

# Mathematik 3

## (Stochastik/Statistik)

**Zusammenfassung**

Fabian Damken

8. November 2023



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

---

# Inhaltsverzeichnis

---

<b>1</b>	<b>Einführung</b>	<b>4</b>
<b>2</b>	<b>Grundbegriffe</b>	<b>5</b>
2.1	Allgemeine Definitionen . . . . .	5
2.2	Messreihen . . . . .	5
2.2.1	Empirische Verteilungsfunktion . . . . .	5
2.2.2	Klassen . . . . .	5
2.2.3	Zweidimensionale Messreihen . . . . .	6
2.3	Maßzahlen . . . . .	6
2.3.1	Lagemaßzahlen . . . . .	6
2.3.2	Streuungsmaße . . . . .	7
2.3.3	Regressionsgerade . . . . .	8
2.4	Zufallsexperimente und Wahrscheinlichkeit . . . . .	8
2.4.1	Zufallsexperimente . . . . .	8
2.4.2	Wahrscheinlichkeit . . . . .	9
2.5	Zufallsvariablen und Verteilungsfunktion . . . . .	11
2.5.1	Zufallsvariablen . . . . .	11
2.5.2	Verteilungsfunktion . . . . .	12
2.5.3	Diskret/Stetig verteilte Zufallsvariablen . . . . .	12
2.5.4	Erwartungswert und Varianz . . . . .	16
2.5.5	Tschebyscheffsche Ungleichung . . . . .	16
2.5.6	Unabhängigkeit . . . . .	17
2.6	Einige Sätze . . . . .	17
2.6.1	Das schwache Gesetz der großen Zahlen . . . . .	17
2.6.2	Zentraler Grenzwertsatz . . . . .	17
2.6.3	Zentralsatz der Statistik . . . . .	18
2.6.4	Anwendungen . . . . .	18
<b>3</b>	<b>Schätzverfahren und Konfidenzintervalle</b>	<b>19</b>
3.1	Grundlagen . . . . .	19
3.1.1	(Asymptotische) Erwartungstreue . . . . .	19
3.1.2	Mittlerer quadratischer Fehler . . . . .	20
3.1.3	Konsistenz . . . . .	20
3.2	Maximum-Likelihood-Schätzer . . . . .	20
3.3	Konfidenzintervalle . . . . .	21
3.3.1	Konstruktion für normalverteilte Zufallsvariablen . . . . .	21
<b>4</b>	<b>Testverfahren</b>	<b>23</b>
4.1	Grundlagen . . . . .	23
4.1.1	Testgröße . . . . .	23

---

4.1.2	Allgemeines Konstruktionsprinzip zum Niveau $\alpha$ . . . . .	24
4.2	Wichtige Tests bei Normalverteilungsannahme . . . . .	24
4.3	Verteilungstests . . . . .	26
4.4	Weitere statische Tests . . . . .	28
<b>5</b>	<b>Robuste Statistik</b>	<b>29</b>
5.1	Median . . . . .	29
5.1.1	Schätzer . . . . .	29
5.2	M-Schätzer . . . . .	30
<b>6</b>	<b>Multivariate Verteilungen und Summen von Zufallsvariablen</b>	<b>31</b>
6.1	Grundlagen . . . . .	31
6.1.1	Multivariate Normalverteilung . . . . .	32
6.1.2	Unabhängigkeit . . . . .	32
6.1.3	Korrelation . . . . .	32
6.2	Verteilung der Summe von Zufallsvariablen . . . . .	33
6.2.1	Faltung . . . . .	33
6.2.2	Diskret verteilte Zufallsvariablen . . . . .	33

# 1 Einführung

---

- Nahezu überall treten unsicherheitsbehaftete Daten, Parameter oder Prozesse auf.  
Beispiele: Messfehler, Materialschwankungen, Rauschen, Inferenz, Nutzerverhalten, ...
- In vielen Bereichen der Informatik sind mathematische Modelle zur Verarbeitung unsicherer Daten eine unerlässliche Basis.  
Beispiele: Signalverarbeitung, Regelungstechnik, Machine Learning, Robotik, ...
- Im allgemeinen lässt sich die Statistik/Stochastik in folgende Bereiche einteilen:
  - Die *Beschreibende Statistik* dient dazu, Beobachtungsdaten darzustellen und zu charakterisieren.
  - In der *Schließenden Statistik* geht es darum, Risiken auf Basis von mathematischen Modellen abzuschätzen und einzustufen.
  - Diese mathematischen Modelle werden in der *Wahrscheinlichkeitstheorie* behandelt.

---

## 2 Grundbegriffe

---

### 2.1 Allgemeine Definitionen

---

**Gamma-Funktion**

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt, \quad x > 0$$

**Beta-Funktion**

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt, \quad \alpha, \beta > 0$$

---

### 2.2 Messreihen

---

Eine *Messreihe* ist eine Reihe von  $n$  Zahlen:

$$x_1, x_2, \dots, x_n$$

Messreihen können in quantitativ-diskrete und quantitativ-stetige Typen eingeordnet werden, wobei die Merkmalsausprägungen bei ersterem ganze Zahlen sind und bei letzterem reelle Zahlen.

Wird eine beliebige Messreihe der Größe nach sortiert, so entsteht eine *geordnete Messreihe*:

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

Sie enthält die gleichen Werte, aber so angeordnet, dass  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  gilt.

---

#### 2.2.1 Empirische Verteilungsfunktion

---

Die *empirische Verteilungsfunktion* zu einer Messreihe  $x_1, x_2, \dots, x_n$  ist die Funktion

$$F(z; x_1, x_2, \dots, x_n) = \frac{\text{Anzahl der } x_i \text{ mit } x_i \leq z}{n} = \frac{\max\{i \mid x_{(i)} \leq z\}}{n}$$

---

#### 2.2.2 Klassen

---

Werden  $r - 1$  Zahlen  $a_1 < a_2 < \dots < a_{r-1}$  gewählt, so entsteht die Unterteilung von  $\mathbb{R}$  in  $r$  Klassen:

$$\mathbb{R} = (-\infty, a_1) \cup (a_1, a_2] \cup \dots \cup (a_{r-1}, a_r] \cup (a_r, \infty)$$

Mit  $F(z) = F(z; x_1, x_2, \dots, x_n)$  ergibt sich die *relative Klassenhäufigkeit* für die  $r$  Klassen mit:

$$F(a_1), \quad F(a_2) - F(a_1), \quad F(a_{r-1}) - F(a_{r-2}), \quad 1 - F(a_{r-1})$$

---

Werden noch zwei zusätzliche Zahlen  $a_0 < \min\{a_1, x_{(1)}\}$  und  $a_r > \max\{a_{r-1}, x_{(n)}\}$  gewählt, so kann die Klassenhäufigkeit als *Histogramm* dargestellt werden, wobei über jedem Intervall  $(a_{j-1}, a_j]$ ,  $j = 1, \dots, r$  ein Rechteck mit der Fläche der jeweiligen Klassenhäufigkeit erstellt wird. Die Gesamtfläche des Histogramms ist somit 1.

---

### 2.2.3 Zweidimensionale Messreihen

---

Werden bei einer statistischen Erhebung zwei Merkmale gleichzeitig ermittelt, entstehen *zweidimensionale Messreihen*:

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

---

## 2.3 Maßzahlen

---

---

### 2.3.1 Lagemaßzahlen

---

---

#### Eindimensional

---

Sei  $x_1, x_2, \dots, x_n$  eine Messreihe mit der dazugehörigen geordneten Messreihe  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ .

#### Arithmetisches Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

#### Empirischer Median

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

#### $p$ -Quantil ( $0 < p < 1$ )

$$x_p = \begin{cases} x_{np} & \text{falls } np \text{ ganzzahlig} \\ x_{\lfloor np \rfloor + 1} & \text{falls } np \text{ nicht ganzzahlig} \end{cases}$$

Das 0.25-Quantil wird *unteres Quantil*, das 0.75-Quantil *oberes Quantil* genannt. Das 0.5-Quantil entspricht dem Median.

#### $\alpha$ -gestutztes Mittel ( $0 < \alpha < 0.5$ )

$$\bar{x}_\alpha = \frac{1}{n - 2k} (x_{(k+1)} + \dots + x_{(n-k)}), \quad k = \lfloor n\alpha \rfloor$$

Anschaulich: Die extremsten  $k$  Messwerte werden ignoriert.

---

#### Zweidimensional

---

Sei  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  eine Messreihe.

---

## Arithmetische Mittel

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

---

## 2.3.2 Streuungsmaße

---

### Eindimensional

Sei  $x_1, x_2, \dots, x_n$  eine Messreihe mit der dazugehörigen geordneten Messreihe  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ .

### Empirische Varianz

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

### Empirische Streuung

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

### Spannweite

$$v = x_{(n)} - x_{(1)}$$

### Quartilsabstand

$$q = x_{0.75} - x_{0.25}$$

---

### Zweidimensional

### Empirische Varianzen

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

### Empirische Streuungen

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

### Empirische Kovarianz

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

### Empirische Korrelationskoeffizient

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Es gilt immer  $|r_{xy}| \leq 1$ . Je näher  $|r_{xy}|$  an 1 liegt, desto stärker korrelieren  $x$  und  $y$ .

