Inhaltsverzeichnis

Wahrscheinlichkeit

W.1 Wahrscheinlichkeiten

W.1.1 Ereignisraum, Grundraum

Ereignisraum: Der *Ereignisraum* oder *Grundraum* Ω ist die Menge aller möglichen Ereignisse eines Zufallexperiments. Die Elemente $\omega \in \Omega$ heissen *Elementarereignisse.*

Ereignis: Ein *Ereignis* $A \subseteq \Omega$ ist eine Teilmenge von Ω.

Die Klasse aller *beobachtbaren Ereignisse* $\mathcal{F}$ ist eine Teilmenge der Potenzmenge $\mathcal{P}(\Omega)$ von Ω.

W.1.2 Wahrscheinlichkeitsmass

Wahrscheinlichkeitsmass: Ein *Wahrscheinlichkeitsmass* $\mathbb{P}$ ist eine Abbildung $\mathbb{P} : \mathcal{F} \to [0, 1]$ mit folgenden Eigenschaften:

i: $\mathbb{P}[\Omega] = 1$.
ii: $\mathbb{P}[A] \geq 0$ für alle $A \in \mathcal{F}$.
iii: $\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ falls $A_i \cap A_j = \emptyset$ für $i \neq j$.

Aus den Axiomen i bis iii folgen direkt die Rechenregeln:

i: $\mathbb{P}[A^C] = 1 - \mathbb{P}[A]$.
ii: $\mathbb{P}[\emptyset] = 0$.
iii: $A \subseteq B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$.
iv: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$ (*Additionsregel*).
v: $\mathbb{P}[B] = \mathbb{P}[B \cap A] + \mathbb{P}[B \cap A^C]$

Endliche Räume: Für *endliche Räume* $\Omega = \{\omega_1, \ldots, \omega_n\}$ gilt: $\mathbb{P}[A] = \sum_{w_i \in A} \mathbb{P}[w_i]$

Laplace-Raum: Im *Laplace-Raum* sind alle Ereignisse gleich wahrscheinlich: $P[A] = \frac{|A|}{|\Omega|}$

W.1.3 Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit: Für Ereignisse A, B mit $\mathbb{P}[A] > 0$ ist die *bedingte Wahrscheinlichkeit* definiert durch:

$$\mathbb{P}[B \mid A] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}$$

Es folgt die *Multiplikationsregel*: $\mathbb{P}[A \cap B] = \mathbb{P}[B \mid A]\mathbb{P}[A]$

Satz (Totale Wahrscheinlichkeit): Sei $A_{1 \leq i \leq n}$ eine Partitionierung von Ω, dann gilt für ein beliebiges Ereignis B:

$$\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \mid A_i]\mathbb{P}[A_i]$$

Satz (Formel von Bayes): Sei $A_1, \ldots, A_n$ eine Partitionierung von Ω mit $\mathbb{P}[A_i] > 0$ für alle i und B ein Ereignis mit $\mathbb{P}[B] > 0$, dann gilt für jedes k:

$$\mathbb{P}[A_k \mid B] = \frac{\mathbb{P}[B \mid A_k]\mathbb{P}[A_k]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B \mid A_k]\mathbb{P}[A_k]}{\sum_{i=1}^{n} \mathbb{P}[B \mid A_i]\mathbb{P}[A_i]}$$

Unabhängigkeit: Zwei Ereignisse A, B heissen *unabhängig* $A \perp B$, wenn $\mathbb{P}[B \mid A] = \mathbb{P}[B]$ gilt. $\Rightarrow \mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$

W.2 Zufallsvariablen in $\mathbb{R}$

Zufallsvariable: Eine (reelwertige) *Zufallsvariable* X auf Ω ist eine Funktion $X : \Omega \to \mathcal{W}(X) \subseteq \mathbb{R}$. Jedes Elementarereignis ω wird auf eine Zahl $X(\omega)$ abgebildet.

Verteilung: Das stochastische Verhalten einer Zufallsvariablen X wird durch ihre *Verteilung* $\mu_X : \mathbb{R} \to [0,1]$ beschrieben

$$\mu_X(B) := \mathbb{P}[X \in B] := \mathbb{P}[\{\omega \mid X(\omega) \in B\}] \quad \text{für } B \subseteq \mathbb{R}.$$

Hinweis: Jedes Wahrscheinlichkeitsmass μ_X erfüllt:
1. $\mu_X(B) \geq 0 \quad$ für alle $B \subseteq \mathbb{R}$
2. $\mu_X(\mathcal{W}(X)) = \mu_X(\mathbb{R}) = 1$

W.2.1 Verteilungsfunktion

Verteilungsfunktion: Die *Verteilungsfunktion* einer Zufallsvariable X ist die Abbildung $F_X : \mathbb{R} \to [0,1]$,

$$F_X(t) := \mathbb{P}[X \leq t] = \mu_X(-\infty, t)$$

Hinweis: F_X hat folgende Eigenschaften:
- i: $a \leq b \Rightarrow F_X(a) \leq F_X(b)$ (monoton wachsend).
- ii: $\lim\limits_{t \to u, t > u} F_X(t) = F_X(u)$ (rechtsstetig).
- iii: $\lim\limits_{t \to -\infty} F_X(t) = 0$ und $\lim\limits_{t \to \infty} F_X(t) = 1$.

W.2.2 Diskrete Zufallsvariablen

Eine Zufallsvariable X heisst *diskret*, falls Ω und somit auch $\mathcal{W}(X) = \{x_1, x_2 \ldots\}$ endlich oder abzählbar ist.

Gewichtsfunktion: Die *Gewichtsfunktion* $p_X : \mathbb{R} \to [0,1]$ ist definiert als

$$p_X(x) := \begin{cases} \mathbb{P}[X = x] & \text{für } x \in \mathcal{W}(X) \\ 0 & \text{sonst} \end{cases}$$

Somit lässt sich die *diskrete Verteilung* μ_X berechnen:

$$\mu_X(B) = P[X \in B] = \sum_{x_i \in B} p_X(x_i)$$

W.2.3 Stetige Zufallsvariablen

Eine Zufallsvariable X heisst *stetig*, falls Ω und somit auch $\mathcal{W}(X)$ überabzählbar ist. Deswegen ist $\mathbb{P}[X = x]$ immer gleich null; es wird eine andere Definition benötigt.

Dichte: Die *Dichte* $f_X : \mathbb{R} \to [0, \infty)$ ist gegeben durch

$$F_X(x) = \int_{-\infty}^{x} f_X(s) \mathrm{d}s \quad \text{für alle } x \in \mathbb{R}.$$

Somit lässt sich die *stetige Verteilung* μ_X berechnen:

$$\mu_X(B) = P[X \in B] = \int_B f_X(x) \mathrm{d}s$$

Hinweis: Es gilt $\frac{\mathrm{d}}{\mathrm{d}t} F_X(t) = f_X(t)$ falls f_X an der Stelle t stetig ist.

W.2.4 Erwartungswert und Momente

Erwartungswert: Sofern die Reihe / das Integral konvergiert, ist der *Erwartungswert* von X definiert als

$$\mathbb{E}[X] := \sum_{x_i \in \mathcal{W}(X)} x_i p_X(x_i) \quad \text{bzw.} \quad \mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$$

Hinweis: Es gelten folgende Rechenregeln
- i: $\mathbb{E}[a] = a$
- ii: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- iii: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, $\quad$ falls $X \perp Y$
- iv: $\mathbb{E}[X] \leq \mathbb{E}[Y] \Leftrightarrow X \leq Y$ *(Monotonie)*

Satz (4.1): : Für eine Zufallsvariable X und $Y = g(X)$ gilt:

$$\mathbb{E}[Y] = \sum_{x_i \in \mathcal{W}(X)} y_i p_X(x_i) \quad \text{bzw.} \quad \mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_X(x) \mathrm{d}x$$

p-tes Moment: Für eine Zufallsvariable X und $p \in \mathbb{R}^+$ gilt:
- Das p-te absolute Moment $M_p := \mathbb{E}[|X|^p] \leq \infty$
- Das p-te Moment $m_p := \mathbb{E}[X^p] < \infty$

Nach Satz 4.1 ist das p-te Moment, sofern existent:

$$m_p = \textstyle\sum_{x_i \in \mathcal{W}(X)} x_i^p p_X(x_i) \quad \text{bzw.} \quad m_p = \int_{-\infty}^{\infty} x^p f_X(x) \mathrm{d}x$$

Momentenerzeugende Funktion: $\mathcal{M}_X(t) := \mathbb{E}[e^{tX}]$

W.2.5 Varianz und Standardabweichung

Varianz: Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$. Die *Varianz* von X ist definiert als

$$\mathrm{Var}[X] := \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Nach Satz 4.1 ist die Varianz einer stetigen Zufallsvariable X:

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) \mathrm{d}x$$

Hinweis: Es gelten folgende Rechenregeln
- i: $\mathrm{Var}[a] = 0$
- ii: $\mathrm{Var}[a + bX] = b^2 \mathrm{Var}[X]$
- iii: $\mathrm{Var}[aX + bY] = a^2 \mathrm{Var}[X] + 2ab\mathrm{Cov}[X,Y] + b^2 \mathrm{Var}[Y]$

Standardabweichung: Die *Standardabweichung* einer Zufallsvariable X ist $\sigma_X := \sqrt{\mathrm{Var}[X]}$.

