

Wahrscheinlichkeit und Statistik

Silvan Metzker
Juli 2024

1 Grundbegriffe

1.1 Wahrscheinlichkeitsraum

Axiome von Kolmogorov

Das Tuple $(\Omega, \mathcal{A}, \mathbb{P})$ ist ein **Wahrscheinlichkeitsraum** mit

- I. **Grundraum** Ω mit $\Omega \neq \emptyset$, wobei $\omega \in \Omega$ ein Elementarereignis ist.
- II. **σ -Algebra** $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ wobei gilt:
 1. $\Omega \in \mathcal{A}$
 2. $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
 3. $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_i A_i \in \mathcal{A}$
- III. **Wahrscheinlichkeitsmass** \mathbb{P} auf (Ω, \mathcal{A}) ist eine Abbildung $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$, wobei gilt:
 1. $\mathbb{P}(\Omega) = 1$
 2. $A_1, A_2, \dots \in \mathcal{A}, \forall i \neq j : A_i \cap A_j = \emptyset \implies \mathbb{P}(\bigcup_i A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

De-Morgan

Sei $(A_i)_{i \geq 1}$ eine Folge von beliebigen Mengen. Dann gilt

$$\left(\bigcup_{i=1}^{\infty} A_i \right)^c = \bigcap_{i=1}^{\infty} (A_i)^c$$

Daraus folgt

1. $A_1, A_2, \dots \in \mathcal{A} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$
 2. $A, B \in \mathcal{A} \implies (A \cup B), (A \cap B) \in \mathcal{A}$
- und für $A, B \in \mathcal{A}$
1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
 2. $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
 3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Sei $A_1, A_2, \dots \in \mathcal{A}$, dann gilt:

Union Bound

$$\mathbb{P} \left(\bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Siebformel

$$\mathbb{P} \left[\bigcup_{i=1}^n A_i \right] = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P} \left[\bigcap_{j=1}^k A_{i_j} \right]$$

Für $n = 2$: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

1.2 Bedingte Wahrscheinlichkeiten

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Bedingte Wahrscheinlichkeit

Sei $A, B \in \mathcal{A}$ und $\mathbb{P}(B) > 0$, dann ist die **bedingte Wahrscheinlichkeit von A gegeben B**

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

Satz der totalen Wahrscheinlichkeit

Sei B_1, \dots, B_N mit $\mathbb{P}[B_n] > 0$ für jedes $1 \leq n \leq N$ eine Partition des Grundraums Ω , d.h. $\bigcup_{n=1}^N B_n = \Omega$ mit $B_n \cap B_m = \emptyset$ für $n \neq m$. Dann gilt für alle $A \in \mathcal{F}$,

$$\mathbb{P}[A] = \sum_{n=1}^N \mathbb{P}[A | B_n] \mathbb{P}[B_n] = \sum_{n=1}^N \mathbb{P}[A \cap B_n]$$

Satz von Bayes

Aus der Definition der bedingten W'keit folgt sofort die Bayessche Formel, welche den Zusammenhang zwischen $\mathbb{P}(A | B)$ und $\mathbb{P}(B | A)$ beschreibt:

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

Mit dem *Satz der totalen W'keit* erhalten wir:

Sei $B_1, \dots, B_N \in \mathcal{F}$ eine **Partition** von Ω mit $\mathbb{P}[B_n] > 0$ für alle n . Für jedes Ereignis A mit $\mathbb{P}[A] > 0$ und jedes $n \in \{1, \dots, N\}$ gilt

$$\mathbb{P}[B_n | A] = \frac{\mathbb{P}[A | B_n] \mathbb{P}[B_n]}{\sum_{k=1}^N \mathbb{P}[A | B_k] \mathbb{P}[B_k]}$$

Intuition Bayessche Statistik

In dieser Form würde man A als das **eingetretene Ereignis** und die B_i als die verschiedene **Hypothesen** ver-

stehen.

In der Bayesschen Statistik versucht man die Hypothese zu finden, so dass $\mathbb{P}(B_i | A)$ **maximiert** wird.
(Wurde in der Vorlesung nicht weiter behandelt)

1.3 Unabhängigkeit

Zwei Ereignisse $A, B \in \mathcal{A}$ heissen (**stochastisch**) **unabhängig**, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Es gilt (*):

- $\mathbb{P}(A) \in \{0, 1\} \implies A$ zu jedem Ereignis unabhängig
- A zu sich selbst unabhängig $\implies \mathbb{P}(A) \in \{0, 1\}$
- A, B unabhängig $\implies A, B^c$ unabhängig

Wenn $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$ gilt:

A, B unabh. $\iff \mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(B|A) = \mathbb{P}(B)$

Eine Kollektion von Ereignissen $(A_i)_{i \in I}$ heisst (**stochastisch**) **unabhängig**, wenn

$$J \subseteq I \text{ endlich} \implies \mathbb{P} \left(\bigcap_{i \in J} A_i \right) = \prod_{i \in J} \mathbb{P}(A_i)$$

2 Zufallsvariablen

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine (reellwertige) **Zufallsvariable** ist eine Abbildung $X : \Omega \rightarrow \mathbb{R}$, sodass für alle $x \in \mathbb{R}$ gilt,

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}.$$

Eine Funktion X ist **messbar** (ZV sind messbar), wenn:

$X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}$ für alle $B \in \mathcal{B}(\mathbb{R})$.

Wobei $\mathcal{B}(\mathbb{R})$ die *borelsche σ -Algebra* auf \mathbb{R} bez. Bsp:

- Alle offenen, abgeschl. und komp. Mengen in \mathbb{R} .
- Alle Intervalle der Form $(a, b), [a, b], (a, b], [a, b), (-\infty, b), (-\infty, b], (a, \infty)$ und $[a, \infty)$ für $a, b \in \mathbb{R}$.

2.1 Verteilungsfunktion

Die **Verteilungsfunktion** ist die Abbildung $F_X : \mathbb{R} \rightarrow [0, 1]$ definiert durch:

$$F_X(t) := \mathbb{P}(X \leq t), \forall t \in \mathbb{R}$$

Die Funktion erfüllt folgende Eigenschaften:

- i) F_X ist monoton wachsend
 - ii) F_X ist rechtsstetig, i.e. $\lim_{h \downarrow 0} F_X(x+h) = F_X(x)$
 - iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ und $\lim_{x \rightarrow \infty} F_X(x) = 1$
- Auch gilt: $\forall a, b \in \mathbb{R}, a > b: \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

Linksstetigkeit

Die Verteilungsfunktion ist nicht immer linksstetig. Sei $F_X(a-) := \lim_{h \downarrow 0} F_X(a-h)$ für $a \in \mathbb{R}$ beliebig. Dann gilt:

$$\mathbb{P}(X = a) = F_X(a) - F_X(a-)$$

Intuitiv folgt daraus, dass wenn...

- F_X in Punkt $a \in \mathbb{R}$ nicht stetig ist, dann ist die "Sprunghöhe" $F_X(a) - F_X(a-)$ gleich der Wahrscheinlichkeit $\mathbb{P}(X = a)$.
- F_X stetig in Punkt $a \in \mathbb{R}$, dann gilt $\mathbb{P}(X = a) = 0$.

Seien X_1, \dots, X_n Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann heißen X_1, \dots, X_n **unabhängig**, falls $\forall x_1, \dots, x_n \in \mathbb{R}$:

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdot \dots \cdot \mathbb{P}(X_n \leq x_n).$$

2.2 Diskrete Zufallsvariablen

Sei $A \in \mathcal{F}$ ein Ereignis.

Wir sagen A tritt **fast sicher (f.s.)** ein, falls $\mathbb{P}(A) = 1$.

Seien $X, Y: \Omega \rightarrow \mathbb{R}$ Zufallsvariablen:

$X \leq Y$ f.s. $\iff \mathbb{P}(X \leq Y) = 1$

Eine Zufallsvariable $X: \Omega \rightarrow \mathbb{R}$ heisst **diskret**, falls eine endliche oder abzählbare Menge $W \subset \mathbb{R}$ existiert, sodass

$$\mathbb{P}(X \in W) = 1$$

Falls Ω endlich oder abzählbar ist, dann ist X immer diskret.

