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.

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*).

W.1.3 Endliche Räume

Für einen endlichen Raum $\Omega = \{\omega_1, \dots, \omega_n\}$ mit $\mathbb{P}[\omega_i] = p_i$ für alle $1 \leq i \leq n$ gilt

$$\mathbb{P}[A] = \sum_{i \text{ mit } w_i \in A} p_i$$

Laplace-Raum: In einem *Laplace-Raum* sind alle Ereignisse $\omega_1, \dots, \omega_n$ gleich wahrscheinlich ($p_i = p_j$ für alle i, j). Es gilt dann

$$P[A] = \frac{|A|}{|\Omega|}$$

W.1.4 Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit:

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

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

Satz (Totale Wahrscheinlichkeit): Sei $A_{1 \leq i \leq n}$ eine disjunkte Zerlegung 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 $A_1, \dots, A_n$ eine disjunkte Zerlegung 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:

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

W.1.5 Unabhängigkeit

Unabhängigkeit: Die Ereignisse $A_1, \dots, A_n$ heissen *unabhängig*, wenn für alle $m \in \mathbb{N}$ und $\{k_1, \dots, k_m\} \subseteq \{1, \dots, n\}$ gilt

$$\mathbb{P}\left[\bigcap_{i=1}^{m} A_{k_i}\right] = \prod_{i=1}^{m} \mathbb{P}[A_{k_i}].$$

Hinweis: Bei unabhängigen Ereignissen A, B hat das Eintreten des einen Ereignisses keinen Einfluss auf die Wahrscheinlichkeit des anderen Ereignisses: $\mathbb{P}[B \mid A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]} = \mathbb{P}[B]$

W.2 Zufallsvariablen

Zufallsvariable: Eine *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.

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

$$F_X(t) := \mathbb{P}[X \leq t] := \mathbb{P}[\{\omega \mid X(\omega) \leq t\}].$$

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

W.2.1 Diskrete Zufallsvariablen

Eine Zufallsvariable heisst *diskret*, falls ihr Wertebereich $\mathcal{W}(X)$ endlich oder abzählbar ist.

Gewichtsfunktion: Die *Wahrscheinlichkeitsfunktion* oder *Gewichtsfunktion* einer diskreten Zufallsvariable X ist gegeben durch

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

Eine Gewichtsfunktion weist folgende Eigenschaften auf:
i: $p_X(x) \in [0, 1]$ für alle x.
ii: $\sum_{x_i \in \mathcal{W}(X)} p_X(x_i) = 1$.

Diskrete Verteilungsfunktion: Die Verteilungsfunktion F_X einer diskreten Zufallsvariable X mit Wertebereich $\mathcal{W}(X) = \{x_1, \dots, x_n\}$ ist die Funktion

$$F_X(t) = \mathbb{P}[X \leq t] = \sum_{\substack{x_k \in \mathcal{W}(X) \\ x_k \leq t}} p_X(x_k)$$

W.2.2 Stetige Zufallsvariablen

Dichte: Eine Zufallsvariable X mit der Verteilungsfunktion $F_X(t)$ heisst stetig mit *Dichte* $f_X : \mathbb{R} \to [0, \infty)$, falls gilt

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

Für eine Dichtefunktion f_X gilt:

i: $f_X(t) \geq 0$ für alle $t \in \mathbb{R}$.
ii: $\int_{-\infty}^{\infty} f_X(s)\mathrm{d}s = 1$.

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.2.3 Transformation von Zufallsvariablen

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

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

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

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

und mit der Kettenregel ergibt sich

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

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

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

und somit

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

W.2.4 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.2.5 Erwartungswert

Diskreter Erwartungswert: Ist X diskrete Zufallsvariable mit Gewichtsfunktion p_X, so ist der *Erwartungswert* von X definiert als

$$\mathbb{E}[X] := \sum_{x_i \in \mathcal{W}(X)} x_i p_X(x_i),$$

sofern diese Reihe konvergiert.

Stetiger Erwartungswert: Falls X eine stetige Zufallsvariable mit Dichte f_X ist, dann ist der *Erwartungswert* von X definiert als

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

falls das Integral konvergiert.