W.2.6 Kovarianz und Korrelation

Kovarianz: Seien X und Y Zufallsvariablen mit $\mathbb{E}[X^2] < \infty$ und $\mathbb{E}[Y^2] < \infty$, dann ist die *Kovarianz* von X und Y gegeben

$$\mathrm{Cov}[X,Y] := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Hinweis: Die Kovarianz ist ein Skalarprodukt:
- i: $\mathrm{Cov}[X, Y + aZ] = \mathrm{Cov}[X,Y] + a\mathrm{Cov}[X,Z]$ ***(bilinear)***
- ii: $\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$. *(symmetrisch)*
- iii: $\mathrm{Cov}[X,X] = \mathrm{Var}[X] \geq 0$, $\quad \mathrm{Cov}[X,X] = 0 \Leftrightarrow X = a$
  $\mathrm{Cov}[X,a] = 0$ für alle $a \in \mathbb{R}$. *(positiv definit)*

$\Rightarrow \mathrm{Cov}[X,Y]^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y]$ *(Cauchy-Schwarz)*

Korrelation: Seien X und Y Zufallsvariablen, dann gilt

$$\mathrm{Corr}[X,Y] := \frac{Cov[X,Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}}$$

Hinweis: Die Korrelation misst die Stärke und Richtung der *linearen Abhängigkeit* zweier Zufallsvariablen X und Y:

$$\text{Corr}[X, Y] = \pm 1 \Leftrightarrow \exists\, a \in \mathbb{R}, b > 0 : Y = a \pm bX$$

unkorreliert: Ist $\text{Corr}[X, Y] = 0$ und somit $\text{Cov}[X, Y] = 0$, dann heissen X und Y *unkorreliert.*

W.3 Wichtige Verteilungen

W.3.1 Diskrete Verteilungen

W.3.1.1 Diskrete Gleichverteilung

Eine diskret gleichverteilte Zufallsvariable $X \sim \mathcal{U}_T$ nimmt alle Werte im Wertebereich $\mathcal{W}(X) = \{x_1, \ldots, x_n\} := T$ mit gleicher Wahrscheinlichkeit an:

$$p_X(x_i) = \frac{1}{n} \quad \text{für } i \in \{1, \ldots, n\}$$

Beispiel (Würfeln): Die Zufallsvariable X gibt die Augenzahl bei einem Würfelwurf an. $\mathcal{W} = \{1, 2, 3, 4, 5, 6\}$, $n = 6$.

W.3.1.2 Bernoulli-Verteilung

Eine bernoulli-verteilte Zufallsvariable $X \sim Be(p)$ mit $p \in [0, 1]$ nimmt die Werte 0 und 1 mit Wahrscheinlichkeiten

$$p_X(1) = p \quad \text{und} \quad p_X(0) = 1 - p$$

an. Eine alternative Schreibweise ist

$$p_X(x) = \begin{cases} p^x(1-p)^{1-x} & x \in \{0, 1\} \\ 0 & \text{sonst.} \end{cases}$$

Erwartungswert : p
Varianz : $p(1 - p)$

Beispiel (Münzwurf): Ein fairer Münzwurf ist bernoulliverteilt mit Parameter $p = \frac{1}{2}$. Für einen Parameter $p \neq \frac{1}{2}$ wäre der Münzwurf unfair.

W.3.1.3 Binomialverteilung

Die Gewichtsfunktion p_X einer binomial-verteilten Zufallsvariable $X \sim Bin(n, p)$ mit Parameter $n \in \mathbb{N}$ und $p \in [0, 1]$ ist

$$p_x(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k \in \{0, \ldots, n\}$$

Erwartungswert : np
Varianz : $np(1 - p)$

X ist die Anzahl der Erfolge k bei n unabhängigen Wiederholungen eines Bernoulli-Experiments.

W.3.1.4 Geometrische Verteilung

Die Gewichtsfunktion p_X einer geometrisch-gleichverteilten Zufallsvariable $X \sim Geom(p)$ mit Parameter $p \in [0, 1]$ ist

$$p_X(k) = p(1-p)^{k-1} \quad \text{für } k \in \{1, 2, \ldots\}$$

Erwartungswert : $\frac{1}{p}$
Varianz : $\frac{1}{p^2}(1 - p)$

Beispiel (Wartezeit): Die Geometrische Verteilung ist die Wahrscheinlichkeitsverteilung der Anzahl X Bernoulli-Versuche, die notwendig sind, um den ersten Erfolg zu erzielen.

W.3.1.5 Negativbinomiale Verteilung

Die Gewichtsfunktion p_X einer negativ-binomial-verteilten Zufallsvariable X mit Parameter $r \in \mathbb{N}$ und $p \in [0, 1]$ ist

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad \text{für } k \in \{r, r+1, \ldots\}$$

Erwartungswert : $\frac{r}{p}$
Varianz : $\frac{r}{p^2}(1 - p)$

X entspricht der Wartezeit auf den r-ten Erfolg. Es gibt $\binom{k-1}{r-1}$ möglichkeiten für $r - 1$ Erfolge bei $k - 1$ Versuchen; der r-te Erfolg tritt ja beim k-ten Versuch ein.

W.3.1.6 Hypergeometrische Verteilung

Die Gewichtsfunktion p_X einer hypergeometrisch-verteilten Zufallsvariable X mit Parameter $r, n, m \in \mathbb{N}$, wobei $r, m \leq n$, ist

$$p_X(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} \quad \text{für } k \in \{0, \ldots, \min\{r, m\}\}$$

Erwartungswert : $m\frac{r}{n}$
Varianz : $m\frac{r}{n}(1 - \frac{r}{n})\frac{n-m}{n-1}$

In einer Urne befinden sich n Gegenstände. Davon sind r Gegenstände vom Typ A und $n - r$ vom Typ B. Es werden m Gegenstände ohne Zurücklegen gezogen. X beschreibt die Wahrscheinlichkeitsverteilung für die Anzahl k der Gegenstände vom Typ A in der Stichprobe.

Beispiel (Lotto): Anzahl Zahlen $n = 45$, richtige Zahlen $r = 6$, meine Zahlen $m = 6$. Die Wahrscheinlichkeit für 4 Richtige ist $p_X(4) \approx 0.00136$.

W.3.1.7 Poisson Verteilung

Die Gewichtsfunktion p_X einer Poisson-verteilten Zufallsvariable $X \sim \mathcal{P}(\lambda)$ mit Parameter λ ist gegeben durch

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{für } k \in \{0, 1, \ldots\}$$

Erwartungswert : λ
Varianz : λ

Die Poisson-Verteilung eignet sich zur Modellierung von seltenen Ereignissen, wie z.B. Versicherungsschäden.

Faustregel: Ab $np^2 \leq 0.05$ ist $Bin(n,p) \overset{\text{approx.}}{\approx} \mathcal{P}(\lambda), \lambda = np$

W.3.2 Stetige Verteilungen

W.3.2.1 Stetige Gleichverteilung

Die Dichte f_X und Verteilungsfunktion F_X einer stetigen und gleichverteilten Zufallsvariable $X \sim \mathcal{U}(a,b)$ mit Parameter $a, b \in \mathbb{R}$ wobei $a < b$ sind gegeben durch

$$f_X(t) = \tfrac{1}{b-a}\mathbb{1}_{t\in[a,b]}$$

$$F_X(t) = \begin{cases} \frac{t-a}{b-a} & \mathbb{1}_{t\in[a,b]} \\ 1 & \text{für } t > b \end{cases}$$

Erwartungswert : $\frac{1}{2}(a+b)$
Varianz : $\frac{1}{12}(a-b)^2$

Beispiel: Zufällige Wahl eines Punktes aus $[a,b]$

W.3.2.2 Exponentialverteilung

Die Dichte f_X und Verteilungsfunktion F_X einer exponentialverteilten Zufallsvariable $X \sim Exp(\lambda)$ mit Parameter $\lambda > 0$ sind

$$f_X(t) = \lambda e^{-\lambda t}\mathbb{1}_{t\geq 0}$$

$$F_X(t) = 1 - e^{-\lambda t}\mathbb{1}_{t\geq 0}$$

Erwartungswert : $\frac{1}{\lambda}$
Varianz : $\frac{1}{\lambda^2}$

Beispiel (Lebensdauer): Die Exponentialverteilung ist eine typische Lebensdauerverteilung. So ist beispielsweise die Lebensdauer von elektronischen Bauelementen häufig annähernd exponentialverteilt.