Die **Verteilungsfunktion** einer diskreten ZV X :

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \in W} p(y) \cdot \mathbb{1}_{y \leq x}$$

Die **Gewichtsfunktion** einer diskreten ZV X :

$$\forall x \in X(\Omega): p(x) = \mathbb{P}(X = x) \text{ wobei } \sum_{x \in X(\Omega)} p(x) = 1$$

2.3 Diskrete Verteilungen

Bernoulli-Verteilung: $X \sim \text{Ber}(p)$

$X(\Omega) = \{0, 1\}$ und die Gewichtsfunktion ist definiert durch

$$p(1) := \mathbb{P}(X = 1) = p \text{ und } p(0) := \mathbb{P}(X = 0) = 1 - p.$$

Binomialverteilung: $X \sim \text{Bin}(n, p)$

Wiederholung von n unabhängigen Bernoulli-Experimenten mit gleichem Parameter p .

$$p(k) := \mathbb{P}(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \forall k \in \{0, 1, \dots, n\}$$

Geometrische Verteilung: $X \sim \text{Geo}(p)$

Warten auf den ersten Erfolg.

$$p(k) := \mathbb{P}(X = k) = (1-p)^{k-1} \cdot p \quad \forall k \in \mathbb{N} \setminus \{0\}$$

Poisson-Verteilung: $X \sim \text{Poisson}(\lambda)$

Grenzwert der Binomialvert. für grosse n und kleine p .

$$p(k) := \mathbb{P}(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \forall k \in \mathbb{N}_0, \lambda > 0$$

1. (★) Für $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$ gilt $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k)$ wobei $Y \sim \text{Poisson}(\lambda)$.
2. (★) Seien $X_1 \sim \text{Poisson}(\lambda_1)$ und $X_2 \sim \text{Poisson}(\lambda_2)$ unabhängig. Dann gilt $(X_1 + X_2) \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

2.4 Stetige Zufallsvariablen

Eine Zufallsvariable $X: \Omega \rightarrow \mathbb{R}$ heisst **stetig**, wenn ihre Verteilungsfunktion F_X wie folgt geschrieben werden kann

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \text{ für alle } x \in \mathbb{R}.$$

wobei $f_X: \mathbb{R} \rightarrow \mathbb{R}^+$ eine nicht-negative Funktion ist. f wird dann als **Dichte** von X benannt.

Wenn $f_X: (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ messbar ist, ist die Zufallsvariable X **absolut stetig**.

Intuition: $f_X(t) dt$ ist die Wahr'keit, dass $X \in [t, t+dt]$.

2.5 Stetige Verteilungen

Gleichverteilung: $X \sim \mathcal{U}([a, b])$

Die Dichte ist auf dem Intervall $[a, b]$ gleich.

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

Exponentialverteilung: $T \sim \text{Exp}(\lambda)$

Lebensdauer oder Wartezeit eines allg. Ereignisses (Stetiges Äquivalent zur Geometrischen Verteilung).

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Normalverteilung: $X \sim \mathcal{N}(\mu, \sigma^2)$

Häufig verwendete Verteilung. Undefiniert für $\sigma = 0$.

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

1. Seien X_1, \dots, X_n **unabhängige** normalverteilte ZV mit Parametern $(\mu_1, \sigma_1^2), \dots, (\mu_n, \sigma_n^2)$, dann ist

$$Z = \mu_0 + \lambda_1 X_1 + \dots + \lambda_n X_n$$

eine normalverteilte ZV mit Parametern $\mu = \mu_0 + \lambda_1 \mu_1 + \dots + \lambda_n \mu_n$ und $\sigma^2 = \lambda_1^2 \sigma_1^2 + \dots + \lambda_n^2 \sigma_n^2$.

2. Sei $Z \sim \mathcal{N}(0, 1)$ eine **standardnormalverteilte** Zufallsvariable. Dann gilt für $X \sim \mathcal{N}(\mu, \sigma^2)$

$$X = \mu + \sigma \cdot Z$$

3. Für $X \sim \mathcal{N}(\mu, \sigma^2)$ gilt $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, also

$$F_X(x) = \mathbb{P}\left[\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

4. $\Phi(-x) = 1 - \Phi(x)$

Gedächtnislosigkeit

Sei $T \sim \text{Geom}(p)$ mit $p \in (0, 1)$. Dann gilt für alle $n \geq 0$ und alle $k \geq 1$

$$\mathbb{P}[T \geq n+k \mid T > n] = \mathbb{P}[T \geq k].$$

(★) Hält auch für $T \sim \text{Exp}(\lambda)$.

Hier noch zum Thema MLE-Schätzer und dessen Eigenschaften, siehe S. 8 für eine Übersicht der Schätzer.

Verteilung	Erwartungstreu	Konsistent
Bernoulli	Ja	Ja
Binomial	Nur p	n und p
Geometrisch	Nein	Ja
Poisson	Ja	Ja
Gleichverteilung	Nein	Ja
Exponentiell	Ja	Ja
Normalverteilung	Nur μ	μ und σ^2

3 Erwartungswert

Erwartungswert (Stetige ZV)

Sei $X : \Omega \rightarrow \mathbb{R}$ eine stetige Zufallsvariable mit Dichte f . Sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Abbildung, sodass $\varphi(X)$ eine Zufallsvariable ist. Dann gilt

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx,$$

falls das Integral wohldef. ist (bei $\varphi = id$ abs. konv.).
Sei X eine stetige ZV mit $X \geq 0$ f.s., dann gilt:

$$\mathbb{E}[X] = \int_0^{\infty} (1 - F_X(x)) dx$$

Der **Allgemeine Erwartungswert** für eine reellwertige ZV X mit $\mathbb{E}[|X|] < \infty$ ist definiert als:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_+] - \mathbb{E}[X_-] \quad \text{mit } X_{\pm} = \max(\pm X, 0) \\ &= \int_0^{\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx \end{aligned}$$

Erwartungswert (Diskrete ZV)

Sei $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable, $W_X := X(\Omega)$ und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Abbildung. Falls die Summe wohldefiniert ist, gilt:

$$\mathbb{E}(\varphi(X)) := \sum_{x \in W_X} \varphi(x) \cdot \mathbb{P}(X = x)$$

Sei X eine nicht-negative Zufallsvariable. Dann gilt $\mathbb{E}[X] \geq 0$. Gleichheit gilt genau dann, wenn $X = 0$ fast sicher ist.

$$\mathbb{E}[X] \geq 0 \iff X \geq 0 \text{ immer}$$

$$\mathbb{E}[X] = 0 \iff X = 0 \text{ fast sicher, d.h. } \mathbb{P}[X \neq 0] = 0$$

3.1 Rechnen mit Erwartungswerten

Linearität des Erwartungswertes:

Seien $X, Y : \Omega \rightarrow \mathbb{R}$ ZV mit $\lambda \in \mathbb{R}$, Falls die Erwartungswerte wohldefiniert sind, gilt:

$$\mathbb{E}(\lambda \cdot X + Y) = \lambda \cdot \mathbb{E}(X) + \mathbb{E}(Y)$$

Falls X, Y **unabhängig**, dann gilt auch:

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

Generell: X_1, X_2, \dots, X_n unabhängig und endlich:

$$\mathbb{E} \left[\prod_{k=1}^n X_k \right] = \prod_{k=1}^n \mathbb{E}[X_k]$$

3.2 Ungleichungen

Monotonie

Seien X, Y ZV mit $X \leq Y$ f.s., dann gilt:

$$\mathbb{E}(X) \leq \mathbb{E}(Y)$$

Markow Ungleichung

Sei X eine ZV und ferner $g : X(\Omega) \rightarrow [0, +\infty)$ eine wachsende Funktion. Für jedes $c \in \mathbb{R}$ mit $g(c) > 0$ gilt dann

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}(g(X))}{g(c)} \stackrel{t \leq 0}{\implies} \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$$

Chebyshev Ungleichung

Sei Y eine ZV mit endlicher Varianz. Für jedes $b > 0$ gilt dann

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq b) \leq \frac{\text{Var}(Y)}{b^2}$$

Jensen Ungleichung

Sei X eine ZV und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion, dann gilt:

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$$

3.3 Varianz

Varianz

Sei X eine ZV, sodass $\mathbb{E}(X^2) < \infty$. Die **Varianz** von X ist definiert durch

$$\mathbb{V}(X) = \sigma_X^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

wobei $m = \mathbb{E}(X)$. Dabei wird σ_X als **Standardabweichung** von X bezeichnet und beschreibt den Erwartungswert für die Distanz von X zu $\mathbb{E}(X)$.