Satz (4.1): Sei X eine diskrete Zufallsvariable mit Gewichtsfunktion p_X und $Y := g(X)$, dann gilt

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{x_i \in \mathcal{W}(X)} g(x_i)p_X(x_i).$$

Falls X eine stetige Zufallsvariable mit Dichte f_X ist, ist analog

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

W.2.6 Varianz und Standardabweichung

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

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

Hinweis: Nach Satz 4.1 lässt sich die Varianz folgendermassen berechnen:

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

Standardabweichung: Die *Standardabweichung* einer Zufallsvariable X ist

$$\sigma_X := \sqrt{\mathrm{Var}[X]}.$$

W.3 Wichtige Verteilungen

W.3.1 Diskrete Verteilungen

W.3.1.1 Diskrete Gleichverteilung

Zufallsvariable X mit Wertebereich $\mathcal{W}(X) = \{x_1, \ldots, x_n\}$ und alle Werte haben die gleiche Wahrscheinlichkeit falls

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

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

W.3.1.2 Bernoulli-Verteilung

Eine bernoulli-verteilte Zufallsvariable $X \sim Be(p)$ mit Parameter $p \in [0,1]$ 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 gegeben durch

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

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

X ist die Anzahl der Erfolge k bei n unabhängigen Wiederholungen eines Bernoulli-Experiments. Es gibt $\binom{n}{k}$ verschiedene Möglichkeiten bei n Versuchen k-mal erfolgreich zu sein. Jeder dieser Möglichkeit hat Wahrscheinlichkeit $p^k(1-p)^{n-k}$.

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 gegeben durch

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

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. Für die Anzahl Würfelfwürfe, die man braucht um eine 6 zu würfeln, ist $p = \frac{1}{6}$.

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 gegeben durch

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

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 gegeben durch

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

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

$$p_X(4) = \frac{\binom{6}{4}\binom{39}{2}}{\binom{45}{6}} \approx 0.00136.$$

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.

Hinweis: Die Binomialverteilung $Bin(n,p)$ kann approximativ durch die Poissonverteilung $\mathcal{P}(\lambda)$ mit $\lambda = np$ berechnet werden. Faustregel: Die Approximation kann für $np^2 \leq 0.05$ benutzt werden.

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

$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{für } t \in [a,b] \\ 0 & \text{für } t \notin [a,b] \end{cases}$$

$$F_X = \begin{cases} 0 & \text{für } t < a \\ \frac{t-a}{b-a} & \text{für } t \in [a,b] \\ 1 & \text{für } t > b \end{cases}$$

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

Beispiel: Rundungsfehler einer Messung.

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 gegeben durch

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

$$F_X(t) = \begin{cases} 1 - e^{-\lambda t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

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*

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

mit $\mu = 0$ und $\sigma^2 = 1$ 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.4 Gemeinsame Verteilungen

Gemeinsame Verteilung: Die *gemeinsame Verteilungsfunktion* von n Zufallsvariablen $X_1, \ldots, X_n$ ist die Abbildung $F : \mathbb{R}^n \to [0,1]$,

$$F(t_1, \ldots, t_n) := \mathbb{P}[X_1 \leq t_1, \ldots, X_n \leq t_n].$$

Gemeinsame Gewichtsfunktion: Falls $X_1, \ldots, X_n$ diskrete Zufallsvariablen sind, ist ihre *gemeinsame Gewichtungsfunktion* $p : \mathbb{R}^n \to [0,1]$ definiert durch

$$p(x_1, \ldots, x_n) := \mathbb{P}[X_1 = x_1, \ldots, X_n = x_n].$$

Gemeinsame Dichte: Seien $X_1, \ldots, X_n$ stetige Zufallsvariablen mit gemeinsamer Verteilungsfunktion $F(t_1, \ldots, t_n)$. Die Funktion $f : \mathbb{R}^n \to [0, \infty)$ heisst *gemeinsame Dichte* von $X_1, \ldots, X_n$, falls für alle $t_i \in \mathbb{R}$ gilt

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f(x_1, \ldots, x_n) \mathrm{d}x_n \ldots \mathrm{d}x_1$$

Falls $X_1, \ldots, X_n$ eine gemeinsame Dichte f haben, so hat diese folgende Eigenschaften:

i: $f(t_1, \ldots, t_n) \geq 0$ für alle $t_i \in \mathbb{R}$.
ii: $\int_{\mathbb{R}^n} f(t_1, \ldots, t_n) \mathrm{d}\mu = 1$.
iii: $\mathbb{P}[(X_1, \ldots, X_n) \in A] = \int_{(t_1,\ldots,t_n)\in A} f(t_1, \ldots, t_n) \mathrm{d}\mu$.
iv: $f(t_1, \ldots, t_n) = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} F(t_1, \ldots, t_n)$, falls definiert.

