.
- Die zentralen Momente $\mu_1, \ldots, \mu_k$ können aus den Momenten um 0 $\mu'_1, \ldots, \mu'_k$ berechnet werden, und umgekehrt.

5.1.1 Zufallsvektoren

Definition (Stetiger Zufallsvektor)

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

Beispiel

Die gleichzeitige Messung von Größe und Gewicht einer Person wird durch einen 2-dimensionalen stetigen Zufallsvektor $\boldsymbol{X} = (X_1, X_2)$ beschrieben.

5.1.1 Zufallsvektoren

Definition (Dichtefunktion eines stetigen Zufallsvektors)

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

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

Definition (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 Dichte von X_i wird als **Randdichte** bezeichnet.

Berechnung der Randverteilung

Es sei $\boldsymbol{X} = (X_1, \dots, X_n)$ ein stetiger Zufallsvektor. Die Dichte der Randverteilung von $\boldsymbol{X}$ bzgl. X_i ist gegeben durch:

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

Definition (Unabhängigkeit)

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

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

5.1.1 Zufallsvektoren

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

Erwartung 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.

Definition (Bedingte Dichte)

Es sei $\boldsymbol{X} = (X_1, X_2)$ ein stetiger Zufallsvektor mit der Dichte $f(x_1, x_2)$ und den Randverteilungsdichten $f_1(x_1)$ bzw. $f_2(x_2)$. Die Funktion $f_{1|2}(x_1|x_2)$, definiert durch:

$$f_{1|2}(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}, \quad f_2(x_2) > 0$$

heißt die durch X_2 **bedingte Dichte** von X_1.

- Die bedingte Dichte ist für festes x_2 die Dichte eine Verteilung, der durch $X_2 = x_2$ bedingten Verteilung von X_1.

5.1.1 Zufallsvektoren

- Der Satz von der totalen Wahrscheinlichkeit und der Satz von Bayes kann auch für bedingte Dichten formuliert werden.

Satz von der totalen Wahrscheinlichkeit

$$f_2(x_2) = \int f_{2|1}(x_2|x_1) f_1(x_1)\,\mathrm{d}x_1$$

Satz von Bayes

$$f_{1|2}(x_1|x_2) = \frac{f_{2|1}(x_2|x_1) f_1(x_1)}{\int f_{2|1}(x_2|x_1) f_1(x_1)\,\mathrm{d}x_1}$$

5.2.1 Dichte und Momente

- Die **Gleichverteilung** $\mathrm{Un}(a, b)$ ordnet allen Werten im Intervall $[a, b]$ die gleiche Wahrscheinlichkeit zu.

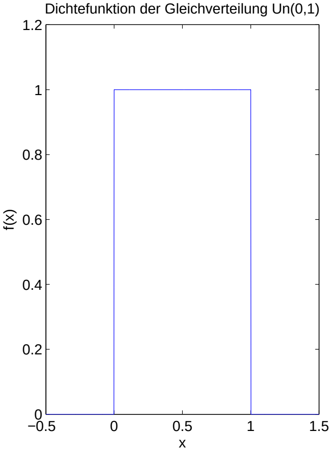
Dichte und Verteilungsfunktion der Gleichverteilung

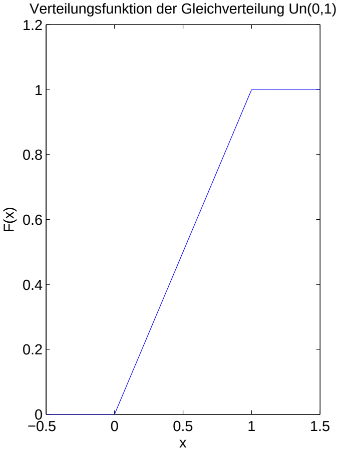
Die Dichte der Gleichverteilung lautet:

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

Die Verteilungsfunktion lautet:

$$F(x|a, b) = \min(1, \max(0, (x-a)/(b-a)))$$


Dichtefunktion der Gleichverteilung Un(0,1)
f(x)
x
1.2
1
0.8
0.6
0.4
0.2
0
−0.5
0
0.5
1
1.5


Verteilungsfunktion der Gleichverteilung Un(0,1)
F(x)
x
1.2
1
0.8
0.6
0.4
0.2
0
−0.5
0
0.5
1
1.5

Erwartung, Varianz und Median der Gleichverteilung

Es sei X gleichverteilt nach $\mathrm{Un}(a, b)$. Dann gilt:

$$\mathsf{E}[X] = \frac{a+b}{2}$$

und

$$\mathsf{var}[X] = \frac{(b-a)^2}{12}$$

Der Median der Verteilung ist gleich dem Mittelwert.

5.3.1 Dichte und Momente

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

Dichte und Verteilungsfunktion der Exponentialverteilung

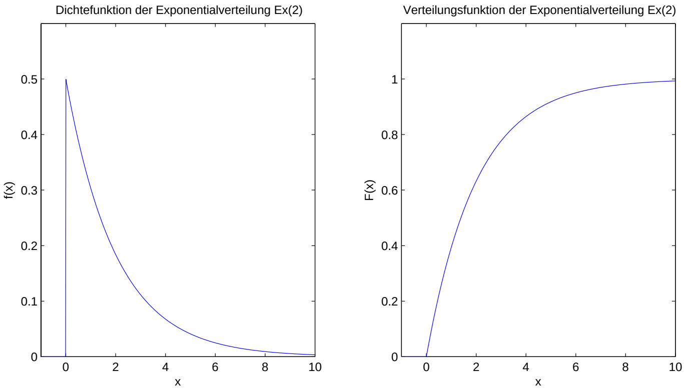
Die Dichte der Exponentialverteilung lautet:

$$f(x|\tau) = \frac{1}{\tau}\, e^{-x/\tau} \cdot I_{[0,\infty)}(x)$$

Die Verteilungsfunktion lautet:

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

5.3.1 Dichte und Momente



Erwartung, Varianz und Median der Exponentialverteilung

Es sei X exponentialverteilt nach $\mathrm{Ex}(\tau)$. Dann gilt:

$$\mathsf{E}[X] = \tau$$

und

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

Der Median der Verteilung ist gleich

$$\mathsf{med}[X] = \tau \ln 2$$

Er wird auch als **Halbwertszeit** bezeichnet.

5.3.1 Der Poissonprozess

- Wir beobachten einen Prozess, bei dem gewisse Ereignisse zu zufälligen Zeitpunkten eintreten.
- Ist die Anzahl der Ereignisse pro Zeiteinheit unabhängig und Poisson-verteilt gemäß $\mathrm{Po}(\lambda)$, sprechen wir von einem **Poissonprozess** mit Intensität λ.

Eigenschaften eines Poissonprozesses

1. Die Anzahl der Ereignisse in einem Zeitintervall der Länge t ist Poisson-verteilt 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 Intensität $\lambda = 1/\tau$.

5.4.1 Dichte und Verteilungsfunktion

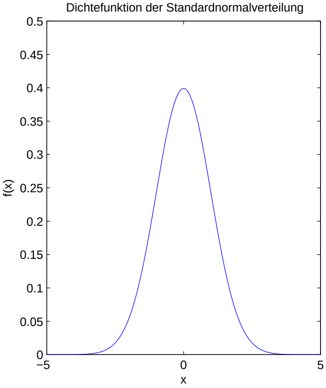
- Die Dichte der Normalverteilung hängt von zwei Parametern μ und σ^2 ab.
- Wir bezeichnen die Verteilung mit $\mathrm{No}(\mu, \sigma^2)$.

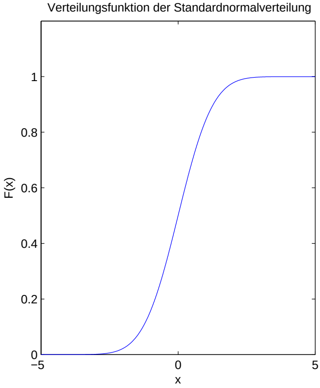
Dichte der Normalverteilung

Die Dichte ist die bekannte Glockenkurve:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

5.4.1 Dichte und Verteilungsfunktion





Erwartung und Varianz

Es sei X normalverteilt nach $\mathrm{No}(\mu, \sigma^2)$. Dann gilt:

$$\mathsf{E}[X] = \mu$$

und

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

- Der Median und das Maximum ist bei $x = \mu$, die Wendepunkte der Dichte bei $x = \mu \pm \sigma$.
- Die halbe Breite auf halber Höhe (HWHM) ist gleich $\sigma\sqrt{2\ln 2} \approx 1,177\sigma$.

5.4.1 Die Standardnormalverteilung

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

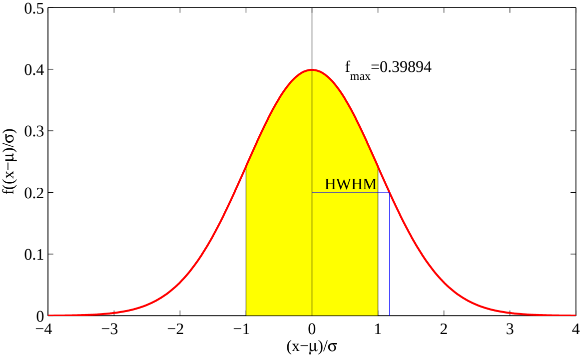
$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

- Ist X normalverteilt gemäß $\mathrm{No}(\mu, \sigma^2)$, so ist

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

  standardnormalverteilt.
- Z wird als das **Standardscore** von X bezeichnet.

5.4.1 Die Standardnormalverteilung



Eindimensionale Normalverteilung, gelber Bereich $= 68.2\%$

5.4.1 Die Standardnormalverteilung

- Ist X normalverteilt gemäß $\mathrm{No}(\mu, \sigma^2)$, gilt:

$$\begin{aligned} W(|X-\mu| \geq \sigma) &= 31.8\% \\ W(|X-\mu| \geq 2\sigma) &= 4.6\% \\ W(|X-\mu| \geq 3\sigma) &= 0.2\% \end{aligned}$$

- 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.

- Ihre Dichte lautet:

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

- $\mathbf{V}$ und $\mathbf{V}^{-1}$ sind symmetrische positiv definite $d \times d$-Matrizen.
- Ist $\boldsymbol{X}$ normalverteilt gemäß $\mathrm{No}(\boldsymbol{\mu}, \mathbf{V})$ und $\mathbf{H}$ eine $m \times d$ Matrix, so ist $\boldsymbol{Y} = \mathbf{H}\boldsymbol{X}$ normalverteilt gemäß $\mathrm{No}(\mathbf{H}\boldsymbol{\mu}, \mathbf{H}\mathbf{V}\mathbf{H}^{\mathrm{T}})$.
- Jede Randverteilung einer Normalverteilung ist wieder eine Normalverteilung. Mittelwert und Matrix der Randverteilung entstehen durch Streichen der Spalten und Zeilen der restlichen Variablen.

- Jede bedingte Verteilung einer Normalverteilung ist wieder eine Normalverteilung.
- Ist $\boldsymbol{X}$ normalverteilt gemäß $\mathrm{No}(\boldsymbol{\mu}, \mathbf{V})$, 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 $\boldsymbol{Y} = \mathbf{D}^{-1}\mathrm{U}(\boldsymbol{X} - \boldsymbol{\mu})$ ist dann standardnormalverteilt. Die Drehung $\mathbf{U}$ heißt **Hauptachsentransformation**.
- Ist $\boldsymbol{Y}$ standardnormalverteilt, so ist $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$ χ^2_n-verteilt mit n Freiheitsgraden.

5.4.1 Die multivariate Normalverteilung

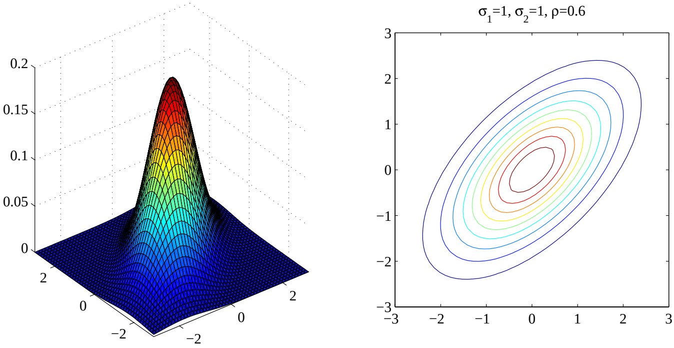
- Für $d = 2$ und $\boldsymbol{\mu} = \mathbf{0}$ kann die Dichte folgendermaßen angeschrieben werden:

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

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

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

- Zwei unkorrelierte normalverteilte Zufallsvariable mit gemeinsamer Normalverteilung sind daher **unabhängig**.


0.2
0.15
0.1
0.05
0
2
0
−2
−2
0
2
σ1=1, σ2=1, ρ=0.6
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

- Die **bedingte Dichte** $f(x_2|x_1)$ ist gegeben durch

$$f(x_2|x_1) = \frac{f(x_2, x_1)}{f(x_1)} =$$

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

- $X_2|X_1 = x_1$ ist also eine normalverteilte Zufallsvariable mit der Erwartung

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

$\mathsf{E}[X_2|X_1]$ heißt die **bedingte Erwartung**.

