Cheatsheet WuS

Amos Herz

FS 2021

Grundbegriffe

Wir definieren einen **Wahrscheinlichkeitsraum** als das Tupel $(\Omega, \mathcal{F}, \mathbb{P})$:

Der **Grundraum** Ω ist eine nicht leere Menge, wobei $\omega \in \Omega$ ein Elementarereignis ist.

Eine **Sigma-Algebra** $\mathcal{F} \subseteq \mathbb{P}(\Omega)$ erfüllt die Bedingungen:

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^{\complement} \in \mathcal{F}$
3. $A_1, A_2, ... \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Ein **Wahrscheinlichkeitsmass** $\mathbb{P}$ auf $(\Omega, \mathcal{F})$ ist eine Abbildung $\mathbb{P} : \mathcal{F} \mapsto [0,1], A \mapsto \mathbb{P}[A]$, so dass:

1. $\mathbb{P}[\Omega] = 1$
2. $\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$, falls $A = \bigcup_{i=1}^{\infty} A_i$

Aus diesen Definitionen ergeben sich folgende nützliche Eigenschaften:

1. $\emptyset \in \mathcal{F}$
2. $A_1, A_2, ... \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
3. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$
4. $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$

und

1. $\mathbb{P}[\emptyset] = 0$
2. $\mathbb{P}[A^{\complement}] = 1 - \mathbb{P}[A]$
3. $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

Daraus ergibt sich, dass wenn $A_1, ..., A_n$ paarweise disjunkt sind:

$$\mathbb{P}[A_1 \cup ... \cup A_n] = \mathbb{P}[A_1] + ... + \mathbb{P}[A_n]$$

Monotonie

Seien $A, B \in \mathcal{F}$ dann gilt

$$A \subseteq B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$$

Union Bound

Seien $A_1, A_2, ... \in \mathcal{F}$ dann gilt

$$\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

Bedingte Wahrscheinlichkeiten

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit $A, B \in \mathcal{F}$ und $\mathbb{P}[B] > 0$. Die bedingte Wahrscheinlichkeit von A gegeben B ist definiert als:

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

Totale Wahrscheinlichkeit

Sei $A_1, ..., A_n \in \mathcal{F}$ eine Partition von Ω mit $\mathbb{P}[A_i] > 0$ für alle $1 \leq i \leq n$. Dann gilt:

$$\forall B \in \mathcal{F}. \quad \mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B|A_i] \cdot \mathbb{P}[A_i]$$

Formel von Bayes

Sei $A_1, ..., A_n \in \mathcal{F}$ eine Partition von Ω mit $\mathbb{P}[A_i] > 0$ für alle $1 \leq i \leq n$. Für jedes $B \in \mathcal{F}$ mit $\mathbb{P}[B] > 0$ gilt:

$$\forall i = 1, ..., n. \quad \mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i] \cdot \mathbb{P}[A_i]}{\sum_{k=1}^{n} \mathbb{P}[B|A_k] \cdot \mathbb{P}[A_k]}$$

Unabhängigkeit

Zwei Ereignisse $A, B \in \mathcal{F}$ sind **unabhängig** falls gilt:

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

Daraus folgt, dass wenn $A \in \{0,1\}$ zu jedem Ereignis B unabhängig ist. Weiter gilt, wenn A, B unabhängig sind, so müssen auch $A, B^{\complement}$ unabhängig sein.

Wir können Unabhängigkeit auch für mehr als zwei Ereignisse definieren. Seien $A_1, ..., A_n \in \mathcal{F}$, so sind die Ereignisse unabhängig falls gilt:

$$\forall I \subseteq \{1, ..., n\}. \quad \mathbb{P}[\bigcap_{i \in I} A_i] = \prod_{i \in I} \mathbb{P}[A_i]$$

Zufallsvariablen

Eine **Zufallsvariable** ist eine messbare Abbildung $X : \Omega \mapsto \mathbb{R}$, wobei Ω die Ereignismenge eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{F}, \mathbb{P})$ ist.

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

Hierbei schreiben wir oftmals nur X.

Verteilungsfunktion

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

$$\forall x \in \mathbb{R}, \; F_X(x) = \mathbb{P}[X \leq x]$$

Aus $a < b$ folgt:

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

Verteilungsfunktion Eingeschaften

Die Verteilungsfunktion F hat folgende Eigenschaften:

1. F ist monoton wachsend
2. F ist rechtsstetig, d.h. $\lim_{t \to 0} F(x + t) = F(x)$.
3. $\lim_{x \to -\infty} F_X(x) = 0$ und $\lim_{x \to \infty} F_X(x) = 1$

Unabhängigkeit von ZV

Unabhängigkeit von ZV

Die ZV $X_1, ..., X_n$ sind unabhängig falls:

$$\forall x_1, ..., x_n \in \mathbb{R}. \quad \mathbb{P}[X_1 \leq x_1, ..., X_n \leq x_n] = \mathbb{P}[X_1 \leq x_1] \cdot ... \cdot \mathbb{P}[X_n \leq x_n]$$

Eine Folge von Zufallsvariablen $X_1, X_2, ...$ ist:

1. unabhängig, falls $\forall n. \quad X_1, ..., X_n$ unabhängig sind
2. unabhängig und identisch verteilt (uiv.), falls sie unabhängig sind und $\forall i, j. \quad F_{X_i} = F_{X_j}$

Transformation von ZV

Sei $\varphi : \mathbb{R} \mapsto \mathbb{R}$ und X ein Zufallsvariable, so ist

$$\varphi(X) = \varphi \circ X$$

auch eine ZV. Seien $X_1, ... X_n$ ZV mit $\phi : \mathbb{R}^n \mapsto \mathbb{R}$, so ist

$$\phi(X_1, ..., X_n) = \phi \circ (X_1, ..., X_n)$$

ebenfalls eine ZV.

Konstruktion einer ZV

Sei eine gültige Verteilungsfunktion F_X gegeben, nun wollen wir eine dazugehörige ZV X konstruieren. Dafür brauchen wir:

Kolmogorov Theorem

$\exists(\Omega, \mathcal{F}, \mathbb{P})$ und $\exists X_1, X_2, ...$ ZV in $(\Omega, \mathcal{F}, \mathbb{P})$ sodass $X_1, X_2, ...$ uiv. Bernoullivariablen mit $p = 0.5$ sind.

Sei $X_1, X_2, ... \sim \text{Ber}(1/2)$ eine unendliche Folge, dann ist

$$U = \sum_{n=1}^{\infty} 2^{-n} \cdot X_n$$

gleichverteilt auf $[0, 1]$.

Aufgrund der Eigenschaften der Verteilungsfunktion F, wissen wir dass eine eindeutige Inverse F^{-1} existiert. wir können die generalisierte Inverse definieren als:

Die generalisierte Inverse

$$\forall \alpha \in [0, 1]. \quad F^{-1}(\alpha) = \inf\{x \in \mathbb{R} \mid F(x) \geq \alpha\}$$

Sei nun F eine Verteilungsfunktion und U eine gleichverteilte ZV in $[0, 1]$. Dann besitzt $X = F^{-1}(U)$ genau die Verteilungsfunktion $F_X = F$.

Diskrete und Stetige ZV

Per Definition ist eine Verteilungsfunktion immer rechtsstetig, analog dazu können wir die Linksstetigkeit definieren:

$$F(x-) = \lim_{t \to 0} F(x - t)$$

Jedoch ist $F(x-) = F(x)$ nicht immer wahr, d.h. nicht jede Verteilungsfunktion ist linksstetig.

$$\forall x \in \mathbb{R}. \quad F(x) - F(x-) = \mathbb{P}[X = x].$$

Daraus:

1. Wenn F in einem Punkt $a \in \mathbb{R}$ nicht stetig ist, dann ist die "Sprünggrösse" $F(a) - F(a-)$ gleich der Wahrscheinlichkeit, dass $X = a$ ist.
2. Wenn F in einem Punkt $a \in \mathbb{R}$ stetig ist, dann ist $P[X = a] = 0$.