W.4.1 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

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

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

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

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.4.2 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.4.3 Unabhängigkeit

Unabhängigkeit: Die Zufallsvariablen $X_1, \ldots, X_n$ heissen *unabhängig*, falls die gemeinsame Verteilungsfunktion das Produkt der Verteilungsfunktionen der Randverteilungen ist:

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

Im diskreten Fall sind $X_1, \ldots, X_n$ unabhängig, genau dann wenn

$$p(x_1, \ldots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$$

gilt und analog im stetigen Fall, falls

$$f(x_1, \ldots, x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

W.4.4 Funktionen von Zufallsvariablen

Ausgehend von den Zufallsvariablen $X_1, \ldots, X_n$ kann man mit einer Funktion $g : \mathbb{R}^n \to \mathbb{R}$ eine neue Zufallsvariable $Y = g(X_1, \ldots, X_n)$ bilden.

Beispiel (Summe, diskret): Für die Gewichtsfunktion p_Z der Summe $Z = X + Y$ zweier diskreten Zufallsvariablen X und Y mit gemeinsamer Gewichtsfunktion p erhält man

$$p_Z(z) = \sum_{x_i \in \mathcal{W}(X)} \mathbb{P}[X = x_i, Y = z - x_i] = \sum_{x_i \in \mathcal{W}(X)} p(x_i, z - x_i)$$

Beispiel (Summe, stetig): Sind X und Y stetige Zufallsvariablen mit gemeinsamer Dichte f, so ist die Verteilungsfunktion F_Z der Summe $Z = X + Y$ gegeben durch

$$F_Z = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) \mathrm{d}y \mathrm{d}x \stackrel{v=x+y}{=} \int_{-\infty}^{z} \int_{-\infty}^{\infty} f(x, v - x) \mathrm{d}x \mathrm{d}v$$

und somit auch die Dichte

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

W.4.4.1 i.i.d Annahme

Die Abkürzung *i.i.d.* kommt vom Englischen *independent and identically distributed*. Die n-fache Wiederholung eines Zufallsexperiments ist selbst wieder ein Zufallsexperiment. Für die Zufallsvariablen X_i der i-ten Wiederholung wird oft aus Gründen der Einfachheit Folgendes angenommen:

i: $X_1, \ldots, X_n$ sind paarweise unabhängig.
ii: Alle X_i haben dieselbe Verteilung.

W.4.4.2 Spezielle Funktionen von Zufallsvariablen

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

1. Wenn $X_i \sim Be(p)$, dann ist $S_n \sim Bin(n, p)$.
2. Wenn $X_i \sim \mathcal{P}(\lambda)$, dann ist $S_n \sim \mathcal{P}(n\lambda)$.
3. Wenn $X_i \sim \mathcal{N}(\mu, \sigma^2)$, dann ist $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$.

Für den Erwartungswert und die Varianz gilt allgemein

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

W.4.5 Erwartungswert

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 $Y := g(X_1, \ldots X_n)$ der Zufallsvariablen $X_1, \ldots, X_n$ gilt im diskreten Fall

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

und analog im stetigen Fall

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

Satz (4.4): Seien $X_1, \ldots, X_n$ Zufallsvariablen mit endlichen Erwartungswerten $\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]$, dann ist

$$\mathbb{E}\left[a + \sum_{i=1}^n b_i X_i\right] = a + \sum_{i=1}^n b_i \mathbb{E}[X_i].$$

W.4.6 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 durch

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

Es gelten folgende Rechenregeln:

i: $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
ii: $\mathrm{Var}[a + bX] = b^2 \mathrm{Var}[X]$.
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: $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]$.
v: $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$.
vi: $\mathrm{Cov}[X, X] = \mathrm{Var}[X]$.
vii: $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.
viii: $\mathrm{Cov}[X, a] = 0$ für alle $a \in \mathbb{R}$.
ix: $\mathrm{Cov}[X, bY] = b\mathrm{Cov}[X, Y]$ für alle $b \in \mathbb{R}$.
x: $\mathrm{Cov}[X, Y + Z] = \mathrm{Cov}[X, Y] + \mathrm{Cov}[X, Z]$.
xi: $\mathrm{Cov}[a + \sum_{i=1}^n b_i X_i, c + \sum_{j=1}^m d_j Y_j]$
$= \sum_{i=1}^n \sum_{j=1}^m b_i d_j \mathrm{Cov}[X_i, Y_j]$
xii: $\mathrm{Cov}[X, Y] = 0$, falls X und Y unabhängig.

Korrelation: Seien X und Y Zufallsvariablen, dann heisst

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

Korrelation von X und Y. Ist $\mathrm{Corr}[X, Y] = 0$, oder äquivalent $\mathrm{Cov}[X, Y] = 0$, dann heissen X und Y unkorreliert.