W.3.2.3 Normalverteilung

Die Dichte f_X einer normalverteilten Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$ mit Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ ist gegeben durch

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

Für die Verteilungsfunktion F_X existiert kein geschlossener Ausdruck. Deshalb werden die Werte der Verteilungsfunktion $\Phi(t)$ der *Standard-Normalverteilung* $\mathcal{N}(0,1)$ tabelliert.
Für allgemeine Normalverteilungen berechnet man dann

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

Erwartungswert : μ
Varianz : σ^2

Beispiel: Streuung von Messwerten um den Mittelwert.

W.3.2.4 Gamma-Verteilung

Die Dichte f_Z einer Zufallsvariablen $Z \sim Ga(\alpha, \lambda)$ mit $\alpha > 0, \lambda > 0$ ist gegeben durch

$$f(z) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha} z^{\alpha-1} e^{-\lambda z} \cdot \mathbb{1}_{x>0}$$

Erwartungswert : $\frac{\alpha}{\lambda}$
Varianz : $\frac{\alpha}{\lambda^2}$

Gammafunktion: $\Gamma(z)$ ist die Erweiterung der Fakultät auf reele und komplexe Argumente $\Gamma(z+1) = z \cdot \Gamma(z), \quad \Gamma(1) = 1$

$$\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}\mathrm{d}t, \qquad \text{für } z \in \mathbb{C}, Re(z) > 0$$

Speziell gilt für $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$ und $\Gamma(1/2) = \sqrt{\pi}$

Hinweis: Die Gamma-Verteilung ist eine Verallgemeinerung der Exponentialverteilung: $Ga(1,\lambda) \Leftrightarrow Exp(\lambda)$

W.3.2.5 Chiquadrat-Verteilung

Die Dichte f_Y einer χ^2_n-verteilten Zufallsvariablen $Y \sim \chi^2_n$ mit n Freiheitsgraden, ist gegeben durch

$$f_Y(y) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{1}{2}y} \cdot \mathbb{1}_{x\geq 0}$$

Erwartungswert : n
Varianz : $2n$

Die χ^2_n-Verteilung mit n Freiheitsgraden beschreibt die Verteilung der Summe $Y = \sum_{i=0}^{n} X_i^2$, für $X_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$

Hinweis: Die χ^2_n-Verteilung ist ein Spezialfall der Gamma-Verteilung: $Ga(\frac{n}{2}, \frac{1}{2}) \Leftrightarrow \chi^2_n$ und somit auch $\chi^2_2 \Leftrightarrow Exp(\frac{1}{2})$

W.3.2.6 t-Verteilung

Die Dichte f_Z einer t_n-verteilten Zufallsvariablen $Z \sim t_n$ mit $n > 0$ Freiheitsgraden, ist gegeben durch:

$$f_Z(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}$$

Erwartungswert : 0 für $n > 1$
Varianz : $\frac{n}{n-1}$ für $n > 2$

Die t-Verteilung mit n Freiheitsgraden beschreibt die Verteilung von $Z = \frac{X}{\sqrt{\frac{1}{n}Y}}$, für $X \perp Y$, $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi^2_n$

Hinweis: Für n=1 ist die t-Verteilung eine *Cauchy-Verteilung*, und für $n \to \infty$ erhält man die $\mathcal{N}(0,1)$-Verteilung

W.3.2.7 Cauchy-Verteilung

Die Cauchy-Verteilung ist die t-Verteilung mit einem Freiheitsgrad: Erwartungswert und Varianz sind nicht definiert. Für X,Y Cauchy-verteilt, ist $\frac{X+Y}{2}$ auch Cauchy-verteilt.

$$f(x) = \frac{1}{\pi}\frac{1}{1+(x-\mu)^2}$$

W.4 Zufallsvariablen in $\mathbb{R}^n$

Verteilung: Das stochastische Verhalten einer Zufallsvariablen $X = (X_1, \dots, X_n)$ über **einem** Wahrscheinlichkeitsraum wird durch ihre *Verteilung* beschrieben. $\mu_X : \mathbb{R}^n \to [0,1]$

$$\begin{aligned}\mu_X(B) &:= \mathbb{P}[(X_1, \dots, X_n) \in B] \\ &:= P[\{\omega \mid (X_1(\omega), \dots, X_n(\omega)) \in B\}] \quad \text{für } B \subseteq \mathbb{R}^n\end{aligned}$$

Hinweis: Jedes Wahrscheinlichkeitsmass μ_X erfüllt:
1. $\mu_X(B) \geq 0$ für alle $B \subseteq \mathbb{R}^n$
2. $\mu_X(\mathcal{W}(X)) = \mu_X(\mathbb{R}^n) = 1$

W.4.1 Gemeinsame Verteilungen

Gemeinsame Verteilung: Die *gemeinsame Verteilungsfunktion* ist die Abbildung $F_X : \mathbb{R}^n \to [0,1]$,

$$\begin{aligned}F_X(t_1, \dots, t_n) &:= \mathbb{P}[X_1 \leq t_1, \dots, X_n \leq t_n] \\ &:= \mu_X((-\infty, t_1] \times \ldots \times (-\infty, t_n])\end{aligned}$$

Gemeinsame Gewichtsfunktion: Im diskreten Fall ist die *gemeinsame Gewichtungsfunktion* $p_X : \mathbb{R}^n \to [0,1]$ definiert:

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

Somit lässt sich die *diskrete Verteilung* μ_X berechnen:

$$\mu_X(B) = \sum_{(x_1,\dots,x_n)_i \in B} p_X((x_1, \dots, x_n)_i)$$

Gemeinsame Dichte: Im stetigen Fall ist die *gemeinsame Dichte* $f_X : \mathbb{R}^n \to [0, \infty)$ definiert, falls für alle $t_i \in \mathbb{R}$ gilt

$$F_X(t_1, \dots, t_n) = \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_n} f_X(x_1, \dots, x_n) \mathrm{d}x_n \dots \mathrm{d}x_1$$

Somit lässt sich die *stetige Verteilung* μ_X berechnen:

$$\mu_X(B) = \int_{(t_1,\dots,t_n)\in B} f_X(t_1, \dots, t_n) \mathrm{d}\mu$$

Hinweis: $f_X(t_1, \dots, t_n) = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} F(t_1, \dots, t_n)$

W.4.2 Randverteilungen

Randverteilung: Seien X und Y Zufallsvariablen mit gemeinsamer Verteilungsfunktion $F_{X,Y}$, dann ist die *Randverteilung* $F_X : \mathbb{R} \to [0,1]$ von X definiert durch

$$F_X(x) := \mathbb{P}[X \leq x] = \mathbb{P}[X \leq x, Y < \infty] = \lim_{y \to \infty} F_{X,Y}(x, y).$$

Randgewichtsfunktion: Für zwei diskrete Zufallsvariablen X und Y mit gemeinsamer Gewichtsfunktion $p_{X,Y}(x,y)$ ist die Gewichtsfunktion der Randverteilung von X gegeben:

$$p_X(x) = \mathbb{P}[X = x] = \sum_j \mathbb{P}[X = x, Y = y_j] = \sum_j p_{X,Y}(x, y_j).$$

Randdichte: Für zwei stetige Zufallsvariablen X und Y mit gemeinsamer Dichte $f_{X,Y}(x,y)$ ist die Randdichte (Dichtefunktion der Randverteilung) von X gegeben durch

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \mathrm{d}y.$$

W.4.3 Bedingte Verteilung

Bedingte Gewichtsfunktion: Seien X und Y diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion $p_{X,Y}(x,y)$, dann ist die *bedingte Gewichtsfunktion* $p_{X|Y}(x \mid y)$ von X gegeben Y definiert durch

$$p_{X|Y}(x \mid y) := \mathbb{P}[X = x \mid Y = y] = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

falls $p_Y(y) > 0$ und 0 falls $p_Y(y) = 0$.