---

### 2.3.3 Regressionsgerade

---

Der Zusammenhang der  $x$ - und  $y$ -Werte lässt sich durch eine *Regressionsgerade* visualisieren.

$$y = \hat{a}x + \hat{b}$$

Die Parameter  $\hat{a}$  und  $\hat{b}$  berechnen sich dabei wie folgt:

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

Der Korrelationskoeffizient gibt den Trend der Abhängigkeiten der  $y$ -Werte von den  $x$ -Werten an:

$$\text{Die Regressionsgerade verläuft} \begin{cases} \text{streng monoton steigend.} & r_{xy} > 0 \\ \text{streng monoton fallend.} & r_{xy} < 0 \\ \text{horizontal.} & r_{xy} = 0 \end{cases}$$

---

### Residuen

---

Die Abweichungen der Punkte  $(x_i, y_i)$  von der Regressionsgerade in vertikaler Richtung

$$r_i = y_i - \hat{a}x_i - \hat{b}, \quad i = 1, \dots, n$$

werden *Residuen* genannt.

Für das *Residuenquadrat* gilt:

$$\sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - \hat{y})^2 (1 - r_{xy}^2)$$

Die vertikale Abweichung von der Regressionsgerade hängt also stark von dem Korrelationskoeffizienten ab. Für Werte von  $|r_{xy}|$ , die nahe an 1 liegen verschwinden die Residuen annähernd (für  $|r_{xy}| = 1$  verschwinden sie vollständig).

---

## 2.4 Zufallsexperimente und Wahrscheinlichkeit

---

---

### 2.4.1 Zufallsexperimente

---

Ein *Zufallsexperiment* ist ein Vorgang, der so genau beschrieben wird, dass er als beliebig oft wiederholbar betrachtet werden kann und dessen Ergebnisse vom Zufall abhängen.

- Die Menge  $\Omega$  heißt *Ergebnismenge*.
- Die Elemente  $\omega \in \Omega$  heißen *Ergebnisse*,
- Teilmengen  $A \subseteq \Omega$  heißen *Ereignisse*. Ein Ereignis  $A$  tritt ein gdw. ein Ergebnis  $\omega \in A$  eintritt.



---

## Ereignisse

---

- Ein *zusammengesetztes Ereignis*  $A \cup B$  tritt ein gdw. ein Ergebnis  $\omega$  mit  $\omega \in A$  oder  $\omega \in B$  eintritt.
- Analog tritt ein Ereignis  $A \cap B$  ein gdw. ein Ergebnis  $\omega$  mit  $\omega \in A$  und  $\omega \in B$  eintritt.
- Das Ereignis  $A^c = \Omega \setminus A$  ist das zu  $A$  *komplementäre Ereignis*.
- Zwei Ereignisse  $A, B$  heißen *unvereinbar* gdw.  $A \cap B = \emptyset$  (d.h. die Ereignisse sind disjunkt).
- Die leere Menge  $\emptyset$  heißt *unmögliches Ereignis* und die Menge  $\Omega$  *sicheres Ereignis*.
- Einelementige Mengen  $\{\omega\}$  heißen *Elementarereignisse*.
- Für Folgen  $A_1, A_2, \dots$  von Ereignissen wird das zusammengesetzte Ereignis  $\bigcup_{i=1}^{\infty} A_i$  definiert, das eintritt gdw. mindestens ein  $A_i$  eintritt. Analog für  $\bigcap_{i=1}^{\infty} A_i$  gdw. alle Ereignisse zugleich eintreten.

**Ereignissysteme** Ein System  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  heißt  $\sigma$  – *Algebra* oder *Ereignissystem* gdw. gilt:

1.  $\Omega \in \mathcal{A}$
2.  $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
3. Für jede Folge  $A_1, A_2, \dots \in \mathcal{A}$  gilt auch  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Aufgrund von 2 und 3 gilt auch:  $A \cap B = (A^c \cup B^c)^c \in \mathcal{A}$ .

- Eine  $\sigma$ -Algebra erlaubt genau die Verknüpfungen von Ereignissen, die in der Praxis nützlich sind.

**Ereignispartition** Mengen  $A_1, \dots, A_n$  werden *Ereignispartition* (oder *vollständige Ereignisdisjunktion*) genannt, wenn die Ereignisse paarweise unvereinbar sind und  $\bigcup_{i=1}^n A_i = \Omega$  gilt.

---

## 2.4.2 Wahrscheinlichkeit

---

---

### Formeln der Kombinatorik

---

Sei  $\Omega$  eine Ereignismenge mit  $n$  Elementen  $k \in \mathbb{N}$ .

**Geordnete Probe mit Wiederholungen** Ein  $k$ -Tupel  $x_1, \dots, x_k$  mit  $x_i \in \Omega$ ,  $i = 1, \dots, k$  heißt *geordnete Probe* von  $\Omega$  vom Umfang  $k$  mit *Wiederholungen*. Dann existieren

$$n^k$$

solcher Proben (für jede Stelle  $x_i$  gibt es  $n$  Möglichkeiten).

**Geordnete Probe ohne Wiederholungen** Ein  $k$ -Tupel  $x_1, \dots, x_k$ ,  $k \leq n$  mit  $x_i \in \Omega$ ,  $i = 1, \dots, k$  und  $x_i \neq x_j$  für  $i \neq j$  heißt *geordnete Probe* von  $\Omega$  vom Umfang  $k$  *ohne Wiederholungen*. Dann existieren

$$n(n-1)(n-2) \cdots (n-k+1)$$

solcher Proben (für die erste Stelle gibt es  $n$  Möglichkeiten, für die zweite  $n-1$ , usw.).

Gilt  $k = n$  wird von *Permutationen* der Menge  $\Omega$  gesprochen, wovon

$$n! = n(n-1)(n-2) \cdots 2 \cdots 1$$

existieren.

---

**Ungeordnete Probe mit Wiederholungen** Eine  $k$ -Sammlung  $x_1, \dots, x_k$ ,  $k \leq n$  mit  $x_i \in \Omega$ ,  $i = 1, \dots, k$  heißt *ungeordnete Probe* von  $\Omega$  vom Umfang  $k$  mit *Wiederholungen*. Dann existieren

$$\frac{(n+k-1)!}{(n-1)! \cdot k!} = \binom{n+k-1}{k}$$

solcher Proben.

**Ungeordnete Probe ohne Wiederholungen** Eine Teilmenge  $x_1, \dots, x_k$ ,  $k \leq n$  mit  $x_i \in \Omega$ ,  $i = 1, \dots, k$  heißt *ungeordnete Probe* von  $\Omega$  vom Umfang  $k$  ohne *Wiederholungen*. Dann existieren

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

solcher Proben (es gibt  $n(n-1)(n-2) \cdots (n-k+1)$  geordnete Proben, aber jeweils  $k!$  bestehen aus den gleichen  $k$  Elementen).

---

## Wahrscheinlichkeiten

Um jedem Ereignis eine Wahrscheinlichkeit zuzuordnen, wird eine Abbildung  $P : \mathcal{A} \rightarrow \mathbb{R}$  betrachtet. Diese Abbildung heißt *Wahrscheinlichkeitsmaß*, wenn sie den *Axiomen von Kolmogorov* genügt:

1.  $\forall A \in \mathcal{A} : P(A) \geq 0$
2.  $P(\Omega) = 1$
3.  $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$  mit paarweise unvereinbaren  $A_1, A_2, \dots \in \mathcal{A}$  (auch für endliche Folgen!)

## Rechenregeln

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- $0 \leq P(A) \leq 1$
- $A \subseteq B \implies P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Ist jedes Elementarereignis gleich Wahrscheinlich (wie z.B. bei einem Würfelwurf), so gilt für beliebige Ereignisse  $A \subseteq \Omega$ :

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \frac{|A|}{n}$$

---

## Bedingte Wahrscheinlichkeit

Seien  $A, B$  zwei Ereignisse mit  $P(A), P(B) > 0$ . In vielen Fällen ist interessant, was die Wahrscheinlichkeit von  $A$  ist unter der Bedingung, dass  $B$  eintritt.

Diese *bedingte Wahrscheinlichkeit* wird als  $P(A|B)$  formuliert (Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$ ) und ist gegeben durch:

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

---

**Ereignispartition** Für eine Ereignispartition  $A_1, \dots, A_n$  mit  $P(A_i) > 0$ ,  $i = 1, \dots, n$  und ein Ereignis  $B$  gilt:

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i)$$

**Formel von Bayes** Seien  $A_1, \dots, A_n$  eine Ereignispartition mit  $P(A_i) > 0$ ,  $i = 1, \dots, n$  und  $B$  ein Ereignis mit  $P(B) > 0$ . Dann gilt für  $i = 1, \dots, n$ :

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{P(B)}$$

**Multiplikationsformel** Seien  $A_1, \dots, A_n$  Ereignisse mit  $P(A_1 \cap A_2 \cap \dots \cap A_n) > 0$ . Dann gilt:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

---

## Unabhängigkeit

---

Zwei Ereignisse  $A, B$  heißen *paarweise unabhängig*, wenn gilt:

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

Mehrere Ereignisse  $A_1, \dots, A_n$  heißen *vollständig unabhängig*, wenn für alle  $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$  gilt:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

**Warning:** Aus der paarweisen Unabhängigkeit mehrerer Ereignisse folgt nicht immer die vollständige Unabhängigkeit!

---

## 2.5 Zufallsvariablen und Verteilungsfunktion

---

Sei  $\Omega$  eine Ereignismenge und  $\mathcal{A}$  ein Ereignissystem bzgl. Wahrscheinlichkeit  $P$ .