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

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

Hinweis: Sind X und Y unabhängig, dann ist $\mathrm{Cov}[X, Y] = 0$ und $\mathrm{Corr}[X, Y] = 0$. Die Umkehrung gilt aber im Allgemeinen nicht.

W.5 Grenzwertsätze

W.5.1 Gesetz der grossen Zahlen

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

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

Das heisst

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

Satz (Starkes GGZ): Für eine Folge $X_1, X_2, \ldots$ 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.5.2 Zentraler Grenzwertsatz

Satz (ZGS): Sei $X_1, X_2, \ldots$ 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:

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

W.5.3 Chebyshev-Ungleichung

Für eine Zufallsvariable Y mit Erwartungswert μ_Y und Varianz σ_Y^2 und jedes $\epsilon > 0$ gilt

$$\mathbb{P}[|Y - \mu_Y| > \epsilon] \leq \frac{\sigma_Y^2}{\epsilon^2}.$$

W.5.4 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.*

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

$$F_n(y) := \frac{1}{n}|\{x_i \mid x_i \leq y\}| = \frac{1}{n}\sum_{i \text{ mit } x_i \leq y} f_i.$$

Empirischer Mittelwert:

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

Empirische Varianz und Standardabweichung:

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

Empirisches Quantil: Das *empirische α-Quantil* zu den geordneten Daten $x_{(1)}, \ldots, x_{(n)}$ ist gegeben durch

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

wobei $k = \lfloor \alpha n \rfloor$ und $\alpha \in (0,1)$. Damit liegt etwa der Anteil α unterhalb des empirischen α-Quantils, und somit etwa der Anteil $1-\alpha$ oberhalb.

Empirischer Median: Der *empirische Median* ist definiert als das 0.5-Quantil.

S.2 Deskriptive Statistik

S.2.1 Histogramm

Bei grossem Stichprobenumfang n werden benachbarte Werte zu einer Klasse zusammengefasst. Der Wertebereich der Daten wird dadurch in disjunkte Intervalle (die Klassen) unterteilt.

i) Die Anzahl der Klassen sollte von der Grössenordnung $\sqrt{n}$ sein.
ii) Die Klassenbreite sollte für alle Klassen gleich sein; als Ausnahme können die Klassen am linken und rechten Rand grösser sein (Ausreisser).

S.2.2 Boxplot

Aus einem Boxplot lässt sich folgendes ablesen:

a: empirischer Median
b: empirisches 0.25-Quantil
c: empirisches 0.75-Quantil
d: kleinster Datenwert x_i mit $b - x_i < 1.5(c-b)$
e: grösster Datenwert x_i mit $x_i - c < 1.5(c-b)$
f: Ausreisser

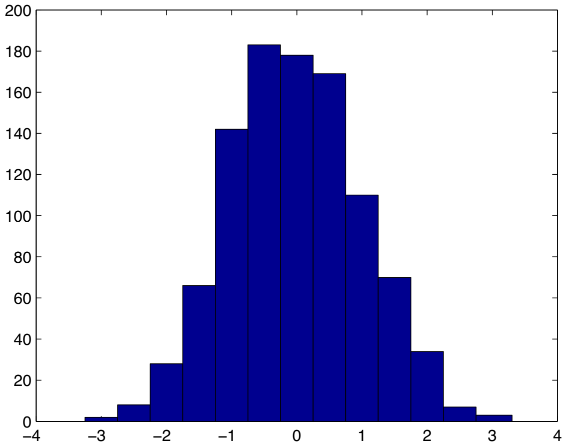

Abb. 1: Histogramm einer standard-norvmalverteilten Zufallsvariable.

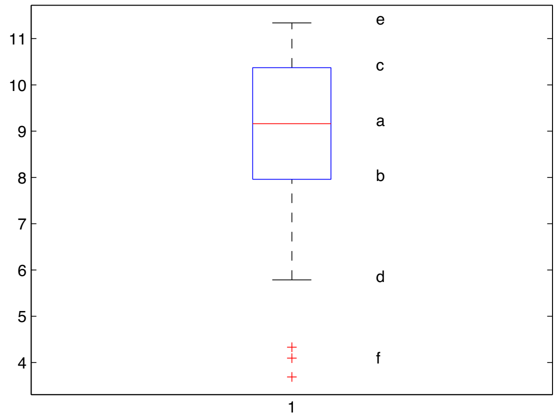

Abb. 2: Boxplot

S.2.3 QQ-Plot

Mit einem *QQ-Plot* (Quantil-Quantil) kann man die Abweichung der Daten von einer gewählten Modell-Verteilung F graphisch überprüfen.
Es werden die empirischen Quantile auf der y-Achse gegenüber den theoretischen Quantilen auf der x-Achse geplottet.