Bedingte Dichte: Für zwei stetige Zufallsvariablen X und Y mit gemeinsamer Dichte $f_{X,Y}(x,y)$ ist die *bedingte Dichte* $f_{X|Y}$ von X gegeben Y definiert durch

$$f_{X|Y}(x \mid y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

falls $f_Y(y) > 0$ und 0 falls $f_Y(y) = 0$.

W.4.4 Unabhängigkeit

Unabhängigkeit: Die Zufallsvariablen $X_1, \dots, X_n$ heissen *unabhängig*, falls gilt:

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

Hinweis: Es gelten folgende Rechenregeln:
- i: $p(x_1, \dots, x_n) = \prod p_{X_i}(x_i)$
- ii: $f(x_1, \dots, x_n) = \prod f_{X_i}(x_i)$
- iii: $\mathcal{M}_{(X_1,\dots,X_n)}(t_1, \dots, t_n) = \prod \mathcal{M}_{X_i}(t_i)$
- iv: $Y_i = f_i(X_i)$ sind für beliebige f_i unabhängig

i.i.d Annahme: Die Abkürzung *i.i.d.* kommt vom Englischen *independent and identically distributed.*

Hinweis: Es gelten die Implikationen:
unabhängig $\Rightarrow$ paarweise unabhängig $\Rightarrow$ unkorreliert

W.4.5 Erwartungswert und Varianz

Hinweis: Der Erwartungswert einer n-dimensionalen Verteilung wird als n-Tupel der Erwartungswerte aller Randverteilungen $\mathbb{E}[X_i]$ angegeben.

Satz (4.2): Für den Erwartungswert $\mathbb{E}[Y]$ einer Funktion $Y := g(X_1, \dots X_n)$ der Zufallsvariablen $X_1, \dots, X_n$ gilt

$$\mathbb{E}[Y] = \sum_{x_1,\dots,x_n} g(x_1, \dots, x_n) p(x_1, \dots, x_n)$$

$$\mathbb{E}[Y] = \underset{\mathbb{R}^n}{\int \dots \int} g(x_1, \dots, x_n) f(x_1, \dots, x_n) \mathrm{d}x_n \dots \mathrm{d}x_1.$$

Hinweis: Es gelten folgende Rechenregeln:
- i: $\mathbb{E}\left[a + \sum_{i=1}^n b_i X_i\right] = a + \sum_{i=1}^n b_i \mathbb{E}[X_i]$
- ii: $\mathbb{E}[\prod_{i=1}^n X_i] = \prod_{i=1}^n X_i \mathbb{E}[X_i] \Leftrightarrow X_i$ unabhängig
- iii: $\mathrm{Var}[a + \sum_{i=1}^n b_i X_i] = \sum_{i=1}^n b_i^2 \mathrm{Var}[X_i]$, für X_i unabhängig
- iv: $\mathrm{Cov}[a + \sum_{i=1}^n b_i X_i, c + \sum_{j=1}^m d_j Y_j]$
  $= \sum_{i=1}^n \sum_{j=1}^m b_i d_j \mathrm{Cov}[X_i, Y_j]$

W.5 Funktionen von Zufallsvariablen

W.5.1 Transformationen

Satz: Sei X eine stetige Zufallsvariable mit Dichte f_X und $f_X(t) = 0$ für $t \notin I \subseteq \mathbb{R}$. Sei $g : \mathbb{R} \to \mathbb{R}$ stetig differenzierbar und streng monoton auf I mit Umkehrfunktion g^{-1}. Dann hat die Zufallsvariable $Y := g(X)$ die Dichte

$$f_Y = \begin{cases} f_X(g^{-1}(t))|\frac{\mathrm{d}}{\mathrm{d}t}g^{-1}(t)| & \text{für } t \in \{g(x) \mid x \in I\} \\ 0 & \text{sonst} \end{cases}$$

Beispiel (Lineare Transformation): Aus $Y := aX+b$ mit $a > 0, b \in \mathbb{R}$ folgt

$$F_Y(t) = \mathbb{P}[aX+b \leq t] = \mathbb{P}\left[X \leq \frac{t-b}{a}\right] = F_X\left(\frac{t-b}{a}\right)$$

$$\Rightarrow f_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_Y(t) = \frac{1}{a}f_X\left(\frac{t-b}{a}\right).$$

Beispiel (Nichtlineare Transformation): $Y := X^2$

$$F_Y(t) = \mathbb{P}[X^2 \leq t] = \mathbb{P}\left[-\sqrt{t} \leq X \leq \sqrt{t}\right] = F_X\left(\sqrt{t}\right) - F_X\left(-\sqrt{t}\right)$$

$$\Rightarrow f_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_Y(t) = \frac{f_X\left(\sqrt{t}\right)+f_X\left(-\sqrt{t}\right)}{2\sqrt{t}}$$

W.5.1.1 Simulation von Verteilungen

Satz: Sei F eine stetige und streng monoton wachsende Verteilungsfunktion mit Umkehrfunktion F^{-1}. Ist $X \sim \mathcal{U}(0,1)$ und $Y := F^{-1}(X)$, so hat Y die Verteilungsfunktion F.

Beispiel: Um die Verteilung $Exp(\lambda)$ zu simulieren bestimmt man zu der Verteilungsfunktion $F(t) = 1 - e^{-\lambda t}$ für $t \geq 0$ die Inverse $F^{-1}(t) = -\frac{\log(1-t)}{\lambda}$. Mit $U \sim \mathcal{U}(0,1)$ erhält man

$$X := F^{-1}(U) = -\frac{\log(1-U)}{\lambda} \sim Exp(\lambda).$$

W.5.2 Funktionen

Ausgehend von der Zufallsvariablen $X = (X_1, \dots, X_n)$ kann mit einer Funktion $g : \mathbb{R}^n \to \mathbb{R}$ eine neue Zufallsvariable $Y = g(X_1, \dots, X_n)$ bilden. Für die Verteilung μ_Y bedeutet dies:

$$\mu_Y(B) = \mu_X(\{\vec{x} \in \mathbb{R}^n \mid g(\vec{x}) \in B\})$$

Danach versuchen wir die rechte Seite auszurechnen, indem wir die genauere Struktur der Transformation g ausnutzen:

Beispiel (Summe): Für die stetige/diskrete Dichte der Summe $Z = X+Y$ zweier Zufallsvariablen X, Y gilt somit:

$$\begin{aligned} p_Z(z) & \\ &= \mu_Z(\{z\}) \\ &= \mu_{X,Y}(\{(x,y) \mid x+y = z\}) \\ &= \sum_{\substack{(x_i,y_i) \\ y_i = z - x_i}} p_{X,Y}(x_i, y_i) \\ &= \sum_{x_i} p_{X,Y}(x_i, z - x_i) \end{aligned} \qquad \begin{aligned} F_Z(z) & \\ &= \mu_Z(\{Z \leq z\}) \\ &= \mu_{X,Y}(\{(x,y) \mid x+y \leq z\}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)\mathrm{d}y\mathrm{d}x \\ &\overset{v=x+y}{=} \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x, v-x)\mathrm{d}x\mathrm{d}v \end{aligned}$$

$$\Rightarrow f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z}F_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\mathrm{d}x.$$

W.5.2.1 Spezielle Funktionen

Wichtige Spezialfälle sind die Summe $S_n = \sum_{i=1}^n X_i$ und das arithmetische Mittel $\overline{X}_n = \frac{S_n}{n}$ für X_i unabhängig.