1. Sei X ein ZV, sodass $\mathbb{E}(X^2) < \infty$ und $a, b \in \mathbb{R}$:

$$\mathbb{V}(a \cdot X + b) = a^2 \cdot \mathbb{V}(X)$$

2. Seien X_1, \dots, X_n paarweise unabhängig. Dann gilt

$$\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n)$$

Kovarianz

Seien X, Y ZV mit $\mathbb{E}(X^2) < \infty, \mathbb{E}(Y^2) < \infty$. Wir definieren die **Kovarianz** zwischen X und Y durch

$$\begin{aligned} \text{cov}(X, Y) &:= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

1. $\text{cov}(X, X) = \mathbb{V}(X)$
2. X, Y unabhängig $\implies \text{cov}(X, Y) = 0$ (\nLeftarrow)
3. $\mathbb{V}(X \pm Y) = \mathbb{V}(X) + \mathbb{V}(Y) \pm 2\text{cov}(X, Y)$
4. $(\star) \text{cov}(\sum_{i=1}^n X_i, \sum_{j=1}^n Y_j) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, Y_j)$

Korrelationen

- $\text{cov}(X, Y) > 0 \implies$ positiv korreliert
- $\text{cov}(X, Y) = 0 \implies$ unkorreliert
- $\text{cov}(X, Y) < 0 \implies$ negativ korreliert/antikorreliert

Es gilt: X_i, X_j unabhängig $\implies X_i, X_j$ unkorreliert

Eigenschaften der Kovarianz

Für X, Y, Z mit $\mathbb{E}[X_i^2] < \infty$ und $a, b, c, d, e, f, g, h \in \mathbb{R}$:

1. Positive Semidefinitheit: $\text{cov}(X, X) \geq 0$
2. Symmetrie: $\text{cov}(X, Y) = \text{cov}(Y, X)$
3. Bilinearität: $\text{cov}(aX + b, cY + d) = ac \text{cov}(X, Y)$ und $\text{cov}(X, (eY + f) + (gZ + h)) = e \text{cov}(X, Y) + g \text{cov}(X, Z)$

4 Mehrere Zufallsvariablen

Die **gemeinsame Verteilungsfunktion** von n Zufallsvariablen X_1, \dots, X_n (stetig oder diskret) ist die Abbildung $F : \mathbb{R}^n \rightarrow [0, 1]$,

$$F(x_1, \dots, x_n) := \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

4.1 Diskreter Fall - Gewichtsfunktion

Für n diskrete ZV X_1, \dots, X_n definieren wir ihre **gemeinsame Gewichtsfunktion** $p : \mathbb{R}^n \rightarrow [0, 1]$ durch

$$p(x_1, \dots, x_n) := \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

Dann ist die **gemeinsame Verteilungsfunktion**:

$$\begin{aligned} F(x_1, \dots, x_n) &= \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} p(y_1, \dots, y_n) \end{aligned}$$

Verteilung des Bildes

Sei $n \geq 1$, $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$, X_1, \dots, X_n **diskrete** ZV mit Werten in W_1, \dots, W_n . Dann ist $Z = \varphi(X_1, \dots, X_n)$ diskret mit Werten in $W = \varphi(W_1 \times \dots \times W_n)$. Die Verteilung von Z für $z \in W$ ist:

$$\mathbb{P}[Z = z] = \sum_{\substack{x_1 \in W_1, \dots, x_n \in W_n \\ \varphi(x_1, \dots, x_n) = z}} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]$$

Randdichte. Seien X_1, \dots, X_n diskrete ZV mit gemeinsamer Gewichtsft. p . Für jedes $k \in \{1, \dots, n\}$ und jedes $x \in W_k$ gilt

$$\mathbb{P}[X_k = x] = \sum_{\substack{x_\ell \in W_\ell \\ \ell \in \{1, \dots, n\} \setminus \{k\}}} p(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n)$$

Der Erw. des Bildes der Funktion $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$ ist

$$\mathbb{E}(\varphi(X_1, \dots, X_n)) = \sum_{x_1, \dots, x_n} \varphi(x_1, \dots, x_n) p(x_1, \dots, x_n)$$

Seien X_1, \dots, X_n diskrete ZV mit gemeinsamer Verteilung $\{p(x_1, \dots, x_n)\}_{x_1 \in W_1, \dots, x_n \in W_n}$. Dann ist **äquivalent**:

- (i) X_1, \dots, X_n sind unabhängig,
- (ii) für alle $x_1 \in W_1, \dots, x_n \in W_n$ gilt

$$p(x_1, \dots, x_n) = \mathbb{P}[X_1 = x_1] \cdot \dots \cdot \mathbb{P}[X_n = x_n],$$

4.2 Stetiger Fall - Gemeinsame Dichte

Gemeinsame Dichte

Falls die gemeinsame Verteilungsfunktion von n Zufallsvariablen X_1, \dots, X_n sich schreiben lässt als

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_n \dots dt_1$$

Randverteilung. Haben X, Y die gemeinsame Verteilungsfunktion $F_{X,Y}$, so ist $F_X: \mathbb{R} \rightarrow [0, 1]$,

$$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq \infty) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

die Vertsft. der Randverteilung von X . Analog für F_Y .

Randdichte. Sei X, Y ZV mit gemeinsamer Dichte $f(x, y)$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{bzw.} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Seien X_1, \dots, X_n ZV mit Dichten f_{X_1}, \dots, f_{X_n} .

Dann sind folgende Aussagen **äquivalent**:

- (i) X_1, \dots, X_n sind unabhängig,
- (ii) X_1, \dots, X_n sind gemeinsam stetig mit gemeinsamer Dichte $f: \mathbb{R}^n \rightarrow \mathbb{R}_+$,
d.h. die gemeinsame Dichtefunktion f ist das Produkt der einzelnen Randdichten f_{X_k} , also
 $f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$.

5 Konvergenz von Wahrheiten

Konvergenz in Verteilung

Seien $(X_n)_{n \in \mathbb{N}}$ und X Zufallsvariablen mit Verteilungsfunktionen $(F_n)_{n \in \mathbb{N}}$ und F . $(X_n)_{n \in \mathbb{N}}$ konvergiert **in Verteilung** gegen X , geschrieben $X_n \xrightarrow{d} X$ für $n \rightarrow \infty$, falls für jeden Stetigkeitspunkt $x \in \mathbb{R}$ von F gilt:

$$\lim_{n \rightarrow \infty} F_n(x) = \mathbb{P}[X_n \leq x] = F(x)$$

Notation: $X_n \xrightarrow{w} X$ oder $X_n \xrightarrow{L} X$, wobei d, w, L für *convergence in distribution*, *weak convergence*, bzw. *convergence in law* stehen. (Nicht in Vorlesung)

Schwaches Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen mit gleichen Erwartungswerten $\mathbb{E}[X_k] = \mu$ und Varianzen $\mathbb{V}[X_k] = \sigma^2$. Sei

$$\bar{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Dann konvergiert \bar{X}_n für $n \rightarrow \infty$ **in Wahrscheinlichkeit** gegen $\mu = \mathbb{E}(X_i)$, d.h. für jedes $\varepsilon > 0$ gilt

$$\mathbb{P}[|\bar{X}_n - \mu| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0.$$

Starkes Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von uiv. Zufallsvariablen. Sei $\mathbb{E}(|X_1|) < \infty$ und $\mu = \mathbb{E}(X_1)$. Für

$$\bar{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$$

gilt dann

$$\bar{X}_n \xrightarrow{n \rightarrow \infty} \mu \quad \mathbb{P}\text{-fast sicher,}$$

das bedeutet,

$$\mathbb{P}\left[\left\{\omega \in \Omega \mid \bar{X}_n(\omega) \xrightarrow{n \rightarrow \infty} \mu\right\}\right] = 1.$$

Sei X eine nicht-negative Zufallsvariable. Dann gilt $\mathbb{E}[X] \geq 0$. Gleichheit gilt genau dann wenn $X = 0$ fast sicher gilt. Also (aus Vorlesung):

$\mathbb{E}[X] \geq 0 \iff X \geq 0$ gilt immer

und

$\mathbb{E}[X] = 0 \iff X = 0$ fast sicher, also $\mathbb{P}[X \neq 0] = 0$

5.1 Zentraler Grenzwertsatz

Zentraler Grenzwertsatz

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von iid. Zufallsvariablen mit $\mathbb{E}(X_i) = \mu < \infty$ und $\text{Var}(X_i) = \sigma^2 < \infty$. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}$$

also

$$\left(\frac{\frac{1}{n} S_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right) \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Bemerkungen:

Man verwendet auch oft die Form für $\bar{X}_n = \frac{1}{n} S_n$ als

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (\star)$$

beziehungsweise

$$S_n \xrightarrow{d} \mathcal{N}(n\mu, n\sigma^2) \quad \text{und} \quad \bar{X}_n \xrightarrow{d} \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right) \quad (\star)$$

Beispielrechnung

Seien $(X_i)_{i \geq 1}, (Y_i)_{i \geq 1}$ und $(Z_i)_{i \geq 1}$ Folgen von iid. ZV mit

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$$

und analog für Y_1 und Z_1 . Wir definieren

$$S_n^{(x)} := \sum_{i=1}^n X_i, \quad S_n^{(y)} := \sum_{i=1}^n Y_i, \quad S_n^{(z)} := \sum_{i=1}^n Z_i$$

Die Folge $((S_n^{(x)}, S_n^{(y)}, S_n^{(z)}))_{n \geq 1}$ wird zufällige Irrfahrt in \mathbb{Z}^3 genannt. Sei $\alpha > \frac{1}{2}$. Zeige, dass

$$\mathbb{P}\left(\left\| (S_n^{(x)}, S_n^{(y)}, S_n^{(z)}) \right\|_2 \leq n^\alpha\right) \rightarrow 1 \text{ für } n \rightarrow \infty,$$

wobei $\|(x, y, z)\|_2 := \sqrt{x^2 + y^2 + z^2}$ die euklidische Norm ist.