---

### 2.5.1 Zufallsvariablen

---

Eine *Zufallsvariable* ist eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}$$

mit der Eigenschaft, dass für jedes Intervall  $I \in \mathbb{R}$  die Urbildmenge

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

zum Ereignissystem  $\mathcal{A}$  gehört. Die Wahrscheinlichkeit, dass  $X$  Werte in diesem Intervall annimmt wird mit  $P(X \in I)$  bezeichnet, woraus sich folgende Schreibweisen ergeben:

$$P(a \leq X \leq b), \quad P(X \leq x), \quad P(X < x), \quad P(|X - a| < b), \quad P(X = b), \quad \text{usw.}$$

---

## Messreihen

---

Eine Messreihe  $x_1, \dots, x_n$  wird als Realisierung der Zufallsvariablen  $X_1, \dots, X_n$  angesehen. Es wird daher angenommen, dass ein Ergebnis  $\omega \in \Omega$  existiert mit:

$$x_1 = X_1(\omega), \quad \dots, \quad x_n = X_n(\omega)$$

---

### 2.5.2 Verteilungsfunktion

---

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable.

Die Abbildung  $F : \mathbb{R} \rightarrow \mathbb{R}$  wird dann *Verteilungsfunktion* von der Zufallsvariable  $X$  genannt:

$$F : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto P(X \leq x)$$

Dabei müssen Verteilungsfunktionen monoton wachsende Funktionen sein mit:

$$F(-\infty) = 0 \quad F(\infty) = 1 \quad F(x+) = F(x), \quad \forall x \in \mathbb{R}$$

Dabei sind die Schreibweisen wie folgt definiert:

$$\begin{aligned} F(x+) &:= \lim_{h \rightarrow 0} F(x+h) & F(x-) &:= \lim_{h \rightarrow 0} F(x-h) \\ F(-\infty) &:= \lim_{x \rightarrow -\infty} F(x) & F(\infty) &:= \lim_{x \rightarrow \infty} F(x) \end{aligned}$$

### Eigenschaften

$$\begin{aligned} P(X = a) &= P(X \leq a) - P(X < a) = F(a) - F(a-) \\ P(a < X \leq b) &= P(X \leq b) - P(X \leq a) = F(b) - F(a) \\ P(a \leq X < b) &= P(X < b) - P(X < a) = F(b-) - F(a-) \\ P(a \leq X \leq b) &= P(X \leq b) - P(X < a) = F(b) - F(a-) \\ P(X > a) &= 1 - P(X \leq a) = 1 - F(a) \end{aligned}$$

---

### Quantile

---

Ist die Verteilungsfunktion  $F$  stetig, so ist das  $p$ -Quantil  $x_p$  gegeben durch die Gleichung

$$F(x_p) = p$$

Die Quantile sind für gängige Verteilungsfunktionen somit tabellierbar und als Tabellen verfügbar.

---

### 2.5.3 Diskret/Stetig verteilte Zufallsvariablen

---

- Eine Zufallsvariable ist *diskret verteilt*, wenn sie nur endlich viele oder abzählbar unendliche viele Werte  $x_1, x_2, \dots$  annehmen kann. Die Verteilungsfunktion ist entsprechend eine monoton wachsende Treppenfunktion, die an den Stellen  $P(X = x_i)$  anspringt.
- Eine Zufallsvariable ist *stetig verteilt mit Dichte  $f$* , wenn die Verteilungsfunktion gegeben ist durch

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R}$$

Die Dichte ist dabei nichtnegativ, die Verteilungsfunktion  $F$  ist stetig und es gilt  $\frac{d}{dx} F = f$ .

---

## Beispiele für diskrete Verteilungen

---

**Geometrische Verteilung** Sei  $0 < p < 1$ .

Eine Zufallsvariable  $X$  mit Wertebereich  $\mathbb{N}^*$  heißt *geometrisch verteilt mit Parameter  $p$* , falls gilt

$$P(X = i) = (1 - p)^{i-1}p, \quad i = 1, 2, \dots$$

**Erwartungswert/Varianz:**

$$E(x) = \frac{1}{p}$$
$$\text{Var}(X) = \frac{1-p}{p^2}$$

**Anwendung** Zufallsexperimente mit Ereignis mit Wahrscheinlichkeit  $p$ . Die Anzahl unabhängiger Wiederholung bis zum Eintreten des Ereignisses kann als geometrisch verteilte Zufallsvariable modelliert werden („Warten auf den ersten Erfolg“).

**Binomialverteilung** Seien  $n \in \mathbb{N}$  und  $0 < p < 1$ .

Eine Zufallsvariable  $X$  mit Wertebereich  $\mathbb{N}_0$  heißt *binomialverteilt mit Parametern  $n, p$*  (kurz:  $B(n, p)$ -verteilt), falls gilt

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n$$

**Erwartungswert/Varianz:**

$$E(x) = np$$
$$\text{Var}(X) = np(1-p)$$

**Anwendung**  $n$ -mal unabhängig wiederholtes Zufallsexperiment mit Ereignis mit Wahrscheinlichkeit  $p$ . Die Anzahl des Ereignis-Eintretens kann als  $B(n, p)$ -verteilte Zufallsvariable modelliert werden („Anzahl der Erfolge bei  $n$  Versuchen“).

**Poissonverteilung** Sei  $\lambda > 0$ .

Eine Zufallsvariable  $X$  mit Wertebereich  $\mathbb{N}_0$  heißt *Poisson-verteilt mit Parameter  $\lambda$* , falls gilt

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, 2, \dots$$

**Erwartungswert/Varianz:**

$$E(X) = \lambda$$
$$\text{Var}(X) = \lambda$$

**Anwendung** Anzahl der in einer Telefonzentrale innerhalb von 10 Minuten eingehenden Anrufe.  $\lambda$  gibt die „mittlere Anzahl“ an eingehenden Anrufen an.

---

## Beispiele für stetige Verteilungen

---

### Rechteckverteilung

Sei  $a < b$ .

Eine stetig verteilte Zufallsvariable heißt *rechteckverteilt im Intervall  $[a, b]$*  (kurz:  $R(a, b)$ -verteilt), falls

$$f(t) = \begin{cases} \frac{1}{b-a} & a \leq t \leq b \\ 0 & \text{sonst} \end{cases}$$

gilt. Dann ergibt sich für die Verteilungsfunktion:

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

### Exponentialverteilung

Sei  $\lambda > 0$ .

Eine stetig verteilte Zufallsvariable heißt *exponentialverteilt mit Parameter  $\lambda$*  (kurz:  $\text{Ex}(\lambda)$ -verteilt), falls

$$f(t) = \begin{cases} 0 & t < 0 \\ \lambda e^{-\lambda t} & t \geq 0 \end{cases}$$

gilt. Dann ergibt sich für die Verteilungsfunktion:

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

**Erwartungswert/Varianz:**

$$E(x) = \frac{1}{\lambda}$$
$$\text{Var}(X) = \frac{1}{\lambda^2}$$

### Normalverteilung

Seien  $\mu \in \mathbb{R}$  und  $\sigma \in \mathbb{R}$ .

Eine stetig verteilte Zufallsvariable heißt *normalverteilt mit Parametern  $\mu, \sigma^2$*  (kurz:  $N(\mu, \sigma^2)$ ), falls

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

gilt. Mit  $\Phi$  aus der Standard-Normalverteilung (siehe 2.5.3) ergibt sich für die Verteilungsfunktion:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

**Erwartungswert/Varianz:**

$$E(x) = \mu$$
$$\text{Var}(X) = \sigma^2$$

**Standard-Normalverteilung** Ist  $\mu = 0$  und  $\sigma^2 = 1$ , wird ist *Standard-Normalverteilung* genannt und die Verteilungsfunktion mit

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

bezeichnet.

Da  $\Phi$  nicht geschlossen angebar ist, muss die Funktion tabelliert oder numerisch ausgewertet werden. Es gilt:

$$\Phi(0) = \frac{1}{2}, \quad \Phi(-x) = 1 - \Phi(x), \quad x \geq 0$$

**Chi-Quadrat-Verteilung** Sei  $r \in \{1, \dots, n\}$ .

Eine Zufallsvariable  $X$  heißt *Chi-Quadrat-verteilt mit Parameter  $r$*  (kurz:  $\chi_r^2$ -verteilt), falls

$$F(x) = P(Z_1^2 + \dots + Z_r^2 \leq x)$$

gilt. Die Dichte ist dabei

$$f(x) = \frac{x^{\frac{r}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{r}{2}} \Gamma(\frac{r}{2})}, \quad x > 0$$

mit der Gamma-Funktion.

**Studentische t-Verteilung** Sei  $r \in \{1, \dots, n-1\}$ .

Eine Zufallsvariable  $X$  heißt *Student-t-verteilt mit Parameter  $r$*  (kurz:  $t_r$ -verteilt), falls

$$F(x) = P\left(\frac{Z_{r+1}}{\sqrt{(Z_1^2 + \dots + Z_r^2)/r}} \leq x\right)$$

gilt. Die Dichte ist dabei

$$f(x) = \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r} \cdot \Gamma(\frac{r}{2})} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}}$$

mit der Gamma-Funktion.