S.3 Schätzer

Für eine Stichprobe $X_1, \dots, X_n$ soll ein passendes Modell gefunden werden. Die Parameter $\vartheta = (\vartheta_1, \dots, \vartheta_m)$ des Modells versucht man mit einem *Schätzer* $T = (T_1, \dots, T_m)$ aufgrund der Stichprobe herauszufinden. Die Schätzer sind Zufallsvariablen der Form $T_j = t_j(X_1, \dots, X_n)$ für eine geeignete Funktion $t_j : \mathbb{R}^n \to \mathbb{R}$. Durch Einsetzen von Daten x_i erhält man *Schätzwertwerte* $t_j(x_1, \dots, x_n)$ für ϑ_j.

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

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

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

Hinweis: Der Grundraum Ω und die Menge der beobachtbaren Ereignisse $\mathcal{F}$ sind fest. Die Wahl des Parameters ϑ aus dem Parameterraum Θ hat aber Einfluss auf das Wahrscheinlichkeitsmass $\mathbb{P}_\vartheta$. Mit $\mathbb{E}_\vartheta$ wird der Erwartungswert unter $\mathbb{P}_\vartheta$ bezeichnet.

S.3.1 Momenten-Methode

Moment: Das *k-te Moment* einer Zufallsvariablen X im Model $\mathbb{P}_\vartheta$ ist

$$\mu_k := \mu_k(\vartheta) := \mathbb{E}_\vartheta[X^k].$$

Stichprobenmoment: Das *k-te Stichprobenmoment* von Zufallsvarialben $X_1, \dots, X_n$ ist

$$\hat{\mu}_k := \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

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

$$\vartheta_j = g_j(\mu_1, \dots, \mu_m) \quad \text{für } j \in \{1, \dots, m\}$$

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

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

Beispiel: Gegeben seien n unabhängige Realisierungen $x_1, \dots, x_n$ 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

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

Es gilt aber auch $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda$. Es kann also auch $g_1(\mu_1, \mu_2) = \mu_2 - \mu_1^2$ gewählt werden. Dadurch erhält man einen anderen Momentenschätzer

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

S.3.2 Maximum-Likelihood

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

$$\mathcal{L}(x_1, \dots, x_n \mid \vartheta) = f(x_1, \dots, x_n \mid \vartheta).$$

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

[1] oder zumindest das stetige Pendant zur Wahrscheinlichkeit.

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 $f(t \mid \vartheta)$, so dass sich die Likelihood-Funktion vereinfacht zu

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

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

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

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

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

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:

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

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.

S.4 Tests

S.4.1 Fehler 1. und 2. Art

Fehler 1. Art: Die Hypothese wird zu Unrecht abgelehnt, d.h. 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 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.4.2 Mögliches Vorgehen

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

1: 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:

$$\begin{aligned} \text{Hypothese } H_0 &: \vartheta \in \Theta_0 \\ \text{Alternative } H_A &: \vartheta \in \Theta_A \end{aligned}$$

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

2. Es wird eine *Teststatistik* $T = t(X_1, \ldots, X_n)$ gewählt, wobei $t : \mathbb{R}^n \to \mathbb{R}$ eine geeignete Funktion ist.
3. Es wird ein *Signifikanzniveau* $\alpha \in (0,1)$ gewählt.
4. Ein *Verwerfungsbereich* $K \subseteq \mathbb{R}$ wird konstruiert, so dass
$$\sup_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta[T \in K] \leq \alpha.$$
Dadurch wird die Wahrscheinlichkeit eines Fehlers 1. Art durch α beschränkt.
5. Die Hypothese wird verworfen, falls der realisierte Wert $t(x_1, \ldots, x_n)$ im Verwerfungsbereich K liegt.

Hinweis: Alternative zu Schritt 4 und 5: Der P-Wert p wird berechnet und die Hypothese verworfen, falls $p \leq \alpha$.

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.

Macht: Die *Macht* eines Tests ist die Funktion

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

Das Maximieren der Macht $\beta(\vartheta)$ entspricht dem Minimieren der Wahrscheinlichkeit für einen Fehler 2. Art $1 - \beta(\vartheta) = \mathbb{P}_\vartheta[T \notin K]$ für $\vartheta \in \Theta_A$.