1. Wenn $X_i \sim Be(p)$, dann ist $S_n \sim Bin(n,p)$.
2. Wenn $X_i \sim Bin(n_i, p)$, dann ist $S_n \sim Bin(\sum n_i, p)$.
3. Wenn $X_i \sim \mathcal{P}(\lambda_i)$, dann ist $S_n \sim \mathcal{P}(\sum \lambda_i)$.
4. Wenn $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, dann ist $S_n \sim \mathcal{N}(\sum \mu_i, \sum \sigma_i^2)$.
5. Wenn $X_i \sim Ga(\alpha_i, \lambda)$, dann ist $S_n \sim Ga(\sum \alpha_i, \lambda)$

Für den Erwartungswert und die Varianz gilt allgemein

$$\mathbb{E}[S_n] = n\mathbb{E}[X_i] \qquad \mathrm{Var}[S_n] = n\mathrm{Var}[X_i]$$

W.6 Grenzwertsätze

W.6.1 Gesetz der grossen Zahlen

Satz (Schwaches GGZ): Für eine Folge $X_1, X_2, \dots$ von unkorrelierten Zufallsvariablen, die alle den Erwartungswert $\mu = \mathbb{E}[X_i]$ und die Varianz $\mathrm{Var}[X_i] = \sigma^2$ haben, gilt

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \quad \overset{n\to\infty}{\longrightarrow} \quad \mu = \mathbb{E}[X_i].$$

Das heisst

$$\mathbb{P}\left[|\overline{X}_n - \mu| > \epsilon\right] \overset{n\to\infty}{\longrightarrow} 0 \quad \forall \epsilon > 0.$$

Satz (Starkes GGZ): Für eine Folge $X_1, X_2, \dots$ unabhängiger Zufallsvariablen, die alle den endlichen Erwartungswert $\mu = \mathbb{E}[X_i]$ haben, gilt

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \quad \overset{n\to\infty}{\longrightarrow} \quad \mu = \mathbb{E}[X_i]. \quad \text{P-fastsicher}$$

Das heisst

$$\mathbb{P}\left[\{\omega \in \Omega \mid \overline{X}_n(\omega) \overset{n\to\infty}{\longrightarrow} \mu\}\right] = 1.$$

W.6.2 Zentraler Grenzwertsatz

Satz (ZGS): Sei $X_1, X_2, \dots$ eine Folge von i.i.d. Zufallsvariablen mit $\mu = \mathbb{E}[X_i]$ und $\sigma^2 = \mathrm{Var}[X_i]$, dann gilt für die Summe $S_n = \sum_{i=1}^n X_i$

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

wobei Φ die Verteilungsfunktion von $\mathcal{N}(0,1)$ ist.

Hinweis: Die Summe S_n hat Erwartungswert $\mathbb{E}[S_n] = n\mu$ und Varianz $\mathrm{Var}[S_n] = n\sigma^2$. Die Grösse

$$S_n^* := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}$$

hat Erwartungswert 0 und Varianz 1. Für grosse n gilt zudem:

$$\begin{aligned} \mathbb{P}[S_n^* \leq x] &\approx \Phi(x) \\ S_n^* &\overset{\text{approx.}}{\sim} \mathcal{N}(0,1) \\ S_n &\overset{\text{approx.}}{\sim} \mathcal{N}(n\mu, n\sigma^2) \end{aligned}$$

W.6.3 Ungleichungen

Markov: Für eine wachsende Funktion $g : \mathbb{R} \to [0, \infty]$ mit $g(c) > 0$ **für alle c** und eine Indikatorvariable I gilt:

$$\left.\begin{array}{rcl} g(c)I_{\{g(c)\leq g(X)\}} & \leq & g(X) \\ g(c)I_{\{X\geq c\}} & \leq & g(X) \\ g(c)\mathbb{E}(I_{\{X\geq c\}}) & \leq & \mathbb{E}(g(X)) \\ g(c)\mathbb{P}[X\geq c] & \leq & \mathbb{E}(g(X)) \end{array}\right\} \quad \mathbb{P}[X \geq c] \leq \frac{\mathbb{E}[g(X)]}{g(c)}$$

Chebychev: Sei $X = |Y - \mathbb{E}[Y]|$ und $g(x) = x^2$, dann folgt:

$$\mathbb{P}[|Y - \mathbb{E}[Y]| \geq c] \quad \overset{\text{Markov}}{\leq} \quad \frac{\mathbb{E}[|Y - \mathbb{E}[Y]|^2]}{c^2} \quad = \quad \frac{\text{Var}[Y]}{c^2}$$

Chernoff: Seien $X_1, \ldots, X_n$ unabhängig mit $X_i \sim Be(p_i)$. Wir wählen $S_n = \sum X_i$ mit $\mathbb{E}[S_n] = \sum p_i$ und $g(c) = exp(ct)$

$$\begin{aligned} \mathbb{P}[S_n \geq c] &\overset{\text{Markov}}{\leq} \frac{1}{g(c)}\mathbb{E}[g(S_n)] \\ &\overset{g(c),S_n}{=} \frac{1}{e^{tc}}\mathbb{E}[exp(t\sum X_i)] \\ &\overset{X_i \perp X_j}{=} \frac{1}{e^{tc}}\prod \mathbb{E}[e^{tX_i}] \\ &\overset{\text{Def } \mathbb{E}[X_i]}{=} \frac{1}{e^{tc}}\prod e^{t\cdot 0}(1-p_i) + e^{t\cdot 1}p_i \\ &\overset{\text{arith.}}{=} \frac{1}{e^{tc}}\prod 1 + p_i(e^t - 1) \\ &\overset{1+z\leq e^z}{\leq} \frac{1}{e^{tc}}\prod exp(p_i(e^t - 1)) \\ &\overset{arith.}{=} \frac{1}{e^{tc}}exp\big(\sum p_i(e^t - 1)\big) \\ &\overset{\mathbb{E}[S_n]}{=} exp\big(\mathbb{E}[S_n](e^t - 1) - tc\big) \end{aligned}$$

Nun setzen wir $\mathbb{E}[S_n] = \mu_n$, $c = (1\pm\delta)\mu_n$ für $\delta > 0$ und wählen $t = log(1 \pm \delta)$, denn so wird die rechte Seite minimiert.

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

W.6.4 Monte Carlo Integration

Das Integral

$$I := \int_0^1 g(x)dx$$

lässt sich als Erwartungswert auffassen, denn mit einer gleichverteilten Zufallsvariable $U \sim \mathcal{U}(0, 1)$ folgt

$$\mathbb{E}[g(U)] = \int_{-\infty}^{\infty} g(x)f_U(x)\mathrm{d}x = \int_0^1 g(x)\mathrm{d}x.$$

Mit einer Folge von Zufallsvariablen $U_1, \ldots, U_n$, die unabhängig gleichverteilt $U_i \sim \mathcal{U}(0, 1)$ sind, lässt sich das Integral approximieren: Nach dem schwachen Gesetz der grossen Zahlen gilt

$$\overline{g(U_n)} = \frac{1}{n}\sum_{i=1}^{n} g(U_i) \overset{n\to\infty}{\longrightarrow} \mathbb{E}[g(U_1)] = I.$$

Statistik

S.1 Grundlagen

Stichprobe: Die Gesamtheit der Beobachtungen $x_1, \ldots, x_n$ oder der Zufallsvariablen $X_1, \ldots, X_n$ wird *Stichprobe* genannt; die Anzahl n heisst *Stichprobenumfang.*

Hinweis: Man muss immer sauber unterscheiden zwischen den *Daten* $x_1, \ldots, x_n$ (konkrete Zahlen) und dem generierenden *Mechanismus* $X_1, \ldots, X_n$ (Zufallsvariablen auf Ω)

Empirische Verteilungsfunktion: Die *empirische Verteilungsfunktion* F_n zu den Messdaten $x_1, \ldots, x_n$ ist definiert

$$F_n(y) := \tfrac{1}{n}|\{x_i \mid x_i \leq y\}| = \tfrac{1}{n}\textstyle\sum_1^n \mathbb{1}_{x_i\leq y}$$

Empirischer Mittelwert: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$

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

k-tes Empirisches Moment: $\hat{m}_k := \frac{1}{n}\sum_{i=1}^n X_i^k$

Empirisches Quantil: Das *empirische α-Quantil* x_α teilt die geordneten Daten $y_{(1)} \leq \ldots \leq y_{(n)}$ so, dass der Anteil α unterhalb des empirischen α-Quantils liegt. Index $k = \lfloor \alpha n \rfloor$

$$x_\alpha = y_{(k)} + \alpha\left(y_{(k+1)} - y_{(k)}\right)$$

Empirischer Median: ist definiert als das 0.5-Quantil.

S.2 Schätzer

Modell: Für eine Stichprobe $X_1, \ldots, X_n$ soll ein passendes Modell gefunden werden. Wir haben also einen Parameterraum $\Theta \subseteq \mathbb{R}^m$ und für jedes $\vartheta \in \Theta$ existiert ein Modell $\mathbb{P}_\vartheta$.

Die Wahl von ϑ bestimmt also das Wahrscheinlichkeitsmass $\mathbb{P}_\vartheta$. Wir wählen nun ein Modell aus, indem wir versuchen die Parameter $\vartheta = (\vartheta_1, \ldots, \vartheta_m)$ aufgrund der Stichprobe mit einem *Schätzer* $T = (T_1, \ldots, T_m)$ herauszufinden.