**Fisher-Verteilung** Seien  $r, s \in \{1, \dots, n-1\}$  mit  $r + s \leq n$ .

Eine Zufallsvariable  $X$  heißt *Fisher-Verteilt mit Parametern  $r, s$*  (kurz:  $F_{r,s}$ -verteilt), falls

$$F(x) = P\left(\frac{(Z_1^2 + \dots + Z_r^2)/r}{(Z_{r+1}^2 + \dots + Z_{r+s}^2)/s} \leq x\right) = \frac{\chi_r^2/r}{\chi_s^2/s}$$

gilt. Die Dichte ist dabei

$$f(x) = m^{\frac{r}{2}} n^{\frac{s}{2}} \cdot \frac{\Gamma(\frac{r}{2} + \frac{s}{2})}{\Gamma(\frac{r}{2}) \cdot \Gamma(\frac{s}{2})} \cdot \frac{x^{\frac{r}{2}-1}}{(rx + s)^{\frac{r+s}{2}}}, \quad x \geq 0$$

mit der Gamma-Funktion.

---

## 2.5.4 Erwartungswert und Varianz

---

### Erwartungswert

---

Sei  $h : \mathbb{R} \rightarrow \mathbb{R}$  eine stückweise stetige Funktion.

Der *Erwartungswert* einer *diskret verteilten* Zufallsvariable  $X$  und einer Funktion  $h(X)$  mit den Werten  $x_1, x_2, \dots$  ist

$$E(X) = \sum_i x_i P(X = x_i) \qquad E(h(X)) = \sum_i h(x_i) P(X = x_i)$$

sofern  $\sum_i |x_i| P(X = x_i)$  konvergiert.

Der *Erwartungswert* einer *stetig verteilten* Zufallsvariable  $X$  und einer Funktion  $h(X)$  mit Dichte  $f$  ist

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \qquad E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx$$

sofern  $\int_{-\infty}^{\infty} |x| f(x) dx$  konvergiert.

**Rechenregeln** Seien  $X, X_1, \dots, X_n$  Zufallsvariablen,  $b, a, a_1, \dots, a_n \in \mathbb{R}$  und  $h_1, h_2 : \mathbb{R} \rightarrow \mathbb{R}$  stückweise stetig. Dann gelten folgende (teilweise redundante) Rechenregeln für den Erwartungswert:

$$\begin{aligned} E(aX + b) &= aE(X) + b \\ E(h_1(X) + h_2(X)) &= E(h_1(X)) + E(h_2(X)) \\ E(a_1X_1 + \dots + a_nX_n + b) &= a_1E(X_1) + \dots + a_nE(X_n) + b \end{aligned}$$

---

### Varianz

---

Die *Varianz* einer Zufallsvariable  $X$  ist der Erwartungswert der quadratischen Abweichung von  $X$  zu ihrem Erwartungswert:

$$\text{Var}(X) = E((X - E(X))^2)$$

Die *Standardabweichung* ist dann definiert durch:  $\sqrt{\text{Var}(X)}$

**Rechenregeln** Sei  $X$  eine Zufallsvariable,  $a, b \in \mathbb{R}$ . Dann gelten folgende Rechenregeln für die Varianz:

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ \text{Var}(aX + b) &= a^2 \text{Var}(X) \end{aligned}$$

Sind Zufallsvariablen  $X_1, \dots, X_n$  unabhängig, dann gilt zusätzlich:

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

---

## 2.5.5 Tschebyscheffsche Ungleichung

---

Sei  $X$  eine Zufallsvariable. Dann gilt nach der *tschebyscheffschen Ungleichung*:

$$P(|X - E(X)| \geq c) \leq \frac{\text{Var}(X)}{c^2}, \quad c > 0$$



---

## 2.5.6 Unabhängigkeit

---

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit den Verteilungsfunktionen  $F_1, \dots, F_n$ . Die *gemeinsame Verteilungsfunktion* ist gegeben durch:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n), \quad (x_1, \dots, x_n)^T \in \mathbb{R}^n$$

Die Zufallsvariablen heißen *unabhängig*, wenn für alle  $(x_1, \dots, x_n)^T \in \mathbb{R}^n$  die Ereignisse

$$\{X_1 \leq x_1\}, \dots, \{X_n \leq x_n\}$$

vollständig unabhängig sind, d.h.

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

oder kurz

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n)$$

---

## 2.6 Einige Sätze

---

---

### 2.6.1 Das schwache Gesetz der großen Zahlen

---

Ist  $X_1, X_2, \dots$  eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit  $\mu = E(X_i)$ ,  $\sigma^2 = \text{Var}(X_i)$ , dann gilt:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) = 0, \quad \forall \varepsilon > 0$$

---

### 2.6.2 Zentraler Grenzwertsatz

---

Sei  $X_1, X_2, \dots$  eine Folge von identisch verteilten unabhängigen Zufallsvariablen mit  $\mu_i = E(X_i)$ ,  $\sigma_i^2 = \text{Var}(X_i)$ ,  $i = 1, 2, \dots$ . Dann gilt für alle  $y \in \mathbb{R}$ :

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - (\mu_1 + \dots + \mu_n)}{\sqrt{\sigma_1^2 + \dots + \sigma_n^2}} \leq y\right) = \Phi(y)$$

Das arithmetische Mittel  $\bar{X}_{(n)} = \frac{1}{n}(X_1 + \dots + X_n)$  ist also für große  $n$  annähernd  $N(\mu, \sigma^2)$  verteilt mit

$$\mu = \frac{1}{n}E(X_1 + \dots + X_n) = \frac{1}{n}(\mu_1 + \dots + \mu_n) \quad \sigma^2 = \frac{1}{n^2}\text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2}(\sigma_1^2 + \dots + \sigma_n^2)$$

Anmerkung: Hat  $X$  den Erwartungswert  $\mu$  und die Varianz  $\sigma^2$ , dann hat  $\frac{X-\mu}{\sigma}$  den Erwartungswert 0 und die Varianz 1.

---

### 2.6.3 Zentralsatz der Statistik

---

$$F_n(z; x_1, \dots, x_n) = \frac{|\{x_i \mid x_i \leq z\}|}{n}$$

Sei  $X_1, X_2, \dots$  eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit der Verteilungsfunktion  $F$  und sei

$$D_n(X_1, \dots, X_n) = \sup_{z \in \mathbb{R}} |F_n(z; X_1, \dots, X_n) - F(z)|$$

die zufällige Maximalabweichung zwischen empirischer und „wahrer“ Verteilungsfunktion. Dann gilt

$$P\left(\lim_{n \rightarrow \infty} D_n(X_1, \dots, X_n) = 0\right) = 1$$

Die zufällige Maximalabweichung konvergiert also mit einer Wahrscheinlichkeit 1 gegen 0.

---

### 2.6.4 Anwendungen

---

Seien  $X_1, \dots, X_n$  unabhängige identisch  $N(\mu, \sigma^2)$ -verteilte Zufallsvariablen mit arithmetischem Mittel und Stichprobenvarianz:

$$\bar{X}_{(n)} = \frac{1}{n} \sum_{i=1}^n X_i \qquad S_{(n)}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_{(n)})^2$$

Dann gilt:

- $\bar{X}_{(n)}$  ist  $N\left(\mu, \frac{\sigma^2}{n}\right)$ -verteilt
- $\frac{n-1}{\sigma^2} S_{(n)}^2$  ist  $\chi_{n-1}^2$ -verteilt
- $\bar{X}_{(n)}$  und  $S_{(n)}^2$  sind unabhängig
- $\sqrt{n} \frac{\bar{X}_{(n)} - \mu}{\sqrt{S_{(n)}^2}}$  ist  $t_{n-1}$ -verteilt

---

## 3 Schätzverfahren und Konfidenzintervalle

---

### 3.1 Grundlagen

---

- Sei im folgenden die Messreihe  $x_1, \dots, x_n$  die Realisierung von unabhängigen identisch wie  $X$  verteilten Zufallsvariablen  $X_1, \dots, X_n$ .
- Außerdem wird angenommen, dass die Verteilungsfunktion  $F$  von  $X$  und aller  $X_i$  einer durch  $k$  Parameter  $\theta \in \Theta \subset \mathbb{R}^k$  parametrisierte Familie  $F_\theta$ ,  $\theta \in \Theta$  von Verteilungsfunktionen angehört.
- Der Parameter oder ein dadurch bestimmter Zahlenwert  $\tau : \Theta \rightarrow \mathbb{R}$  sei unbekannt und soll geschätzt werden.
- Ein *Schätzverfahren* ist wie folgt definiert:  
Ein *Schätzverfahren* (oder eine *Schätzfunktion* oder ein *Schätzer*) ist eine Abbildung

$$T_n : \mathbb{R}^n \rightarrow \mathbb{R}$$

die einer Messreihe  $x_1, \dots, x_n$  einen Schätzwert  $T_n(x_1, \dots, x_n)$  für den Wert  $\tau(\theta)$  zuordnet. Die Zufallsvariable  $T_n(X_1, \dots, X_n)$  wird *Schätzvariable* genannt.

- Der Erwartungswert und die Varianz der Schätzvariablen und allen  $X_i$  hängen von der Verteilungsfunktion  $F_\theta$  ab. Zur Verdeutlichung dieses Umstandes wird der Parameter  $\theta$  an sämtliche Funktionen geschrieben:

$$\begin{array}{lll} E_\theta(T_n(X_1, \dots, X_n)), & E_\theta(X_1), & \dots \\ \text{Var}_\Theta(T_n(X_1, \dots, X_n)), & \text{Var}_\theta(X_i), & \dots \\ P_\theta(a \leq T_n(X_1, \dots, X_n) \leq b), & P_\theta(a \leq X_1 \leq b), & \dots \end{array}$$

---

#### 3.1.1 (Asymptotische) Erwartungstreue

---

Ein Schätzer  $T_n : \mathbb{R}^n \rightarrow \mathbb{R}$  heißt *erwartungstreu* für  $\tau : \Theta \rightarrow \mathbb{R}$ , wenn gilt:

$$E_\theta(T_n(X_1, \dots, X_n)) = \tau(\theta) \quad \text{für alle } \theta \in \Theta$$

Eine Folge von Schätzern  $T_n : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $n = 1, 2, \dots$  heißt *asymptotisch erwartungstreu* für  $\tau : \Theta \rightarrow \mathbb{R}$ , wenn gilt:

$$\lim_{n \rightarrow \infty} E_\theta(T_n(X_1, \dots, X_n)) = \tau(\theta) \quad \text{für alle } \theta \in \Theta$$

Das heißt, ein Schätzer liefert bei genügender Anzahl an Stichproben ein erwartungstreu Ergebnis.