5.4.1 Die multivariate Normalverteilung

- 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 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

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

5.5.1 Gammaverteilung

5.5.1 Gammaverteilung

- Die Exponentialverteilung ist ein Spezialfall einer allgemeineren Familie von Verteilungen, der **Gammaverteilung**.
- Die Dichte der Gammaverteilung $\mathrm{Ga}(a, b)$ lautet:

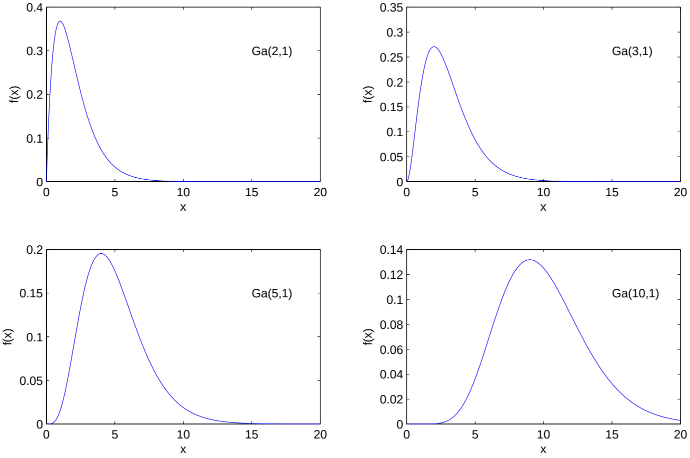
$$f(x|a,b) = \frac{x^{a-1} e^{-x/b}}{b^a \Gamma(a)} \cdot I_{[0,\infty)}(x)$$

- Ihre Verteilungsfunktion ist die regularisierte unvollständige Gammafunktion:

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

- Die Erwartung ist gleich ab, die Varianz gleich ab^2.

5.5.1 Gammaverteilung



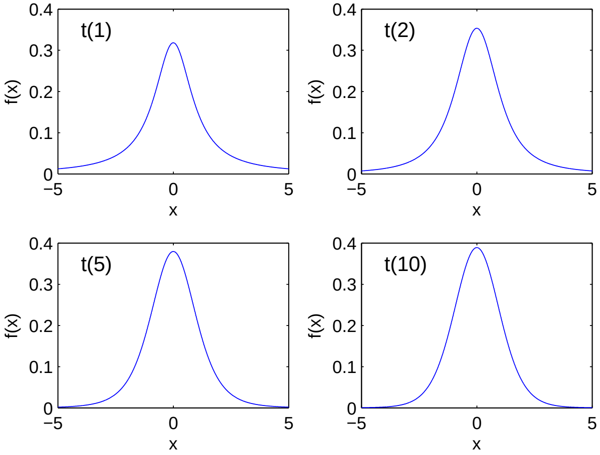
5.5.1 T-Verteilung

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

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

- Die Erwartung ist gleich 0 für $n > 1$, die Varianz ist gleich $n/(n-2)$ für $n > 2$.
- Die $\mathrm{T}(1)$-Verteilung ist auch als Cauchy-Verteilung oder Breit-Wigner-Verteilung bekannt. Sie ist das klassische Beispiel einer Verteilung ohne Momente.

5.5.1 T-Verteilung



5.5.1 χ^2-Verteilung

5.5.1 χ^2-Verteilung

- Die Dichte der $\chi^2(n)$-Verteilung mit n Freiheitsgraden lautet:

$$f(x|n) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\, x^{n/2-1} e^{-x/2} \cdot I_{[0,\infty)}(x)$$

- Sie ist die Gammaverteilung $\mathrm{Ga}(n/2, 2)$.
- Ist X standardnormalverteilt, so ist $Y = X^2$ χ^2-verteilt mit einem Freiheitsgrad.
- Die Erwartung ist gleich n, die Varianz ist gleich $2n$.

5.6.1 Funktionen von Zufallsvariablen

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

Satz

Es sei X eine diskrete Zufallsvariable und $Y = h(X)$. Dann gilt:

$$W_Y(k) = W_X(h^{-1}(k))$$

Beispiel

Wir ordnen den geraden Augenzahlen des Würfels die Zahl 0 zu, den ungeraden die Zahl 1:

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

Die Verteilung von Y ist dann gegeben durch

$$W_Y(0) = W(h^{-1}(0)) = W_X(\{2,4,6\}) = \frac{1}{2}$$
$$W_Y(1) = W(h^{-1}(1)) = W_X(\{1,3,5\}) = \frac{1}{2}$$

Beispiel

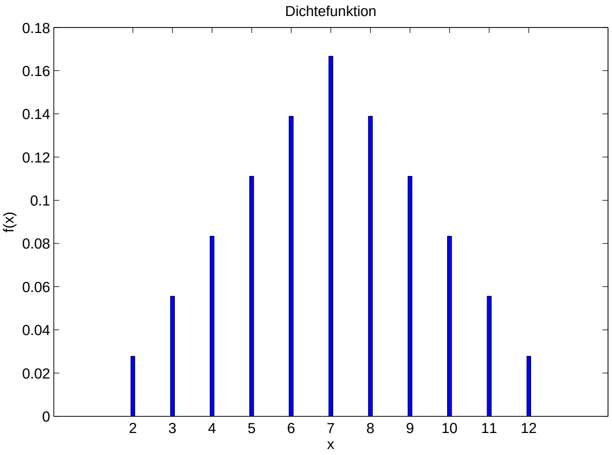
Wir ordnen dem Ergebnis $\boldsymbol{X} = (X_1, X_2)$ eines Doppelwurfs die Summe der Augenzahlen zu:

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

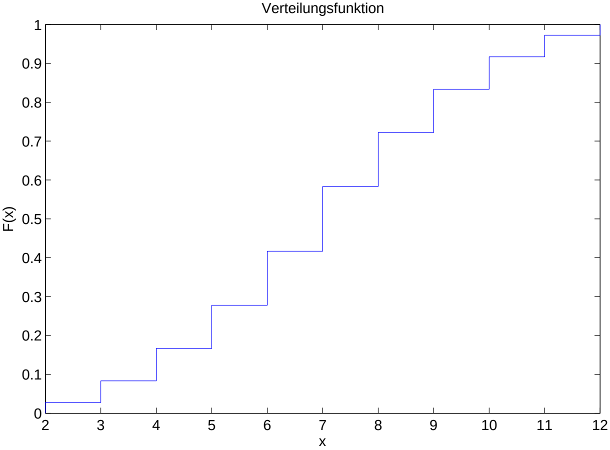
Die Werte von Y sind die natürlichen Zahlen von 2 bis 12. Die Verteilung von Y ist dann gegeben durch

$$W_Y(k) = W_X(h^{-1}(k)) = \sum_{i+j=k} W_X(\{(i,j)\}) = \begin{cases} \dfrac{k-1}{36}, & k \le 7 \\ \dfrac{13-k}{36}, & k \ge 7 \end{cases}$$

- Die Dichte der Zufallsvariablen $Y = X_1 + X_2$:



- Die Verteilungsfunktion der Zufallsvariablen $Y = X_1 + X_2$:



Satz

Es sei X eine stetige Zufallsvariable und $Y = h(X)$, h differenzierbar und eindeutig umkehrbar. Dann gilt:

$$f_Y(y)\,\mathrm{d}y = f_X(x)\,\mathrm{d}x$$

oder

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

5.6.1 Funktionen von Zufallsvariablen

Beispiel

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

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

und $Y = \cos X$. Dann ist

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

Y ist also gleichverteilt in $[-1, 1]$.

5.6.1 Funktionen von Zufallsvariablen

Beispiel

Es sei X eine exponentialverteilte Zufallsvariable mit der Dichte

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

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 Rayleigh-Verteilung oder Maxwell-Boltzmann-Verteilung in 2 Dimensionen.

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

Beispiel

Es sei X eine stetige Zufallsvariable mit der Dichte $f_X(x)$ 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

Es sei X eine standardnormalverteilte Zufallsvariable mit der Dichte

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

und $Y = X^2$. Dann ist

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

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

Satz

Es sei $\boldsymbol{X}$ eine stetiger Zufallsvektor und $\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{X})$, $\boldsymbol{h}$ differenzierbar und eindeutig umkehrbar. 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

Es sei $\boldsymbol{X}$ eine stetiger Zufallsvektor mit der Dichte $f_{\boldsymbol{X}}(\boldsymbol{x})$ und $\boldsymbol{Y} = \mathbf{A}\boldsymbol{X} + \boldsymbol{b}$. Dann ist

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

Beispiel (Transformation in Polarkoordinaten)

Es sei $\boldsymbol{X}$ ein standardnormalverteilter Zufallsvektor der Dimension 3 mit der Dichte

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)$$

Transformation in Polarkoordinaten:

$$r = \sqrt{x_1^2 + x_2^2 + x_3^2}$$
$$\varphi = \arctan(x_2/x_1)$$
$$\theta = \arccos(x_3/r)$$

Die Determinante der Jacobi-Matrix der Umkehrtransformation ist bekanntlich $r^2 \sin\theta$.

Beispiel (Fortsetzung)

Damit ergibt sich die gemeinsame Dichte des transformierten Zufallsvektors (R, Φ, Θ) zu:

$$\begin{aligned} g(r, \varphi, \theta) &= \frac{1}{(2\pi)^{3/2}} r^2 \sin\theta \exp\left(-\frac{r^2}{2}\right) \\ &= \frac{1}{2\pi} \cdot \frac{1}{2} \sin\theta \cdot \sqrt{\frac{2}{\pi}} r^2 \exp\left(-\frac{r^2}{2}\right) \end{aligned}$$

R, Φ, Θ sind also unabhängig. Φ ist gleichverteilt in $[0, 2\pi]$, $\cos\Theta$ ist gleichverteilt in $[-1, 1]$, und R ist Maxwell-Boltzmann-verteilt.

5.6.1 Faltung, Messfehler und Fehlerfortpflanzung

Definition (Faltung)

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

Satz

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

$$g(x) = \int_{-\infty}^{\infty} f_1(x - x_2) \cdot f_2(x_2)\,\mathrm{d}x_2 = \\ = \int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(x - x_1)\,\mathrm{d}x_1$$

- Die Dichte g wird als **Faltungsprodukt** von f_1 und f_2 bezeichnet:

$$g = f_1 * f_2$$

Beispiel (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 Dichte:

$$g(x) = \int_{-\infty}^{\infty} f_1(x - x_2) f_2(x_2)\, \mathrm{d}x_2 =$$

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

$$= \frac{1}{\tau^2}\, x e^{-x/\tau}$$

Die Verteilung von X heißt Erlangverteilung und ist wie die Exponentialverteilung ein Spezialfall der Gammaverteilung.

Beispiel (Faltung von zwei Normalverteilungen)

Es seien X_1 und X_2 unabhängig und normalverteilt mit Mittel μ_1 bzw. μ_2 und Varianz σ_1^2 bzw. σ_2^2. Die Summe $X = X_1 + X_2$ ist dann wieder normalverteilt mit Mittel $\mu = \mu_1 + \mu_2$ und Varianz $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Beispiel (Faltung von zwei Poissonverteilungen)

Es seien X_1 und X_2 unabhängig und Poisson-verteilt mit Mittel λ_1 bzw. λ_2. Die Summe $X = X_1 + X_2$ ist dann wieder Poisson-verteilt mit Mittel $\lambda = \lambda_1 + \lambda_2$.

Beispiel (Faltung von zwei χ^2-Verteilungen)

Es seien X_1 und X_2 unabhängig und χ^2-verteilt mit n_1 bzw. n_2 Freiheitsgraden. Die Summe $X = X_1 + X_2$ ist dann wieder χ^2-verteilt mit $n = n_1 + n_2$ Freiheitsgraden.

Beispiel (Der Messfehler)

Es wird eine Zufallsgröße X beobachtet. Der Messfehler wird durch eine bedingte Dichte $g(y|x)$ beschrieben, die die Wahrscheinlichkeit angibt, dass y bei der Messung registriert wird, wenn X den Wert x annimmt. Für die gemessene Verteilung der Messung 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. Dies ist genau dann der Fall, wenn der Messfehler und der gemessene Wert unabhängig sind.

- Wir untersuchen nun, wie sich Varianzen und Kovarianzen von Zufallsvariablen unter Variablentransformationen verhalten.
- Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ eine n-dimensionaler Zufallsvektor und $\mathbf{H}$ eine $m \times n$ - Matrix. Dann ist $\boldsymbol{Y} = (Y_1, \ldots Y_m) = \mathbf{H}\boldsymbol{X}$ ein Zufallsvektor der Dimension m.
- Es gilt **exakt**:

Lineare Transformation von Erwartung und Varianz

1. $\mathsf{E}[\boldsymbol{Y}] = \mathbf{H} \cdot \mathsf{E}[\boldsymbol{X}]$
2. $\mathsf{Cov}[\boldsymbol{Y}] = \mathbf{H} \cdot \mathsf{Cov}[\boldsymbol{X}] \cdot \mathbf{H}^{\mathrm{T}}$

Varianz einer Linearkombination

- Es sei $\boldsymbol{X} = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor und $Y = \sum_{i=1}^n a_i X_i$. Dann ist

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

- Sind alle X_i unkorreliert, gilt daher

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

und insbesondere

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

- Es soll nun statt der linearen Abbildung $\mathbf{H}$ eine allgemeine Funktion $\boldsymbol{h} = (h_1, \ldots, h_m)$ betrachtet werden.
- Es wird angenommen, dass $\boldsymbol{h}$ in jenem Bereich, in dem die Dichte von $\boldsymbol{X}$ signifikant von 0 verschieden ist, genügend gut durch eine lineare Funktion angenähert werden kann.
- Entwickelt man $\boldsymbol{h}$ an der Stelle $\mathsf{E}[\boldsymbol{X}]$, so gilt in 1. Näherung:

Lineare Fehlerfortpflanzung

1. $\mathsf{E}[\boldsymbol{Y}] \approx \boldsymbol{h}(\mathsf{E}[\boldsymbol{X}])$
2. $\mathsf{Cov}[\boldsymbol{Y}] \approx \left(\dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right) \cdot \mathsf{Cov}[\boldsymbol{X}] \cdot \left(\dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}$

5.6.1 Systematische Fehler

- Kann der Messfehler durch eine Zufallsvariable mit Mittel 0 beschrieben werden, hat die Messung nur einen **statistischen Fehler**.
- Die Messung kann jedoch durch eine falsche Kalibration (z.B. Skalenfehler oder Nullpunktfehler) des Messgeräts verfälscht sein. Solche Fehler werden **systematische Fehler** genannt.
- Systematische Fehler werden durch Vergrößerung der Stichprobe **nicht kleiner**!
- Die Korrektur von systematischen Fehlern erfordert sorgfältige Kalibration der Messaparatur, Überprüfung von theoretischen Annahmen, etc.
- Das Gesetz der Fehlerfortpflanzung gilt nicht für systematische Fehler!