Schritt 1: $\forall \alpha > 1/2$ zeigen wir $\mathbb{P}(|S_n^{(x)}| \leq n^\alpha) \xrightarrow{n \rightarrow \infty} 1$.

Da $\mathbb{E}(X_i) = 0$ und $\text{Var}(X_i) = 1$ folgt für $a \in \mathbb{R}$ beliebig per ZGS

$$\mathbb{P}\left(S_n^{(x)} \leq a\sqrt{n}\right) = \mathbb{P}\left(\frac{S_n^{(x)}}{\sqrt{n}} \leq a\right) \xrightarrow{n \rightarrow \infty} \Phi(a)$$

und somit auch

$$\begin{aligned} \mathbb{P}\left(|S_n^{(x)}| \leq a\sqrt{n}\right) &= \mathbb{P}\left(S_n^{(x)} \leq a\sqrt{n}\right) - \mathbb{P}\left(S_n^{(x)} \leq -a\sqrt{n}\right) \\ &\xrightarrow{n \rightarrow \infty} \Phi(a) - \Phi(-a) = 2\Phi(a) - 1 \end{aligned}$$

Sei $\alpha = 1/2 + \beta, \beta > 0$. Dann instanzieren wir mit $a = n^\beta$.

$$\mathbb{P}(|S_n^{(x)}| \leq n^\alpha) = \mathbb{P}(|S_n^{(x)}| \leq n^\beta \sqrt{n}) \rightarrow \lim_{n \rightarrow \infty} (2\Phi(n^\beta) - 1) = 1$$

Dies gilt analog für $S_n^{(y)}$ und $S_n^{(z)}$.

Schritt 2: $\forall \alpha > 1/2, \mathbb{P}\left(\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\|_2 \leq n^\alpha\right) \xrightarrow{n \rightarrow \infty} 1$

Sei $\alpha' \in (1/2, \alpha)$. Dann folgt

$$\begin{aligned} &\left\{ |S_n^{(x)}| \leq n^{\alpha'} \wedge |S_n^{(y)}| \leq n^{\alpha'} \wedge |S_n^{(z)}| \leq n^{\alpha'} \right\} \\ &\subseteq \left\{ \left\| (S_n^{(x)}, S_n^{(y)}, S_n^{(z)}) \right\|_2 \leq \sqrt{3} \cdot n^{\alpha'} \right\} \end{aligned}$$

Da $n^\alpha \geq \sqrt{3}n^{\alpha'}$ für grosse n , folgt

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}\left(\left\| (S_n^{(x)}, S_n^{(y)}, S_n^{(z)}) \right\|_2 \leq n^\alpha\right) \\ &\geq \lim_{n \rightarrow \infty} \mathbb{P}\left(\left\| (S_n^{(x)}, S_n^{(y)}, S_n^{(z)}) \right\|_2 \leq \sqrt{3} \cdot n^{\alpha'}\right) \\ &\geq \lim_{n \rightarrow \infty} \mathbb{P}\left(|S_n^{(x)}| \leq n^{\alpha'}, |S_n^{(y)}| \leq n^{\alpha'}, |S_n^{(z)}| \leq n^{\alpha'}\right) = 1 \end{aligned}$$

Momenterzeugende Funktion

Die momenterzeugende Funktion einer Zufallsvariablen X ist für $t \in \mathbb{R}$ definiert durch

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx.$$

Immer wohldef. in $[0, \infty]$, kann aber $+\infty$ werden.

Chernoff-Ungleichung

Seien X_1, \dots, X_n i.i.d. Zufallsvariablen, für welche $M_X(t)$ für alle $t \in \mathbb{R}$ endlich ist. Für jedes $b \in \mathbb{R}$ gilt

$$\mathbb{P}[S_n \geq b] \leq \exp\left(\inf_{t \in \mathbb{R}} (n \log M_X(t) - tb)\right).$$

Chernoff-Schranke

Seien X_1, \dots, X_n unabhängig mit $X_k \sim \text{Ber}(p_k)$ und sei $S_n = \sum_{k=1}^n X_k, \mu_n = \mathbb{E}[S_n] = \sum_{k=1}^n p_k$ und $\delta > 0$, dann gilt

$$\mathbb{P}[S_n \geq (1 + \delta)\mu_n] \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}}\right)^{\mu_n}.$$

6 Schätzer

Wir treffen folgende Annahmen:

- Parameterraum $\vartheta \subset \mathbb{R}^m$
- Familie von Wahrscheinlichkeitsmassen $(\mathbb{P}_\vartheta)_{\vartheta \in \vartheta}$ auf (Ω, \mathcal{F}) ; für jedes Element im Parameterraum existiert ein Modell / Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P}_\vartheta)$.
- Zufallsvariablen X_1, \dots, X_n auf (Ω, \mathcal{F})

Wir nennen die Gesamtheit der beobachteten Daten x_1, \dots, x_n (wobei $x_i = X_i(\omega)$) und die ZV X_1, \dots, X_n Stichprobe.

Ein **Schätzer** ist eine Zufallsvariable der Form

$$T_\ell = t_\ell(X_1, \dots, X_n)$$

Die **Schätzfunktionen** $t_\ell : \mathbb{R}^n \rightarrow \mathbb{R}$ müssen gewählt werden. Einsetzen von Daten $x_k = X_k(\omega), k = 1, \dots, n$, liefert **Schätzwerte** $T_\ell(\omega) = t_\ell(x_1, \dots, x_n)$ für $\vartheta_\ell, \ell = 1, \dots, m$. Kurz: $T = (T_1, \dots, T_m)$ und $\vartheta = (\vartheta_1, \dots, \vartheta_m)$.

Ein Schätzer T ist **erwartungstreu**, falls für alle $\vartheta \in \Theta$ gilt:

$$\mathbb{E}_\vartheta[T] = \vartheta$$

Sei $\vartheta \in \vartheta$ und T ein Schätzer. Der **Bias** (erwartete Schätzfehler) von T im Modell \mathbb{P}_ϑ ist definiert als:

$$\mathbb{E}_\vartheta[T] - \vartheta$$

Der mittlere quadratische Schätzfehler (MSE) von T im Modell \mathbb{P}_ϑ ist definiert als:

$$\begin{aligned} \text{MSE}_\vartheta[T] &= \mathbb{E}_\vartheta[(T - \vartheta)^2] \\ &= \text{Var}_\vartheta(T) + (\mathbb{E}_\vartheta[T] - \vartheta)^2 \end{aligned}$$

Eine Folge von Schätzern $T^{(n)}, n \in \mathbb{N}$, heisst **konsistent** für ϑ , falls $T^{(n)}$ für $n \rightarrow \infty$ in \mathbb{P}_ϑ -Wahrscheinlichkeit gegen ϑ konvergiert, d.h. für jedes $\vartheta \in \Theta$ und jedes $\varepsilon > 0$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}_\vartheta[|T^{(n)} - \vartheta| > \varepsilon] = 0$$

6.1 Maximum-Likelihood-Methode

6.1.1 Likelihood-Funktion, ML-Schätzer

Die Likelihood-Funktion ist definiert als

$$L(x_1, \dots, x_n; \vartheta) = \begin{cases} p(x_1, \dots, x_n; \vartheta) & \text{falls diskret} \\ f(x_1, \dots, x_n; \vartheta) & \text{falls stetig} \end{cases}$$

Wenn X_k unter \mathbb{P}_ϑ i.i.d. gilt (analog mit $f_{\vec{x}}$ und f_X):

$$p_{\vec{x}}(x_1, \dots, x_n; \vartheta) = \prod_{k=1}^n p_X(x_k; \vartheta)$$

Für jedes $x_1, \dots, x_n \in W$ sei $t_{ML}(x_1, \dots, x_n)$ der Wert, welcher die Funktion $\vartheta \mapsto L(x_1, \dots, x_n; \vartheta)$ maximiert. Ein Maximum-Likelihood-Schätzer ist dann definiert als

$$T_{ML} = t_{TM}(X_1, \dots, X_n) \in \arg \max_{\vartheta \in \vartheta} L(X_1, \dots, X_n; \vartheta)$$

Notiz: Nicht vergessen zu zeigen, dass es ein *Maxima* ist.