Fast sichere Ereignisse

Ein Ereignis $A \in \mathcal{F}$ tritt **fast sicher** (f.s./a.s.) ein, falls $\mathbb{P}[A] = 1$. Seien X, Y ZV, so schreiben wir: $X \leq Y$ f.s. $\Leftrightarrow \mathbb{P}[X \leq Y] = 1$.

Diskrete ZV

Diskret ZV

Eine ZV X heisst **diskret**, falls $\exists W \subset \mathbb{R}$ endlich oder abzählbar, so dass $X \in W$ f.s.

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

Die **Verteilungsfunktion** einer diskreten ZV ist definiert als:

$$(p(x))_{x \in W} \quad \text{wobei} \quad \sum_{x \in W} p(x) = 1$$

Die **Gewichtsfunktion** einer diskreten ZV ist definiert als:

$$\forall x \in W. \quad p(x) = P[X = x]$$

Diskrete Verteilungen

Bernoulli-Verteilung Eine Bernoulli-Verteilte ZV kennt nur die Ereignisse $\{0, 1\}$, wir schreiben auch $X \sim \text{Ber}(p)$. Sie wird definiert durch:

Bernoulli-Verteilung

$$\mathbb{P}[X = 0] = 1 - p \quad \text{und} \quad \mathbb{P}[X = 1] = p$$

Binomialverteilung Dies beschreibt die Wiederholung von Bernoulli-Experimenten. Wir schreiben $X \sim \text{Bin}(n, p)$ und definieren:

Binomialverteilung

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

Sei $0 \leq p \leq 1, n \in \mathbb{N}$.
Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-Zufallsvariablen mit dem Parameter p.
Dann ist $S_n := X_1 + \ldots + X_n$ eine binomial Zufallsvariable mit den Parametern n und p.

Insbesondere ist die Verteilung $\text{Bin}(1, p)$ die gleiche wie die Verteilung $\text{Ber}(p)$. Man kann auch prüfen, dass wenn $X \sim \text{Bin}(m, p)$ und $Y \sim \text{Bin}(n, p)$ und X, Y unabhängig sind, dann $X + Y \sim \text{Bin}(m + n, p)$.

Geometrische Verteilung Eine Geometrische Verteilung beschriebt das erste Auftreten eines Erfolges. Wir schreiben $X \sim \text{Geom}(p)$ und definieren:

Geometrische Verteilung

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

Sei $X_1, X_2, \ldots$ eine Folge von unendlich vielen unabhängigen Bernoulli ZV mit dem Parameter p. Dann $T := \min\{n \geq 1 : X_n = 1\}$ eine geometrische Zufallsvariable mit dem Parameter p.

Abwesenheit der Erinnerung von der geometrische Verteilung

Sei $T \sim \text{Geom}(p)$ für einige $0 < p < 1$. Dann $\forall n \geq 0 \forall k \geq 1 \mathbb{P}[T \geq n + k | T > n] = P[T \geq k]$.

Poisson-Verteilung Diese Verteilung ist eine Annäherung an die Binomialverteilung für grosse n und kleine p. Sie nimmt nur Werte in $\mathbb{N}$ an. Wir schreiben $X \sim \text{Poisson}(\lambda)$ und definieren:

Poisson-Verteilung

$$\forall k \in \mathbb{N}, \lambda > 0. \quad \mathbb{P}[X = k] = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

Poisson-Approximation der Binomialform

Es sei $\lambda > 0$. Für jedes $n \geq 1$, betrachten wir eine Zufallsvariable $X_n \sim \text{Bin}(n, \lambda/n)$. Dann $\forall k \in \mathbb{N} \lim_{n \to \infty} \mathbb{P}[X_n = k] = P[N = k]$, wobei N eine Poisson-Zufallsvariable mit dem Parameter λ ist.

Stetige ZV

Eine ZV X heisst **stetig**, wenn ihre Verteilungsfunktion F_X wie folgt geschrieben werden kann:

$$\forall x \in \mathbb{R}. \quad F_X(x) = \int_{-\infty}^{x} f(y) dy$$

Hierbei ist $f : \mathbb{R} \mapsto \mathbb{R}_+$ die **Dichte** von X. Für die Dichte gilt:

$$\int_{-\infty}^{+\infty} f(y) dy = 1$$

Es gelten folgende Eigenschaften:

1. $\mathbb{P}[a \leq x \leq b] = \mathbb{P}[a < x < b]$
2. $\mathbb{P}[X = x] = 0$
3. $\mathbb{P}[X \in [a,b]] = \mathbb{P}[X \in (a,b)]$

Stetige Verteilungen

Gleichverteilung Dies beschreibt die Situation wobei jedes Ereignis die gleiche Wahrscheinlichkeit hat. Wir schreiben $X \sim \mathcal{U}([a,b])$ und definieren:

Gleichverteilung

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

Eigenschaften einer Gleichverteilte ZV X in $[a,b]$:

- Die Wahrscheinlichkeit, in ein Intervall $[c, c+l] \subset [a,b]$ zu fallen, hängt nur von seiner Länge l ab:

$$\mathbb{P}[X \in [c, c+l]] = \frac{l}{b-a}$$

- Die Verteilungsfunktion von X ist gleich

$$F_X(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases}$$

Exponentialverteilung Dies ist das stetige Pendant zur Geometrischen Verteilung. Wir schreiben $X \sim \mathrm{Exp}(\lambda)$ und definieren:

Exponentialverteilung

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

Eigenschaften einer Exponentialverteilte ZV T mit Parameter λ

- Die Wartezeitwahrscheinlichkeit ist exponentiell klein:

$$\forall t \geq 0 \mathbb{P}[T > t] = e^{\lambda t}$$

- Abwesenheit der Erinnerung:

$$\forall t, s \geq 0 \mathbb{P}[T > t + s | T > t] = \mathbb{P}[T > s]$$

Normalverteilung Wir schreiben $X \sim \mathcal{N}(m, \sigma^2)$ und definieren:

Normalverteilung

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

Eigenschaften einer Normalverteilte ZV

- Wenn $X_1, \dots, X_n$ sind unabhängige Zufallsvariablen mit Parametern $(m_1, \sigma_1^2), \dots, (m_n, \sigma_n^2)$ jeweils, dann

$$Z = m_0 + \lambda_1 X_1 + \ldots + \lambda_n X_n$$

  eine normale Zufallsvariable mit den Parametern $m = m_0 + \lambda_1 m_1 + \ldots + \lambda_n m_n$ und $\sigma^2 = m_0 + \lambda_1^2 \sigma_1^2 + \ldots + \lambda_n^2 \sigma_n^2$ ist.
- Insbesondere, wenn $X \sim \mathcal{N}(0,1)$ ist (in diesem Fall sagen wir, dass X eine standardnormale Zufallsvariable), dann

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

  eine normale Zufallsvariable mit den Parametern m und σ^2 ist.
- Wenn X eine normale Zufallsvariable mit den Parametern m und σ^2 ist, dann liegt die gesamte "Wahrscheinlichkeit Masse" hauptsächlich im Intervall $[m - 3\sigma, m + 3\sigma]$. Wir haben nämlich

$$\mathbb{P}[|X - m| \geq 3\sigma] \leq 0.0027$$

Standard Normalverteilung $\mathcal{N}(0,1)$. Weder für die zugehörige Dichte $\varphi(t)$ noch die Verteilungsfunktion $\Phi(t)$ gibt es geschlossene Ausdrücke, aber die Verteilung

Standard Normalverteilung

$$\mathbb{P}[X \leq t] = \Phi(t) = \int_{-\infty}^{t} \varphi(s) ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{s^2}{2}}$$

ist tabelliert. Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Erwartungswert

Sei $X : \Omega \mapsto \mathbb{R}_+$ eine ZV mit nicht-negativen Werten. Dann ist

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

der **Erwartungswert** von X.