5.6.1 Systematische Fehler

Beispiel

Wir messen zwei Spannungen X_1, X_2 mit dem gleichen Messgerät. Durch fehlerhafte Kalibration misst das Gerät statt der wahren Spannung U die Spannung $X_\mathrm{m} = aU + b + \varepsilon$, mit $a = 0.99, b = 0.05, \sigma[\varepsilon] = 0.03\,\mathrm{V}$. Der Mittelwert $\overline{X}$ der beiden Spannungen hat dann einen statistischen Fehler von $0.02\,\mathrm{V}$. Der systematische Fehler des Mittelwerts wird beschrieben durch $\overline{X}_\mathrm{m} = a\overline{X} + b$, ist also der der Einzelmessung. Der systematische Fehler der Differenz ΔX wird beschrieben durch $\Delta X_\mathrm{m} = a\Delta X$. Der Nullpunktfehler ist verschwunden.

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 endlicher Erwartung 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 ist $U = \lim_{n\to\infty} U_n$ standardnormalverteilt.

Zentraler Grenzwertsatz für beliebig verteilte Folgen von Zufallsvariablen

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit beliebigen Verteilungen. $\mu_i = \mathsf{E}[X_i]$ und $\sigma_i^2 = \mathsf{var}[X_i]$ seien endlich für alle $i \in \mathbb{N}$. Definiert man für jedes $n \in \mathbb{N}$ U_i und Y_n durch:

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

so ist $Y = \lim_{n\to\infty} Y_n$ standardnormalverteilt.

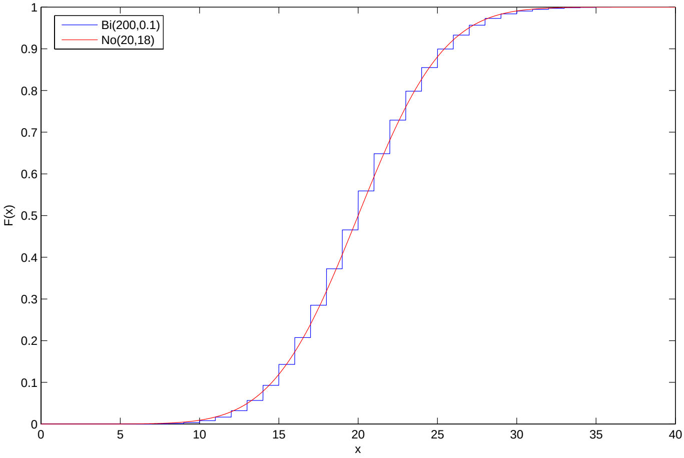
- Der zentrale Grenzwertsatz erklärt, warum die Normalverteilung in der Natur eine so bedeutende Rolle spielt, etwa bei der Verteilung der Impulskomponenten von Gasmolekülen, die das Ergebnis von zahlreichen Stößen ist.

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

Beispiel (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 $n \to \infty$ gegen eine Normalverteilung streben. Die Abbildung 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

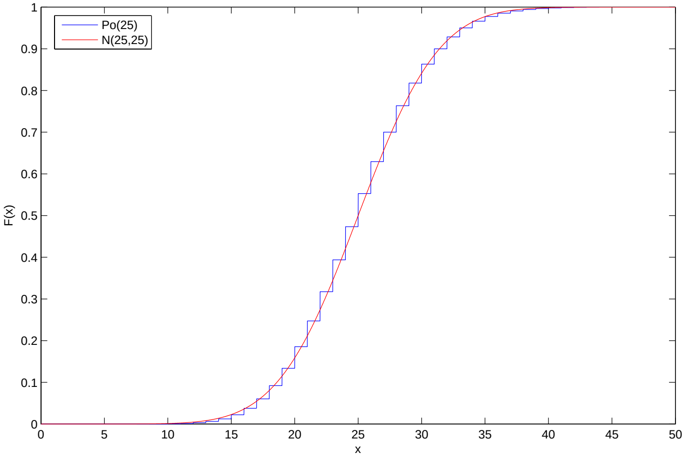
$$\mu = np = 20, \quad \sigma^2 = np(1-p) = 18$$


Bi(200,0.1)
No(20,18)
F(x)
x
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30
35
40

Beispiel (Poissonverteilung für großes n)

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

$$\mu = \lambda = 25, \quad \sigma^2 = \lambda = 25$$


Po(25)
N(25,25)
F(x)
x
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30
35
40
45
50

Teil 6

Schätzung

6.1.1 Grundbegriffe

- $X_1, \ldots, X_n$ seien unabhängige Zufallsvariable, die alle die gleiche Verteilungsfunktion F haben.
- Sie bilden dann eine **zufällige Stichprobe** der Verteilung F.
- Eine Zufallsvariable
$$Y = h(X_1, \ldots, X_n)$$
heißt eine **Stichprobenfunktion**.
- In vielen Fällen sind Momente oder die Verteilung von Y zu bestimmen.

Definition (Stichprobenmittel)

Das **Stichprobenmittel** $\overline{X}$ der Stichprobe $X_1, \ldots, X_n$ ist definiert durch

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Momente des Stichprobenmittels

Hat F das Mittel 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.

Zentraler Grenzwertsatz

1. Hat F das Mittel μ und die Varianz σ^2, so konvergiert die Verteilung von
$$U = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$
gegen die Standardnormalverteilung.
2. Ist F eine Normalverteilung, ist U für alle n standardnormalverteilt.

6.1.1 Stichprobenvarianz

Definition (Stichprobenvarianz)

Die **Stichprobenvarianz** S^2 der Stichprobe $X_1, \dots, X_n$ ist definiert durch

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Erwartung der Stichprobenvarianz

Hat F die Varianz σ^2, gilt:

$$\mathsf{E}[S^2] = \sigma^2$$

6.1.1 Stichprobenvarianz

Satz

Ist F eine Normalverteilung mit Mittel 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. Die Varianz von S^2 ist gegeben durch
$$\mathsf{var}[S^2] = \frac{2\sigma^4}{n-1}$$
4. Die Größe
$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$
ist T-verteilt mit $n-1$ Freiheitsgraden.

6.1.1 Stichprobenmedian

Definition (Stichprobenmedian)

Der **Stichprobenmedian** $\widetilde{X}$ der Stichprobe $X_1, \ldots, X_n$ ist definiert durch

$$\widetilde{X} = \begin{cases} X_{((n+1)/2)}, & n \text{ ungerade} \\ \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right), & n \text{ gerade} \end{cases}$$

Momente des Stichprobenmedians

Hat F den Median m und die Dichte 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 ist eine Stichprobenfunktion, die einen möglichst genauen Näherungswert für einen unbekannten Verteilungsparameter ϑ liefern soll:

$$T = t(X_1, \ldots, X_n)$$

- Die Funktion $t(x_1, \ldots, x_n)$ wird die Schätzfunktion genannt.
- 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.

6.2.1 Eigenschaften von Punktschätzern

Definition (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:

$$\lim_{n\to\infty} \mathsf{E}_\vartheta[T] = \vartheta$$

- Ist der unbekannte Parameter gleich ϑ, dann ist die Erwartung des Punktschätzers gleich ϑ.
- Ein unverzerrter Punktschätzer hat zwar zufällige Abweichungen vom wahren Wert ϑ, aber keine systematische Verzerrung.

6.2.1 Eigenschaften von Punktschätzern

Definition (MSE)

Die **mittlere quadratische Abweichung (mean squared error, MSE)** eines Punktschätzers T für den Parameter ϑ ist definiert durch:

$$\mathsf{MSE}[T] = \mathsf{E}_\vartheta[(T-\vartheta)^2]$$

Definition (MSE-Konsistenz)

Ein Punktschätzer T für den Parameter ϑ heißt **konsistent im quadratischen Mittel (MSE-konsistent)**, wenn gilt:

$$\lim_{n\to\infty} \mathsf{MSE}[T] = 0$$

Definition (MSE-Effizienz)

Ein Punktschätzer T_1 heißt **MSE-effizienter** als der Punktschätzer T_2, wenn für alle zulässigen ϑ gilt:

$$\mathsf{MSE}[T_1] \leq \mathsf{MSE}[T_2]$$

Definition (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.

Definition (Fisher-Information)

Es sei $X_1, \ldots, X_n$ eine Stichprobe mit der gemeinsamen Dichte $g(x_1, \ldots, x_n|\vartheta)$. Die Erwartung

$$I_\vartheta = \mathsf{E}\left[-\frac{\partial^2 \ln g(X_1, \ldots, X_n|\vartheta)}{\partial\vartheta^2}\right]$$

heißt die **Fisher-Information** der Stichprobe.

Satz von Rao und Cramèr

Es sei $X_1, \ldots, X_n$ eine Stichprobe mit der gemeinsamen Dichte $g(x_1, \ldots, x_n|\vartheta)$. Die Varianz eines unverzerrten Schätzers T für den Parameter ϑ ist nach unten beschränkt durch:

$$\mathsf{var}[T] \geq 1/I_\vartheta$$

Beispiel

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$. Die gemeinsame Dichte 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:

$$\ln g(x_1, \ldots, x_n | \tau) = -\, n \ln \tau - \sum_{i=1}^{n} x_i / \tau$$

$$\frac{\partial^2}{\partial \tau^2} \ln g(x_1, \ldots, x_n | \tau) = \frac{n}{\tau^2} - \frac{2 \sum_{i=1}^{n} x_i}{\tau^3}$$

$$\mathsf{E}\left[\frac{\partial^2}{\partial \tau^2} \ln g(X_1, \ldots, X_n | \tau) \right] = \frac{n}{\tau^2} - \frac{2\, n\tau}{\tau^3} - = -\frac{n}{\tau^2}$$

Beispiel (Fortsetzung)

Die Fisher-Information ist also gleich

$$I_\tau = \frac{n}{\tau^2}$$

Für jeden unverzerrten Schätzer T von τ gilt folglich:

$$\mathsf{var}[T] \geq \frac{\tau^2}{n}$$

Satz

1. Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit Erwartung μ. Dann ist das Stichprobenmittel $\overline{X}$ ein unverzerrter Punktschätzer von μ.
2. Hat F die endliche Varianz σ^2, so ist $\overline{X}$ MSE-konsistent.

Beispiel

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 $I_\mu = n/\sigma^2$ ist, ist $\overline{X}$ effizient für μ.

Beispiel

Ist F die Exponentialverteilung $\mathrm{Ex}(\tau)$, so ist $\overline{X}$ Gammaverteilt mit Mittel τ und Varianz τ^2/n. Da die Fisher-Information für τ gleich $I_\tau = n/\tau^2$ ist, ist $\overline{X}$ effizient für τ.

6.2.1 Schätzung des Mittelwerts

Beispiel

Ist F die Poissonverteilung $\mathrm{Po}(\lambda)$, hat $\overline{X}$ Mittel λ und Varianz λ/n. Da die Fisher-Information für λ gleich $I_\lambda = n/\lambda$ ist, ist $\overline{X}$ effizient für λ.

Beispiel

Ist F die Alternativverteilung $\mathrm{Al}(p)$, hat $\overline{X}$ Mittel p und Varianz $p(1-p)/n$. Da die Fisher-Information für p gleich $I_p = n/[p(1-p)]$ ist, ist $\overline{X}$ effizient für p.

6.2.1 Schätzung der Varianz

6.2.1 Schätzung der Varianz

Satz

1. Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit Erwartung μ und Varianz σ^2. Dann ist die Stichprobenvarianz S^2 ein unverzerrter Punktschätzer von σ^2.
2. Hat F das endliche vierte zentrale Moment μ_4, so ist

$$\mathrm{var}(S^2) = \frac{\mu_4}{n} - \frac{(n-3)\mu_2^2}{n(n-1)}$$

3. In diesem Fall ist S^2 MSE-konsistent.

Beispiel

Ist F die Normalverteilung $\mathrm{No}(\mu, \sigma^2)$, so ist $(n-1)S^2/\sigma^2$ χ^2-verteilt mit $n-1$ Freiheitsgraden. 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

$$I_\sigma^2 = \frac{n}{2\sigma^4}$$

S^2 ist also ein asymptotisch effizienter Punktschätzer für σ^2.

Satz

1. Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der stetigen Verteilung F mit Median m und Dichte f. Dann ist der Stichprobenmedian $\tilde{X}$ ein asymptotisch unverzerrter Punktschätzer von m.
2. Für symmetrisches F ist $\widetilde{X}$ unverzerrt.
3. Der Stichprobenmedian $\widetilde{X}$ hat asymptotisch die Varianz

$$\mathsf{var}[\widetilde{X}] \approx \frac{1}{4\,nf(m)^2}$$

4. Der Stichprobenmedian ist MSE-konsistent, sofern $f(m) > 0$.

Beispiel

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 gleich

$$\mathsf{var}[\widetilde{X}] = \frac{2\,\pi\sigma^2}{4\,n} \approx 1.57\frac{\sigma^2}{n}$$

Sie ist also um mehr als 50 Prozent größer als die Varianz von $\overline{X}$.

Beispiel

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der T-Verteilung $\mathrm{T}(3)$. Die Varianz von $\overline{X}$ ist gleich

$$\mathsf{var}[\overline{X}] = \frac{3}{n}$$

Die Varianz von $\widetilde{X}$ ist für großes n gleich

$$\mathsf{var}[\widetilde{X}] = \frac{1}{4\,n f(0)^2} = \frac{1.8506}{n} \approx 0.62\frac{3}{n}$$

Sie ist also fast um 40 Prozent kleiner als die Varianz von $\overline{X}$.