S.4.3 Likelihood-Quotienten Test

Als Teststatistik wird der *Likelihood-Quotient* $\mathcal{R}$ gewählt, wobei $\mathcal{L}$ die Likelihood-Funktion ist:

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

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.

Hinweis: Sind Hypothese und Alternative beide einfach, so ist der Test optimal (nach Neyman-Pearson-Lemma).

S.4.4 z-Test

Seien $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\vartheta_0, \sigma^2)$ unter $\mathbb{P}_{\vartheta_0}$ mit *bekannter* Varianz σ^2. Es soll die Hypothese $H_0 : \vartheta = \vartheta_0$ getestet werden. Mögliche Alternativen H_A sind $\vartheta > \vartheta_0$, $\vartheta < \vartheta_0$ (einseitig) oder $\vartheta \neq \vartheta_0$ (zweiseitig). Die Teststatistik ist

$$T := \frac{\overline{X}_n - \vartheta_0}{\sigma_X / \sqrt{n}} \sim \mathcal{N}(0,1)$$

unter dem Modell $\mathbb{P}_{\vartheta_0}$. Der Verwerfungsbereich ist von der Form $(c_>, \infty)$, bzw. $(-\infty, c_<)$, bzw. $(-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$. Zum Beispiel liefert die Bedingung

$$\alpha = \mathbb{P}_{\vartheta_0}[T \in K_>] = \mathbb{P}_{\vartheta_0}[T > c_>] = 1 - \Phi(c_>),$$

dass $c_> = \Phi^{-1}(1-\alpha)$, also das $(1-\alpha)$-Quantil der $\mathcal{N}(0,1)$-Verteilung, sein muss.

S.4.5 t-Test

Seien $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_0, \sigma^2)$ unter $\mathbb{P}_\vartheta$ wobei $\vartheta = (\mu, \sigma^2)$ und insbesondere die Varianz σ^2 *unbekannt* ist. Die Hypothese $H_0 : \mu = \mu_0$ soll getestet werden. Die unbekannte Varianz σ^2 wird durch den Schätzer $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ (empirische Varianz) ersetzt. Danach kann mit der Teststatistik

$$T := \frac{\overline{X}_n - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

gleich wie beim z-Test vorgegangen werden.

S.4.6 Gepaarter Zweistichprobentest

Seien $X_{1\leq i\leq n} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ und $Y_{1\leq i\leq n} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ unter $\mathbb{P}_\vartheta$. Falls man eine natürliche Paarbildung zwischen X_i und Y_i hat, lässt der Test zum Vergleich von μ_X und μ_Y auf eine Stichprobe zurückführen:

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

S.4.7 Ungepaarter Zweistichprobentest

Seien $X_{1\leq i\leq n} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ und $Y_{1\leq i\leq m} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ unter $\mathbb{P}_\vartheta$.

a) Ist σ^2 bekannt, so ist die Teststatistik

$$T := \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1).$$

b) Ist σ^2 unbekannt, berechnet man

$$s^2 := \frac{1}{m+n-2}((n-1)s_X^2 + (m-1)s_Y^2)$$

und wählt für die Teststatistik

$$T := \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

S.4.8 Konfidenzbereiche

Konfidenzbereich: Ein *Konfidenzbereich* für ϑ zu 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

$$\mathbb{P}_\vartheta[\vartheta \in C(X_1, \dots, X_n)] \geq 1 - \alpha$$

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

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

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.

$$\begin{aligned}
\int a \,\mathrm{d}x &= ax \\
\int x^a \,\mathrm{d}x &= \tfrac{1}{a+1} x^{a+1}, && a \neq -1 \\
\int (ax+b)^c \,\mathrm{d}x &= \tfrac{1}{a(c+1)} (ax+b)^{c+1}, && c \neq -1 \\
\int \tfrac{1}{x} \,\mathrm{d}x &= \log|x|, && x \neq 0 \\
\int \tfrac{1}{ax+b} \,\mathrm{d}x &= \tfrac{1}{a} \log|ax+b| \\
\int \tfrac{1}{x^2+a^2} \,\mathrm{d}x &= \tfrac{1}{a} \arctan \tfrac{x}{a}
\end{aligned}$$

$$\begin{aligned}
\int e^{ax} \,\mathrm{d}x &= \tfrac{1}{a} e^{ax} \\
\int x e^{ax} \,\mathrm{d}x &= \tfrac{e^{ax}}{a^2}(ax-1) \\
\int x^2 e^{ax} \,\mathrm{d}x &= e^{ax} \left(\tfrac{x^2}{a} - \tfrac{2x}{a^2} + \tfrac{2}{a^3} \right)
\end{aligned}$$