---

### 3.1.2 Mittlerer quadratischer Fehler

---

Um die Güte eines Schätzers zu beurteilen dient der *Mittlere quadratische Fehler* (Mean squared error, MSE):

$$\text{MSE}_\theta(T) := E_\theta((T - \tau(\theta))^2)$$

Es gilt  $T$  erwartungstreu  $\implies \text{MSE}_\theta(T) = \text{Var}_\theta(T)$ .

Seien  $T_1$  und  $T_2$  Schätzer für  $\tau$ , dann heißt  $T_1$  *effizienter* als  $T_2$ , wenn gilt

$$\text{MSE}_\theta(T_1) \leq \text{MSE}_\theta(T_2) \quad \forall \theta \in \Theta$$

Sind  $T_1, T_2$  erwartungstreu, dann heißt dies

$$\text{Var}_\theta(T_1) \leq \text{Var}_\theta(T_2) \quad \forall \theta \in \Theta$$

---

### 3.1.3 Konsistenz

---

Eine Folge an Schätzern  $T_1, T_2, \dots$  heißt *konsistent* für  $\tau$ , wenn für alle  $\varepsilon > 0$  und alle  $\theta \in \Theta$  gilt

$$\lim_{n \rightarrow \infty} P_\theta(|T_n(X_1, \dots, X_n) - \tau(\theta)| > \varepsilon) = 0$$

Die Folge heißt *konsistent im quadratischen Mittel* für  $\tau$ , wenn für alle  $\theta \in \Theta$  gilt

$$\lim_{n \rightarrow \infty} \text{MSE}_\theta(T_n) = 0$$

**Sätze** Ist  $T_1, T_2, \dots$  eine Folge von Schätzern, die für  $\tau$  erwartungstreu sind und gilt

$$\lim_{n \rightarrow \infty} \text{Var}_\theta(T_n(X_1, \dots, X_n)) = 0 \quad \forall \theta \in \Theta$$

dann ist die Folge von Schätzern konsistent für  $\tau$ .

Allgemeiner gilt: Ist  $T_1, T_2, \dots$  eine Folge von Schätzern, die konsistent im quadratischen Mittel für  $\tau$  sind, dann ist die Folge konsistent für  $\tau$ .

---

## 3.2 Maximum-Likelihood-Schätzer

---

- Bei gegebener Verteilungsklasse  $F_\theta$ ,  $\theta \in \Theta$  lassen sich Schätzer für den Parameter  $\theta$  oftmals mit der Maximum-Likelihood-Methode gewinnen.
- Sind die Zufallsvariablen  $X_1, \dots, X_n$  stetig mit einer Dichte verteilt, hängt die Dichte ebenfalls von den Parametern ab ( $f_\theta$ ).
- Für diskrete Zufallsvariablen sei  $f_\theta(x) = P_\theta(X = x)$  für alle  $x$  aus dem Wertebereich. Sei dieser Wertebereich  $\mathbb{X}$ .

Für eine Messreihe  $x_1, \dots, x_n$  heißt die Funktion  $L(\cdot; x_1, \dots, x_n)$  mit

$$L(\theta; x_1, \dots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n)$$

die zu der Messreihe gehörige *Likelihood-Funktion*.

Eine Parameterschätzung  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$  mit

$$L(\hat{\theta}; x_1, \dots, x_n) \geq L(\theta; x_1, \dots, x_n) \quad \forall \theta \in \Theta$$

heißt *Maximum-Likelihood-Schätzwert*. Existiert ein solcher Schätzwert für jede mögliche Messreihe, dann ist

$$T_n : \mathbb{X}^n \rightarrow \Theta : (x_1, \dots, x_n) \mapsto \hat{\theta}(x_1, \dots, x_n)$$

ein *Maximum-Likelihood-Schätzer*.

---

### 3.3 Konfidenzintervalle

---

- Wie beim Schätzen wird eine Messreihe  $x_1, \dots, x_n$  beobachtet und es sollen Ober- und Unterschranken für  $\tau(\theta)$  ermittelt werden.
- Dabei wird durch ein Paar  $U : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $O : \mathbb{R}^n \rightarrow \mathbb{R}$  von Schätzern mit  $U(x_1, \dots, x_n) \leq O(x_1, \dots, x_n)$  ein „zufälliges“ Intervall  $I(X_1, \dots, X_n) = [U(X_1, \dots, X_n), O(X_1, \dots, X_n)]$  definiert.
- Dieses Intervall heißt *Konfidenzintervall* für  $\tau(\theta)$  zum *Konfidenzniveau*  $1 - \alpha$ , falls gilt:

$$P_\theta(U(X_1, \dots, X_n) \leq \tau(\theta) \leq O(X_1, \dots, X_n)) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

- Gehört das Intervall zu einer bestimmten Messreihe, heißt es *konkretes Schätzintervall* für  $\tau(\theta)$ .
- Dann enthält ein konkretes Schätzintervall den Wert  $\tau(\theta)$  mit einer Wahrscheinlichkeit von  $1 - \alpha$ .

---

#### 3.3.1 Konstruktion für normalverteilte Zufallsvariablen

---

Seien  $X_1, \dots, X_n$  unabhängig und identisch normalverteilte Zufallsvariablen. Dann ist die Verteilungsfunktion  $F_\theta$  durch den Parameter  $\theta = (\mu, \sigma^2)$  bestimmt:

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

Das Konfidenzniveau ist dabei immer  $1 - \alpha$ .

**Für  $\mu$  bei bekannter Varianz** Es ist  $\Theta = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}\}$  und  $\tau(\theta) = \mu$ .

Das Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$  lautet dann

$$I(X_1, \dots, X_n) = \left[ \bar{X}_{(n)} - u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \quad \bar{X}_{(n)} + u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \right]$$

mit dem  $1 - \frac{\alpha}{2}$ -Quantil  $u_{1-\frac{\alpha}{2}}$  der  $N(0, 1)$ -Verteilung, also

$$\Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

---

**Für  $\mu$  bei unbekannter Varianz** Es ist  $\Theta = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$  und  $\tau(\theta) = \mu$ .

Das Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$  lautet dann

$$I(X_1, \dots, X_n) = \left[ \bar{X}_{(n)} - t_{n-1; 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_{(n)}^2}{n}}, \quad \bar{X}_{(n)} + t_{n-1; 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_{(n)}^2}{n}} \right]$$

mit dem  $1 - \frac{\alpha}{2}$ -Quantil  $t_{n-1; 1-\frac{\alpha}{2}}$  der  $t_{n-1}$ -Verteilung.

**Für  $\sigma^2$  bei bekanntem Erwartungswert** Es ist  $\Theta = \{(\mu_0, \sigma^2) \mid \sigma^2 > 0\}$  und  $\tau(\theta) = \sigma^2$ .

Das Konfidenzintervall für  $\sigma^2$  zum Niveau  $1 - \alpha$  lautet dann

$$I(X_1, \dots, X_n) = \left[ \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n; 1-\frac{\alpha}{2}}^2}, \quad \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n; \frac{\alpha}{2}}^2} \right]$$

mit dem  $1 - \frac{\alpha}{2}$ -Quantil  $\chi_{n; 1-\frac{\alpha}{2}}^2$  und dem  $\frac{\alpha}{2}$ -Quantil  $\chi_{n; \frac{\alpha}{2}}^2$  der  $\chi_n^2$ -Verteilung.

**Für  $\sigma^2$  bei unbekanntem Erwartungswert** Es ist  $\Theta = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$  und  $\tau(\theta) = \sigma^2$ .

Das Konfidenzintervall für  $\sigma^2$  zum Niveau  $1 - \alpha$  lautet dann

$$I(X_1, \dots, X_n) = \left[ \frac{(n-1) S_{(n)}^2}{\chi_{n-1; 1-\frac{\alpha}{2}}^2}, \quad \frac{(n-1) S_{(n)}^2}{\chi_{n-1; \frac{\alpha}{2}}^2} \right]$$

mit dem  $1 - \frac{\alpha}{2}$ -Quantil  $\chi_{n-1; 1-\frac{\alpha}{2}}^2$  und dem  $\frac{\alpha}{2}$ -Quantil  $\chi_{n-1; \frac{\alpha}{2}}^2$  der  $\chi_{n-1}^2$ -Verteilung.

---

## 4 Testverfahren

---

**Warning:** Sämtliche Tests, die auf einer approximierten Verteilung und empirischen Daten basieren, sind nur für große Anzahl an Werten (großes  $n$ ) anwendbar!