Definition (ML-Schätzer)

1. Es sei $X_1, \ldots, X_n$ eine Stichprobe mit der gemeinsamen Dichte $g(x_1, \ldots, x_n|\vartheta)$. Die Funktion
$$L(\vartheta|X_1, \ldots, X_n) = g(X_1, \ldots, X_n|\vartheta)$$
heißt die **Likelihoodfunktion** der Stichprobe.
2. Der **plausible** oder **Maximum-Likelihood-Schätzer** $\hat{\vartheta}$ ist jener Wert von ϑ, der die Likelihoodfunktion der Stichprobe maximiert.

- Oft wird statt der Likelihoodfunktion ihr Logarithmus, die Log-Likelihoodfunktion $\ell(\vartheta) = \ln L(\vartheta)$ maximiert.

Beispiel (ML-Schätzung eines Bernoulli-Parameters)

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Alternativverteilung $\mathrm{Al}(p)$. Die gemeinsame Dichte 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-Likelihoodfunktion 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:

$$\frac{\partial \ell(p)}{\partial p} = \frac{1}{p} \sum_{i=1}^{n} X_i - \frac{1}{1-p} \left(n - \sum_{i=1}^{n} X_i \right)$$

Beispiel (Fortsetzung)

Nullsetzen der Ableitung und Auflösen nach p ergibt:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

Der ML-Schätzer ist unverzerrt und effizient.

Beispiel (ML-Schätzung eines Poisson-Parameters)

Es sei $X_1, \dots, X_n$ eine Stichprobe aus der Poissonverteilung $\mathrm{Po}(\lambda)$. Die gemeinsame Dichte lautet:

$$g(x_1, \dots, x_n | \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Die Log-Likelihoodfunktion ist daher:

$$\ell(\lambda) = \sum_{i=1}^{n} [X_i \ln \lambda - \lambda - \ln(X_i!)]$$

Ableiten nach λ ergibt:

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} X_i - n$$

Beispiel (Fortsetzung)

Nullsetzen der Ableitung und Auflösen nach λ ergibt:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Der ML-Schätzer ist unverzerrt und effizient.

Beispiel (ML-Schätzung einer mittleren Lebensdauer)

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$. Die gemeinsame Dichte lautet:

$$g(x_1, \ldots, x_n|\tau) = \prod_{i=1}^{n} \frac{e^{-x_i/\tau}}{\tau}$$

Die Log-Likelihoodfunktion ist daher:

$$\ell(\tau) = \sum_{i=1}^{n} [-\ln \tau - \frac{1}{\tau} \sum_{i=1}^{n} X_i]$$

Ableiten nach τ ergibt:

$$\frac{\partial \ell(\tau)}{\partial \tau} = -\frac{n}{\tau} + \frac{1}{\tau^2} \sum_{i=1}^{n} X_i$$

Beispiel (Fortsetzung)

Nullsetzen der Ableitung und Auflösen nach τ ergibt:

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

Der ML-Schätzer ist unverzerrt und effizient.

Beispiel (ML-Schätzung der Parameter einer Normalverteilung)

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Normalverteilung $\mathrm{No}(\mu, \sigma^2)$. Die gemeinsame Dichte 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-Likelihoodfunktion ist daher:

$$\ell(\mu, \sigma^2) = \sum_{i=1}^{n} \left[-\ln\sqrt{2\pi} - \frac{1}{2}\ln\sigma^2 - \frac{(X_i - \mu)^2}{2\,\sigma^2}\right]$$

Ableiten nach μ und σ^2 ergibt:

$$\frac{\partial\ell(\mu, \sigma^2)}{\partial\mu} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma^2}, \quad \frac{\partial\ell(\mu, \sigma^2)}{\partial\sigma^2} = \sum_{i=1}^{n} \left[-\frac{1}{2\,\sigma^2} + \frac{(X_i - \mu)^2}{2\,\sigma^4}\right]$$

Statistik
R. Frühwirth
6.1 Stichprobenfunktionen
6.2 Punktschätzer
6.2.1 Eigenschaften von Punktschätzern
6.2.1 Schätzung des Mittelwerts
6.2.1 Schätzung der Varianz
6.2.1 Schätzung des Medians
6.2.1 Maximum-Likelihood-Schätzer
6.3 Intervallschätzer

Beispiel (Fortsetzung)

Nullsetzen der Ableitungen und Auflösen nach μ und σ^2 ergibt:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n-1}{n}S^2$$

Der ML-Schätzer von μ ist unverzerrt und effizient. Der ML-Schätzer von σ^2 ist asymptotisch unverzerrt und asymptotisch effizient.

- Der ML-Schätzer hat die folgende wichtige Eigenschaft:

Satz

Existieren die ersten beiden Ableitungen von $L(\vartheta)$, existiert die Fisher-Information $I_g(\vartheta)$ für alle ϑ und ist $E\left[(\ln L)'\right] = 0$, so ist die Likelihoodschätzung $\hat{\vartheta}$ asymptotisch normalverteilt mit Mittel ϑ und Varianz $1/I_g(\vartheta)$. $\hat{\vartheta}$ ist daher asymptotisch unverzerrt und asymptotisch effizient.

- Daraus folgt sofort die nächste Eigenschaft:

Satz

Der Likelihoodschätzer $\hat{\vartheta}$ ist (unter den selben Voraussetzungen) konsistent.

Beispiel (ML-Schätzung des Lageparameters einer Cauchyverteilung)

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Cauchyverteilung $\mathrm{T}(1)$ mit Lageparameter μ. Die gemeinsame Dichte lautet:

$$g(x_1, \ldots, x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\pi[1 + (x_i - \mu)^2]}$$

Die Log-Likelihoodfunktion ist daher:

$$\ell(\mu) = -n \ln \pi - \sum_{i=1}^{n} \ln[1 + (X_i - \mu)^2]$$

Der Wert $\hat{\mu}$, der $\ell(\mu)$ maximiert, muss numerisch gefunden werden.

MATLAB: `make_ML_cauchy`

6.2.1 Maximum-Likelihood-Schätzer

Beispiel (Fortsetzung)

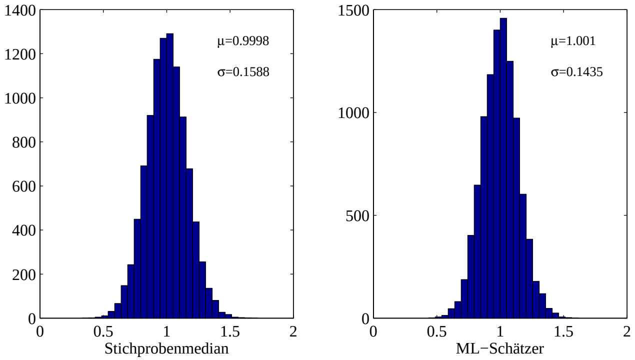
Man kann zeigen, dass die Fisher-Information der Stichprobe gleich

$$I_\mu = \frac{n}{2}$$

ist. Für große Stichproben muss daher die Varianz des ML-Schätzers $\hat{\mu}$ ungefähr gleich $2/n$ sein.

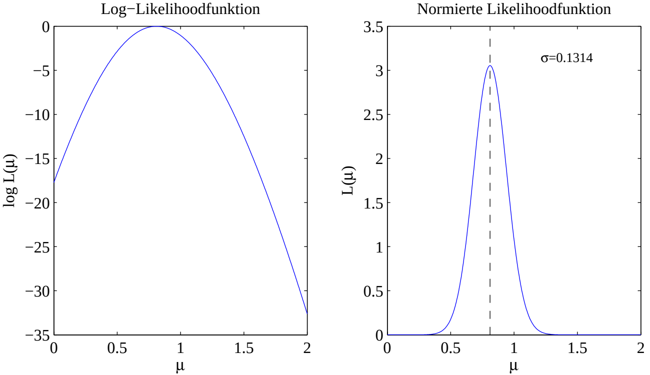
Der Stichprobenmedian $\widetilde{X}$ ist ebenfalls ein konsistenter Schätzer für μ. Seine Varianz ist asymptotisch gleich $\pi^2/(4n) \approx 2.47/n$. Sie ist also um etwa 23 Prozent größer als die Varianz des ML-Schätzers.

- Simulation von 10000 Stichproben der Größe $n = 100$:



- Die Korrelation zwischen $\widetilde{X}$ und $\hat{\mu}$ ist etwa 90%.

- Die Standardabweichung des ML-Schätzers kann näerungsweise aus der normierten Likelihoodfunktion einer Stichprobe abgelesen werden:



Beispiel (ML-Schätzung des Obergrenze einer Gleichverteilung)

Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Gleichverteilung $\mathrm{Un}(0, b)$ mit Obergrenze b. Die gemeinsame Dichte lautet:

$$g(x_1, \ldots, x_n|b) = \frac{1}{b^n}, 0 \le x_1, \ldots, x_n \le b$$

Der größte Wert der Likelihoodfunktion ist daher an der Stelle

$$\hat{b} = \max_i X_i$$

Da ein Randmaximum vorliegt, gelten die üblichen asymptotischen Eigenschaften nicht.

Beispiel (Fortsetzung)

Die Dichte von $\hat{b} = \max_i X_i$ lautet:

$$f(x) = \frac{nx^{n-1}}{b^n}$$

Daraus können Erwartung und Varianz berechnet werden:

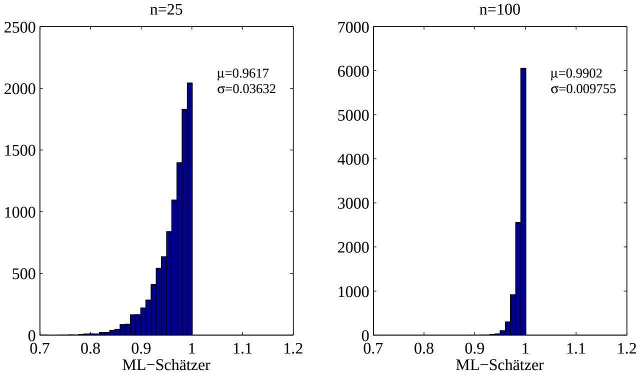
$$\mathsf{E}[\hat{b}] = \frac{n}{n+1}, \quad \mathsf{var}[\hat{b}] = \frac{b^2 n}{(n+2)(n+1)^2}$$

Der Schätzer ist asymptotisch unverzerrt, die Varianz geht aber wie $1/n^2$ gegen Null! Der Schätzer ist auch nicht asymptotisch normalverteilt.

MATLAB: `make_ML_uniform`

6.2.1 Maximum-Likelihood-Schätzer

- Simulation von 10000 Stichproben $(b = 1)$ der Größe $n = 25$ bzw. $n = 100$:



6.3.1 Grundbegriffe

6.3.1 Grundbegriffe

- Neben dem Schätzwert selbst ist auch seine Streuung um den wahren Wert von Interesse.
- Wir wollen aus einer Stichprobe ein Intervall bestimmen, das den wahren Wert mit einer gewissen Wahrscheinlichkeit enthält.

Definition (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.

6.3.1 Grundbegriffe

- 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:

$$W(\vartheta \le G_1) = W(\vartheta \ge G_2)$$

- Ein einseitiges Konfidenzintervall liegt vor, wenn gilt:

$$W(\vartheta \le G_2) = 1-\alpha \qquad \text{oder} \qquad W(\vartheta \ge G_1) = 1-\alpha$$

Mittelwert einer Alternativverteilung

- Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Alternativverteilung $\mathrm{Al}(p)$.
- Für genügend großes n ist $\hat{p} = \overline{X}$ annähernd normalverteilt gemäß $\mathrm{No}(p, p(1-p)/n)$.
- Das Standardscore
$$Z = \frac{\hat{p} - p}{\sigma[\hat{p}]}$$
ist dann annähernd standardnormalverteilt.
- Aus
$$W(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$$
folgt
$$W(\hat{p} - z_{1-\alpha/2}\sigma[\hat{p}] \leq p \leq \hat{p} + z_{1-\alpha/2}\sigma[\hat{p}]) = 1 - \alpha$$

6.3.1 Konfidenzintervall für den Mittelwert

- Da p nicht bekannt ist, muss $\sigma[\hat{p}]$ näherungsweise bestimmt werden.
- **Bootstrap-Verfahren**: p wird durch $\hat{p}$ angenähert.
- **Robustes Verfahren**: p wird so gewählt, dass $\sigma[\hat{p}]$ maximal ist, also $p = 0.5$.

Beispiel

Angabe: Bei einer Umfrage unter $n = 400$ Personen geben $k = 157$ Personen an, ein gewisse Marke zu kennen. Wir suchen ein 95%-Konfidenzintervall für den Bekanntheitsgrad p.

Lösung: Es gilt $\hat{p} = 0.3925$ und $z_{1-\alpha/2} = 1.96$. Mit dem Bootstrap-Verfahren ergibt sich $\sigma[\hat{p}] = 0.0244$. Die Grenzen des Konfidenzintervalls sind daher

$$G_1 = 0.3925 - 1.96 \cdot 0.0244 = 0.3446$$
$$G_2 = 0.3925 + 1.96 \cdot 0.0244 = 0.4404$$

Beispiel (Fortsetzung)

Mit dem robusten Verfahren ergibt sich $\sigma[\hat{p}] = 0.025$ und die Grenzen

$$G_1 = 0.3925 - 1.96 \cdot 0.025 = 0.3435$$
$$G_2 = 0.3925 + 1.96 \cdot 0.025 = 0.4415$$

Das robuste Intervall ist in diesem Fall nur unwesentlich länger als das Bootstrap-Intervall.