6.1.2 Anwendung der Methode

Die Maximum-Likelihood-Methode ist ein Weg, um systematisch einen Schätzer zu bestimmen.

1. Gemeinsame Dichte/Verteilung der ZV finden
2. Bestimme davon die Log-Likelihood-Funktion $f(\vartheta) := \ln(L(x_1, \dots, x_n; \vartheta))$
3. $f(\vartheta)$ nach ϑ ableiten
4. Nullstelle von $f'(\vartheta)$ finden
5. $f''(\vartheta) < 0$ oder anderes Argument, dass wir das Maximum gefunden haben (evtl. Randstellen überprüfen!).

6.2 Momentenmethode /-schätzer:

1. Sei X_1, \dots, X_n iid. eine Stichprobe.
2. Sei ϑ ein m -dimensionaler Parameterraum.
3. Stelle für $\vartheta = (\vartheta_1, \dots, \vartheta_m)$ ein Gleichungssystem auf, in dem das k -te empirische Moment dem k -ten Moment gleichgesetzt wird:

$$\hat{m}_k(x_1, \dots, x_n) = g_k(\vartheta_1, \dots, \vartheta_m), \quad k \in \{1, \dots, m\}$$
4. Der Vektor $\hat{\vartheta}(X_1, \dots, X_n)$ heißt Momentenschätzer des Parameters ϑ .

Momentenschätzer. Der Schätzer $T = (T_1, T_2)$ ist allgemein in jedem Modell \mathbb{P}_ϑ , in dem X_1, \dots, X_n i.i.d. sind, der sogenannte Momentenschätzer für

$$(\mathbb{E}_\vartheta[X], \mathbb{V}_\vartheta[X])$$

Dieser Schätzer ist allerdings nicht erwartungstreu für $(\mathbb{E}_\vartheta[X], \mathbb{V}_\vartheta[X])$. Es gilt zwar

$$\mathbb{E}_\vartheta[T_1] = \mathbb{E}_\vartheta[\bar{X}_n] = \mathbb{E}_\vartheta[X]$$

aber

$$\mathbb{E}_\vartheta[(\bar{X}_n)^2] = \frac{1}{n} \mathbb{E}_\vartheta[X^2] + \frac{n-1}{n} \mathbb{E}_\vartheta[X]^2$$

Daraus folgt

$$\mathbb{E}_\vartheta[T_2] = \frac{n-1}{n} \mathbb{V}_\vartheta[X] \neq \mathbb{V}_\vartheta[X]$$

Um einen erwartungstreuen Schätzer T' für $(\mathbb{E}_\vartheta[X], \mathbb{V}_\vartheta[X])$ zu erhalten, verwendet man

$$T'_1 = \bar{X}_n$$

$$\begin{aligned} T'_2 &= \frac{n}{n-1} T_2 \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \\ &= \frac{1}{n-1} \sum_{k=1}^n X_k^2 - \frac{n}{n-1} (\bar{X}_n)^2 \end{aligned}$$

Für T'_2 schreibt man oft

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

Man nennt S^2 die (korrigierte) empirische Varianz.

Gammafunktion

Die Funktion Γ nennt man (Eulersche) Gammafunktion und sie ist für $x \geq 0$ definiert durch

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

Γ hat eine grundlegende Verbindung zur Fakultätsfunktion, denn

$$\Gamma(n+1) = n! \text{ für } n \in \mathbb{N}_0.$$

Studentsche t -Verteilung: $X \sim t_m$

Eine stetige Zufallsvariable X heisst t -verteilt mit m Freiheitsgraden falls ihre Dichte für $x \in \mathbb{R}$ gegeben ist durch

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}$$

Entstehung der t -Verteilung: Sind $X \sim \mathcal{N}(0, 1)$ und $Y \sim \chi_m^2$ unabhängig, so ist der Quotient

$$\frac{X}{\sqrt{\frac{1}{m}Y}} \sim t_m$$

1. Für $m = 1$ ergibt sich eine Cauchy-Verteilung.
2. Für $m \rightarrow \infty$ erhält man asymptotisch eine $\mathcal{N}(0, 1)$ -Verteilung.
3. Die t -Verteilung ist symmetrisch um 0, aber langschwänziger als die $\mathcal{N}(0, 1)$ -Verteilung; die Dichte geht langsamer gegen 0, je kleiner m ist.

7 Tests

Die **Nullhypothese** H_0 und die **Alternativhypothese** H_A sind zwei Teilmengen $\Theta_0 \subseteq \Theta, \Theta_A \subseteq \Theta$ wobei $\Theta_0 \cap \Theta_A = \emptyset$.

Falls keine explizite Alternativhypothese spezifiziert ist, so hat man $\Theta_A = \Theta \setminus \Theta_0$.

Eine Hypothese heisst *einfach*, falls die Teilmenge aus einem einzelnen Wert besteht; sonst *zusammengesetzt*.

Definition Test

Ein Test ist ein Paar (T, K) , wobei:

- $T = t(X_1, \dots, X_n)$ die Teststatistik ist, mit einer messbaren Funktion $t: \mathbb{R}^n \rightarrow \mathbb{R}$.
- $K \subseteq \mathbb{R}$ der kritische Bereich oder Verwerfungsbereich ist.

Wir wollen nun anhand der Daten $(X_1(\omega), \dots, X_n(\omega))$ entscheiden, ob die Nullhypothese akzeptiert oder verworfen wird. Zuerst berechnen wir die Teststatistik $T(\omega) = t(X_1(\omega), \dots, X_n(\omega))$ und gehen dann wie folgt vor:

- Die Hypothese H_0 wird *verworfen*, falls $T(\omega) \in K$.
- Die Hypothese H_0 wird *akzeptiert*, falls $T(\omega) \notin K$.

Ein **Fehler 1. Art** ist, wenn H_0 fälschlicherweise verworfen wird, obwohl sie richtig ist.

$$\mathbb{P}_\theta(T \in K), \quad \theta \in \Theta_0$$

Ein **Fehler 2. Art** ist, wenn H_0 fälschlicherweise akzeptiert wird, obwohl sie falsch ist.

$$\mathbb{P}_\theta(T \notin K) = 1 - \mathbb{P}_\theta(T \in K), \quad \theta \in \Theta_A$$

Bemerkung: Da T eine ZV und somit bezüglich dem Mass $\mathbb{P}_\theta: \mathcal{F} \rightarrow [0, 1]$ messbar ist, gilt $\{T \in K\} \in \mathcal{F}$ und somit ist $\mathbb{P}_\theta(T \in K)$ wohldefiniert.

7.1 Signifikanzniveau und Macht

Ein Test hat Signifikanzniveau $a \in [0, 1]$ falls

$$\forall \theta \in \Theta_0 \quad \mathbb{P}_\theta(T \in K) \leq a$$

Es ist meist unser primäres Ziel, die Fehler 1. Art zu minimieren.

Das sekundäre Ziel ist, Fehler 2. Art zu vermeiden. Hierfür definieren wir die Macht eines Tests als Funktion:

$$\beta: \Theta_A \mapsto [0, 1], \quad \theta \mapsto \mathbb{P}_\theta(T \in K)$$

Zu beachten ist, dass eine kleine Wahrscheinlichkeit für einen Fehler 2. Art einem *grossen* β entspricht.

7.2 Konstruktion von Tests

Wir nehmen an, dass X_1, \dots, X_n diskret oder gemeinsam stetig unter \mathbb{P}_{θ_0} und \mathbb{P}_{θ_A} sind, wobei $\Theta_0 \cap \Theta_A = \emptyset$ einfach sind ($\theta_0 \in \Theta_0 \wedge \theta_A \in \Theta_A$).