---

### 4.1 Grundlagen

---

- Die *Nullhypothese*  $H_0$  ist die zu prüfende Annahme.
- Ein Verfahren zur Prüfung, ob eine Messreihe der Nullhypothese genügt, wird *Test* genannt.
- Die Tests sind dabei durch Angabe von *kritischen Bereichen*  $K \subset \mathbb{R}^n$  vollständig beschrieben.

$$\begin{cases} \text{Lehne } H_0 \text{ ab} & \text{falls } (x_1, \dots, x_n) \in K \\ \text{Akzeptierte } H_0 & \text{sonst} \end{cases}$$

- Die wichtigen Fehlermöglichkeiten eines solchen Tests sind:
  - Fehler 1. Art (False Negative)**  $H_0$  wird abgelehnt, obwohl  $H_0$  zutrifft.
  - Fehler 2. Art (False Positive)**  $H_0$  wird akzeptiert, obwohl  $H_0$  nicht zutrifft.
- Ziel
  - Wahrscheinlichkeit für den Fehler 1. Art klein.
  - Dazu wird ein Testniveau  $\alpha$  vorgegeben.
  - Dann muss gelten:

$$\text{Unter Nullhypothese gilt } P((X_1, \dots, X_n) \in K) \leq \alpha \iff P((K_1, \dots, K_n) \in K \mid H_0) \leq \alpha$$

---

#### 4.1.1 Testgröße

---

Der kritische Bereich wird durch eine passende Funktion, genannt *Testgröße*:

$$T : \mathbb{R}^n \rightarrow \mathbb{R}$$

und kritische Schranken  $c$  bzw.  $c_1, c_2$  beschrieben.

Damit können bspw. folgende Möglichkeiten an Tests spezifiziert werden, wobei diese immer die allgemeine Form

$$K = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \psi(T(x_1, \dots, x_n)) \right\}$$

für die Definition der kritischen Bereiche annehmen mit unterschiedlichen Prädikaten  $\psi(t)$ :

- $\psi(t) := (|t| > c)$       Betragsmäßig große Werte sprechen gegen  $H_0$ .
- $\psi(t) := (t < c_1 \vee t > c_2)$       Betragsmäßig kleine Werte sprechen gegen  $H_0$ .
- $\psi(t) := (t > c)$       Große Werte sprechen gegen  $H_0$ .
- $\psi(t) := (t < c)$       Kleine Werte sprechen gegen  $H_0$ .

---

#### 4.1.2 Allgemeines Konstruktionsprinzip zum Niveau $\alpha$

---

1. Verteilungsannahmen formulieren.
2. Nullhypothese  $H_0$  formulieren.
3. Testgröße  $T$  wählen und die Verteilung dieser unter  $H_0$  bestimmen.
4.  $I \subseteq \mathbb{R}$  so wählen, dass unter  $H_0$  gilt  $P(T(X_1, \dots, X_N) \in I) \leq \alpha$ .

Dabei wird  $I$  durch die kritischen Schranken festgelegt und ist bspw. von der Form:

$$I = \mathbb{R} \setminus [-c, c] \quad I = \mathbb{R} \setminus [c_1, c_2] \quad I = (c, \infty) \quad I = (-\infty, c)$$

Für das Niveau  $\alpha$  wird oft 0.1, 0.05 oder 0.01 gewählt.

---

## 4.2 Wichtige Tests bei Normalverteilungsannahme

---

**Für  $\mu_0$  bei bekannter Varianz (Gauß-Test)** Seien  $X_1, \dots, X_n$  unabhängig und identisch  $N(\mu, \sigma_0^2)$ -verteilt, sei  $\sigma_0^2$  bekannt und  $\mu_0$  zu testen.

1. Wählen der Nullhypothese:

- A)  $H_0: \mu = \mu_0$
- B)  $H_0: \mu \leq \mu_0$
- C)  $H_0: \mu \geq \mu_0$

2. Berechnen der Testgröße:

$$T(X_1, \dots, X_n) = \frac{\sqrt{n}}{\sigma_0} (\bar{X}_{(n)} - \mu_0)$$

3. Ablehnung von  $H_0$ , falls:

- A)  $|T| > u_{1-\frac{\alpha}{2}}$
- B)  $T > u_{1-\alpha}$
- C)  $T < u_\alpha$



---

**Für  $\mu_0$  bei unbekannter Varianz (t-Test)** Seien  $X_1, \dots, X_n$  unabhängig und identisch  $N(\mu, \sigma^2)$ -verteilt und  $\mu_0$  zu testen.

1. Wählen der Nullhypothese:

- A)  $H_0: \mu = \mu_0$
- B)  $H_0: \mu \leq \mu_0$
- C)  $H_0: \mu \geq \mu_0$

2. Berechnen der Testgröße:

$$T(X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X}_{(n)} - \mu_0}{\sqrt{S_{(n)}^2}}$$

3. Ablehnung von  $H_0$ , falls:

- A)  $|T| > t_{n-1; 1-\frac{\alpha}{2}}$
- B)  $T > t_{n-1; 1-\alpha}$
- C)  $T < t_{n-1; \alpha}$

**Für  $\sigma^2$  bei bekanntem Erwartungswert** Seien  $X_1, \dots, X_n$  unabhängig und identisch  $N(\mu, \sigma^2)$ -verteilt, sei  $\mu$  bekannt und  $\sigma_0^2$  zu testen.

1. Wählen der Nullhypothese:

- A)  $H_0: \sigma^2 = \sigma_0^2$
- B)  $H_0: \sigma^2 \leq \sigma_0^2$
- C)  $H_0: \sigma^2 \geq \sigma_0^2$

2. Berechnen der Testgröße:

$$T(X_1, \dots, X_n) = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2$$

3. Ablehnung von  $H_0$ , falls:

- A)  $T < \chi_{n; \frac{\alpha}{2}}^2$  oder  $T > \chi_{n; 1-\frac{\alpha}{2}}^2$
- B)  $T > \chi_{n; 1-\alpha}^2$
- C)  $T < \chi_{n; \alpha}^2$

**Für  $\sigma_0^2$  bei unbekanntem Erwartungswert ( $\chi^2$ -Test)** Seien  $X_1, \dots, X_n$  unabhängig und identisch  $N(\mu, \sigma^2)$ -verteilt und  $\sigma_0^2$  zu testen.

1. Wählen der Nullhypothese:

- A)  $H_0: \sigma^2 = \sigma_0^2$
- B)  $H_0: \sigma^2 \leq \sigma_0^2$
- C)  $H_0: \sigma^2 \geq \sigma_0^2$

2. Berechnen der Testgröße:

$$T(X_1, \dots, X_n) = \frac{n-1}{\sigma_0^2} S_{(n)}^2$$

3. Ablehnung von  $H_0$ , falls:

- A)  $T < \chi_{n; \frac{\alpha}{2}}^2$  oder  $T > \chi_{n; 1-\frac{\alpha}{2}}^2$
- B)  $T > \chi_{n; 1-\alpha}^2$
- C)  $T < \chi_{n; \alpha}^2$

---

## 4.3 Verteilungstests

---

**$\chi^2$ -Anpassungstest** Der  $\chi^2$ -Anpassungstest dient zur Prüfung, ob die empirische Verteilung einer Zufallsvariable einer erwarteten Verteilung entspricht.

Seien  $X_1, \dots, X_n$  unabhängig und identisch verteilt mit unbekannter Verteilung  $F$  und sei  $x_1, \dots, x_n$  eine realisierende Messreihe.

1.  $H_0$ :  $F = F_0 \iff X_i \sim F_0$
2. Partitionierung von  $R$  in  $k$  Intervalle ( $z_1 < z_2 < \dots < z_{k-1}$ ):

$$A_1 = (-\infty, z_1], \quad A_2 = (z_1, z_2], \quad \dots, \quad A_k = (z_{k-1}, \infty)$$

3. Bestimmen der Häufigkeiten:

$$h_j = |\{i \mid x_i \in A_j\}|, \quad j = 1, \dots, k$$

4. Unter  $H_0$  gilt (mit  $z_0 := -\infty$  und  $z_k := \infty$ ):

$$p_j := P(X \in A_j) \stackrel{H_0}{=} F_0(z_j) - F_0(z_{j-1}) \approx \frac{h_j}{n}, \quad j = 1, \dots, k$$

5. Berechnen der Testgröße:

$$T(X_1, \dots, X_n) = \sum_{j=1}^k \frac{(H_j - np_j)^2}{np_j}$$

6. Ablehnung von  $H_0$ , falls:

$$T > \chi_{k-1; 1-\alpha}^2$$

**$\chi^2$ -Test auf Unabhängigkeit (Kontingenztest)** Der  $\chi^2$ -Kontingenztest dient zur Prüfung, ob Zufallsvariablen unabhängig sind.

Seien  $(X_1, Y_1), \dots, (X_n, Y_n)$  unabhängig und identisch wie  $(X, Y)$  verteilt und sei  $(x_1, y_1), \dots, (x_n, y_n)$  eine realisierende Messreihe.