MATLAB: `make_KI_alternative`

Mittelwert einer 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

$$W(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$$

folgt

$$W(\overline{X} - z_{1-\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha$$

6.3.1 Konfidenzintervall für den Mittelwert

- 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

$$W(-t^{n-1}_{1-\alpha/2} \leq T \leq t^{n-1}_{1-\alpha/2}) = 1 - \alpha$$

folgt

$$W(\overline{X} - t^{n-1}_{1-\alpha/2} S/\sqrt{n} \leq \mu \leq \overline{X} + t^{n-1}_{1-\alpha/2} S/\sqrt{n}) = 1 - \alpha$$

6.3.1 Konfidenzintervall für den Mittelwert

Beispiel

Eine Stichprobe vom Umfang $n = 50$ aus der Standardnormalverteilung hat das Stichprobenmittel $\overline{X} = 0.0540$ und die Stichprobenvarianz $S^2 = 1.0987$. Wird die Varianz als bekannt vorausgesetzt, lautet das symmetrische 95%-Konfidenzintervall für μ:

$$G_1 = 0.0540 - 1.96/\sqrt{50} = -0.2232$$

$$G_2 = 0.0540 + 1.96/\sqrt{50} = 0.3312$$

Wird die Varianz als unbekannt angenommen, lautet das symmetrische 95%-Konfidenzintervall für μ:

$$G_1 = 0.0540 - 2.01 \cdot 1.0482/\sqrt{50} = -0.2439$$

$$G_2 = 0.0540 + 2.01 \cdot 1.0482/\sqrt{50} = 0.3519$$

MATLAB: `make_KI_normal`

Mittelwert einer Exponentialverteilung

- Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung $\mathrm{Ex}(\tau)$.
- $T = \sum_{i=1}^{n} X_i$ hat die folgende Dichte:

$$f(t) = \frac{t^{n-1}}{\tau^n \Gamma(n)} \exp\left(-\frac{t}{\tau}\right)$$

- T ist also Gammaverteilt gemäß $\mathrm{Ga}(n, \tau)$, und T/τ ist verteilt gemäß $\mathrm{Ga}(n, 1)$.
- Aus

$$W\left(\gamma_{\alpha/2,n,1} \le \frac{T}{\tau} \le \gamma_{1-\alpha/2,n,1}\right) = 1 - \alpha$$

folgt

$$W\left(\frac{T}{\gamma_{1-\alpha/2,n,1}} \le \tau \le \frac{T}{\gamma_{\alpha/2,n,1}}\right) = 1 - \alpha$$

Mittelwert einer beliebigen Verteilung

- Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Verteilung F mit Mittel μ und Varianz σ^2.
- Aufgrund 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 **approximativ**:

$$W(\overline{X} - z_{1-\alpha/2} S/\sqrt{n} \leq \mu \leq \overline{X} + z_{1-\alpha/2} S/\sqrt{n}) = 1 - \alpha$$

6.3.1 Konfidenzintervall für den Mittelwert

Beispiel

Für exponentialverteilte Stichproben vom Umfang n gibt die folgende Tabelle die Sicherheit des 95%-Konfidenzintervall 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

MATLAB: `make_KI_exponential`

6.3.1 Konfidenzintervall für die Varianz

Varianz einer 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

$$W\left(\chi^2_{\alpha/2,n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha$$

folgt

$$W\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right) = 1-\alpha$$

6.3.1 Konfidenzintervall für die Varianz

Beispiel

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:

$$G_1 = 49 \cdot 4.3949/70.2224 = 3.0667$$

$$G_2 = 49 \cdot 4.3949/31.5549 = 6.8246$$

Werden die Quantile der χ^2-Verteilung $\chi^2(n-1)$ durch die Quantile der Normalverteilung $\mathrm{No}(n-1, 2(n-1))$ ersetzt, lautet das Konfidenzintervall:

$$G_1 = 49 \cdot 4.3949/68.4027 = 3.1483$$

$$G_2 = 49 \cdot 4.3949/29.5973 = 7.2760$$

MATLAB: `make_KI_normal_varianz.m`

Teil 7

Testen von Hypothesen

Teil 7: Testen von Hypothesen

7.1 Einleitung

7.1 Einleitung

- Wir beobachten eine Stichprobe $X_1, \ldots, X_n$ aus einer Verteilung F.
- 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**.
- Der Test entscheidet, ob die Stichprobe mit der Hypothese vereinbar ist, nicht ob die Hypothese richtig ist!

Statistik

R. Frühwirth

7.1 Einleitung

7.2 Parametrische Tests

7.3 Anpassungstests

Allgemeine Vorgangsweise

- Aus der Stichprobe wird eine Testgröße (Teststatistik) T berechnet.
- Der Wertebereich von T wird, in Abhängigkeit von H_0, in einen **Ablehnungsbereich** (kritischen Bereich) 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 Daten mit der Hypothese vereinbar sind.

Signifikanz und Güte

- Bei jedem Testverfahren sind zwei Arten von Fehlern möglich.
  1. **Fehler 1. Art**: Die Hypothese H_0 wird abgelehnt, obwohl sie zutrifft.
  2. **Fehler 2. Art**: Die Hypothese H_0 wird beibehalten, obwohl sie nicht zutrifft.
- Die Verteilung von T unter Annahme von H_0 wird bestimmt.
- Der Ablehnungsbereich wird so festgelegt, dass die Wahrscheinlichkeit eines Fehlers 1. Art maximal gleich einem Wert α ist.
- α heißt das **Signifikanzniveau** des Tests. Gängige Werte sind $\alpha = 0.05, 0.01, 0.005$.

7.1 Einleitung

- 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** des Tests für H_1.
- Die Güte sollte nie kleiner als α sein.
- Ist die Güte nie kleiner als α, heißt der Test **unverzerrt**.
- Ein Ziel der Testtheorie ist es, unverzerrte Tests mit maximaler Güte (UMPU) zu konstruieren.

7.2.1 Grundlagen

- Wir betrachten eine Stichprobe $X_1, \ldots, X_n$ aus einer Verteilung F, die bis auf einen oder mehrere Parameter spezifiziert ist.
- Tests von Hypothesen über F heißen **parametrisch**.
- Eine Nullhypothese H_0 kann als eine Teilmenge des Parameterraums Θ aufgefasst werden.
- Der Test entscheidet, ob die Stichprobe mit der Hypothese vereinbar ist.
- Vor der Anwendung ist zu klären, ob die angenommene parametrische Form plausibel ist.

7.2.1 Grundlagen

- Zunächst wird die Teststatistik T und das Signifikanzniveau α gewählt.
- Dann wird der kritische Bereich C so festgelegt, dass

$$W(T \in C | \vartheta \in H_0) \leq \alpha$$

- Zu einer Nullhypothese H_0 kann eine **Gegenhypothese** H_1 formuliert werden.
- H_1 kann ebenfalls als Teilmenge des Parameterraums Θ aufgefasst werden.
- Ist das Signifikanzniveau α festgelegt, kann für jedes $\vartheta \in H_1$ die Güte berechnet werden:

$$1 - \beta(\vartheta) = W(T \in C | \vartheta \in H_1)$$

- $1 - \beta(\vartheta)$ heißt die **Gütefunktion** des Tests.

Beispiel mit Exponentialverteilung

- $X_1, \ldots, X_n$ ist eine exponentialverteilte Stichprobe aus $\mathrm{Ex}(\tau)$.
- Die Hypothese $H_0 : \tau = \tau_0$ soll anhand der Stichprobe getestet werden.
- Als Teststatistik T wählen wir das Stichprobenmittel: $T = \overline{X}$.
- Unter Annahme von H_0 hat T die folgende Dichte:

$$f(t) = \frac{t^{n-1}}{(\tau_0/n)^n \Gamma(n)} \exp\left(-\frac{t}{\tau_0/n}\right)$$

- 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.

7.2.1 Grundlagen

- Ein Verwerfungsbereich mit Signifikanzniveau α ist die Menge

$$C = [0, Q_{\alpha/2}] \cup [Q_{1-\alpha/2}, \infty[$$

wo Q_p das Quantil der $\mathrm{Ga}(n, \tau_0/n)$-Verteilung zum Niveau p ist.

- Die Gütefunktion für einen Wert τ ergibt sich durch:

$$1 - \beta(\tau) = W(T \in C) = G(Q_{\alpha/2}) + 1 - G(Q_{1-\alpha/2})$$

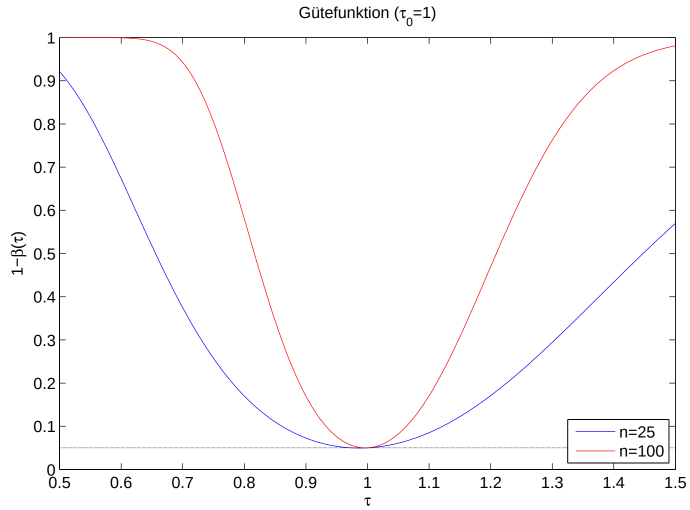
wo G die Verteilungsfunktion der $\mathrm{Ga}(n, \tau/n)$-Verteilung ist.

- Der Test ist nicht unverzerrt, da z.B. für $\tau_0 = 1$ und $n = 25$

$$1 - \beta(0.986) = 0.0495 < \alpha$$

- MATLAB: `make_test_exponential_mean.m`

7.2.1 Grundlagen



7.2.1 Tests für normalverteilte Stichproben

Erwartungswert bei bekannter Varianz

- $X_1, \ldots, X_n$ ist eine normalverteilte Stichprobe aus $\mathrm{No}(\mu, \sigma^2)$ mit bekanntem σ^2.
- Die Hypothese $H_0 : \mu = \mu_0$ soll anhand der Stichprobe gegen die Alternativhypothese $H_1 : \mu \neq \mu_0$ getestet werden.
- Als Teststatistik T wählen wir das Standardscore des Stichprobenmittels:

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}$$

- Unter Annahme von H_0 ist T verteilt gemäß $\mathrm{No}(0, 1)$.
- H_0 wird abgelehnt, wenn T von seinem Erwartungswert „weit entfernt“, also relativ klein oder relativ groß ist.

7.2.1 Tests für normalverteilte Stichproben

- Ein Verwerfungsbereich mit Signifikanzniveau α ist die Menge

$$C =]-\infty, z_{\alpha/2}] \cup [z_{1-\alpha/2}, \infty[$$

  wo z_p das Quantil der Standardnormalverteilung zum Niveau p ist.

- Die Hypothese H_0 wird also abgelehnt, wenn

$$|T| = \frac{\sqrt{n}\,\left|\overline{X} - \mu_0\right|}{\sigma} > z_{1-\alpha/2}$$

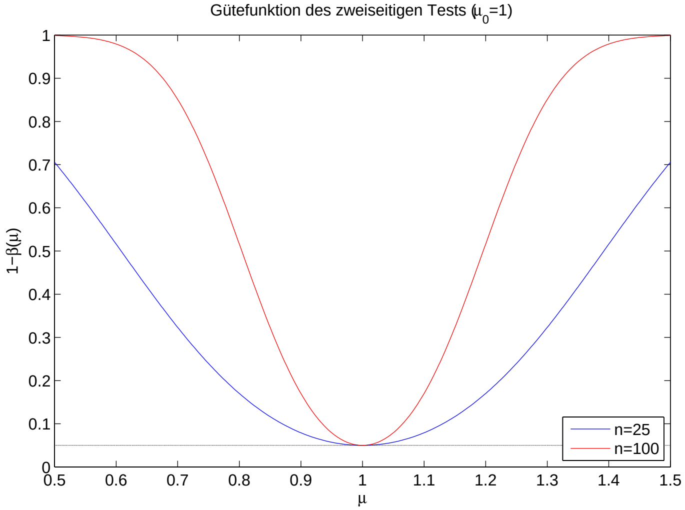
- Die Gütefunktion für einen Wert μ ergibt sich durch:

$$1 - \beta(\mu) = W(T \in C) = G(z_{\alpha/2}) + 1 - G(z_{1-\alpha/2})$$

  wo G die Verteilungsfunktion der $\mathrm{No}(\sqrt{n}(\mu - \mu_0)/\sigma, 1)$-Verteilung ist.

- Der Test ist unverzerrt.