$$\begin{aligned}
\int \log|x| \,\mathrm{d}x &= x(\log|x| - 1) \\
\int \log_a|x| \,\mathrm{d}x &= x(\log_a|x| - \log_a e) \\
\int x^a \log x \,\mathrm{d}x &= \tfrac{x^{a+1}}{a+1} \left(\log x - \tfrac{1}{a+1} \right), && a \neq -1, x > 0 \\
\int \tfrac{1}{x} \log x \,\mathrm{d}x &= \tfrac{1}{2} \log^2 x, && x > 0
\end{aligned}$$

$$\begin{aligned}
\int \sin(ax+b) \,\mathrm{d}x &= -\tfrac{1}{a} \cos(ax+b) \\
\int \cos(ax+b) \,\mathrm{d}x &= \tfrac{1}{a} \sin(ax+b) \\
\int \tan x \,\mathrm{d}x &= -\log|\cos x| \\
\int \tfrac{1}{\sin x} \,\mathrm{d}x &= \log|\tan \tfrac{x}{2}| \\
\int \tfrac{1}{\cos x} \,\mathrm{d}x &= \log|\tan(\tfrac{x}{2} + \tfrac{\pi}{4})| \\
\int \sin^2 x \,\mathrm{d}x &= \tfrac{1}{2}(x - \sin x \cos x) \\
\int \cos^2 x \,\mathrm{d}x &= \tfrac{1}{2}(x + \sin x \cos x) \\
\int \tan^2 x \,\mathrm{d}x &= \tan x - x
\end{aligned}$$

$$\int \tfrac{f'(x)}{f(x)} \,\mathrm{d}x = \log|f(x)|$$

A.3 Beispiele

A.3.1 Regeln

- Die Dichte kann bei stetigen ZV > 1 sein.
- $Kov[X,Y] = 0 \Leftrightarrow$ X,Y sind unabhängig $\Leftrightarrow f_X(x) f_Y(y) = f_{X,Y}(x,y)$.
- X und Y sind unkorreliert $\to E[XY] = E[X]E[Y]$
- Mit der gemeinsamen Verteilung von X und Y kann man die Dichte berechnen
- $P[a \leq X \geq b] = F_X(a) - F_X(b)$
- $0 \leq F_X(x) \geq 1$ aber $f_X(x)$ kann größer 1 sein.
- Wenn X - $Exp(\lambda) : P[X > t+s | X > s] = P[X > t] \forall t, s \in \mathbb{R}^+$.
- $f(x_1, ..., x_n) = f(x_1)...f(x_n) \Rightarrow$ Die gemeinsame Dichte ist das Produkt der Randdichten.

A.3.2 Berechnung des Median

Für den Median muss gelten : $F(m) = 0.5$
Dann einfach Gleichung der Verteilungsfunktion (= 0.5) für m auflösen.

A.3.3 Verteilungsfunktion mit der Dichte

Dichtefunktion integrieren und als Grenzen der Integrale die Grenzen des Wahrscheinlichkeitsraums benutzen.
$\to$ Dichte anhand Verteilung : Verteilung ableiten

A.3.4 Erwartungswert mit Dichte

$E[X] = \int_*^* x f(x) dx$, mit * die Grenzen des Definitionsbereich

A.3.5 Verteilungsfunktion Beweis

- Monotonie
- Rechtsstetig
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

A.3.6 P[X $\leq \alpha$]

$P[X \leq \alpha] = F_X(\alpha) = \int_*^\alpha f_X(x) dx$, mit * die untere Grenze des Wahr.Raums.

A.3.7 Erwartungswert aus Dichtegraph lesen

Wo die Dichte symmetrisch zuläuft liegt der E.Wert.

A.3.8 Beweis lognormale ZV

Sei R lognormal und $(R\ N(\mu, \sigma^2))$.
$V = \alpha R \beta \to log(V) = log(\alpha R \beta) = log(\alpha\beta) + log(R^\alpha)$
Das heisst V ist auch lognormal verteilt mit Parametern : $V log N(x\mu + log(\alpha\beta), x^2\sigma^2)$

A.3.9 Doppelte Integration

$$\begin{aligned}
F(x,y) &= \int_0^{1/4} \int_0^1 (1 + 16xy + 6y^2) dy dx \\
&= \int_0^{1/4} ([y + 8x^2 + 2y^3]_0^1) dx \\
&= \int_0^{\frac{1}{4}} (8x + 3 - 0) dx \\
&= [3x + 4x^2]_0^{1/4} \\
&= 1
\end{aligned}$$