1.  $H_0$ :  $X$  und  $Y$  sind unabhängig
2. Partitionierung in  $k$  bzw.  $l$  Intervalle ( $z_1 < z_2 < \dots < z_{k-1}$ ,  $\tilde{z}_1 < \tilde{z}_2 < \dots < \tilde{z}_{l-1}$ ):
  - x-Achse:  $A_1 = (-\infty, z_1], \quad A_2 = (z_1, z_2], \quad \dots, \quad A_k = (z_{k-1}, \infty)$

- y-Achse:  $B_1 = (-\infty, \tilde{z}_1]$ ,  $B_2 = (\tilde{z}_1, \tilde{z}_2]$ ,  $\dots$ ,  $B_l = (\tilde{z}_{l-1}, \infty)$

3. Bestimmen der Häufigkeiten:

$$h_{ij} = \left| \left\{ r \in \{1, \dots, n\} \mid (x_r, y_r) \in A_i \times B_j \right\} \right|, \quad i = 1, \dots, k, j = 1, \dots, l$$

und der Randhäufigkeiten:

$$h_{i.} = h_{i1} + \dots + h_{il}, \quad i = 1, \dots, k \quad h_{.j} = h_{1j} + \dots + h_{kj}, \quad j = 1, \dots, l$$

4. Seien unter  $H_0$ :

$$\frac{h_{i.}}{n} \approx P(X \in A_i) \quad \frac{h_{.j}}{n} \approx P(Y \in B_j)$$

$$\frac{\tilde{h}_{ij}}{n} := \frac{h_{i.} h_{.j}}{n^2} \approx P(X \in A_i) \cdot P(Y \in B_j) \stackrel{H_0}{=} P(X \in A_i, Y \in B_j) \approx \frac{h_{ij}}{n}$$

5. Berechnen der Testgröße (mit Zufallsvariablen  $H_{ij}$  und  $\tilde{H}_{ij}$  für  $h_{ij}$  und  $\tilde{h}_{ij}$ ):

$$T(X_1, \dots, X_n) = \sum_{i=1}^k \sum_{j=1}^l \frac{(H_{ij} - \tilde{H}_{ij})^2}{\tilde{H}_{ij}}$$

6. Ablehnung von  $H_0$ , falls:

$$T > \chi_{(k-1)(l-1); 1-\alpha}^2$$

**$\chi^2$ -Homogenitätstest** Der  $\chi^2$ -Homogenitätstest dient zur Prüfung, ob die mehrere Zufallsvariablen identisch verteilt sind.

Seien  $X_1^{(i)}, \dots, X_{n_i}^{(i)}$  unabhängig und identisch verteilt mit einer Verteilungsfunktion  $F_i$  für alle  $i = 1, \dots, k$ .

1.  $H_0$ :  $F_1 = F_2 = \dots = F_k$

2. Partitionierung in  $m$  Intervalle ( $z_1 < z_2 < \dots < z_{m-1}$ ):

$$A_1 = (-\infty, z_1], \quad A_2 = (z_1, z_2], \quad \dots, \quad A_m = (z_{m-1}, \infty)$$

3. Bestimmen der Häufigkeiten:

$$H_{ij} = \left| \{X_j^{(i)} \mid X_j^{(i)} \in A_j\} \right|, \quad i = 1, \dots, k, j = 1, \dots, m$$

und der summierten Häufigkeiten:

$$H_{.j} = H_{1j} + \dots + H_{kj}, \quad j = 1, \dots, m$$

4. Unter  $H_0$  gilt daher für die relativen Häufigkeiten (mit  $n = n_1 + \dots + n_k$ ):

$$\frac{h_{ij}}{n_i} \approx \frac{h_{.j}}{n} \iff h_{ij} - \frac{n_i h_{.j}}{n} \approx 0$$

5. Berechnen der Testgröße (mit Zufallsvariablen  $X_{ij}$  und  $H_{.j}$  für  $h_{ij}$  und  $h_{.j}$ ):

$$T(X_1^{(1)}, \dots, X_{n_k}^{(k)}) = \sum_{i=1}^k \sum_{j=1}^m \frac{\left( H_{ij} - \frac{n_i H_{.j}}{n} \right)^2}{\frac{n_i H_{.j}}{n}}$$

6. Ablehnung von  $H_0$ , falls:

$$T > \chi_{(k-1)(m-1); 1-\alpha}^2$$

---

**Wilcoxon Vorzeichen-Rang-Test** Der *Wilcoxon Vorzeichen-Rang-Test* dient zur Prüfung, ob zwei Algorithmen in der gleichen Zeit laufen, d.h. gleich schnell sind.

Seien  $X, Y$  Zufallsvariablen, die die Laufzeit der beiden Algorithmen angeben und seien  $x_1, \dots, x_n, y_1, \dots, y_n$  zwei realisierende Messreihen.

1.  $H_0$ : Beide Algorithmen sind gleich schnell.
2. Berechne  $q_1 = \frac{x_1}{y_1}, \dots, q_n = \frac{x_n}{y_n}$  und entferne alle Quotienten nahe 1 und sei  $N$  die Anzahl noch verbleibender Messwerte.
3. Ersetze alle  $0 \leq q_i < 1$  durch  $-\frac{1}{q_i}$  ( $i = 1, \dots, N$ ) und erhalte das Ergebnis  $\tilde{q}_1, \dots, \tilde{q}_N$ .
4. Sortiere  $|\tilde{q}_1| < |\tilde{q}_2| < \dots < |\tilde{q}_N|$ .
5. Berechne und setze

$$G_1 := \{i \mid \tilde{q}_i < -1\} \quad (\text{Algorithmus 1 schneller})$$

$$G_2 := \{i \mid \tilde{q}_i > 1\} \quad (\text{Algorithmus 2 schneller})$$

$$r_1 := \sum_{i \in G_1} i$$

$$r_2 := \sum_{i \in G_2} i$$

6. Berechnen der Testgröße:

$$T = \frac{\min\{r_1, r_2\} - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$$

7. Ablehnung von  $H_0$ , falls:

$$|T| > u_{1-\frac{\alpha}{2}}$$

---

## 4.4 Weitere statische Tests

---

**Parametrische Tests** Verteilung bekannt, Parameter unbekannt  
ANOVA (mehrere Stichproben, normalverteilt, gleiche Varianz), ...

**Nichtparametrische Tests** Verteilung soll getestet werden  
Kolmogorow-Smirnow-Test (Verteilungstyp), Wilcoxon-Mann-Whitney-Test (Lage von Stichproben), Kruskal-Wallis-Tests ( $\geq 3$  Gruppen von Stichproben), ...

**Unabhängigkeitstests** McNemar-Test (zwei abhängige Stichproben), ...

**Tests zu Regressionsmethoden**  $t$ -Test: Regressionskoeffizient, ...  
...

---

## 5 Robuste Statistik

---

Ausreißer innerhalb einer Messreihe können die geschätzten statistischen Parameter stark verfälschen. Diesem Phänomen soll die robuste Statistik mit bestimmten Methodiken entgegenwirken.

---

### 5.1 Median

---

Sei  $X$  eine Zufallsvariable. Dann ist jede Zahl  $\mu_m$  ein *robuster Median* mit:

$$P(X \leq \mu_m) \geq \frac{1}{2} \quad \text{und} \quad P(X \geq \mu_m) \geq \frac{1}{2}$$

mit der Verteilungsfunktion  $F$  von  $X$  gilt gleichbedeutend:

$$F(\mu_m) \geq \frac{1}{2} \quad \text{und} \quad F(\mu_m-) \leq \frac{1}{2}$$

#### Eigenschaften

- Der Median ist so nur eindeutig, wenn  $F(x) = \frac{1}{2}$  genau eine Lösung besitzt.
- Wenn der Median eindeutig ist und die Verteilung  $F$  symmetrisch ist (d.h. es gilt  $\forall x \in \mathbb{R} : F(\mu_m + x) = q - F(\mu_m - x)$ ), dann gleicht der Median dem Erwartungswert.

---

#### 5.1.1 Schätzer

---

Sei  $T(x_1, \dots, x_n) := \tilde{x}$  der empirische Median, dann kann ein Schätzer  $\tilde{X}_{(n)}$  für  $\mu_m$  mit diesem konstruiert werden:

$$\tilde{X}_{(n)} := \begin{cases} X_{(\frac{n}{2})}(\omega) & \text{falls } n \text{ gerade} \\ X_{(\frac{n+1}{2})}(\omega) & \text{falls } n \text{ ungerade} \end{cases}$$

für  $\omega \in \Omega$  und der geordneten Messreihe  $X_{(1)}(\omega), \dots, X_{(n)}(\omega)$ .

**Erwartungstreue** Seien  $X_1, \dots, X_n$  unabhängig und identisch verteilt mit Verteilungsfunktion  $F_\theta$ . Sei außerdem der Median  $\mu_m = \tau(\theta)$  eindeutig und  $F_\theta$  symmetrisch.

Dann ist  $\mu_m = \mu$  und  $\tilde{X}_{(n)}$  ein erwartungstreu für  $\mu_m = \mu = \tau(\theta)$ .

**Vergleich Median/Arithmetisches Mittel** Seien  $\bar{X}_{(n)}$  und  $\tilde{X}_{(n)}$  Schätzer, wobei  $\bar{X}_{(n)}$  auf dem arithmetischen Mittel und  $\tilde{X}_{(n)}$  auf dem empirischen Median basiert.

Für die beiden Schätzer gilt (asymptotisch):

$$\begin{aligned}\text{MSE}_\theta(\bar{X}_{(n)}) &= \frac{\sigma^2}{n} \\ \text{MSE}_\theta(\tilde{X}_{(n)}) &= \frac{\pi\sigma^2}{2n}\end{aligned}$$

Somit ist der empirische Median um den Faktor  $\frac{2}{\pi} \approx 0.64$  weniger effizient als das arithmetische Mittel. Anders ausgedrückt: Der Median ist bei 100 Beobachtungen genauso verlässlich wie das arithmetische Mittel bei 64 Beobachtungen als Schätzer für den Erwartungswert

---

## 5.2 M-Schätzer

---

- Seien  $X_1, \dots, X_n$  unabhängige identisch symmetrisch verteilte Zufallsvariablen mit der Realisierung  $x_1, \dots, x_n$ .
- **Ziel:** Erstellung eines Schätzers für den Erwartungswert  $\mu$ .
- Der *M-Schätzer* bildet ein allgemeines Prinzip zur Konstruktion für Schätzer des Erwartungswertes.
- Sei  $\Phi : [0, \infty) \rightarrow \mathbb{R}$  eine monoton wachsende Straffunktion und betrachte

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

- Existiert ein eindeutiges Minimum  $\mu_M(x_1, \dots, x_n)$ , ist dies der zu  $\Phi$  gehörige M-Schätzer.