- MATLAB: `make_test_normal_mean.m`


Gütefunktion des zweiseitigen Tests (μ_0=1)
1−β(μ)
μ
n=25
n=100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5

Einseitiger Test

- Die Hypothese $H_0 : \mu = \mu_0$ soll mit der Teststatistik T gegen die Alternativhypothese $H_1 : \mu > \mu_0$ getestet werden.
- H_0 wird abgelehnt, wenn T „zu groß" ist.
- Ein Verwerfungsbereich mit Signifikanzniveau α ist die Menge

$$C = [z_{1-\alpha}, \infty[$$

- Die Hypothese H_0 wird also abgelehnt, wenn

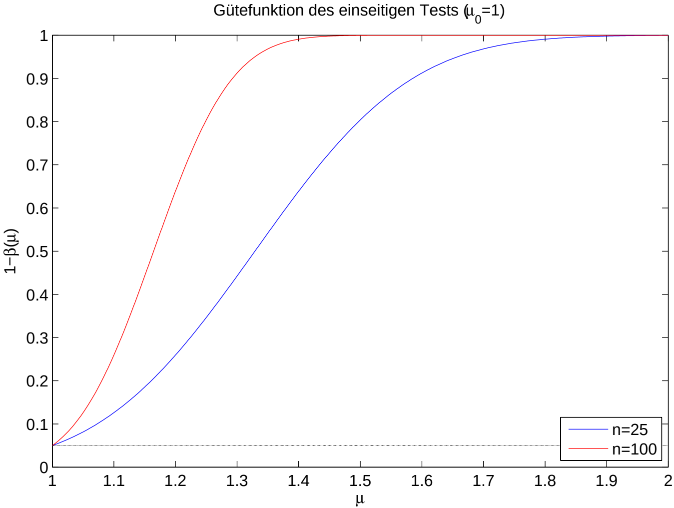
$$T = \frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma} > z_{1-\alpha}$$

7.2.1 Tests für normalverteilte Stichproben

- Die Gütefunktion für einen Wert $\mu > \mu_0$ ergibt sich durch:

$$1 - \beta(\mu) = W(T \in C) = 1 - G(z_{1-\alpha})$$

  wo G die Verteilungsfunktion der $\mathrm{No}(\sqrt{n}(\mu - \mu_0)/\sigma, 1)$-Verteilung ist.
- Analog verläuft der Test mit $H_1 : \mu < \mu_0$.
- Matlab: `make_test_normal_mean.m`


Gütefunktion des einseitigen Tests (μ0=1)
1−β(μ)
μ
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2
n=25
n=100

Erwartungswert bei unbekannter Varianz: T-Test

- $X_1, \ldots, X_n$ ist eine normalverteilte Stichprobe aus $\mathrm{No}(\mu, \sigma^2)$ mit unbekanntem σ^2.
- Die Hypothese $H_0 : \mu = \mu_0$ soll anhand der Stichprobe gegen die Alternativhypothese $H_1 : \mu \neq \mu_0$ getestet werden.
- Als Teststatistik T wählen wir das Standardscore des Stichprobenmittels, unter Benützung der Stichprobenvarianz S^2:

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S}$$

- Unter Annahme von H_0 ist T verteilt gemäß $\mathrm{T}(n-1)$.

- H_0 wird abgelehnt, wenn T von seinem Erwartungswert „weit entfernt“, also relativ klein oder relativ groß ist.
- Ein Verwerfungsbereich mit Signifikanzniveau α ist die Menge

$$C =]-\infty, t^{n-1}_{\alpha/2}] \cup [t^{n-1}_{1-\alpha/2}, \infty[$$

wo t^{n-1}_p das Quantil der T-Verteilung mit $n-1$ Freiheitsgraden zum Niveau p ist.

- Die Hypothese H_0 wird also abgelehnt, wenn

$$|T| = \frac{\sqrt{n}\,\left|\overline{X} - \mu_0\right|}{S} > t^{n-1}_{1-\alpha/2}$$

- Die Gütefunktion für einen Wert μ ergibt sich durch:

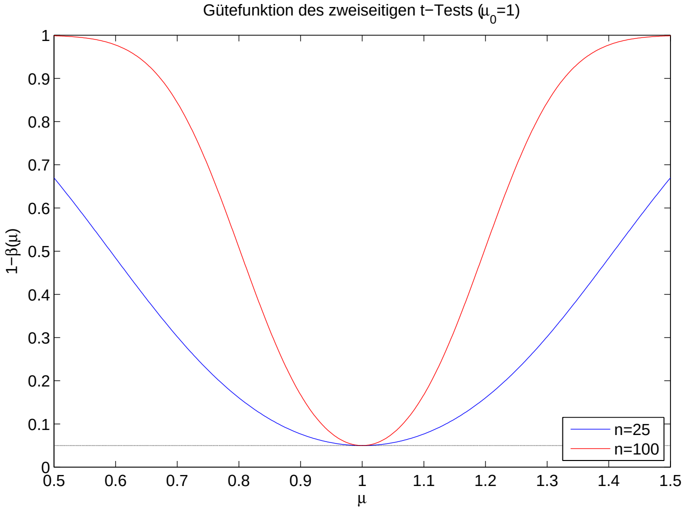
$$1 - \beta(\tau) = W(T \in C) = G(z_{\alpha/2}) + 1 - G(z_{1-\alpha/2})$$

  wo G die Verteilungsfunktion der nichtzentralen $\mathrm{T}(n-1, \delta)$-Verteilung mit

$$\delta = \sqrt{n}(\mu - \mu_0)/S$$

  ist.
- Der Test ist unverzerrt.
- MATLAB: `make_test_normal_mean.m`

7.2.1 Tests für normalverteilte Stichproben



Gleichheit von zwei Erwartungswerten

- $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ sind zwei unabhängige normalverteilte Stichprobe aus $\mathrm{No}(\mu_x, \sigma_x^2)$ bzw. $\mathrm{No}(\mu_y, \sigma_y^2)$.
- Die Hypothese $H_0 : \mu_x = \mu_y$ soll anhand der Stichproben gegen die Alternativhypothese $H_1 : \mu_x \neq \mu_y$ getestet werden.
- Sind die **Varianzen bekannt**, wählen wir als Teststatistik T die Differenz der Stichprobenmittel:

$$T = \overline{X} - \overline{Y}$$

- Unter Annahme von H_0 ist T verteilt gemäß $\mathrm{No}(0, \sigma_x^2/n + \sigma_y^2/m)$.

- 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
$$\frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} > z_{1-\alpha/2}$$

- Sind die **Varianzen unbekannt und gleich**, kann die Varianz aus der kombinierten („gepoolten“) 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

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S^2(1/n + 1/m)}}$$

  T-verteilt mit $n+m-2$ Freiheitsgraden.

- Die Hypothese H_0 wird also abgelehnt, wenn

$$|T| > t_{1-\alpha/2}^{n+m-2}$$

  wo $t_{1-\alpha/2}^{n+m-2}$ das Quantil der T-Verteilung mit $n+m-2$ Freiheitsgraden ist.

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.
- Wir nehmen an, dass $W_1, \ldots, W_n$ normalverteilt mit Mittel μ_w und unbekannter Varianz σ_w^2 ist.
- Die Hypothese $H_0 : \mu_w = 0$ (keine Wirkung der Intervention) soll anhand der Stichprobe gegen die Alternativhypothese $H_1 : \mu_w \neq 0$ getestet werden.
- Dies erfolgt mit dem T-Test für einzelne Stichproben.

Test der Varianz

- $X_1, \ldots, X_n$ ist eine normalverteilte Stichprobe mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2.
- Die Hypothese $H_0 : \sigma^2 = \sigma_0^2$ soll anhand der Stichprobe gegen die Alternativhypothese $H_1 : \sigma^2 \neq \sigma_0^2$ 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.

7.2.1 Tests für normalverteilte Stichproben

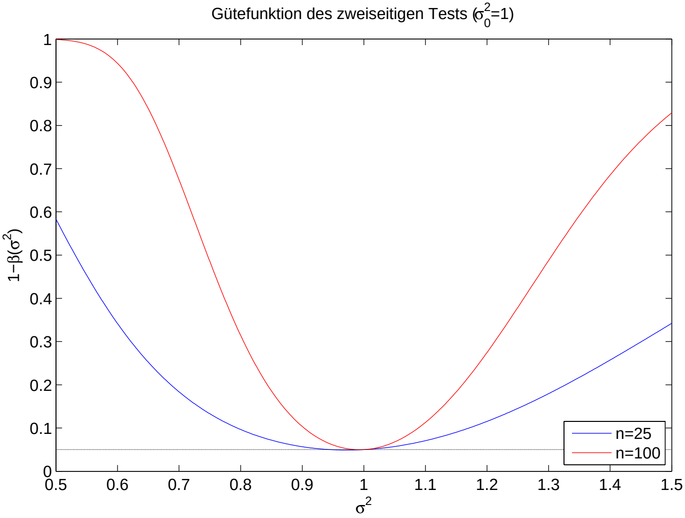
- Die Hypothese H_0 wird also abgelehnt, wenn

$$T < \chi^2_{\alpha/2,n-1} \quad \text{oder} \quad T > \chi^2_{1-\alpha/2,n-1}$$

  wo $\chi^2_{p,k}$ das Quantil der χ^2-Verteilung mit k Freiheitsgraden zum Niveau p ist.
- Die Gütefunktion für einen Wert σ^2 ergibt sich durch:

$$1 - \beta(\sigma^2) = G(\sigma_0^2/\sigma^2 \cdot \chi^2_{\alpha/2}) + 1 - G(\sigma_0^2/\sigma^2 \cdot \chi^2_{1-\alpha/2})$$

  wo G die Verteilungsfunktion der $\chi^2(n-1)$-Verteilung ist.
- Der Test ist nicht unverzerrt.
- MATLAB: `make_test_normal_variance.m`


Gütefunktion des zweiseitigen Tests (σ_0^2=1)
1−β(σ²)
σ²
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
n=25
n=100

Gleichheit von zwei Varianzen

- $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ sind zwei unabhängige normalverteilte Stichprobe aus $\mathrm{No}(\mu_x, \sigma_x^2)$ bzw. $\mathrm{No}(\mu_y, \sigma_y^2)$.
- Die Hypothese $H_0 : \sigma_x^2 = \sigma_y^2$ soll anhand der Stichproben gegen die Alternativhypothese $H_1 : \sigma_x^2 \neq \sigma_y^2$ 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 $\mathrm{F}(n-1, m-1)$.

7.2.1 Tests für normalverteilte Stichproben

- Die Hypothese H_0 wird also abgelehnt, wenn

$$T < F_{\alpha/2} \quad \text{oder} \quad T > F_{1-\alpha/2}$$

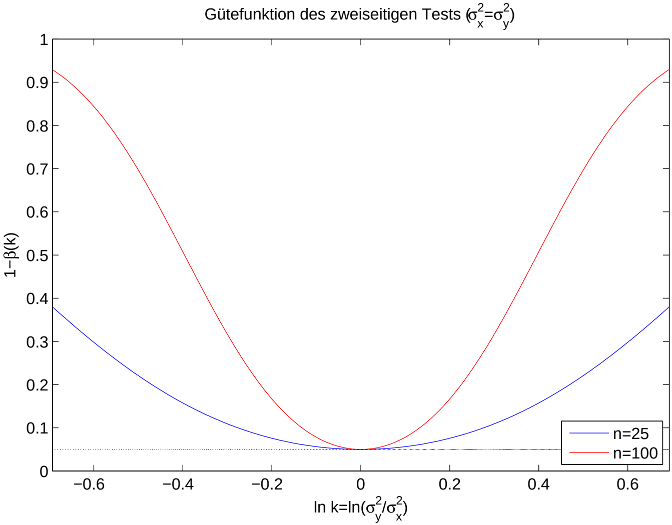
wo F_p das Quantil der F-Verteilung mit $n-1$ bzw. $m-1$ Freiheitsgraden zum Niveau p ist.

- Ist $\sigma_y^2 = k\sigma_x^2$, ergibt sich die Gütefunktion für einen Wert k durch:

$$1 - \beta(\tau) = G(\sigma_0^2/\sigma^2 \cdot F_{\alpha/2}) + 1 - G(\sigma_0^2/\sigma^2 \cdot F_{1-\alpha/2})$$

wo G die Verteilungsfunktion der $\mathrm{F}(n-1, m-1)$-Verteilung ist.

- Der Test ist unverzerrt.
- MATLAB: `make_test_normal_variance.m`


Gütefunktion des zweiseitigen Tests ($\sigma_x^2=\sigma_y^2$)
1−β(k)
ln k=ln(σ_y^2/σ_x^2)
n=25
n=100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
−0.6
−0.4
−0.2
0
0.2
0.4
0.6

Einseitiger Test auf Erwartungswert

- $X_1, \ldots, X_n$ ist eine alternativverteilte Stichprobe aus $\mathrm{Al}(p)$.
- Die Hypothese $H_0 : p \le p_0$ soll anhand der Stichprobe gegen die Alternativhypothese $H_1 : p > p_0$ getestet werden.
- Als Teststatistik T wählen wir die Anzahl der Ergebnisse, die 1 sind:

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

- T ist binomialverteilt gemäß $\mathrm{Bi}(n, p)$.
- H_0 wird abgelehnt, wenn T „zu groß“ ist.

7.2.1 Tests für alternativverteilte Stichproben

- Ist $p \leq p_0$, gilt

$$W(T \geq k) \leq \sum_{i=k}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}$$

- Die Hypothese H_0 wird abgelehnt, wenn

$$\sum_{i=T}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$$

Beispiel

Ein Hersteller behauptet, dass nicht mehr als 2 Prozent 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 Prozent nicht widerlegen.

MATLAB: `make_test_alternative_mean.m`

7.2.1 Tests für alternativverteilte Stichproben

Näherung durch Normalverteilung

- Ist n genügend groß, kann die Verteilung von T durch eine Normalverteilung $\mathrm{No}(np, np(1-p))$ angenähert werden.
- H_0 wird abgelehnt, wenn das Standardscore größer als das $(1-\alpha)$-Quantil der Standardnormalverteilung ist:

$$Z = \frac{T - np_0}{\sqrt{np(1-p_0)}} \geq z_{1-\alpha}$$

Beispiel

Mit der Angabe des letzten Beispiels ergibt die Näherung:

$$Z = 1.2372 < z_{0.95} = 1.6449$$

Die Hypothese kann also nicht abgelehnt werden.