Erwartungswert diskreter ZV

Sei $X : \Omega \mapsto \mathbb{R}$ eine diskrete ZV mit $X \in W$ f.s. Sei $\phi : \mathbb{R} \mapsto \mathbb{R}$ eine Abbildung. Falls die Summe wohldefiniert ist, gilt:

$$\mathbb{E}[\phi(X)] = \sum_{x \in W} \phi(x) \cdot \mathbb{P}[X = x]$$

Sei $\phi = \mathrm{id}$, gilt

$$\mathbb{E}[X] = \sum_{x \in W} x \cdot \mathbb{P}[X = x]$$

Erwartungswert stetiger ZV

Sei X eine stetige ZV mit Dichtefunktion f. Sei $\phi : \mathbb{R} \mapsto \mathbb{R}$ eine Abbildung, sodass $\phi(X)$ eine Zufallsvariable ist. Sofern das Integral wohldefiniert ist, gilt:

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) f(x) dx$$

Auch hier können wir analog den Erwartungswert für $\phi = \mathrm{id}$ definieren.

Rechnen mit Erwartungswerten

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

$$\mathbb{E}[\lambda \cdot X + Y] = \lambda \cdot \mathbb{E}[X] + \mathbb{E}[Y]$$

Wir nennen dies auch die **Linearität** des Erwartungswertes.

Falls zwei ZV X, Y **unabhängig** sind, dann gilt auch:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Dies gilt nicht für die Division, hier müssen wir wie folgt vorgehen

$$\mathbb{E}\Big[\frac{X}{Y}\Big] = \mathbb{E}[X] \cdot \mathbb{E}\Big[\frac{1}{Y}\Big]$$

und dabei $\mathbb{E}[1/Y]$ individuell berechnen. Daraus ergibt sich dann die folgende Eigenschaft:

Seien $X_1, ..., X_n$ diskrete ZV. Dann sind die folgenden Aussagen äquivalent:

1. $X_1, ..., X_n$ sind unabhängig
2. Für jedes $\phi_1 : \mathbb{R} \mapsto \mathbb{R}, ..., \phi_n : \mathbb{R} \mapsto \mathbb{R}$ (messbar) beschränkt gilt:

$$\mathbb{E}[\phi_1(X_1) \cdot ... \cdot \phi_n(X_n)] = \mathbb{E}[\phi_1(X_1)] \cdot ... \cdot \mathbb{E}[\phi_n(X_n)]$$

Extremwertformel

Sei X eine diskrete ZV mit Werten in $\mathbb{N}$. Dann gilt folgende Identität, auch **Tailsum-Formel** genannt (**Achtung!** $n = 1$):

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}[X \geq n]$$

Sei X eine stetige ZV mit $X \geq 0$ f.s., dann gilt:

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}[X > x] dx$$

Ungleichungen

Monotonie

Seien X, Y ZV sodass $X \leq Y$ f.s. dann gilt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Markov Ungleichung

Sei X eine ZV mit $X \geq 0$ f.s. dann gilt für jedes $a > 0$:

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Jensen Ungleichung

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

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$$

Varianz

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

Varianz

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

wobei $m = \mathbb{E}[X]$. Dabei wird σ_X auch die **Standardabweichung** von X genannt und beschreibt die typische Distanz eines Wertes $x \in X$ zu $\mathbb{E}[X]$.

Chebychev Ungleichung

Sei X eine ZV sodass $\mathbb{E}[X^2] < \infty$. Dann gilt für jedes $a \geq 0$:

$$\mathbb{P}[|X - m| \geq a] \leq \frac{\sigma_X^2}{a^2}$$

1. Sei X ein ZV sodass $\mathbb{E}[X^2] < \infty$ und $\lambda \in \mathbb{R}$:

$$\text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X]$$

2. $S = X_1 + ... + X_n$, wobei $X_1, ..., X_n$ paarweise unabhängig sind, so gilt:

$$\text{Var}[S] = \text{Var}[X_1] + ... + \text{Var}[X_n]$$

Kovarianz

Die **Kovarianz** kann verwendet werden, um die Abhängigkeit zweier ZV zu messen.

Covarianz

Seien X, Y zwei ZV mit $\mathbb{E}[X^2] < \infty, \mathbb{E}[Y^2] < \infty$, dann ist die **Kovarianz** zwischen X, Y definiert als:

$$\text{Cov}[X, Y] = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Gemeinsame Verteilungen

Diskrete gemeinsame Verteilungen

gemeinsame Verteilung

Seien $X_1, ..., X_n$ diskrete ZV mit $X_i \in W_i$ f.s. für $W_i \subset \mathbb{R}$. Die **gemeinsame Verteilung** (GV) von $X_1, ..., X_n$ ist die Familie $p = (p(x_1, ..., x_n))_{x_1 \in W_1, ..., x_n \in W_n}$ definiert durch:

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

Daraus ergibt sich folgende Eigenschaft. Sei $\phi : \mathbb{R}^n \mapsto \mathbb{R}$, so ist $Z = \phi(X_1, ..., X_n)$ eine diskrete ZV mit Werten in $W = \phi(W_1 \times ... \times W_n)$ und folgender Verteilung:

$$\forall z \in W. \quad \mathbb{P}[Z = z] = \sum_{\substack{x_1 \in W_1, ... x_n \in W_n \\ \phi(x_1, ..., x_n) = z}} \mathbb{P}[X_1 = x_1, ..., X_n = x_n]$$

Randverteilung

$$\forall z \in W_i.$$

$$\mathbb{P}[X_i = z] = \sum_{x_1, ..., x_{i-1}, x_{i+1}, ..., x_n} p(x_1, ..., x_{i-1}, z, x_{i+1}, ..., x_n)$$

Erwartungswert

Der Erwartungswert definiert als:

$$\mathbb{E}[\phi(X_1, ..., X_n)] = \sum_{x_1, ..., x_n} \phi(x_1, ..., x_n) \cdot p(x_1, ..., x_n)$$

Satz. Folgenden Aussagen sind äquivalent:

1. $X_1, ..., X_n$ sind unabhängig
2. Für alle $x_1 \in W_1, ..., x_n \in W_n$ gilt:

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

Stetige gemeinsame Verteilungen

gemeinsame Verteilung

Seien $X_1, ..., X_n$ stetige ZV, so haben sie eine **gemeinsame Verteilung**, falls eine Funktion $f : \mathbb{R}^n \mapsto \mathbb{R}_+$ existiert, die für jedes $a_1, ..., a_n \in \mathbb{R}$ folgende Eigenschaft erfüllt:

$$\mathbb{P}[X_1 \leq a_1, ..., X_n \leq a_n] = \int_{-\infty}^{a_1} ... \int_{-\infty}^{a_n} f(x_1, ..., x_n) dx_n ... dx_1$$

Wir nennen f die **gemeinsame Dichte**.