Schätzer: Schätzer sind Zufallsvariablen T_j, die eine Berechnungsmethode zur Schätzung der Parameters ϑ_j angeben. Dabei wird nur angenommen, dass die X_i nach dem Modell $\mathbb{P}_\vartheta$ verteilt sind.

$$T_j = t_j(X_1, \ldots, X_n)$$

für eine geeignete Funktion $t_j : \mathbb{R}^n \to \mathbb{R}$.

Schätzwerte: Durch Einsetzen von Daten x_i erhält man *Schätzwertwerte* $t_j(x_1, \ldots, x_n)$ für ϑ_j.

Erwartungstreu: Ein Schätzer T heisst *erwartungstreu* für ϑ, falls $\mathbb{E}_\vartheta[T] = \vartheta$ (im Mittel wird richtig geschätzt).

mean-squared error: $\text{MSE}_\vartheta[T] := \mathbb{E}_\vartheta[(T - \vartheta)^2]$.

Konsistent: Eine Folge von Schätzern $T^{(n)}, n \in \mathbb{N}$ heisst *konsistent* für ϑ, falls $T^{(n)}$ für $n \to \infty$ im Modell $\mathbb{P}_\vartheta$ gegen ϑ konvergiert. Das heisst für jedes $\vartheta \in \Theta$ und $\epsilon > 0$ gilt

$$\lim_{n\to\infty} \mathbb{P}_\vartheta[\,|T^{(n)} - \vartheta| > \epsilon\,] = 0.$$

S.2.1 Momenten-Methode

Die Parameter ϑ_i der theoretischen Verteilung werden als Funktion der Momente m_k angegeben.

$$\vartheta_j = g_j(m_1, \ldots, m_m) \quad \text{für } j \in \{1, \ldots, m\}$$

Den *Momentenschätzer* für $\vartheta = (\vartheta_1, \ldots, \vartheta_m)$ erhält man, indem man die Stichprobenmomente in die Funktionen der Momente einsetzt; der Schätzer ist also $T = (T_1, \ldots, T_m)$ mit

$$T_j := g_j(\hat{m}_1, \ldots, \hat{m}_m) \quad \text{für } j \in \{1, \ldots, m\}$$

Beispiel: Gegeben seien n unabhängige Realisierungen $x_1, \ldots, x_n$ einer Zufallsvariablen $X \sim \mathcal{P}(\lambda)$. Es gilt $\mathbb{E}[X] = \lambda$. Für die Funktion g_1 kann also die Idendität gewählt werden. Der Momentenschätzer ist somit

$$\lambda_{\mathrm{MM}} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Es gilt aber auch $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda$. Es kann also auch $g_1(m_1, m_2) = m_2 - m_1^2$ gewählt werden. Dadurch erhält man einen anderen Momentenschätzer

$$\lambda_{\mathrm{MM}} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

S.2.2 Maximum-Likelihood

Es wird von Zufallsvariablen $X_1, \ldots, X_n$ ausgegangen, deren gemeinsame Dichte $f(t_1, \ldots, t_n; \vartheta)$ von einem Parameter ϑ abhängt. Die *Likelihood-Funktion* $\mathcal{L}$ ist gegeben durch

$$\mathcal{L}(x_1, \ldots, x_n; \vartheta) = f(x_1, \ldots, x_n; \vartheta).$$

Anschaulich ist das die Wahrscheinlichkeit[1], dass im Modell $\mathbb{P}_\vartheta$ die Stichprobe $X_1, \ldots, X_n$ die Werte $x_1, \ldots, x_n$ liefert. Um eine möglichst gute Anpassung des Modells an die Daten zu erreichen, wird der Likelihood-Schätzer als Funktion von ϑ maximiert.

Hinweis: Im diskreten Fall wird lediglich die Dichte f durch die Gewichtsfunktion p ersetzt.

Oft sind die Zufallsvariablen X_i unter $\mathbb{P}_\vartheta$ i.i.d. mit Dichtefunktion $f(t; \vartheta)$, so dass sich die Likelihood-Funktion vereinfacht zu

$$\mathcal{L}(x_1, \ldots, x_n; \vartheta) = \prod_{i=1}^{n} f(x_i; \vartheta).$$

Aufgrund der Monotonie des Logarithmus kann dann die logarithmierte Likelihood-Funktion verwendet werden, ohne dass sich dadurch das Maximum der Funktion verschiebt.

$$\log \mathcal{L}(x_1, \ldots, x_n; \vartheta) = \sum_{i=1}^{n} \log f(x_i; \vartheta)$$

Beispiel: Gegeben seien n unabhängige Realisierungen $x_1, \ldots, x_n$ einer Zufallsvariable $X \sim Exp(\lambda)$ mit Dichte $f(t) = \lambda e^{-\lambda t}\mathbb{1}_{[0,\infty)}(t)$ und unbekanntem Parameter λ. Für die Likelihood-Funktion erhält man

$$\mathcal{L}(\lambda) := \mathcal{L}(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

und durch logarithmieren

$$\log \mathcal{L}(\lambda) = \sum_{i=1}^{n} \log \lambda e^{-\lambda x_i} = n \log \lambda - \lambda \sum_{i=1}^{n} x_i.$$

Zur Bestimmung des Maximums wird die Ableitung nullgesetzt:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \log \mathcal{L}(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \overset{!}{=} 0 \quad \Rightarrow \quad \lambda^* = \frac{n}{\sum_{i=1}^{n} x_i}$$

Aus $\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathcal{L}(\lambda) = -\frac{n}{\lambda^2} < 0$ für $\lambda > 0$ folgt, dass es sich auch tatsächlich um ein Maximum handelt. Der ML-Schätzer T für den unbekannten Parameter λ ist also gegeben durch

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

S.2.3 Verteilungsaussagen

In vielen Situation ist es nützlich oder nötig, die *Verteilung* eines Schätzers unter $\mathbb{P}_\vartheta$ (für alle $\vartheta \in \Theta$) zu kennen.

Verteilung: Das stochastische Verhalten eines Schätzers T unter $\mathbb{P}_\vartheta$ wird durch seine *Verteilung* $\mu_{T,\vartheta}$ beschrieben

$$\mu_{T,\vartheta}(B) := \mathbb{P}_\vartheta[T \in B] \qquad \text{für } B \subseteq \Theta.$$

bei Summen: Oft ist ein Schätzer T eine Funktion einer Summe $\sum X_i$, wobei die X_i im Modell $\mathbb{P}_\vartheta$ i.i.d. sind. Nach dem Zentralen Grenzwertsatz gilt dann für grosse n

$$\sum_{i=1}^{n} X_i \overset{\text{approx.}}{\sim} \mathcal{N}(n\mu, n\sigma) \qquad \mu = \mathbb{E}_\vartheta[X_i], \sigma = \mathrm{Var}_\vartheta[X_i]$$

Satz (7.1): Für $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma)$ gilt:

1. Sei die Zufallsvariable $U = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$.
   Aus $\overline{X}_n \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$, folgt $U \sim \mathcal{N}(0,1)$
2. Sei die Zufallsvariable $V = \frac{n-1}{\sigma^2}S^2 = \sum_{i=1}^{n}(\frac{X_i - \overline{X}_n}{\sigma})^2$ und
   $$W = \sum(\tfrac{X_i-\mu}{\sigma})^2 = \sum(\tfrac{X_i - \overline{X}_n}{\sigma} + \tfrac{\overline{X}_n - \mu}{\sigma})^2 = V + U^2$$
   Da $W \sim \chi^2_n$, $U^2 \sim \chi^2_1$ und $U^2 \perp V$, gilt $V \sim \chi^2_{n-1}$
3. Unabhängigkeiten: $\overline{X}_n \perp S^2$ und $U \perp V$
   Da $\overline{X}_n \perp (.., X_i - \overline{X}_n, ..)$ und $f(\overline{X}_n) \perp g(.., X_i - \overline{X}_n, ..)$
4. Sei die Zufallsvariable $Z = \frac{\overline{X}_n - \mu}{S/\sqrt{n}}$. Nun gilt:
   $$Z = \frac{\overline{X}_n-\mu}{S/\sqrt{n}} = \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\frac{1}{\frac{1}{\sigma}S} = U\frac{1}{\sqrt{\frac{1}{n-1}\frac{n-1}{\sigma^2}S^2}} = \frac{U}{\sqrt{\frac{1}{n-1}V}}$$

Da $U \sim \mathcal{N}(0,1)$, $V \sim \chi^2_{n-1}$ und $U \perp V$, ist $Z \sim t_{n-1}$ verteilt. Formal werden Schritte 2 und 3 via $\mathcal{M}_X(t)$ bewiesen.