Der Likelihood-Quotient ist somit wohldefiniert:

$$R(x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n; \theta_A)}{L(x_1, \dots, x_n; \theta_0)}$$

(Falls $L(x_1, \dots, x_n; \theta_0) = 0$ setzen wir $R(x_1, \dots, x_n) = +\infty$.)

Für zusammengesetzte Θ_0 und Θ_A können wir den verallg. Likelihood-Quotient definieren:

$$R(x_1, \dots, x_n) := \frac{\sup_{\theta \in \Theta_A} L(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta_0} L(x_1, \dots, x_n; \theta)}$$

Wenn $R \gg 1$, so gilt $H_A > H_0$ und analog $R \ll 1 \implies H_A < H_0$.

Der **Likelihood-Quotient-Test (LQ-Test)** mit Parameter $c \geq 0$ ist definiert durch:

$$T = R(X_1, \dots, X_n) \quad \text{und} \quad K = (c, \infty]$$

Neyman-Pearson-Lemma

Sei $\Theta_0 = \{\vartheta_0\}$ und $\Theta_A = \{\vartheta_A\}$. Sei (T, K) ein Likelihood-Quotienten-Test mit Parameter c und Signifikanzniveau $\alpha^* := \mathbb{P}_{\vartheta_0}[T \in K]$. Ist (T', K') ein anderer Test mit Signifikanzniveau $\alpha := \mathbb{P}_{\vartheta_0}[T' \in K'] \leq \alpha^*$, so gilt:

$$\mathbb{P}_{\vartheta_A}[T' \in K'] \leq \mathbb{P}_{\vartheta_A}[T \in K].$$

Das bedeutet, jeder andere Test mit kleinerem Signifikanzniveau hat auch geringere Macht bzw. eine größere Wahrscheinlichkeit für einen Fehler 2. Art.

7.3 T-Test / Gauss Test

Zuerst berechnen wir T :

$$T = \frac{\bar{X}_n - \theta_0}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)$$

mit dem erwartungstreuen Schätzer:

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

Dann unterscheiden wir zwischen den folgenden Fällen:

- 1. Einseitiger Test $H_A: \theta > \theta_0$
 - Obere Grenze: $c = \Phi^{-1}(1 - \alpha)$
 - Verwerfungsbereich: Verwerfe H_0 falls $T > c$
- 2. Einseitiger Test $H_A: \theta < \theta_0$
 - Untere Grenze: $c = \Phi^{-1}(\alpha)$
 - Verwerfungsbereich: Verwerfe H_0 falls $T < c$

3. Beidseitiger Test $H_A: \theta \neq \theta_0$

- Untere Grenze: $c_1 = \Phi^{-1}(\alpha/2)$
- Obere Grenze: $c_2 = \Phi^{-1}(1 - \alpha/2)$
- Verwerfe H_0 falls $T < c_1$ oder $T > c_2$

Notiz: Restliche Tests nicht in Vorlesung behandelt. (für restliche Tests siehe Nicolas Wehrli's Cheat Sheet)

8 Aufgaben

Integrale(r) Ratgeber:

- *Complete The Square*, e.g. umgekehrtes Bin. Theorem, evtl. notwendig um nächsten Punkt zu erreichen.
- Integral über Verteilungsfunktion $\int_{-\infty}^{\infty} f_X(x) dx = 1$
- Gaußsche Glockenkurve
- Substitution
- Partielle Integration

Häufige Formen:

$$\mathbb{P}[a < X \leq b] = \mathbb{P}[X \leq b] - \mathbb{P}[X \leq a] = F_X(b) - F_X(a)$$

$$\mathbb{P}[X > Y] = \sum_{i=1}^n \mathbb{P}[X > Y \mid Y = i] \mathbb{P}[Y = i]$$

$$\mathbb{P}[X > Y] = \int_{-\infty}^{\infty} \mathbb{P}[X > Y \mid Y = y] f_Y(y) dy$$

$$\begin{aligned} \mathbb{P}[\max(X, Y) \leq z] &= \mathbb{P}[X \leq z, Y \leq z] \\ &= F_X(z) \cdot F_Y(z) \quad (X, Y \text{ unabh.}) \\ \mathbb{P}[\min(X, Y) \leq z] &= 1 - \mathbb{P}[\min(X, Y) > z] \\ &= 1 - \mathbb{P}[X > z, Y > z] \\ &= 1 - \mathbb{P}[X > z] \mathbb{P}[Y > z] \quad (X, Y \text{ unabh.}) \\ &= 1 - (1 - F_X(z))(1 - F_Y(z)) \end{aligned}$$

$$\mathbb{P}[X + Y = t] = \int_{k=0}^t f_X(k) f_Y(t - k) dk \quad (t \geq 0)$$

Für $L = \min(X_1, \dots, X_n)$ und $M = \max(X_1, \dots, X_n)$:

$$\begin{aligned} \mathbb{P}[M < m, L \leq l] &= \mathbb{P}[M < m] - \mathbb{P}[M < m, L > l] \\ &= \mathbb{P}[M < m] - \mathbb{P}[l < X_1 < m, \dots, l < X_n < m] \\ &= (\mathbb{P}[X_1 < m])^n - (\mathbb{P}[l < X_1 < m])^n \quad (\text{iid.}) \end{aligned}$$

Sei X_1, \dots, X_n iid. mit $X_1 \sim \mathcal{U}([a, b])$:

$$\mathbb{P}[X_1 > X_2, X_1 > X_3, \dots, X_1 > X_n] = \frac{(n-1)!}{n!}$$

8.1 Multiple Choice Aufgaben

Seien X, Y zwei ZV mit gemeinsamer Dichte $f_{X,Y}$. Welche Aussage ist korrekt?

✓ X, Y sind immer stetig

□ Die ZV sind nicht notwendigerweise stetig.

Sei Y eine stetige Zufallsvariable. Für alle $s, t \in \mathbb{R}^+$:

$$\exists \lambda > 0. Y \sim \text{Exp}(\lambda) \iff \mathbb{P}(Y > s) = \mathbb{P}(Y > s+t \mid Y > t)$$

✓ wahr.

□ falsch.

Seien $(X_i)_{i=1}^n$ uiv. mit Verteilungsfunktion $F_{X_i} = F$. Was ist die Verteilungsfunktion von $M = \max(X_1, \dots, X_n)$?

✓ $F_M(a) = F(a)^n$

□ $F_M(a) = 1 - F(a)^n$

□ $F_M(a) = (1 - F(a))^n$

Seien X, Y unabhängig und lognormalverteilt ($\ln X, \ln Y$ sind normalverteilt). Welche Aussage ist korrekt?

✓ XY ist lognormalverteilt

□ XY ist normalverteilt

□ e^{X+Y} ist normalverteilt

Tabelle der Standard-Normalverteilungsfunktion

$\Phi(z) = P[Z \leq z]$ mit $Z \sim \mathcal{N}(0, 1)$ und $\Phi(-x) = 1 - \Phi(x)$.

z	.00	.02	.05	.07	.09
.0	0.5000	0.5080	0.5199	0.5279	0.5359
.1	0.5398	0.5478	0.5596	0.5675	0.5753
.2	0.5793	0.5871	0.5987	0.6064	0.6141
.3	0.6179	0.6255	0.6368	0.6443	0.6517
.4	0.6554	0.6628	0.6736	0.6808	0.6879
.5	0.6915	0.6985	0.7088	0.7157	0.7224
.6	0.7257	0.7324	0.7422	0.7486	0.7549
.7	0.7580	0.7642	0.7734	0.7794	0.7852
.8	0.7881	0.7939	0.8023	0.8078	0.8133
.9	0.8159	0.8212	0.8289	0.8340	0.8389
1.0	0.8413	0.8461	0.8531	0.8577	0.8621
1.1	0.8643	0.8686	0.8749	0.8790	0.8830
1.2	0.8849	0.8888	0.8944	0.8980	0.9015
1.3	0.9032	0.9066	0.9115	0.9147	0.9177
1.4	0.9192	0.9222	0.9265	0.9292	0.9319
1.5	0.9332	0.9357	0.9394	0.9418	0.9441
1.6	0.9452	0.9474	0.9505	0.9525	0.9545
1.7	0.9554	0.9573	0.9599	0.9616	0.9633
1.8	0.9641	0.9656	0.9678	0.9693	0.9706
1.9	0.9713	0.9726	0.9744	0.9756	0.9767