MATLAB: `make_test_alternative_mean.m`

Einseitiger Test auf Erwartungswert

- $X_1, \ldots, X_n$ ist eine Poissonverteilte Stichprobe aus $\mathrm{Po}(\lambda)$.
- Die Hypothese $H_0 : \lambda \leq \lambda_0$ soll anhand der Stichprobe gegen die Alternativhypothese $H_1 : \lambda > \lambda_0$ getestet werden.
- Als Teststatistik T wählen wir die Stichprobensumme:

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

- T ist Poissonverteilt gemäß $\mathrm{Po}(n\lambda)$.
- H_0 wird abgelehnt, wenn T „zu groß“ ist, also wenn

$$T > P_{1-\alpha,n\lambda}$$

  wo $P_{1-\alpha,n\lambda}$ das $(1-\alpha)$-Quantil der Poissonverteilung $\mathrm{Po}(n\lambda)$ ist.

Beispiel

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 lässt sich also auf einem Signifikanzniveau von 1 Prozent oder mehr widerlegen.

MATLAB: `make_test_poisson_mean.m`

Näherung durch Normalverteilung

- Ist n genügend groß, kann die Verteilung von T durch eine Normalverteilung $\mathrm{No}(n\lambda, n\lambda)$ angenähert werden.
- H_0 wird abgelehnt, wenn das Standardscore größer als das $(1-\alpha)$-Quantil der Standardnormalverteilung ist:

$$Z = \frac{T - n\lambda_0}{\sqrt{n\lambda_0}} \geq z_{1-\alpha}$$

Beispiel

Mit der Angabe des letzten Beispiels ergibt die Näherung:

$$Z = 2.5938 > z_{0.99} = 1.6449$$

Die Hypothese kann also auf einem Signifikanzniveau von 1 Prozent abgelehnt werden.

MATLAB: `make_test_poisson_mean.m`

7.3 Anpassungstests

- Ein Test, der die Hypothese überprüft, ob die Daten 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.

Der Chiquadrat-Test für diskrete Beobachtungen

- Die Stichprobe $X_1, \ldots, X_n$ entstammt einer diskreten Verteilung mit Wertebereich $\{1, \ldots, k\}$.
- Wir testen die Hypothese H_0, dass die Dichte f die Werte $f(j) = p_j, j = 1, \ldots, k$ hat:

$$H_0 : W(X_i = j) = p_j, j = 1, \ldots, k$$

gegen

$$H_1 : W(X_i = j) \neq p_j, \text{für ein } j$$

- Es sei Y_j die Zahl der Beobachtungen, die gleich j sind.
- Unter der Nullhypothese ist $Y_1, \ldots, Y_k$ multinomial verteilt gemäß $\mathrm{Mu}(n, p_1, \ldots, p_k)$ und $\mathsf{E}[Y_j] = np_j$.

7.3.1 Der Chiquadrat-Test

- 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 kritische Bereich kann nach dem folgenden Ergebnis bestimmt werden:

Satz

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 H_0 abgelehnt, wenn

$$T \geq \chi^2_{1-\alpha,k-1}$$

  wo $\chi^2_{1-\alpha,k}$ das Quantil der χ^2-Verteilung mit $k-1$ Freiheitsgraden zum Niveau $1-\alpha$ ist.
- 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 $np_j > 5, j = 1, \ldots, k$.
- Ist das nicht erfüllt, sollte der Ablehnungsbereich durch Simulation bestimmt werden.

Beispiel

Wir testen anhand einer Stichprobe vom Umfang 50, ob ein Würfel symmetrisch ist, d.h. ob die Augenzahl X folgende Verteilung hat:

$$W(X=1)=\ldots=W(X=6)=\frac{1}{6}$$

Eine Simulation von $N=100000$ Stichproben ergibt:

$$\overline{T}=5.000, \quad S_T^2=9.789$$

Das 0.95-Quantil der χ^2-Verteilung mit fünf Freiheitsgraden ist $\chi^2_{0.95,5}=11.07$, und

$$W(T\geq 11.07)=0.048$$

MATLAB: `make_chi2test_wuerfel.m`

Der Chiquadrat-Test für stetige Beobachtungen

- Die Stichprobe $X_1, \ldots, X_n$ entstammt einer stetigen Verteilung F.
- Wir testen die Hypothese $H_0 : F(x) = F_0(x)$.
- Dazu wird der Wertebereich von X in k Gruppen $G_1, \ldots, G_k$ eingeteilt.
- Es sei Y_j die Zahl der Beobachtungen in Gruppe G_j.
- Unter der Nullhypothese ist $Y_1, \ldots, Y_k$ multinomial verteilt gemäß $\mathrm{Mu}(n, p_1, \ldots, p_k)$ und $\mathsf{E}[Y_j] = np_j$, mit

$$p_j = W(X \in G_j | H_0)$$

- Der Test verläuft weiter wie im diskreten Fall.

Unbekannte Parameter

- Die Nullhypothese muss nicht vollständig spezifiziert sein. Wir betrachten den Fall, dass die p_j noch von unbekannten Parametern $\boldsymbol{\vartheta}$ abhängen:

$$W(X \in G_j) = p_j(\boldsymbol{\vartheta})$$

- Die Statistik T ist nun 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 Minimierung von T:

$$\tilde{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} T(\boldsymbol{\vartheta})$$

- Der kritische Bereich kann nach dem folgenden Ergebnis bestimmt werden.

Satz

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

$$T \geq \chi^2_{1-\alpha,k-1-m}$$

wo $\chi^2_{1-\alpha,k}$ das Quantil der χ^2-Verteilung mit $k-1-m$ Freiheitsgraden zum Niveau $1-\alpha$ ist.

Beispiel

Angabe: Die Zahl der Arbeitsunfälle wurde in einem großen Betrieb über 30 Wochen erhoben. Es ergaben sich folgende Werte:

$$\boldsymbol{X} = \{8,0,0,1,3,4,0,2,12,5,1,8,0,2,0,\\ 1,9,3,4,5,3,3,4,7,4,0,1,2,1,2\}$$

Es soll die Hypothese überprüft werden, dass die Beobachtungen Poisson-verteilt gemäß $\mathrm{Po}(\lambda)$ sind.

Lösung: Die Beobachtungen werden in fünf Gruppen eingeteilt:

Gruppe	1	2	3	4	5
X	0	1	2–3	4–5	>5

Die Häufigkeiten der Gruppen sind:

$$Y_1 = 6, Y_2 = 5, Y_3 = 8, Y_4 = 6, Y_5 = 5$$

Beispiel (Fortsetzung)

Der Schätzwert für λ ist das Stichprobenmittel:

$$\tilde{\lambda} = 3.1667$$

Die Erwartungswerte der Y_j unter Annahme von $H_0 = \mathrm{Po}(\tilde{\lambda})$ sind:

j	1	2	3	4	5
$\mathsf{E}[Y_j]$	1.2643	4.0037	13.0304	8.6522	3.0493

Die Testgröße T ist gleich

$$T = 21.99$$

Das 99%-Quantil der χ^2-Verteilung mit drei Freiheitsgraden ist gleich $\chi^2_{0.99,3} = 11.35$. Die Hypothese, dass die Beobachtungen Poisson-verteilt sind, ist also abzulehnen.

MATLAB: `make_chi2test_poisson.m`

Eine Stichprobe

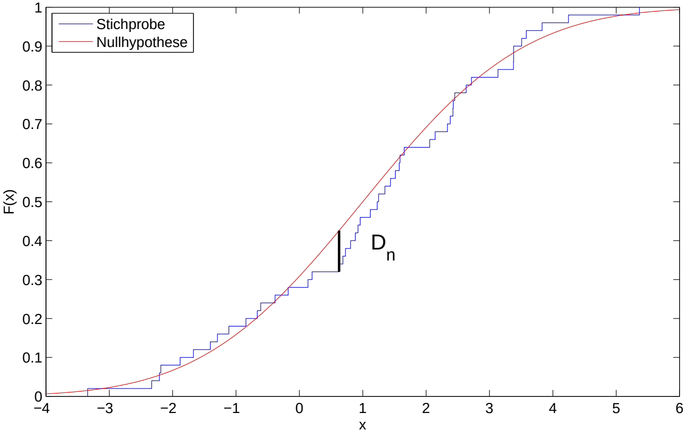
- Die Stichprobe $X_1, \ldots, X_n$ ist aus der stetigen Verteilung mit Verteilungsfunktion F.
- Wir testen die Hypothese $H_0 : F(x) = F_0(x)$.
- Die Testgröße D_n ist die maximale absolute Abweichung der empirischen Verteilungsfunktion $F_n(x)$ der Stichprobe von der hypothetischen Verteilungsfunktion $F_0(x)$:

$$D_n = \max_x |F_n(x) - F_0(x)|$$

- Für Stichproben aus F_0 ist die Verteilung von D_n unabhängig von F_0!
- Für Stichproben aus F_0 strebt die Verteilungsfunktion von $\sqrt{n}D_n$ für $n \to \infty$ gegen:

$$K(x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2x^2}$$

7.3.1 Der Kolmogorov-Smirnov-Test



7.3.1 Der Kolmogorov-Smirnov-Test

- 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.
- MATLAB: Funktion `kstest`

7.3.1 Der Kolmogorov-Smirnov-Test

Zwei Stichproben

- Wir testen, ob zwei Stichproben vom Umfang n bzw. m aus der gleichen Verteilung F stammen.
- Die Testgröße ist die maximale absolute Differenz der empirischen Verteilungsfunktionen:

$$D_{n,m} = \max_x |F_n^1(x) - F_m^2(x)|$$

- Die Nullhypothese wird abgelehnt, wenn

$$\sqrt{\frac{nm}{n+m}} D_{n,m} > K_{1-\alpha}$$

- MATLAB: Funktion `kstest2`

Teil 8

Regressionsanalyse

Teil 8: Regressionsanalyse

8.1 Einleitung

- Regressionsanalyse untersucht die Abhängigkeit der Beobachtungen von diversen bekannten (nicht zufälligen) Variablen.
- Einflussvariable (unabhängige Variable) $\boldsymbol{x} = (x_1, \ldots, x_r)$.
- Ergebnisvariable (abhängige Variable) Y.
- Regressionsmodell:

$$Y = f(\boldsymbol{\beta}, \boldsymbol{x}) + \varepsilon$$

  mit Regressionskoeffizienten $\boldsymbol{\beta}$ und Fehlerterm ε.
- Ziel ist die Schätzung von $\boldsymbol{\beta}$ anhand von Beobachtungen $Y_1, \ldots, Y_n$.
- Eine Einflussvariable: einfache Regression;
  Mehrere Einflussvariable: mehrfache (multiple) Regression.

8.2.1 Lineare Regression

- Das einfachste Regressionsmodell ist eine Gerade:

$$Y = \alpha + \beta x + \varepsilon, \quad \mathsf{E}[\varepsilon] = 0, \ \mathsf{var}[\varepsilon] = \sigma^2$$

- Es seien nun $Y_1, \ldots, Y_n$ die Ergebnisse für die Werte $x_1, \ldots, x_n$ der Einflussvariablen x.
- Die Schätzung von α und β kann nach dem Prinzip der kleinsten Fehlerquadrate erfolgen.
- Die folgende Zielfunktion wird minimiert:

$$SS = \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2$$

- Gradient von SS:

$$\frac{\partial SS}{\partial \alpha} = -2 \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i), \ \frac{\partial SS}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (Y_i - \alpha - \beta x_i)$$

- Nullsetzen des Gradienten gibt die **Normalgleichungen**:

$$\sum_{i=1}^{n} Y_i = n\alpha + \beta \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i Y_i = \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2$$

- Die geschätzten Regressionskoeffizienten lauten:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i - \bar{x} \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\bar{x}$$

- 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

$$R_i = Y_i - \hat{Y}_i, \quad \hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$$

- Kovarianzmatrix der geschätzten Regressionkoeffizienten:

$$\mathsf{Cov}[\hat{\alpha}, \hat{\beta}] = \sigma^2 \begin{pmatrix} \dfrac{\sum x_i^2}{n\left(\sum x_i^2 - n\bar{x}^2\right)} & -\dfrac{\sum x_i}{n\left(\sum x_i^2 - n\bar{x}^2\right)} \\[2ex] -\dfrac{\sum x_i}{n\left(\sum x_i^2 - n\bar{x}^2\right)} & \dfrac{1}{\sum x_i^2 - n\bar{x}^2} \end{pmatrix}$$

- Ist $\beta = 0$, hängt das Ergebnis überhaupt nicht von den Einflussvariablen ab.
- Ein Test der Nullhypothese $H_0 : \beta = 0$ gegen $H_1 : \beta \neq 0$ beruht auf dem folgenden Satz:

Satz

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

$$\hat{\sigma}_{\hat{\alpha}}^2 = \frac{\hat{\sigma}^2 \sum x_i^2}{n \left(\sum x_i^2 - n\bar{x}^2 \right)}, \quad \hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{\sum x_i^2 - n\bar{x}^2}$$

- Die Nullhypothese $H_0 : \beta = 0$ wird abgelehnt, wenn die Testgröße

$$T = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$$

relativ klein oder relativ groß ist, also wenn

$$\frac{|\hat{\beta}|}{\hat{\sigma}_{\hat{\beta}}} > t^{n-2}_{1-\alpha/2}$$

wo t_p^{n-2} das Quantil der T-Verteilung mit $n-2$ Freiheitsgraden zum Niveau p ist.

- Ein analoger Test kann für die Nullhypothese $H_0 : \alpha = 0$ durchgeführt werden.