[1] oder zumindest das stetige Pendant zur Wahrscheinlichkeit.

S.3 Tests

S.3.1 Grundlagen

Ausgangspunkt ist eine Stichprobe $X_1, \dots, X_n$ in einem Modell $\mathbb{P}_\vartheta$ mit unbekanntem Parameter $\vartheta \in \Theta$.

Modellklassen: Aufgrund einer Vermutung, wo sich der richtige Parameter ϑ befindet, werden eine *Hypothese* $\Theta_0 \subseteq \Theta$ und eine *Alternative* $\Theta_A \subseteq \Theta$ mit $\Theta_0 \cap \Theta_A = \emptyset$ formuliert:

$$\text{Hypothese } H_0 : \vartheta \in \Theta_0$$
$$\text{Alternative } H_A : \vartheta \in \Theta_A$$

Hinweis: Die Hypothese (bzw. Alternative) heisst *einfach*, falls sie nur aus einem einzelnen Wert besteht, z.B. $\Theta_0 = \{\vartheta_0\}$

Test: Eine Interpretation der Übereinstimmung zwischen den Daten $x_1, \dots, x_n$ und einem vermutetem Modell.

1. *Entscheidungsregel:* $\mathbb{1}_{t(x_1,\dots,x_n)\in K}$
2. *Teststatistik:* $T = t(X_1, \dots, X_n)$
3. *Verwerfungsbereich:* $K \subseteq \mathbb{R}$

Fehler 1. Art: Die Hypothese wird abgelehnt, obwohl sie richtig ist. Die Wahrscheinlichkeit für einen Fehler 1. Art ist

$$\mathbb{P}_\vartheta[T \in K] \quad \text{für } \vartheta \in \Theta_0.$$

Fehler 2. Art: Die Hypothese ϑ wird akzeptiert, obwohl sie falsch ist. Die Wahrscheinlichkeit für einen Fehler 2. Art ist

$$\mathbb{P}_\vartheta[T \notin K] = 1 - \mathbb{P}_\vartheta[T \in K] \quad \text{für } \vartheta \in \Theta_A.$$

S.3.2 Konstruktion von Tests (T, K)

1. Wähle eine *Hypothese* Θ_0 (normalerweise $\{\vartheta_0\} = \Theta_0$).
2. Wähle ein *Signifikanzniveau* $\alpha \in (0, 1)$. Dadurch werden Fehler 1. Art von α kontrolliert.
$$\sup_{\vartheta\in\Theta_0} \mathbb{P}_\vartheta[T \in K] \leq \alpha.$$
3. Maximiere die *Macht* β des Tests. Dadurch werden Fehler 2. Art minimiert.
$$\beta(\vartheta) := \mathbb{P}_\vartheta[T \in K] \qquad \text{für } \vartheta \in \Theta_0$$
4. Falls der realisierte Wert $T(\omega)$ im Verwerfungsbereich K_α liegt, wird die Nullhypothese auf dem Niveau α verworfen

Hinweis: Die Hypothese zu verwerfen, ist schwieriger. Es wird häufig die Negation der gewünschten Aussage benutzt.

P-Wert: Der P-Wert ist die Wahrscheinlichkeit, dass unter der Nullhypothese H_0 ein zufälliger Versuch mindestens so extrem ausfällt, wie der beobachtete Wert $T(\omega)$.

$$\textbf{p-Wert} = \begin{cases} \mathbb{P}_{\vartheta_0}[T \leq T(\omega)] & \text{für} \quad H_A : \vartheta < \vartheta_0 \\ \mathbb{P}_{\vartheta_0}[T \geq T(\omega)] & \text{für} \quad H_A : \vartheta > \vartheta_0 \end{cases}$$

Likelihood-Quotienten Test: Als Teststatistik wird der *Likelihood-Quotient* $\mathcal{R}$ gewählt,

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

Ist dieser Quotient klein, sind die Beobachtungen im Modell $\mathbb{P}_{\Theta_A}$ deutlich wahrscheinlicher als im Modell $\mathbb{P}_{\Theta_0}$. Der Verwerfungsbereich $K := [0, c)$ wird so gewählt, dass der Test das gewünschte Signifikanzniveau einhält.

Satz (Neyman-Pearson): Sind Hypothese und Alternative beide einfach, so ist der Test optimal bzgl. Fehler 1. & 2. Art

S.3.3 Einstichprobentests

Neyman-Pearson gibt uns eine Systematische Methode um 'gute' Tests zu konstruieren. Hier ein paar Beispiele:

z-Test: Test für Erwartungswert bei bekannter Varianz σ^2

1. Stichproben $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\vartheta, \sigma^2)$
2. Nullhypothese $H_0 : \vartheta = \vartheta_0$
3. Teststatistik: $T := \frac{\overline{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$ unter $\mathbb{P}_{\vartheta_0}$
4. Signifikanzniveau: wir wählen ein $\alpha \in (0, 1)$
5. Wir konstruieren den Verwerfungsbereich K_α
$$K_\alpha = \begin{cases} (-\infty, z_\alpha) & \text{für} \quad H_A : \vartheta < \vartheta_0 \\ (z_{1-\alpha}, \infty) & \text{für} \quad H_A : \vartheta > \vartheta_0 \\ (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty) & \text{für} \quad H_A : \vartheta \neq \vartheta_0 \end{cases}$$
6. Falls $T(\omega) \in K_\alpha$, wird H_0 auf dem Niveau α verworfen

t-Test: Test für Erwartungswert bei unbekannter Varianz σ^2
Die Varianz S^2 wird geschätzt

1. Stichproben $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$
2. Nullhypothese $H_0 : \mu = \mu_0$
3. Teststatistik: $T := \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ unter $\mathbb{P}_{\mu_0}$
4. Signifikanzniveau: wir wählen ein $\alpha \in (0, 1)$
5. Wir konstruieren den Verwerfungsbereich K_α
$$K_\alpha = \begin{cases} (-\infty, t_{n-1,\alpha}) & \text{für } \mu < \mu_0 \\ (t_{n-1,1-\alpha}, \infty) & \text{für } \mu > \mu_0 \\ (-\infty, t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty) & \text{für } \mu \neq \mu_0 \end{cases}$$
6. Falls $T(\omega) \in K_\alpha$, wird H_0 auf dem Niveau α verworfen

S.3.4 Zweistichprobentest

gepaart: Seien $(X_1, Y_1) \dots (X_n, Y_n)$ natürliche Paare von ZV mit $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \dots, Y_n \overset{i.i.d}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ Dann lässt sich der Test zum Vergleich von μ_X und μ_Y auf eine Stichprobe zurückführen:

$$Z_i := X_i - Y_i \overset{i.i.d}{\sim} \mathcal{N}(\mu_x - \mu_y, 2\sigma^2)$$

ungepaart: Test mit bekannten Varianzen $\sigma_X, \sigma_Y > 0$

1. Stichproben
$X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu_X, \sigma_X^2), \quad Y_1, \dots, Y_m \overset{i.i.d}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$
2. Nullhypothese $H_0 : \mu_X - \mu_Y = \mu_0$
3. Teststatistik: $T := \frac{\overline{X}_n - \overline{Y}_n - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}(0, 1)$ unter $\mathbb{P}_{\mu_0}$
4. Signifikanzniveau: wir wählen ein $\alpha \in (0, 1)$
5. Wir konstruieren den Verwerfungsbereich K_α
$$K_\alpha = \begin{cases} (-\infty, z_\alpha) & \text{für } \mu_X - \mu_Y < \mu_0 \\ (z_{1-\alpha}, \infty) & \text{für } \mu_X - \mu_Y > \mu_0 \\ (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty) & \text{für } \mu_X - \mu_Y \neq \mu_0 \end{cases}$$
6. Falls $T(\omega) \in K_\alpha$, wird H_0 auf dem Niveau α verworfen

ungepaart: Test bei unbekannter, gleicher Varianz $\sigma > 0$
Die Varianz $S^2 = \frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$ wird geschätzt