9 Tabellen & Diverses

9.1 Grenzwerte

$\lim_{x \rightarrow \infty} \frac{e^x}{x^m} = \infty$	$\lim_{x \rightarrow -\infty} x e^x = 0$
$\lim_{x \rightarrow \infty} (1+x)^{\frac{1}{x}} = 1$	$\lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}} = e$
$\lim_{x \rightarrow \infty} (1+\frac{1}{x})^b = 1$	$\lim_{x \rightarrow \infty} n^{\frac{1}{n}} = 1$
$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1$	$\lim_{x \rightarrow \infty} (1 - \frac{1}{x})^x = \frac{1}{e}$
$\lim_{x \rightarrow \pm \infty} (1 + \frac{k}{x})^{mx} = e^{km}$	$\lim_{x \rightarrow \infty} (\frac{x}{x+k})^x = e^{-k}$
$\lim_{x \rightarrow 0} \frac{\log 1-x}{x} = -1$	$\lim_{x \rightarrow 0} x \log x = 0$
$\lim_{x \rightarrow 0} \frac{e^{ax} - 1}{x} = a$	$\lim_{x \rightarrow 0} \frac{\ln(x+1)}{x} = 1$
$\lim_{x \rightarrow 1} \frac{\ln(x)}{x-1} = 1$	$\lim_{x \rightarrow \infty} \frac{\log(x)}{x^a} = 0$

Partielle Integration

$$\int f'(x)g(x) \, dx = f(x)g(x) - \int f(x)g'(x) \, dx$$

- Meist gilt: Polynome ableiten ($g(x)$), wo das Integral periodisch ist (\sin, \cos, e^x, \dots) integrieren ($f'(x)$)
- Teils: mit 1 multiplizieren, um partielle Integration anwenden zu können (z.B. im Fall von $\int \log(x) \, dx$)

Substitution

Um $\int_a^b f(g(x)) \, dx$ zu berechnen: Ersetze $g(x)$ durch u und integriere $\int_{g(a)}^{g(b)} f(u) \frac{du}{g'(x)}$.

- $g'(x)$ muss sich herauskürzen, sonst nutzlos.
- Grenzen substituieren nicht vergessen.
- Man kann auch das Theorem in die andere Richtung anwenden:

$$\int_a^b f(u) \, du = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(x))g'(x) \, dx$$

- Sei \mathcal{X}, Y kompakt, $f : Y \subset \mathbb{R}^n \rightarrow \mathbb{R}$ stetig.
Sei $\gamma : \mathcal{X} \rightarrow Y$ mit $\mathcal{X} = \mathcal{X}_0 \cup B, Y = Y_0 \cup C$ (B, C Rand von \mathcal{X}, Y).

Wenn $\gamma : \mathcal{X}_0 \rightarrow Y_0$ bijektiv und C^1 mit $\det(J_\gamma(x)) \neq 0, \forall x \in \mathcal{X}_0$, dann gilt

$$\int_Y f(y) \, dy = \int_{\mathcal{X}} f(\gamma(x)) |\det(J_\gamma(x))| \, dx$$

Polarkoordinaten		
$x = r \cos \theta$	$0 \leq r < \infty$	$dx dy = r \, dr d\theta$
$y = r \sin \theta$	$0 \leq \theta < 2\pi$	
oder: $x^2 + y^2 = r^2$		

Gamma-Verteilung

Die Gamma-Verteilung ist eine stetige Verteilung mit der Dichtefunktion

$$f(z) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha z^{\alpha-1} e^{-\lambda z} \text{ für } z \geq 0, \alpha > 0, \lambda > 0$$

1. Wir schreiben $Z \sim Ga(\alpha, \lambda)$ für eine gamma-verteilte ZV Z mit Parametern α und λ .
2. Die Summe von $n \in \mathbb{N}$ unabhängigen $Exp(\lambda)$ -verteilten Zufallsvariablen ist $Ga(n, \lambda)$ -verteilt.
3. Die χ^2 -Verteilung mit k Freiheitsgraden ist $Ga(\frac{k}{2}, \frac{1}{2})$ -verteilt.

Sei $(X_i)_{i \geq 1} \sim \mathcal{N}(0, 1)$ iid. eine Folge von Zufallsvariablen.

1. $\sum_{i=1}^n X_i^2 \sim \chi_n^2$
2. $\frac{1}{n} \sum_{i=1}^n X_i \sim \chi_1^2$
3. $X_1^2 + X_2^2 \sim Exp(\frac{1}{2})$
4. Sei $Y \sim \chi_m^2$ unabhängig von $\mathcal{N}(0, 1)$. Dann gilt

$$\frac{X}{\sqrt{\frac{1}{m}Y}} \sim t_m$$

5. Es gilt $\lim_{m \rightarrow \infty} t_m \sim \mathcal{N}(0, 1)$ verteilt, für endliche m ist t_m langschwänziger als $\mathcal{N}(0, 1)$.

Seien X_1, \dots, X_n iid. $\sim \mathcal{N}(\mu, \sigma^2)$. Wir erinnern uns an die Notationen für Stichprobenmittel \overline{X}_n und Stichprobenvarianz $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$.

1. $\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$
2. \overline{X}_n und S^2 sind unabhängig.

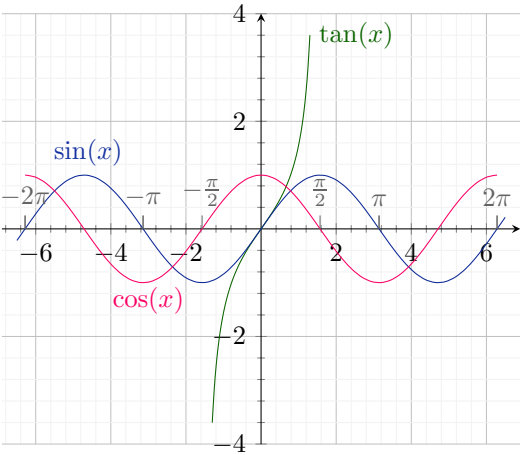
3.

$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}$$

9.2 MLE Schätzer

- $X_1, \dots, X_n \sim Exp(\theta)$ iid.: $T = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\overline{X}_n}$
- $X_1, \dots, X_n \sim Geo(\theta)$ iid.: $T = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\overline{X}_n}$
- $X_1, \dots, X_n \sim Bin(N, \theta)$ iid.: $T = \frac{1}{N} \sum_{i=1}^n X_i$
- $X_1, \dots, X_n \sim Ber(p)$ iid.: $T = \frac{\sum_{i=1}^n X_i}{n}$
- $X_1, \dots, X_n \sim P(\theta)$ iid.: $T = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}_n$
- $X_1, \dots, X_n \sim \mathcal{U}([\theta_1, \theta_2])$ iid.: $T_{\theta_1} = \max(X_i), T_{\theta_2} = \min(X_i)$
- $X_1, \dots, X_n \sim \mathcal{N}(\theta_1, \theta_2)$ iid. : $T_{\theta_1} = \overline{X}_n, T_{\theta_2} = S^2$

Wichtige Werte						
deg	0°	30°	45°	60°	90°	180°
rad	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π
cos	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	-1
sin	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	0
tan	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	$+\infty$	0



9.3 Ableitungen

$\mathbf{F(x)}$	$\mathbf{f(x)}$	$\mathbf{f'(x)}$
$\frac{x^{-a+1}}{-a+1}$	$\frac{1}{x^a}$	$\frac{a}{x^{a+1}}$
$\frac{x^{a+1}}{a+1}$	$x^a \ (a \neq -1)$	$a \cdot x^{a-1}$
$\frac{1}{k \ln(a)} a^{kx}$	a^{kx}	$ka^{kx} \ln(a)$
$\ln x $	$\frac{1}{x}$	$-\frac{1}{x^2}$
$\frac{2}{3}x^{3/2}$	\sqrt{x}	$\frac{1}{2\sqrt{x}}$
$\frac{n}{n+1}x^{\frac{1}{n}+1}$	$\sqrt[n]{x}$	$\frac{1}{n}x^{\frac{1}{n}-1}$
$-\cos(x)$	$\sin(x)$	$\cos(x)$
$\sin(x)$	$\cos(x)$	$-\sin(x)$
$\frac{1}{2}(x - \frac{1}{2}\sin(2x))$	$\sin^2(x)$	$2\sin(x)\cos(x)$
$\frac{1}{2}(x + \frac{1}{2}\sin(2x))$	$\cos^2(x)$	$-2\sin(x)\cos(x)$
$-\ln \cos(x) $	$\tan(x)$	$\frac{1}{\cos^2(x)}$
		$1 + \tan^2(x)$
$\cosh(x)$	$\sinh(x)$	$\cosh(x)$
$\log(\cosh(x))$	$\tanh(x)$	$\frac{1}{\cosh^2(x)}$
$\ln \sin(x) $	$\cot(x)$	$-\frac{1}{\sin^2(x)}$
$\frac{1}{c} \cdot e^{cx}$	e^{cx}	$c \cdot e^{cx}$
$x(\ln x - 1)$	$\ln x $	$\frac{1}{x}$
$\frac{1}{2}(\ln(x))^2$	$\frac{\ln(x)}{x}$	$\frac{1-\ln(x)}{x^2}$
$\frac{x}{\ln(a)}(\ln x - 1)$	$\log_a x $	$\frac{1}{\ln(a)x}$