Seien $X_1, ..., X_n$ stetige ZV mit einer gemeinsamen Dicht f und $\phi : \mathbb{R}^n \mapsto \mathbb{R}$. Dann ist der Erwartungswert definiert als:

$$\mathbb{E}[\phi(X_1, ..., X_n)] = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \phi(x_1, ..., x_n) \cdot f(x_1, ..., x_n) dx_n ... dx_1$$

Falls $X_1, ..., X_n$ eine gemeinsame Dichte f besitzt. Dann ist die Randverteilung:

$$f_i(z) = \int_{(x_1, ..., x_{i-1}, x_{i+1}, ... x_n) \in R^{n-1}} f(x_1, ..., x_{i-1}, z, x_{i+1}, ..., x_n) dx_n ... dx_{i+1} dx_{i-1} ... dx_1$$

Seien $X_1, ..., X_n$ stetige ZV mit Dichten $f_1, ..., f_n$, dann sind die folgenden Aussagen äquivalent:

1. $X_1, ..., X_n$ sind unabhängig
2. $X_1, ..., X_n$ sind stetig mit gemeinsamer Dichte:

$$f(x_1, ..., x_n) = f_1(x_1) \cdot ... \cdot f_n(x_n)$$

3. Für alle $\phi_1 ..., \phi_n : \mathbb{R} \mapsto \mathbb{R}$ gilt:

$$\mathbb{E}[\phi_1(x_1) \cdot ... \cdot \phi_n(x_n)] = \mathbb{E}[\phi_1(x_1)] \cdot ... \cdot \mathbb{E}[\phi_n(x_n)]$$

Grenzwertsätze

Gegeben eine unendliche Sequenz an uiv. ZV $X_1, X_2, ...$, für jedes n betrachten wir die Teilsumme $S_n = X_1 + ... + X_n$.

Gesetz der grossen Zahlen

Sei $\mathbb{E}[|X_1|] < \infty$ und $m = \mathbb{E}[X_1]$, so gilt:

$$\lim_{n\to\infty} \frac{X_1 + ... + X_n}{n} = m \quad \text{f.s.}$$

Da die ZV uiv. sind, gilt auch $\mathbb{E}[|X_i|] < \infty$ und $m = \mathbb{E}[X_i]$ für alle i.

Konvergenz in Verteilung

Seien $(X_n)_{n\in\mathbb{N}}$ und X ZV. Wir schreiben

$$X_n \approx X \qquad \text{für } n \to \infty$$

falls für jedes $x \in \mathbb{R}$ gilt:

$$\lim_{n\to\infty} \mathbb{P}[X_n \leq x] = \mathbb{P}[X \leq x]$$

Zentraler Grenzwertsatz

Sei $\mathbb{E}[X_1^2] < \infty$ und wohldefiniert. Weiter sei $m = \mathbb{E}[X_1]$ und $\sigma^2 = \text{Var}(X_1)$, so gilt:

$$\mathbb{P}[\frac{S_n - n \cdot m}{\sqrt{\sigma^2 \cdot n}} \leq a] \xrightarrow[n\to\infty]{} \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$$

Der Zentrale Grenzwertsatz besagt, dass die Verteilung einer ZV

$$Z_n = \frac{S_n - n \cdot m}{\sqrt{\sigma^2 \cdot n}}$$

wie die Verteilung von $\mathcal{N}(0,1)$ aussieht. Es gilt also

$$Z_n \approx Z \qquad \text{für } n \to \infty$$

wobei $Z \sim \mathcal{N}(0,1)$. Für normalverteilte ZV $X_1, ..., X_n$, ist Z_n immer Standardnormalverteilt.

Statistik

Schätzer

Wir nehmen an, dass wir wie einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und Zufallsvariablen X_i haben. Zudem haben wir einen Parameterraum $\Theta \subset \mathbb{R}$, wobei $(\mathbb{P}_\theta)_{\theta\in\Theta}$ eine Familie von Wahrscheinlichkeitsmassen ist. $\mathbb{P}_\theta$ wird auch Modell genannt. Die Gesamtheit der beobachteten Daten nennen wir Stichprobe und die Anzahl n den Stichprobenumfang.

Schätzer

Ein **Schätzer** ist eine Zufallsvariable $T : \Omega \mapsto \mathbb{R}$, von der Form:

$$T = t(X_1, ..., X_n), \quad t : \mathbb{R}^n \mapsto \mathbb{R}$$

Ein Schätzer T heisst **erwartungstreu** für den Modellparameter, falls für alle $\theta \in \Theta$ gilt:

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

Der **Bias** (erwarteter Schätzfehler) von T im Modell $\mathbb{P}_\theta$ ist definiert als:

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

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

$$\text{MSE}_\theta[T] = \mathbb{E}_\theta[(T - \theta)^2]$$

$$\text{MSE}_\theta[T] = \text{Var}_\theta(T) + (\mathbb{E}_\theta[T] - \theta)^2$$

Maximum-Likelihood-Methode

Nehmen wir an $X_1, ..., X_n \in W$ sind ZV unter $\mathbb{P}_\theta$. Die **Likelikhood Funktion** ist definiert durch:

Likelikhood Funktion

$$L(x_1, ..., x_n; \theta) = \begin{cases} p(x_1, ..., x_n; \theta) & \text{im diskreten Fall} \\ f(x_1, ..., x_n; \theta) & \text{im stetigen Fall} \end{cases}$$

wobei p respektive f die gemeinsame Gewichtsfunktion / Dichtefunktion ist. Falls die X_i uiv. sind unter $\mathbb{P}_\theta$, so gilt:

$$p_X(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} p_{X_i}(x_i; \theta)$$

$$f_X(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

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

$$T_{\text{ML}} = t_{\text{ML}}(X_1, ..., X_n)$$

Die **Log-Likelihood** Funktion hat den Vorteil, dass sie durch eine Summe anstelle eines Produkts gegeben ist und daher oftmals einfach zu berechnen ist.

Konfidenzintervalle

Wir haben nun Methoden für Schätzer von unbekannten Parameter kennengelernt. Nun wollen wir wissen wie weit diese Schätzer vom effektiven Wert p weg liegen.

Konfidenzintervall

Sei $\alpha \in [0,1]$. Ein **Konfidenzintervall** für θ mit Niveau $1-\alpha$ ist ein Zufallsintervall $I = [A, B]$, sodass gilt

$$\forall \theta \in \Theta. \quad \mathbb{P}_\theta[A \leq \theta \leq B] \geq 1 - \alpha$$

wobei A, B ZV der Form $A = a(X_1, ..., X_n)$ und $B = b(X_1, ..., X_n)$, mit $a, b : \mathbb{R}^n \mapsto \mathbb{R}$, sind.

Wenn wir nun einen $T = T_{\text{ML}} \sim \mathcal{N}(m, 1/n)$ haben, (dz.Bsp. T_{ML} mit $X_1, ..., X_n$ uiv. $\mathcal{N}(m, 1)$ oder $T = T_{ML} = \frac{X_1 + ... + X_n}{n}$) suchen wir einen Konfidenzinterfall der Form:

$$I = [T - c/\sqrt{n}, T + c/\sqrt{n}]$$

Hierbei gilt:

$$\begin{aligned}\mathbb{P}_\theta[T - c/\sqrt{n} \leq m \leq T + c/\sqrt{n}] &= \mathbb{P}_\theta[-c \leq Z \leq c] \\ &= \mathbb{P}_\theta[Z \leq c] - \mathbb{P}_\theta[Z < -c] \\ &= \mathbb{P}_\theta[Z \leq c] - (1 - \mathbb{P}_\theta[Z \leq c]) \\ &= 2\phi(c) - 1\end{aligned}$$

wobei $Z = \sqrt{n}(T - m) = \frac{X_1 + ... + X_n - mn}{\sqrt{n}}$ eine standardnormalverteilte ZV ist.