1. Stichproben
   $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu_X, \sigma^2), \quad Y_1, \dots, Y_m \overset{i.i.d}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$
2. Nullhypothese $H_0 : \mu_X - \mu_Y = \mu_0$
3. Teststatistik: $T := \frac{\overline{X}_n - \overline{Y}_n - \mu_0}{S\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$ unter $\mathbb{P}_{\mu_0}$
4. Signifikanzniveau: wir wählen ein $\alpha \in (0,1)$
5. Wir konstruieren den Verwerfungsbereich K_α

$$K_\alpha = \begin{cases} (-\infty, t_{n+m-2,\alpha}) & \text{für } \mu_X - \mu_Y < \mu_0 \\ (t_{n+m-2,1-\alpha}, \infty) & \text{für } \mu_X - \mu_Y > \mu_0 \\ (-\infty, t_{n+m-2,\alpha/2}) \cup (t_{n+m-2,1-\alpha/2}, \infty) & \text{für } \mu_X - \mu_Y \neq \mu_0 \end{cases}$$

6. Falls $T(\omega) \in K_\alpha$, wird H_0 auf dem Niveau α verworfen

S.4 Konfidenzbereiche

Konfidenzbereich: Ein *Konfidenzbereich* für einen Schäetzwert ϑ von den Stichproben $X_1, \dots, X_n$ ist eine Menge $C(X_1, \dots, X_n) \subseteq \Theta$. In den meisten Fällen ist das ein Intervall, dessen Endpunkte von $X_1, \dots, X_n$ abhängen.
C heisst ein Konfidenzbereich zum *Niveau* $1-\alpha$, falls gilt

$$1 - \alpha \leq \mathbb{P}_\vartheta[\vartheta \in C(X_1, \dots, X_n)] \quad \text{für alle } \vartheta \in \Theta$$

Beispiel: Konfidenzintervall des Stichprobenmittels
Annahme: $C(X_1, \dots, X_n) = [\,\vartheta - \delta, \vartheta + \delta\,]$
$1 - \alpha \leq \mathbb{P}_\vartheta\big[|\overline{X}_n - \mu| \leq \delta\big] \leq \mathbb{P}_\vartheta\big[|\frac{\overline{X}_n - \mu}{S/\sqrt{n}}| \leq \frac{\delta}{S/\sqrt{n}}\big]$
Satz 7.1 $\Rightarrow \delta = t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$

Beispiel: Konfidenzintervall der Stichprobenvarianz
$1 - \alpha = \mathbb{P}_\vartheta\big[\chi^2_{n-1,\frac{\alpha}{2}} \leq \sigma \leq \chi^2_{n-1,1-\frac{\alpha}{2}}\big]$
$= \mathbb{P}_\vartheta\big[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \leq \sigma \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\big] \Rightarrow C = \big[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\big]$

Anhang

A.1 Kombinatorik

Ziehen von k Elementen aus einer Menge mit n Elementen

	geordnet	ungeordnet
mit zurücklegen	n^k	$\binom{n+k-1}{k}$
ohne zurücklegen	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

A.2 Reihen und Integrale

$$\begin{aligned}
\textstyle\sum_{k=1}^n k &= \tfrac{n(n+1)}{2} \\
\textstyle\sum_{k=1}^n k^2 &= \tfrac{n(n+1)(2n+1)}{6} \\
\textstyle\sum_{k=0}^n a_0 q^k &= a_0 \tfrac{1-q^{n+1}}{1-q} \\
\textstyle\sum_{k=0}^\infty a_0 q^k &= \tfrac{a_0}{1-q} \\
\textstyle\sum_{k=0}^\infty \frac{k}{a^k} &= \tfrac{a}{(a-1)^2}, \qquad |a| > 1 \\
\textstyle\sum_{k=0}^\infty \frac{x^k}{k!} &= e^x
\end{aligned}$$

Partielle Integration:

$$\int_a^b f'(x)g(x)\mathrm{d}x = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)\mathrm{d}x$$

Substitutionsregel:

$$\int_a^b f(g(x))g'(x)\mathrm{d}x \overset{t=g(x)}{=} \int_{g(a)}^{g(b)} f(t)\mathrm{d}t$$

Bei den folgenden Integralen wurden die Integrationskonstanten weggelassen.

$$\begin{aligned}
\int a \,\mathrm{d}x &= ax \\
\int x^a \,\mathrm{d}x &= \tfrac{1}{a+1}x^{a+1}, && a \neq -1 \\
\int (ax+b)^c \,\mathrm{d}x &= \tfrac{1}{a(c+1)}(ax+b)^{c+1}, && c \neq -1 \\
\int \tfrac{1}{x} \,\mathrm{d}x &= \log|x|, && x \neq 0 \\
\int \tfrac{1}{ax+b} \,\mathrm{d}x &= \tfrac{1}{a}\log|ax+b| \\
\int \tfrac{1}{x^2+a^2} \,\mathrm{d}x &= \tfrac{1}{a}\arctan\tfrac{x}{a} \\
\int e^{ax} \,\mathrm{d}x &= \tfrac{1}{a}e^{ax} \\
\int xe^{ax} \,\mathrm{d}x &= \tfrac{e^{ax}}{a^2}(ax-1) \\
\int x^2e^{ax} \,\mathrm{d}x &= e^{ax}\left(\tfrac{x^2}{a} - \tfrac{2x}{a^2} + \tfrac{2}{a^3}\right) \\
\int \log|x| \,\mathrm{d}x &= x(\log|x| - 1) \\
\int \log_a|x| \,\mathrm{d}x &= x(\log_a|x| - \log_a e) \\
\int x^a \log x \,\mathrm{d}x &= \tfrac{x^{a+1}}{a+1}\left(\log x - \tfrac{1}{a+1}\right), && a \neq -1, x > 0 \\
\int \tfrac{1}{x}\log x \,\mathrm{d}x &= \tfrac{1}{2}\log^2 x, && x > 0 \\
\int \sin(ax+b) \,\mathrm{d}x &= -\tfrac{1}{a}\cos(ax+b) \\
\int \cos(ax+b) \,\mathrm{d}x &= \tfrac{1}{a}\sin(ax+b) \\
\int \tan x \,\mathrm{d}x &= -\log|\cos x| \\
\int \tfrac{1}{\sin x} \,\mathrm{d}x &= \log|\tan\tfrac{x}{2}| \\
\int \tfrac{1}{\cos x} \,\mathrm{d}x &= \log|\tan(\tfrac{x}{2} + \tfrac{\pi}{4})| \\
\int \sin^2 x \,\mathrm{d}x &= \tfrac{1}{2}(x - \sin x\cos x) \\
\int \cos^2 x \,\mathrm{d}x &= \tfrac{1}{2}(x + \sin x\cos x) \\
\int \tan^2 x \,\mathrm{d}x &= \tan x - x \\
\int \tfrac{f'(x)}{f(x)} \,\mathrm{d}x &= \log|f(x)|
\end{aligned}$$

A.3 Verteilungs-/Momentenerzeugende Funktionen

Verteilung X	$F_X(x)$	$\mathcal{M}_X(t)$
disk. Unif	$\frac{1}{n}\lvert\{i \mid k_i \leq n\}\rvert$	$\frac{e^t - e^{(n+1)t}}{n(1-e^t)}$
Bernoulli	$(1-p)\mathbb{1}_{x\in[0,1]} + \mathbb{1}_{x\in[0,\infty)}$	$1 - p + pe^t$
Binomial	$\sum_{i=0}^{\lfloor n \rfloor} \binom{n}{i} p^i(1-p)^{n-i}$	$(1-p+pe^t))^n$
Geometrisch	$1-(1-p)^{\lfloor x+1 \rfloor}$	$\frac{pe^t}{1-(1-p)e^t}$
Hypergeom.	$\sum_{i=max(0,m-n)}^{\lfloor x \rfloor} \frac{\binom{r}{i}\binom{n-r}{m-i}}{\binom{n}{m}}$	kompliziert!
Poisson	$\sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!}e^{-\lambda}$	$exp(\lambda(e^\lambda - 1)$
Exponential	$1 - e^{-\lambda t}\mathbb{1}_{t\geq 0}$	$\frac{\lambda}{\lambda - t}$ für $t < \lambda$
Normal	$\Phi(x)$, siehe z_α-Quantile	$exp(\mu t + \frac{\sigma^2 t^2}{2})$
Gamma	unbestimmtes Integral	$(\frac{\lambda}{\lambda-1})^\alpha$
Chiquadrat	siehe $\chi_{n,1-\alpha}$-Quantile	$\frac{1}{(1-2t)^{n/2}}$
t-Vert.	siehe $t_{n,1-\alpha}$-Quantile	existiert nicht
Cauchy	$\frac{1}{2} + \frac{1}{\pi}\arctan(x-\mu)$	existiert nicht