9.4 Weitere Ableitungen

$\mathbf{F(x)}$	$\mathbf{f(x)}$
$\frac{1}{a \cdot (n+1)}(ax+b)^{n+1}$	$(ax+b)^n$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$\frac{-1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$\operatorname{arcsinh}(x)$	$\frac{1}{\sqrt{1+x^2}}$
$\operatorname{arccosh}(x)$	$\frac{1}{\sqrt{x^2-1}}$
$\operatorname{arctanh}(x)$	$\frac{1}{1-x^2}$
$x^x \ (x > 0)$	$x^x \cdot (1 + \ln x)$
$\log_a x $	$\frac{1}{x \ln a} = \log_a(e) \frac{1}{x}$
$\frac{(ax+b)^{n+2}}{a^2(n+1)(n+2)}$	$\frac{(ax+b)^{n+1}}{a \cdot (n+1)}$
$\sqrt{1-x^2} + x \cdot \arcsin(x)$	$\arcsin(x)$
$x \cdot \arccos(x) - \sqrt{1-x^2}$	$\arccos(x)$
$x \cdot \arctan(x) - \frac{1}{2} \log(x^2+1)$	$\arctan(x)$
$x \cdot \operatorname{arcsinh}(x) - \sqrt{x^2+1}$	$\operatorname{arcsinh}(x)$
$\frac{x \cdot \operatorname{arccosh}(x) - \sqrt{x^2-1}\sqrt{x^2+1}}{\sqrt{x^2-1}\sqrt{x^2+1}}$	$\operatorname{arccosh}(x)$
$\frac{1}{2} \log(1-x^2) + x \cdot \operatorname{arctanh}(x)$	$\operatorname{arctanh}(x)$
$\frac{\alpha}{\gamma} \log \gamma x + \beta $	$\frac{\alpha}{\gamma x + \beta}$

9.5 Integrale

$\mathbf{f(x)}$	$\mathbf{F(x)}$
$\int f'(x)f(x) \, dx$	$\frac{1}{2}(f(x))^2$
$\int \frac{f'(x)}{f(x)} \, dx$	$\ln f(x) $
$\int_{-\infty}^{\infty} e^{-x^2} \, dx$	$\sqrt{\pi}$
$\int (ax+b)^n \, dx$	$\frac{1}{a(n+1)}(ax+b)^{n+1}$
$\int x(ax+b)^n \, dx$	$\frac{(ax+b)^{n+2}}{(n+2)a^2} - \frac{b(ax+b)^{n+1}}{(n+1)a^2}$
$\int (ax^p+b)^n x^{p-1} \, dx$	$\frac{(ax^p+b)^{n+1}}{ap(n+1)}$
$\int (ax^p+b)^{-1} x^{p-1} \, dx$	$\frac{1}{ap} \ln ax^p+b $
$\int \frac{ax+b}{cx+d} \, dx$	$\frac{ax}{c} - \frac{ad-bc}{c^2} \ln cx+d $
$\int \frac{1}{x^2+a^2} \, dx$	$\frac{1}{a} \arctan \frac{x}{a}$
$\int \frac{1}{x^2-a^2} \, dx$	$\frac{1}{2a} \ln \left \frac{x-a}{x+a} \right $
$\int \sqrt{a^2+x^2} \, dx$	$\frac{x}{2}f(x) + \frac{a^2}{2} \ln(x+f(x))$

9.6 Definite Integrale

$$\int_0^{2\pi} \sin(x) \, dx = \int_0^{2\pi} \cos(x) \, dx = 0,$$

$$\int_0^{2\pi} \sin^2(x) \, dx = \int_0^{2\pi} \cos^2(x) \, dx = \pi$$

Gaußsche Glockenkurve

Für das uneigentliche Integral über die *gaußsche Glockenkurve* gilt

$$\int_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2}} \, dx = \sqrt{2\pi\sigma^2}$$

9.7 Verteilungen

Verteilung	Notation	Parameter	$\mathbb{E}[X]$	$\text{Var}(X)$	$p_X(t)/f_X(t)$	$F_X(t)$
Gleichverteilung	unbekannt	n : Anzahl Ereignisse $(x_i$: Ereignisse)	$\frac{1}{n} \sum_{i=1}^n x_i$	$\frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n^2} (\sum_{i=1}^n x_i)^2$	$\frac{1}{n}$	$\frac{ \{k: x_k \leq t\} }{n}$
Bernoulli	$\text{Ber}(p)$	p : ErfolgsW'keit	p	$p \cdot (1 - p)$	$p^t(1 - p)^{1-t}$	$1 - p$ für $0 \leq t < 1$
Binomial	$\text{Bin}(n, p)$	n : Anzahl Versuche p : ErfolgsW'keit	np	$np(1 - p)$	$\binom{n}{t} p^t(1 - p)^{n-t}$	$\sum_{k=0}^t \binom{n}{k} p^k(1 - p)^{n-k}$
Geometrisch	$\text{Geo}(p)$	p : ErfolgsW'keit $(t$: Anzahl Versuche)	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$p(1 - p)^{t-1}$	$1 - (1 - p)^t$
Poisson	$\text{Poisson}(\lambda)$	λ : Erwartungswert und Varianz	λ	λ	$\frac{\lambda^t}{t!} e^{-\lambda}$	$e^{-\lambda} \sum_{k=0}^t \frac{\lambda^k}{k!}$
Gleichverteilung (im Intervall)	$U \sim \mathcal{U}([0, 1])$	$[a, b]$: Intervall	$\frac{a+b}{2}$	$\frac{1}{12}(b - a)^2$	$\begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$	$\begin{cases} 0 & x \leq a \\ \frac{t-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$
Exponentialv.	$\text{Exp}(\lambda)$	$\lambda : \frac{1}{\mathbb{E}[X]}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$	$\begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$
Normalverteilung	$\mathcal{N}(\mu, \sigma^2)$	$\mu : \mathbb{E}[X]$ σ^2 : Varianz	μ	σ^2	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} dy$
χ^2 -Verteilung	χ_m^2	n : Freiheitsgrad	n	$2n$	$\frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}$ für $t > 0$	$P(\frac{n}{2}, \frac{t}{2})$
t-Verteilung	t_m	n : Freiheitsgrad	$\begin{cases} 0 & n > 1 \\ \text{undef.} & \text{sonst} \end{cases}$	$\begin{cases} \frac{\frac{n}{n-2}} & n > 2 \\ \infty & 1 < n \leq 2 \\ \text{undef.} & \text{sonst} \end{cases}$	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \cdot \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$	zu kompliziert
Negativbinomial	$\text{NBin}(r, p)$	$r \in \mathbb{N}, p \in [0, 1]$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\binom{k-1}{r-1} p^r (1 - p)^{k-r}$	zu kompliziert
Cauchy-Verteilung	Cauchy (x_0, γ)	$x_0 \in \mathbb{R}, \gamma > 0$	Existiert nicht	Existiert nicht	$\frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-x_0)^2}$	$\frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-x_0}{\gamma}\right)$
Hypergeometrisch	$H(n, r, m)$	$n \in \mathbb{N}, m, r \in \{1, \dots, n\}$	$m \frac{r}{n}$	$m \frac{r}{n} \left(1 - \frac{r}{n}\right) \frac{n-m}{n-1}$	$\frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$	$\sum_{y=0}^k \frac{\binom{r}{y} \binom{n-r}{m-y}}{\binom{n}{m}}$

Binomischer Lehrsatz

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \quad \text{mit:} \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Geometrische Reihe

Für $\alpha < 1$:

$$\sum_{n=0}^{\infty} \alpha^n = \frac{1}{1 - \alpha}$$

Cauchy Produkt

Falls $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ absolut konv, dann

$$\sum_{n=0}^{\infty} \sum_{k=0}^n a_k b_{n-k} = \sum_{n=0}^{\infty} \sum_{i+j=n} a_i b_j = \left(\sum_{n=0}^{\infty} a_n\right) \cdot \left(\sum_{k=0}^{\infty} b_k\right)$$