Verteilungsaussagen

χ^2-Verteilung

Eine stetige ZV heisst **χ^2-verteilt** mit m Freiheitsgrade, falls ihre Dichte gegeben ist durch

$$f_X(x) = \frac{1}{2^{\frac{m}{2}} \Gamma(\frac{m}{2})} x^{\frac{m}{2} - 1} e^{-\frac{1}{2}x} \mathbb{1}_{x > 0},$$

wobei

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

Wir schreiben dann $X \sim \chi^2_m$.
Für natürliche Zahlen gilt $\Gamma(n) = (n-1)!$. Ein Spezialfall ist $m = 2$, hierbei erhalten wir $X \sim \text{Exp}(\frac{1}{2})$

Für ZV $X_1, ..., X_m$ u.i.v. mit $X_i \sim \mathcal{N}(0,1)$ ist die Summe

$$Y = \sum_{i=1}^m X_i^2 \sim \chi^2_m$$

t-Verteilung

Eine stetige ZV X heisst ***t*-verteilt** mit m Freiheitsgrade, falls ihre Dichte gegeben ist durch

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

Für X, Y unabhängig mit $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi^2_m$, ist der Quotient

$$Z = \frac{X}{\sqrt{\frac{Y}{m}}} \sim t_m.$$

Normalverteilung mit μ, σ^2 unbekannt

Seien $X_1, ..., X_n$ uiv. $\sim \mathcal{N}(\mu, \sigma^2)$. Dann sind

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

unabhängig.

Nun haben wir $X_1, ..., X_n$ ZV, die alle unter $\mathbb{P}_\theta$ uiv. $\sim \mathcal{N}(\mu, \sigma^2)$ sind. Die offensichtlichen Schätzer für μ, σ^2 sind das Stichprobenmittel $\overline{X}_n$ und die Stichprobenvarianz S^2. Für jedes $\theta \in \Theta$ gilt:

$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1} \text{ unter } \mathbb{P}_\theta$$

Also wollen wir:

$$1 - \alpha \leq \mathbb{P}_\theta \left[|\frac{\overline{X}_n - \mu}{S/\sqrt{n}}| \leq \frac{...}{S/\sqrt{n}} \right]$$

Um eine kleines Intervall zu erhalten, wollen wir die Bedingung mit Gleichheit erfüllen und nehmen $\frac{...}{S/\sqrt{n}} = t_{n-1,1-\alpha/2}$, somit erhalten wir für μ folgendes Konfidenzintervall:

$$\left[\overline{X}_n - t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}}, \overline{X}_n + t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}}\right]$$

Um ein Konfidenzintervall für σ^2 zu konstruieren brauchen wir:

$$\frac{1}{\sigma^2}(n-1)S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1} \text{ unter } \mathbb{P}_\theta$$

Mit der Notation $\chi^2_{m,\gamma}$ für das γ-Quantil einer χ^2_m Verteilung wollen wir:

$$1 - \alpha = \mathbb{P}_\theta \left[\chi^2_{n-1,\alpha/2} \leq \frac{1}{\sigma^2}(n-1)S^2 \leq \chi^2_{n-1,1-\alpha/2} \right]$$

Somit erhalten wir das Konfidenzintervall:

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \right]$$

Fazit: Das wichtigste Tool zur Bestimmung von Konfidenzintervallen sind Verteilungsaussagen über Schätzer. Die ist im Allgemeinen aber schwierig / nicht möglich.

Approximative Konfidenzintervalle

Zur Erinnerung der zentrale Grenzwertsatz besagt, dass wenn X_i in $\mathbb{P}_\theta$ uiv. sind, dann ist $\sum_{i=1}^n X_i$ approximativ normalverteilt mit $\mu = n\mathbb{E}[X_i]$ und $\sigma^2 = n\text{Var}_\theta[X_i]$. Insbesondere können wir daraus auch eine standard normalverteilte ZV erhalten (siehe Grenzwertsatz).

Tests

Null- und Alternativhypothese

Null- und Alternativhypothese

Die **Nullhypothese** H_0 und die **Alternativhypothese** H_A sind zwei Teilmengen $\Theta_0 \subseteq \Theta, \Theta_A \subseteq \Theta$ wobei $\Theta_0 \cap \Theta_A = \emptyset$. Eine Hypothese heisst **einfach**, falls die Teilmenge aus einem einzelnen Wert besteht; sonst zusammengesetzt.

Test und Entscheidung

Test

Ein **Test** ist ein (T, K), wobei T eine ZV der Form $T = t(X_1, ..., X_n)$ ist und $K \subseteq \mathbb{R}$ ist eine deterministische Teilmenge von $\mathbb{R}$. Man nennt T die **Teststatistik** und K den **Verwerfungsbereich** oder kritischen Bereich.

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

- die Hypothese H_0 wird verworfen, falls $T(\omega) \in K$
- die Hypothese H_0 wird nicht verworfen bzw. angenommen, falls $T(\omega) \notin K$

Fehler 1. Art

Ein **Fehler 1. Art** ist, wenn die H_0 verworfen wird, obschon sie richtig ist.

$$\mathbb{P}_\theta[T \in K], \quad \theta \in \Theta_0$$

Fehler 2. Art

Ein **Fehler 2. Art** ist, wenn die H_0 akzeptiert wird, obschon sie falsch ist.

$$\mathbb{P}_\theta[T \notin K] = 1 - \mathbb{P}_\theta[T \in K] \quad \theta \in \Theta_A$$

Signifikanzniveau und Macht

Bei der Auswahl eines geeigneten Tests ist insbesondere die Minimierung von Fehlern 1. Art entscheidend.

Signifikanzniveau

Sei $\alpha \in [0,1]$. Ein Test hat nun **Signifikanzniveau** α falls:

$$\forall \theta \in \Theta_0. \quad \mathbb{P}_\theta[T \in K] \leq \alpha$$

Das Sekundäre Ziel ist es den Fehler 2. Art zu vermeiden.

Macht

Die **Macht** eines Tests wird definiert als Funktion:

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

Bem. α klein entspricht einer kleine Wahrscheinlichkeit für Fehler 1. Art, während β gross einer kleinen Wahrscheinlichkeit für einen Fehler 2. Art entspricht.

Das obige asymmetrische Vorgehen macht es schwieriger, die Nullhypothese zu verwerfen als sie beizubehalten. Ein guter Test wird deshalb als Nullhypothese immer die Negation der eigentlich gewünschten Aussage benutzen.

Bem. $\beta = 1 - \mathbb{P}_{H_A}[H_0$ angenommen$]$. Das heisst, whenn zum Beispiel $\alpha = \mathbb{P}_{H_A}[T \leq 4]$ dann $\beta = 1 - \mathbb{P}[T > 4]$.

Konstruktion von Tests

Wir nehmen an, dass $X_1, ..., X_n$ diskret oder gemeinsam stetig sind unter $\mathbb{P}_{\theta_0}$ und $\mathbb{P}_{\theta_A}$, wobei $\theta_0 \neq \theta_A$ von der einfachen Form sind.

Likelihood-Quotient

Der **Likelihood-Quotient** ist somit wohldefiniert:

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

Wobei wir $R(x_1, ..., x_n) = +\infty$ definieren falls $L(x_1, ..., x_n; \theta_0) = 0$ sein sollte. Daraus ergibt sich, dass $R \gg 1 \Rightarrow H_A > H_0$ und $R \ll 1 \Rightarrow H_A < H_0$.