**Typische M-Schätzer** Üblicherweise wird  $\Phi(s) = s^p$  mit  $p > 0$  gewählt. Dann liefert:

$p = 1$  den Median  $\tilde{x}$ . Er minimiert die Abstandssumme:

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

$p = 2$  das arithmetische Mittel. Es minimiert die quadratische Abstandssumme:

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

$p \rightarrow \infty$  die Midrange

$$\frac{\max\{x_1, \dots, x_n\} + \min\{x_1, \dots, x_n\}}{2}$$

Kleinere Werte für  $p$  liefern dabei robustere M-Schätzer, da Ausreißer weniger stark bestraft werden.  
Eine weitere übliche Straffunktion ist z.B. die Lorentz-Strafffunktion

$$\Phi(s) = \ln\left(1 + \frac{s^2}{2}\right)$$

die robuster ist als die übliche quadratische Straffunktion.

---

## 6 Multivariate Verteilungen und Summen von Zufallsvariablen

---

Oftmals ist es nötig, Zufallsvariablen zu betrachten, die voneinander abhängig, also nicht unabhängig sind. Dazu wird hier die gemeinsame Verteilung des Zufallsvektors  $X = (X_1, \dots, X_n)^T$  und die Summe der Zufallsvariablen betrachtet.

---

### 6.1 Grundlagen

---

**Gemeinsame Verteilungsfunktion** Seien  $X_1, \dots, X_n$  Zufallsvariablen mit Verteilungsfunktionen  $F_1, \dots, F_n$ . Dann ist die *gemeinsame Verteilungsfunktion* (oder Verteilung des Zufallsvektors  $XI(X_1, \dots, X_n)^T$ ) gegeben durch:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

**Gemeinsame Dichte** Eine Funktion  $f : \mathbb{R}^n \rightarrow [0, \infty)$  heißt *gemeinsame Dichte* von  $X_1, \dots, X_n$ , wenn für alle  $(x_1, \dots, x_n) \in \mathbb{R}^n$  gilt:

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(s_1, \dots, s_n) ds_1 \dots ds_n$$

**Erwartungswertvektor** Der Vektor  $\mu = (E(X_1), \dots, E(X_n))^T$  heißt (sofern er existiert) *Erwartungswertvektor* von  $X$ .

**Kovarianzmatrix** Die *Kovarianzmatrix* ist (sofern sie existiert) eine Matrix  $\Sigma \in \mathbb{R}^{n \times n}$  der folgenden Form:

$$\Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Var}(X_n) \end{pmatrix}$$

wobei die *Kovarianz* von zwei Zufallsvariablen gegeben ist durch

$$\text{Cov}(X_i, X_j) := E\left((X_i - E(X_i)) \cdot (X_j - E(X_j))\right)$$

sofern  $\text{Var}(X_i)$  existiert, d.h.  $< \infty$  ist ( $i = 1, \dots, n$ ).

Es gilt weiterhin  $\text{Var}(X_i) = \text{Cov}(X_i, X_i)$  und  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ .

---

**Rechenregeln** Seien  $X, Y, Z$  Zufallsvariablen und  $a, b, c, d \in \mathbb{R}$ .

Dann gelten:

- $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$
- $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$
- $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

### Interpretationen

- Unabhängigkeit impliziert  $\text{Cov}(X, Y) = 0$ .
- Ist  $\text{Cov}(X, Y) > 0$ , so wird  $X$  erhöht, wenn  $Y$  erhöht wird und umgekehrt.
- Ist  $\text{Cov}(X, Y) < 0$ , so wird  $X$  verringert, wenn  $Y$  erhöht wird und umgekehrt.
- Der *Korrelationskoeffizient*

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} \in [-1, 1]$$

ist skalierungsunabhängig.

- Der empirische Korrelationskoeffizient ist ein erwartungstreuer Schätzer für  $\rho$ .

---

### 6.1.1 Multivariate Normalverteilung

- Dies ist die wichtigste multivariate Verteilung:  $N_n(\mu, \Sigma)$
- Sei  $X = (X_1, \dots, X_n)^T$  ein Vektor mit normalverteilten Zufallsvariablen mit Erwartungswert  $\mu = (E(X_1), \dots, E(X_n))^T$  und Kovarianzmatrix  $\Sigma$ .
- Dann ist die multivariate Normalverteilungsdichte gegeben durch:

$$f(x) = \frac{1}{\sqrt{2\pi}^n \sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}, \quad x \in \mathbb{R}^n$$

---

### 6.1.2 Unabhängigkeit

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit Dichten  $f_1(x_1), \dots, f_n(x_n)$ . Die Zufallsvariable sind unabhängig gdw. für die gemeinsame Verteilungsfunktion  $f$  gilt:

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

---

### 6.1.3 Korrelation

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit  $\text{Var}(X_i) < \infty$ ,  $i = 1, \dots, n$ . Die heißen *paarweise unkorreliert*, wenn gilt:

$$\text{Cov}(X_i, X_j) = 0 \quad \forall i \neq j$$

Unabhängigkeit impliziert Unkorreliertheit, da für unabhängige Zufallsvariablen  $X_1, \dots, X_n$  die obige Bedingung immer gilt.

Sind die Zufallsvariablen gemeinsam normalverteilt, folgt aus der Unkorreliertheit sogar die Unabhängigkeit.



---

## 6.2 Verteilung der Summe von Zufallsvariablen

---

Seien  $X_1, \dots, X_n$  unabhängige Zufallsvariablen, die  $N(\mu_i, \sigma_i^2)$ -verteilt sind. Dann ist  $X = X_1 + \dots + X_n$   $N(\mu, \sigma^2)$ -verteilt mit

$$\mu = \mu_1 + \dots + \mu_n, \quad \sigma^2 = \sigma_1^2 + \dots + \sigma_n^2$$

---

### 6.2.1 Faltung

---

Falls für die Funktionen  $f, g: \mathbb{R} \rightarrow \mathbb{R}$  das Integral

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y) dy$$

für alle  $x \in \mathbb{R}$  existiert, dann heißt  $f * g$  die *Faltung* von  $f$  und  $g$ .

Für  $f = (f_i)_{i \in \mathbb{Z}}$  und  $g = (g_i)_{i \in \mathbb{Z}}$  ist die *diskrete Faltung* von  $f$  und  $g$  gegeben durch:

$$(f * g)_i := \sum_{j \in \mathbb{Z}} f_{i-j} g_j$$

Sind  $X_1, X_2$  unabhängige stetige Zufallsvariablen mit Dichten  $f_1(x_1)$  und  $f_2(x_2)$ , dann hat  $X_1 + X_2$  die Dichte  $f_1 * f_2$ .

---

### 6.2.2 Diskret verteilte Zufallsvariablen

---

Seien  $X_1, X_2$  unabhängige diskrete  $\mathbb{Z}$ -wertige Zufallsvariablen und seien

$$f_{X_1} := (P(X_1 = i))_{i \in \mathbb{Z}}$$
$$f_{X_2} := (P(X_2 = i))_{i \in \mathbb{Z}}$$

Dann ist  $f_{X_1+X_2} = (P(X_1 + X_2 = i))_{i \in \mathbb{Z}}$  gegeben durch

$$f_{X_1+X_2} = f_{X_1} * f_{X_2}$$

---

### Binomialverteilten Zufallsvariablen

---

Seien  $X, Y$  jeweils  $B(n, p)$  bzw.  $B(m, p)$  verteilt. Dann ist  $X + Y$   $B(n + m, p)$ -verteilt.

---

### Poissonverteilte Zufallsvariablen

---

Seien  $X, Y$  jeweils Poisson-verteilt mit Parameter  $\lambda_1$  bzw.  $\lambda_2$ . Dann ist  $X + Y$  Poisson-verteilt mit Parameter  $\lambda_1 + \lambda_2$ .

---

### Poisson Verteilung und bedingte Wahrscheinlichkeit

---

Seien  $X, Y$  jeweils Poisson-verteilt mit Parameter  $\lambda_1$  bzw.  $\lambda_2$ . Sei außerdem  $Z = X + Y$ . Dann ist  $Z$  ebenfalls Poisson-verteilt mit Parameter  $\lambda_1 + \lambda_2$ .

Betrachte nun  $X|Z$ . Aufgrund der Unabhängigkeit folgt, dass  $X|Z$  mit  $P(X = x | Z = z)$  Binomialverteilt ist mit  $B(z, \frac{\lambda_1}{\lambda_1 + \lambda_2})$ .

---

## Geometrische Verteilung und bedingte Wahrscheinlichkeit

---

Sei  $X$  geometrisch verteilt mit Parameter  $p$ .

Betrachte nun  $Y_k = X - k \mid X > k$ , d.h. die Anzahl der Versuche, bis das Ereignis eintritt, unter der Voraussetzung, dass es in den vorherigen  $k$  Versuchen nicht eingetreten ist. Die Zufallsvariable  $Y_k$  ist dann wieder geometrisch verteilt mit Parameter  $p$ .