- Die symmetrischen Konfidenzintervalle mit 95% Sicherheit lauten:

$$\hat{\alpha} \pm \hat{\sigma}_{\hat{\alpha}} \cdot t^{n-2}_{1-\alpha/2}, \quad \hat{\beta} \pm \hat{\sigma}_{\hat{\beta}} \cdot t^{n-2}_{1-\alpha/2}$$

- 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 mittels Fehlerfortpflanzung:

$$\mathsf{var}[\mathsf{E}[Y_0]] = \sigma^2 \left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n\bar{x}^2} \right]$$

- Da Y_0 um seinen Erwartungswert mit Varianz σ^2 streut, ergibt sich:

$$\mathsf{var}[Y_0] = \sigma^2 \left[\frac{n+1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n\bar{x}^2} \right]$$

- Das symmetrische Prognoseintervall für Y_0 mit Sicherheit α ist daher gleich:

$$\hat{\alpha} + \hat{\beta} x_0 \pm t^{n-2}_{1-\alpha/2} \hat{\sigma} \sqrt{\frac{n+1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n\bar{x}^2}}$$

8.2.1 Tests, Konfidenz- und Prognoseintervalle

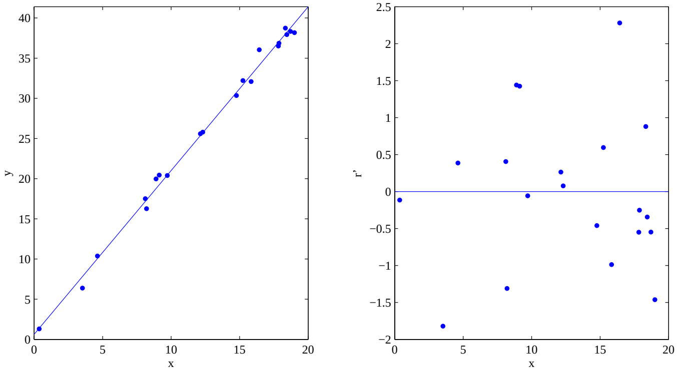
- Die Angemessenheit des Modells kann durch Untersuchung der studentisierten Residuen (Restfehler) überprüft werden.
- Das Residuum R_k hat die Varianz

$$\mathsf{var}[R_k] = \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_k - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}\right]$$

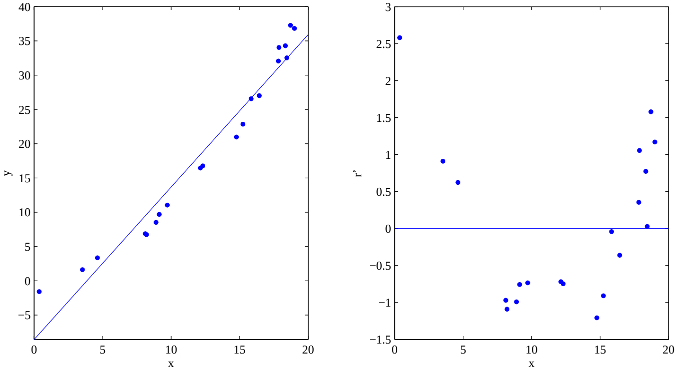
- Das studentisierte Residuum ist dann

$$R'_k = \frac{R_k}{\hat{\sigma}\sqrt{1 - \frac{1}{n} - \frac{(x_k - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}}}$$

- Es hat Erwartung 0 und Varianz 1.

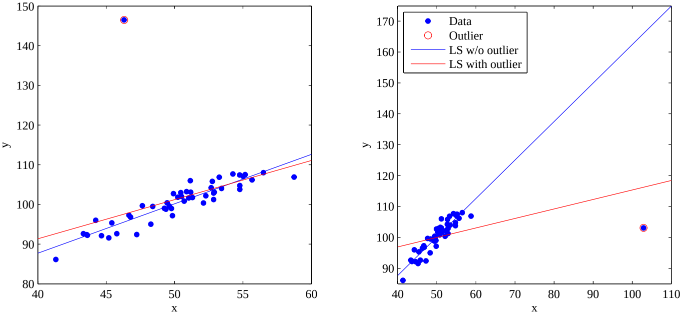


Regressionsgerade und studentisierte Residuen



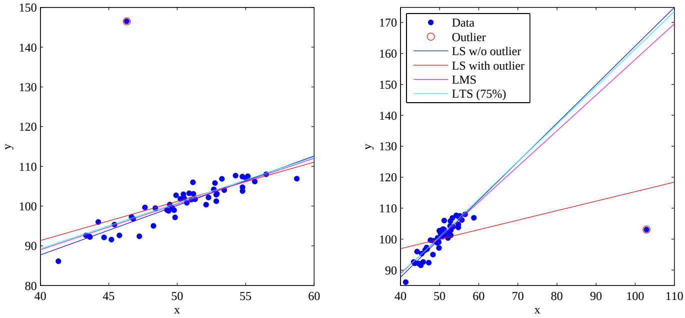
Regressionsgerade und studentisierte Residuen

- Als LS-Schätzer ist die Regressionsgerade nicht robust, d.h. empfindlich gegen Ausreißer.



Lineare Regression mit Ausreißern

- **LMS (Least Median of Squares)**: Anstatt der Summe der Fehlerquadrate wird der **Median** der Fehlerquadrate minimiert.
- "Exact fit property": Die LMS-Gerade geht durch zwei Datenpunkte.
- Berechnung kombinatorisch.
- **LTS (Least Trimmed Squares)**: Es wird die Summe einer festen Anzahl $h \leq n$ von Fehlerquadraten minimiert.
- Berechnung iterativ (FAST-LTS).
- Beide Methoden gehen auf P. Rousseeuw zurück.



Robuste Regression mit Ausreißern

8.2.1 Polynomiale Regression

- Ist der Zusammenhang zwischen x und Y nicht annähernd linear, kann man versuchen, ein Polynom anzupassen.
- Das Modell lautet dann:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_r x^r + \varepsilon, \quad \mathsf{E}[\varepsilon] = 0,\ \mathsf{var}[\varepsilon] = \sigma^2$$

- Es seien wieder $Y_1, \ldots, Y_n$ die Ergebnisse für die Werte $x_1, \ldots, x_n$ der Einflussvariablen x.
- In Matrix-Vektor-Schreibweise:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

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:

$$SS = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

- Gradient von SS:

$$\frac{\partial SS}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

- Nullsetzen des Gradienten gibt die **Normalgleichungen**:

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}$$

- Die Lösung lautet:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}$$

8.2.1 Polynomiale Regression

- Die Varianz des Fehlerterms wird erwartungstreu geschätzt durch:

$$\hat{\sigma}^2 = \frac{1}{n-r-1} \sum_{i=1}^{n} R_i^2$$

mit

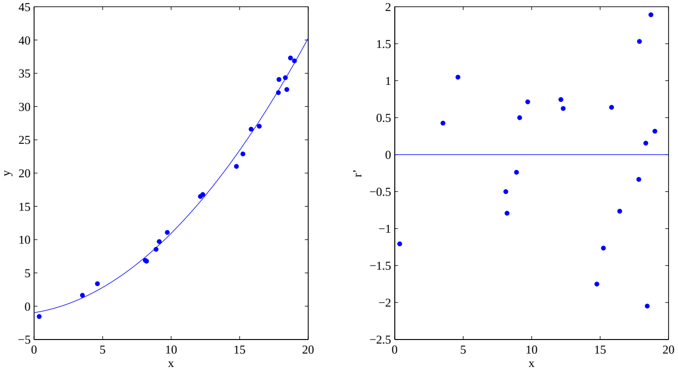
$$\boldsymbol{R} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}, \quad \hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

- Kovarianzmatrix der geschätzten Regressionkoeffizienten:

$$\mathsf{Cov}[\hat{\boldsymbol{\beta}}] = \sigma^2 \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

- Kovarianzmatrix der Residuen $\boldsymbol{R}$:

$$\mathsf{Cov}[\boldsymbol{R}] = \sigma^2 \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]$$



Regressionsparabel und studentisierte Residuen

8.3.1 Das lineare Modell

- Hängt das Ergebnis Y von mehreren Einflussvariablen ab, lautet das einfachste lineare Regressionmodell:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1 + \cdots + \beta_r x_r + \varepsilon, \quad \mathsf{E}[\varepsilon] = 0,\ \mathsf{var}[\varepsilon] = \sigma^2$$

- 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:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

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 folgende Zielfunktion wird minimiert:

$$SS = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

- Gradient von SS:

$$\frac{\partial SS}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

- Nullsetzen des Gradienten gibt die **Normalgleichungen**:

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{Y} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}$$

- Die Lösung lautet:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{Y}$$

8.3.1 Schätzung, Tests und Prognoseintervalle

- Die Varianz des Fehlerterms wird erwartungstreu geschätzt durch:

$$\hat{\sigma}^2 = \frac{1}{n-r-1} \sum_{i=1}^{n} R_i^2$$

mit

$$\boldsymbol{R} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}, \quad \hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

- Kovarianzmatrix der geschätzten Regressionskoeffizienten:

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}] = \sigma^2 \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

- Kovarianzmatrix der Residuen $\boldsymbol{R}$:

$$\mathrm{Cov}[\boldsymbol{R}] = \sigma^2 \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\right]$$

- Ist $\beta_k = 0$, hängt das Ergebnis überhaupt nicht von den Einflussvariablen x_k ab.
- Ein Test der Nullhypothese $H_0 : \beta_k = 0$ gegen $H_1 : \beta_k \neq 0$ beruht auf dem folgenden Satz:

Satz

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 $H_0 : \beta_k = 0$ wird abgelehnt, wenn die Testgröße

$$T = \frac{\hat{\beta}_k}{\hat{\sigma}_{\hat{\beta}_k}}$$

relativ klein oder relativ groß ist, also wenn

$$\frac{|\hat{\beta}_k|}{\hat{\sigma}_{\hat{\beta}_k}} > t^{n-r-1}_{1-\alpha/2}$$

wo t^{n-2}_p das Quantil der T-Verteilung mit $n-2$ Freiheitsgraden zum Niveau p ist.

- Das symmetrische Konfidenzintervall für β_k mit 95% Sicherheit lautet:

$$\hat{\beta}_k \pm \hat{\sigma}_{\hat{\beta}_k} \cdot t^{n-r-1}_{1-\alpha/2}$$

8.3.1 Schätzung, Tests und Prognoseintervalle

- 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.
- Wir erweitern $\boldsymbol{x}_0$ um den Wert 1: $\boldsymbol{x}_+ = (1, x_{01}, \ldots, x_{0r})$.
- 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 Fehlerfortpflanzung:

$$\mathsf{var}[\mathsf{E}[Y_0]] = \sigma^2 \boldsymbol{x}_+ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} {\boldsymbol{x}_+}^{\mathrm{T}}$$

- Da Y_0 um seinen Erwartungswert mit Varianz σ^2 streut, ergibt sich:

$$\text{var}[\text{E}[Y_0]] = \sigma^2 \left[1 + \boldsymbol{x}_+ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \boldsymbol{x}_+{}^{\mathrm{T}}\right]$$

- Das symmetrische Prognoseintervall für Y_0 mit Sicherheit α ist daher gleich:

$$\boldsymbol{x}_+ \cdot \hat{\boldsymbol{\beta}} \pm t^{n-k-1}_{1-\alpha/2}\hat{\sigma}\sqrt{1 + \boldsymbol{x}_+ \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \boldsymbol{x}_+{}^{\mathrm{T}}}$$

8.3.1 Gewichtete Regression

- Im allgemeinen Fall können die Fehlerterme eine beliebige Kovarianzmatrix haben:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathsf{Cov}[\boldsymbol{\varepsilon}] = \mathbf{V}$$

- Ist $\mathbf{V}$ bekannt, lautet die Zielfunktion:

$$SS = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}), \quad \mathbf{G} = \mathbf{V}^{-1}$$

- Gradient von SS:

$$\frac{\partial SS}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})$$

- Nullsetzen des Gradienten gibt die **Normalgleichungen**:

$$\mathbf{X}^{\mathrm{T}}\mathbf{G}\boldsymbol{Y} = \mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\boldsymbol{\beta}$$

- Die Lösung lautet:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}\boldsymbol{Y}$$

8.3.1 Gewichtete Regression

- Kovarianzmatrix der geschätzten Regressionskoeffizienten:

$$\mathsf{Cov}[\hat{\boldsymbol{\beta}}] = \left(\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right)^{-1}$$

- Kovarianzmatrix der Residuen $\boldsymbol{R}$:

$$\mathsf{Cov}[\boldsymbol{R}] = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}$$

- Tests und Prognoseintervalle können entsprechend modifizert werden.

8.3.1 Nichtlineare Regression

- In der Praxis ist die Abhängigkeit der Ergebnisse von den Regressionskoeffizienten oft nichtlinear:

$$\boldsymbol{Y} = \boldsymbol{h}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \quad \mathsf{Cov}[\boldsymbol{\varepsilon}] = \mathbf{V}$$

- Ist $\mathbf{V}$ bekannt, lautet die Zielfunktion:

$$SS = [\boldsymbol{Y} - \boldsymbol{h}(\boldsymbol{\beta})]^{\mathrm{T}} \mathbf{G} [\boldsymbol{Y} - \boldsymbol{h}(\boldsymbol{\beta})], \quad \mathbf{G} = \mathbf{V}^{-1}$$

- SS kann mit dem Gauß-Newton-Verfahren minimiert werden.
- Dazu wird $\boldsymbol{h}$ an einer Stelle $\boldsymbol{\beta}_0$ linearisiert:

$$\boldsymbol{h}(\boldsymbol{\beta}) \approx \boldsymbol{h}(\boldsymbol{\beta}_0) + \mathbf{H}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \boldsymbol{c} + \mathbf{H}\boldsymbol{\beta}, \quad \mathbf{H} = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0}$$

- Die Schätzung von $\boldsymbol{\beta}$ lautet:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{G}(\boldsymbol{Y} - \boldsymbol{c})$$

- $\boldsymbol{h}$ wird neuerlich an der Stelle $\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}$ linearisiert.
- Das Verfahren wird iteriert, bis die Schätzung sich nicht mehr wesentlich ändert.
- Es gibt noch viele andere Methoden zur Minimierung von SS.