Likelihood-Quotient-Test

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

$$T = R(x_1, ..., x_n) \quad \text{und} \quad K = (c, \infty]$$

Der LQ-Test ist optimal, da jeder andere Test mit einem kleineren Signifikanzniveau auch eine kleinere Macht hat:

Neyman-Paerson-Lemma

Sei $c \geq 0$ und (T, K) der LQ-Test mit Parameter c und $H_A = \theta_A, H_0 = \theta_0$.

$$\alpha^* = \mathbb{P}_{\theta_0}[T \in K]$$

Sei (T', K') ein Test mit S-Niveau $\alpha \leq \alpha^*$ so gilt:

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

p-Wert

Wir wollen eine Hypothese $H_0 : \theta = \theta_0$ gegen eine Alternativhypothese $H_A : \theta \in \Theta_A$ testen. Eine Familie von Tests $(T, (K_t)_{t \geq 0})$ heisst **geordnet** bzgl. T falls $K_t \subset \mathbb{R}$ und $s \leq t \Rightarrow K_t \subset K_s$ gilt. Typische Beispiele dafür sind $K_t = (t, \infty)$ (rechtsseitiger Test), $K_t = (-\infty, -t)$ (linksseitiger Test) und $K_t = (-\infty, -t) \cup (t, \infty)$ (beidseitiger Test).

p-Wert

Sei $H_0 : \theta = \theta_0$ eine einfache Nullhypothese und $(T, K_t)_{t \geq 0}$ eine geordnete Familie von Tests. Der p-**Wert** ist definiert als ZV $G(t)$, wobei:

$$G : \mathbb{R}_+ \mapsto [0,1], \quad G(t) = \mathbb{P}_{\theta_0}[T \in K_t]$$

Der p-Wert hat folgende Eigenschaften:

1. Sei T stetig und $K_t = (t, \infty)$, so ist der p-Wert unter $\mathbb{P}_{\theta_0}$ auf $[0,1]$ gleichverteilt.
2. Für einen p-Wert γ gilt, dass alle Tests mit Signifikanzniveau $\alpha > \gamma$ die Nullhypothese verwerfen.

Wir können zusammenfassend sagen:

p-Wert ist klein $\implies$ H_0 wird wahrscheinlich verworfen

Testbeispiele

Normalverteilung, Test für Erwartungswert bei bekannter Varian

Variabeln: $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\theta, \sigma^2)$ unter $\mathbb{P}_\theta$ mit bekannter Varianz σ^2.
Hypothese: $H_0 : \theta = \theta_0$
Alternativ: $H_A : \theta > \theta_0$ oder $\theta < \theta_0$ (*einseitig*), oder $\theta \neq \theta_0$ (*zweiseitig*).
Teststatistik:

$$T := \frac{\overline{X}_n - \theta_0}{\sigma / \sqrt{n}}$$

Kritische Bereich: $(c_>, \infty)$ wenn $H_A : \theta > \theta_0$, bzw. $(-\infty, c_<)$, bzw. $(-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$. Im zweiseitigen Fall verwirft man H_0 also zugunsten der Alternative $H_A : \theta \neq \theta_0$, falls $|T| > c_{\neq}$ ist. Um $c_>, c_<, c_{\neq}$ zu bestimmen:

$$\alpha = \mathbb{P}_{\theta_0}[T \in K_>] = \mathbb{P}_{\theta_0}[T > c_>] = 1 - \mathbb{P}_{\theta_0}[T \leq c_>] = 1 - \Phi(c_>)$$

$$\Rightarrow c_> = \Phi^{-1}(1 - \alpha) =: z_{1-\alpha}$$

verwirft man also H_0, falls

$$\overline{X}_n > \theta_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

Analog ist $c_< = z_\alpha = -z_{1-\alpha}$ und $c_{\neq} = z_{1-\frac{\alpha}{2}}$.

Normalverteilung, Test für Erwartungswert bei unbekannter Var

Variabeln: $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\theta, \sigma^2)$ unter $\mathbb{P}_\theta$ wobei $\theta = (\mu, \sigma^2)$.

Hypothese: $H_0 : \mu = \mu_0$
Alternativ: $H_A : \theta > \theta_0$ oder $\theta < \theta_0$ (*einseitig*), oder $\theta \neq \theta_0$ (*zweiseitig*).
Teststatistik:

$$T := \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \text{ unter } \mathbb{P}_{\theta_0}$$

wir ersetzen also die unbekannte Varianz durch den Schätzer

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

Kritische Bereich: die kritischen Werte hier sind $c_> = t_{n-1,1-\alpha}$, bzw. $c_< = t_{n-1,\alpha} = -t_{n-1,1-\alpha}$, bzw. $c_f = t_{n-1,1-\frac{\alpha}{2}}$. Für $t_{m,\gamma}$ es gilt $P\left[X \leq t_{m,\gamma}\right] = \gamma$ für X-verteilt mit m Freiheitsgraden, d.h. $X \sim t_m$.

Gepaarter Zweistichproben-Test bei Normalverteilung

Variabeln: $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_n$ i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$ unter $\mathbb{P}_\theta$ wobei $\theta = (\mu, \sigma^2)$.
Die Differenzen $Z_i := X_i - Y_i$ sind nämlich unter $\mathbb{P}_\theta$ i.i.d. $\sim \mathcal{N}\left(\mu_X - \mu_Y, 2\sigma^2\right)$. Damit kann man die bisherigen Tests in leicht angepasster Form benutzen, sowohl für bekannte wie für unbekannte Varianz σ^2. Die resultierenden Tests heissen dann nicht überraschend gepaarter Zweistichproben-z-Test (bei bekanntem σ^2) bzw. gepaarter Zweistichproben- t-Test (bei unbekanntem σ^2).

Ungepaarter Zweistichproben-Test bei Normalverteilung

Variabeln: $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_m$ i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$ unter $\mathbb{P}_\theta$ wobei $\theta = (\mu, \sigma^2)$.

1. σ^2 bekannt (*ungepaarte Zweistichproben- z-Test.*):
   Teststatistik:
   $$T := \frac{\left(\bar{X}_n - \bar{Y}_m\right) - \left(\mu_X - \mu_Y\right)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1)$$
   unter jedem $\mathbb{P}_\theta$. Dabei ist σ nach Annahme bekannt, und $\mu_X - \mu_Y$ muss sich aus der gewünschten Hypothese H_0 als bekannt ergeben. Die
   Kritischen Werte: Wie oben geeignete Quantile der $\mathcal{N}(0,1)$-Verteilung, je nach Alternative.
2. σ^2 unbekannt *Zweistichproben-t-Test*:
   $$S_X^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2,$$
   $$S_Y^2 := \frac{1}{m-1} \sum_{j=1}^{m} \left(Y_j - \bar{Y}_m\right)^2.$$

Mit

$$S^2 := \frac{1}{m+n-2}\left((n-1)S_X^2 + (m-1)S_Y^2\right)$$
$$= \frac{1}{m+n-2}\left(\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 + \sum_{j=1}^{m}\left(Y_j - \bar{Y}_m\right)^2\right)$$

ist dann die Teststatistik

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

unter jedem $\mathbb{P}_\theta$. Der Rest geht dann analog wie oben.