A.3.10 Bedingter E. Wert

$$E[Y|X=x] = \int_a^b y f_{Y|X=x}(y)dy$$
$$= \int_a^b y \frac{f_{X,Y}(x,y)}{f_X(x)} dy$$

A.3.11 Dichtefunktion Beweis

- $f_{X,Y}(x,y) \geq 0, \forall(x, y = \in \mathbb{R}^2$
- $\int_{\mathbb{R}^2} f_{X,Y}(x,y)dxdy = 1$ Bzw. in D. Bereich.

A.3.12 Randdichten von zusammengesezten ZV berechnen

Gegeben $f_{X,Y}(x,y) = 1+16xy+6y^2$ und $D = [0,1/4] \times [0,1]$.

- $f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = 3+8x, x \in [0,1/4]$
- $f_Y(y) = \int_0^{1/4} f_{X,Y}(x,y)dx = \frac{1}{4} + \frac{1}{2}y + \frac{3}{2}y^2, x \in [0,1]$

A.3.13 Kovarianz berechnen

* Gleiche Funktion und ZV wie obige Bsp.
$Cov[X,Y] = E[XY] - E[X]E[Y]$

$$E[XY] = \int_0^{1/4} \int_0^1 xy(1+16xy+6y^2)dxdy$$
$$= \int_0^{1/4} (2x + \frac{1}{6}3x^2 dx$$
$$= \frac{13}{144}$$

And

$$E[X] * E[Y] = \int_0^{1/4} (3x+8x^2)dx * \int_0^1 \frac{1}{4}y\frac{1}{2}y^2\frac{3}{2}y^3$$
$$= 2/3 * 13/96$$
$$= 13/144$$

Hence : Cov[X,Y] = 0 => X und Y sind unabhängig.

A.3.14 Dichte einer uniform verteilten ZV

N uniform auf $[a,b]$ verteilt.
$\rightarrow f_X(x) = \frac{1}{(b-a)}$

A.3.15 Zur standartisierten ZV Z übergehen

Sei $X\ N(\mu,\sigma^2)$ dann gilt $Z = \frac{(X-\mu)}{\sigma}$
z.B : $P[X \leq \alpha] = P[Z \leq \frac{\alpha-\mu}{\sigma}]$

A.3.16 SummenRechnen

Konstante C bestimmen :

$$1 = \sum_{k=2}^{\infty}\sum_{j=1}^{k-1} C(\frac{1}{2})^k$$
$$= C\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}(\frac{1}{2})^k$$
$$= C\sum_{j=1}^{\infty}(\frac{1}{2})^{j+1}\sum_{k=0}^{\infty}(\frac{1}{2})^k$$
$$=$$

A.3.17 Bedingte Gewichtsfunktion berechnen

$$p_{X|Y}(j|k) = \frac{p_{X,Y}(j,k)}{p_Y(k)}$$

A.3.18 Normalverteilung

Für $X : N(\mu,\sigma^2)$ gilt dass $((X-\mu)/\sqrt{\sigma^2}) : N(0,1)$

A.3.19 X N(1,2) berechnen

$$P[E[X]-1 \leq X \leq E[X]+1 = P[0 \leq X \leq 2]$$
$$= P[-1/2 \leq \frac{X-1}{2} \leq 1/2]$$
$$= \Phi(1/2) - \Phi(-1/2)$$
$$= \Phi(1/2) - (1-\Phi(1/2))$$
$$= 2\Phi(1/2) - 1$$

$\Rightarrow$ **in der Tabelle ablesen**

A.3.20 Funktion von unabhängigen ZV

Gegeben $X_n = min\{X_1...X_k\}$, Dichtefunktion berechnen :
$F(t) = 1 - P[X_n > t] = 1 - P[X_1...X_k] = 1-(1-F(t))^n$
Daher gilt : $f(t) = \frac{d}{dt}F(t) = n(1-F(t))^{n-1}f(t)$

A.3.21 Interval Normalverteilte ZV

Gegeben : $X - N(\mu,\sigma^2)$, und Gesucht : $P[40 < X < 60]$

$$P[|X-50| < 10] = P[40 < X < 60]$$
$$= P[X < 60] - P[X \leq 40]$$
$$= P[\frac{X-50}{5} < \frac{60-50}{5}] - P[\frac{X-50}{5} \leq \frac{40-50}{5}]$$
$$= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1...$$