Sonstiges

Nützliche Formeln

$$ax^2 + bx + c = 0 \Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Exponentialfunktion / Logarithmus

Für die Exponentialfunktion gilt:

(1) $\exp(x)\exp(y) = \exp(x+y)$
(2) $\exp(x) > 1, \quad \forall x > 0$
(3) $x^a = \exp(a \cdot \ln(x))$ und $x^0 = 1$
(4) $\exp(iz) = \cos(z) + i \cdot \sin(z)$
(5) $\exp\left(i \cdot \frac{\pi}{2}\right) = i$, $\exp(i\pi) = -1$ und $\exp(2i\pi) = 1$

Der natürliche Logarithmus $\ln :]0, \infty[\to \mathbb{R}$ bildet die Umkehrfunktion zu exp und ist streng monoton wachsend und stetig. Für den natürliche Logarithmus gilt:

(1) $\ln(1) = 0$
(2) $\ln(e) = 1$
(3) $\ln(a \cdot b) = \ln(a) + \ln(b)$
(4) $\ln(a/b) = \ln(a) - \ln(b)$
(5) $\ln(x^a) = a \cdot \ln(x)$
(6) $x^a \cdot x^b = x^{a+b}$
(7) $(x^a)^b = x^{a \cdot b}$

Im Allgemeinen gilt $\log_b(a) = \frac{\ln(a)}{\ln(b)}$.

Ableitungs und Integrations Regeln

$F(x)$	$F'(x) = f(x)$
Summenregel	$(f(x) + g(x))' = f'(x) + g'(x)$
Produktregel	$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$
Quotientenregel	$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}$ wenn $g(x) \neq 0$
Kettenregel	$(f(g(x)))' = f'(g(x)) \cdot g'(x)$
Part. Integration	$\int f'(x) \cdot g(x)dx = f(x) \cdot g(x) - \int f(x) \cdot g'(x)d$
Substitution	$\int_{\varphi(a)}^{\varphi(b)} f(x)\, dx = \int_a^b f(\varphi(t)) \cdot \varphi'(t)\, dt$
Logarithmus	$\int \frac{f'(t)}{f(t)}\, dt = \log(\lvert f(x) \rvert)$

Bsp. Substitution: Wir wollen $\int \cos\left(x^2\right)2x\; dx$ berechnen dabei gehen wir wie folgt vor:

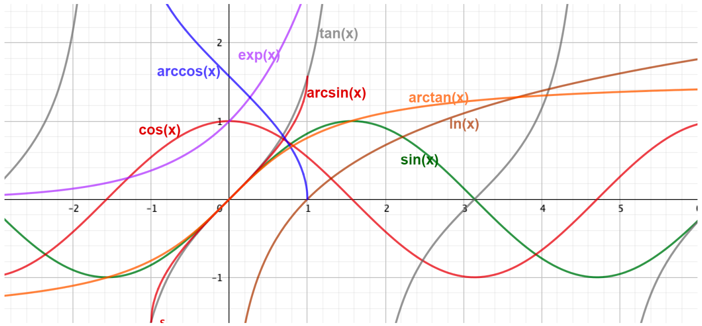
1. Zuerst bestimmen wir die Substitution: $u = x^2$
2. Nun berechnen wir die Umkehrfunktion: $x = \sqrt{u}$
3. Dann brauchen wir noch: $\frac{du}{dx} = \frac{dx^2}{dx} = \frac{2x}{1} = 2x \implies dx = \frac{du}{2x}$
4. Zuletzt können wir dies im Integral einsetzen und erhalten:

$$\int \cos\left(x^2\right)2x\; dx = \int \cos(u)2x\, \frac{du}{2x} = \int \cos(u)\; du = \sin(u)$$

Typische Ableitungen und Stammfunktionen

$F(x)$	$F'(x) = f(x)$
c	0
x^a	$a \cdot x^{a-1}$
$\frac{1}{a+1}x^{a+1}$	x^a
$\frac{1}{a\cdot(n+1)}(ax+b)^{n+1}$	$(ax+b)^n$
$\frac{x^{\alpha+1}}{\alpha+1}$	$x^\alpha, \alpha \neq -1$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\sqrt[n]{x}$	$\frac{1}{n}x^{\frac{1}{n}-1}$
$\frac{2}{3}x^{\frac{3}{2}}$	$\sqrt{x}$
$\frac{n}{n+1}x^{\frac{1}{n}+1}$	$\sqrt[n]{x}$
e^x	e^x
$\ln(\lvert x\rvert)$	$\frac{1}{x}$
$\log_a \lvert x\rvert$	$\frac{1}{x \ln a} = \log_a(e)\frac{1}{x}$
a^{cx}	$a^{cx} \cdot c \ln a$
x^x	$x^x \cdot (1+\ln x) \quad {\scriptstyle x>0}$
$x \cdot (\ln\lvert x\rvert - 1)$	$\ln\lvert x\rvert$

Funktionen



Aufgaben

0.1 Multiple Choice

Seien $A_1, A_2, A_3 \in \mathcal{F}$ paarweise unabhängige Ereignisse, welche Aussage ist korrekt?

- ☑ Die Ereignisse A_1, A_2, A_3 sind nicht zwangsläufig unabhängig
- ☐ Die Ereignisse A_1, A_2, A_3 sind zwangsläufig unabhängig

---

Es gilt $\mathbb{P}[X > t+s \mid X > s] = \mathbb{P}[X > t]$ für alle $t, s \geq 0$, falls

- ☐ $X \sim \mathcal{U}([a,b])$
- ☐ $X \sim \text{Poisson}(\lambda)$
- ☑ $X \sim \text{Exp}(\lambda)$ (Gedächnislosigkeit)

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

- ☑ X, Y sind immer stetig
- ☐ Die ZV sind nicht notwendigerweise stetig.

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

- ☑ $F_M(a) = F(a)^n$
- ☐ $F_M(a) = 1 - F(a)^n$
- ☐ $F_M(a) = (1 - F(a))^n$

0.2 Sonstige Aufgaben

Aufgabe Es werden ein blauer und ein grüner Würfel geworfen. Wir wählen $\Omega = \{1, ..., 6\}^2$. Wir betrachten die Algebra:

$$\mathcal{F}_{\text{sym}} = \{A \subseteq \Omega \mid \forall(\omega_1,\omega_2) \in \Omega, (\omega_1,\omega_2) \in A \Leftrightarrow (\omega_2,\omega_1) \in A\}$$

Zeige, dass $\mathcal{F}_{\text{sym}}$ eine σ-Algebra ist.

Wir müssen 3 Eigenschaften überprüfen.

1. $\forall(\omega_1,\omega_2) \in \Omega \Leftrightarrow (\omega_2,\omega_1) \in \Omega$ daher gilt $\Omega \in \mathcal{F}_{\text{sym}}$.
2. Sei $A \in \mathcal{F}_{\text{sym}}$. Somit gilt für jedes $(\omega_1,\omega_2) \in \Omega$
$$(\omega_1,\omega_2) \in A \Leftrightarrow (\omega_2,\omega_1) \in \Omega$$
was äquivalent ist zu
$$(\omega_1,\omega_2) \in A^\complement \Leftrightarrow (\omega_2,\omega_1) \in A^\complement$$
somit ist $A^\complement \in \mathcal{F}_{\text{sym}}$.
3. Seien $A_1, A_2, ... \in \mathcal{F}_{\text{sym}}$ gilt:
$$\begin{aligned}(\omega_1,\omega_2) \in \cup_{i=1}^\infty A_i &\Leftrightarrow \exists i.(\omega_1,\omega_2) \in A_i\\ &\Leftrightarrow \exists i.(\omega_2,\omega_1) \in A_i\\ &\Leftrightarrow (\omega_2,\omega_1) \in \cup_{i=1}^\infty A_i\end{aligned}$$
Somit folgt $\cup_{i=1}^\infty A_i \in \mathcal{F}_{\text{sym}}$.

Aufgabe Sei $(X_i)_{i\geq 1}$ eine unendliche Folge von unabhängig Ber(1/2)-verteilten ZV. Wir betrachten folgenden Algorithmus:

$i \leftarrow 1$
while $X_i = X_{i+1} = 1$ **do**
  $i = i+2$
end while
return $Z = X_i + 2 \cdot X_{i+1}$

Zeige, dass der Algorithmus immer nach endlich vielen Schritten terminiert (1). Zeige, dass Z eine gleichverteilte ZV in $\{0,1,2\}$ ist (2). Konstruiert einen Algorithmus, der eine Ber(1/5)-verteilten ZV ausgibt (3).

(1) Wir definieren $A_j := \{\text{While-Schlaufe wir j-Mal durchlaufen}\}$ und berechnen:

$$\begin{aligned}\mathbb{P}[A_j] &= \mathbb{P}\left[\bigcap_{i=1}^{2j}\{X_i = 1\} \cap (\{X_{2j+1} = 0\} \cup \{X_{2j+2} = 0\})\right]\\ &= \left(\Pi_{i=1}^{2j}\mathbb{P}[X_i = 1]\right) \cdot \mathbb{P}[X_{2j+1} = 0] \cdot \mathbb{P}[X_{2j+2} = 0]\\ &= \left(\frac{1}{2}\right)^{2j} \cdot \frac{3}{4}\end{aligned}$$

Wenn wir nun über alle A_j summieren, sehen wir, dass der Algorithmus immer in endlich Schritten terminieren wird.

(2) Wir wissen, dass alle A_j disjunkt sind.

$$\mathbb{P}[Z=0] = \mathbb{P}[\{Z=0\} \cap A] + \mathbb{P}[\{Z=0\} \cap A^\complement] = \sum_{j=0}^\infty \mathbb{P}[\{Z=0\} \cap A_j]$$

$$= \sum_{j=0}^\infty \left(\frac{1}{2}\right)^{2j+2} = \frac{1}{4}\sum_{j=0}^\infty\left(\frac{1}{4}\right)^j = \frac{1}{4} \cdot \frac{1}{1-\frac{1}{4}} = \frac{1}{3}$$

Dies können wir nun auch für $1, 2$ machen uns sehen, dass Z gleichverteilt sein muss.

(3) Wir betrachten folgenden Algorithmus:

$i \leftarrow 1$
while $X_i = X_{i+2} = 1$ or $X_{i+1} = X_{i+2} = 1$ **do**
  $i = i+3$
end while
return $Z = X_i + 2 \cdot X_{i+1} + 4 \cdot X_{i+2} = 4\ ?\ 1 : 0$

Es ist leicht wie in $(1), (2)$ zu zeigen, dass er alle Eigenschaften erfüllt.

Aufgabe Sei $T \sim \text{Exp}(\lambda)$. Berechne die Dichte von $T' = c \cdot T^2$ und den Erwartungswert von T'.

Sei $\phi : \mathbb{R} \mapsto \mathbb{R}$ messbar und beschränkt. Wir definieren $\psi(x) = \phi(c \cdot x^2)$. Somit erhalten wir:

$$\mathbb{E}[\phi(T')] = \mathbb{E}[\phi(c \cdot T^2)] = \mathbb{E}[\psi(T)] = \int_{-\infty}^{\infty} \psi(x) \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0} dx$$
$$= \int_0^{\infty} \phi(c \cdot x^2) \lambda e^{-\lambda x} dx = \int_0^{\infty} \phi(y) \lambda e^{-\lambda \sqrt{y/c}} \frac{dy}{2\sqrt{cy}}$$

Wobei wir die Dichte der Exponentialverteilung verwendet haben. daraus folgt:

$$f_{T'}(y) = \frac{\lambda}{2\sqrt{cy}} e^{-\lambda \sqrt{y/c}}$$

Für den Erwartungswert gilt $\mathbb{E}[c \cdot T^2] = c \cdot \mathbb{E}[T^2]$:

$$\mathbb{E}[T^2] = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

Somit erhalten wir $\mathbb{E}[T'] = \frac{2c}{\lambda^2}$.

Aufgabe Sei X eine ZV mit Verteilungsfunktion F_X. Zeige, dass X diskret ist.

$$F_X(a) = \begin{cases} 0, & a < 1 \\ 1/5, & 1 \leq a < 4 \\ 3/4, & 4 \leq a < 6 \\ 1, & 6 \leq a \end{cases}$$

Wir stellen fest, dass $\mathbb{P}[X = x] = 0$ für alle $x \notin \{1, 4, 6\}$. Da die Menge $\{1, 4, 6\}$ endlich ist, ist die ZV diskret.

Aufgabe Sei T eine ZV mit Verteilungsfunktion F_T. Zeige, dass t stetig ist.

$$F_T(a) = \begin{cases} 0, & a < 0 \\ 1 - e^{-2a}, & a <\geq 0 \end{cases}$$

Wir stellen fest, dass F_T stückweise stetig differenzierbar ist (auf $(-\infty, 0)$ und $(0, \infty)$). Somit folgt, dass T eine stetige ZV ist.

Diskrete Verteilungen

Verteilung	Parameter	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	$p_X(t)$	$F_X(t)$
Gleichverteilung	n: Anzahl Ereignisse x_i: Ereignisse	$\frac{1}{n}\sum_{i=1}^n x_i$	$\frac{1}{n}\sum_{i=1}^n x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^n x_i\right)^2$	$\frac{1}{n}$	$\frac{\lvert\{k:x_k\leq t\}\rvert}{n}$
Bernoulli	p: ErfolgsWK	p	$p\cdot(1-p)$	$p^t(1-p)^{1-t}$	$1-p$ für $0\leq t<1$
Binomial	p: ErfolgsWK n: Anzahl Versuche	np	$np(1-p)$	$\binom{n}{t}p^t(1-p)^{n-t}$	$\sum_{k=0}^t\binom{n}{k}p^k(1-p)^{n-k}$
Geometrisch	p: ErfolgsWK t: Anzahl Versuche	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$p(1-p)^{t-1}$	$1-(1-p)^t$
Poisson	λ: Erwartungswert und Varianz	λ	λ	$\frac{\lambda^t}{t!}e^{-\lambda}$	$e^{-\lambda}\sum\limits_{k=0}^{t}\frac{\lambda^k}{k!}$

1 Stetige Verteilungen

Verteilung	Parameter	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	$f_X(t)$	$F_X(t)$
Gleichverteilung	$[a,b]$: Intervall	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$	$\begin{cases}\frac{1}{b-a} & \text{for } t\in\{a,b\}\\ 0 & \text{otherwise}\end{cases}$	$\begin{cases}0 & \text{for } t<a\\ \frac{t-a}{b-a} & \text{for } t\in\{a,b\}\\ 1\text{for } t>b\end{cases}$
Exponentialverteilung	$\lambda:\frac{1}{\mathbb{E}[X]}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\begin{cases}\lambda e^{-\lambda t} & t\geq 0\\ 0 & t<0\end{cases}$	$\begin{cases}1-e^{-\lambda t} & t>0\\ 0 & t\leq 0\end{cases}$
Normalverteilung	σ^2: Varianz $\mu:\mathbb{E}[X]$	μ	σ^2	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\quad -\infty<t<\infty$	$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^t e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\mathrm{d}y$
χ^2-Verteilung	n: Freiheitsgrad	n	$2n$	$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}t^{\frac{n}{2}-1}e^{-\frac{t}{2}}\quad t>0$	$\mathrm{Gamma}(\frac{n}{2},\frac{t}{2})$
t-Verteilung	n: Freiheitsgrad	$\begin{cases}0 & n>1\\ \text{undef.} & \text{sonst}\end{cases}$	$\begin{cases}\frac{n}{n-2} & n>2\\ \infty & 1<n\leq 2\\ \text{undef.} & \text{sonst}\end{cases}$	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$	oof