Cheatsheet WuS

Nicolas Wehrli

June 2023

1 Grundbegriffe

1.1 Wahrscheinlichkeitsraum

Axiome von Kolmogorov

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

I. **Grundraum** Ω mit $\Omega \neq \emptyset$, wobei $\omega \in \Omega$ ein Elementarereignis ist.

II. **σ-Algebra** $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ wobei gilt:

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \implies A^{\complement} \in \mathcal{A}$
3. $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_i A_i \in \mathcal{A}$

III. **Wahrscheinlichkeitsmass** $\mathbb{P}$ auf $(\Omega, \mathcal{A})$ ist eine Abbildung $\mathbb{P} : \mathcal{A} \mapsto [0,1]$, wobei gilt:

1. $\mathbb{P}(\Omega) = 1$
2. $A_1, A_2, \dots \in \mathcal{A}, \forall i \neq j : A_i \cap A_j = \emptyset$
$\implies \mathbb{P}(\bigcup_i A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

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

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

Daraus folgt

1. $A_1, A_2, \dots \in \mathcal{A} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$
2. $A, B \in \mathcal{A} \implies (A \cup B), (A \cap B) \in \mathcal{A}$

und für $A, B \in \mathcal{A}$

1. $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$
2. $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

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

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

Siebformel

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

Atome
Sei Ω nicht leer und diskret. Sei $\mathcal{F}$ eine beliebige σ-Algebra über Ω. Eine nichtleere Menge $A \in \mathcal{F}$ heisst **atomare** Menge von $\mathcal{F}$ falls für alle $B \in \mathcal{F}$ gilt:

$$B \subseteq A \implies B = \emptyset \vee B = A$$

(Intuitiv: A ist die kleinste nichtleere Menge bezüglich der Inklusion in $\mathcal{F}$)
Die Menge der atomaren Mengen von $\mathcal{F}$ bezeichnen wir mit Atom($\mathcal{F}$). Jedes Element von $\mathcal{F}$ lässt sich als abzählbare Vereinigung von Elementen aus Atom($\mathcal{F}$) schreiben.

1.2 Bedingte Wahrscheinlichkeiten

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

Bedingte Wahrscheinlichkeit

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

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

Satz der totalen Wahrscheinlichkeit
Sei $(B_i)_{i\in I}$ eine Partition von Ω. Dann gilt für jedes beliebige $A \in \mathcal{A}$

$$\mathbb{P}(A) = \sum_{i:\ \mathbb{P}(B_i)>0} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

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

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

Mit dem **Satz der totalen W'keit** können wir $\mathbb{P}(A)$ umschreiben und kommen auf folgende Form:
Sei $(B_i)_{i\in I}$ eine **Partition** von Ω. Dann gilt für jedes beliebige $A \in \mathcal{A}, \mathbb{P}(A) > 0$

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{j:\ \mathbb{P}(B_j)>0} \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}$$

Intuition Bayessche Statistik
In dieser Form würde man A als das **eingetretene Ereignis** und die B_i als die verschiedene **Hypothesen** verstehen.
In der Bayesschen Statistik versucht man die Hypothese zu finden, so dass $\mathbb{P}(B_i|A)$ **maximiert** wird.
(Wurde in der Vorlesung nicht weiter behandelt)

1.3 Unabhängigkeit

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

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

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

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

Wir können die Definition der Unabhängigkeit auf beliebige Mengen von Ereignissen erweitern.

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

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

2 Zufallsvariablen

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

Zufallsvariable

Eine (reellwertige) **Zufallsvariable** auf Ω ist eine messbare Funktion $X : \Omega \to \mathbb{R}$.

$X : \Omega \to \mathbb{R}$ **messbar** $\iff \forall B \subset \mathbb{R}$ closed. $X^{-1}(B) \in \mathcal{A}$

Die Eigenschaft **messbar** ist bezüglich dem Wahrscheinlichkeitsmass $\mathbb{P} : \mathcal{A} \to [0,1]$ relevant. Dann ist $\mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\})$ wohldefiniert.

Bei diskretem Ω, können wir die rechte Seite vom ' $\iff$ ' durch $\forall x \in \mathbb{R} : X^{-1}(\{x\}) \in \mathcal{A}$ (1) ersetzen.
Für die Messbarkeit von X ist nur $X(\Omega) \subseteq \mathbb{R}$ entscheidend und jede Teilmenge $A \subseteq X(\Omega)$ ist abzählbar (da Ω abzählbar). Somit kann $X^{-1}(A)$ als abzählbare Vereinigung von $\bigcup_{x\in A} X^{-1}(\{x\})$ geschrieben werden.
(1) $\implies X^{-1}(A) \in \mathcal{A}$ per Def. σ-Algebra

2.1 Verteilungsfunktion

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

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

Die Funktion erfüllt folgende Eigenschaften:

1. F_X ist monoton wachsend
2. F_X ist rechtsstetig, i.e. $\lim_{h\downarrow 0} F_X(x+h) = F_X(x)$

3. $\lim_{x\to-\infty} F_X(x) = 0$ und $\lim_{x\to\infty} F_X(x) = 1$
4. $\forall a, b \in \mathbb{R}, a < b : \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

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

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

Intuitiv folgt daraus

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

Seien $X_1, ..., X_n$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann heissen $X_1, ..., X_n$ **unabhängig**, falls

$$\forall x_1, ..., x_n \in \mathbb{R} :$$
$$\mathbb{P}(X_1 \leq x_1, ..., X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdot ... \cdot \mathbb{P}(X_n \leq x_n).$$

2.2 Diskrete Zufallsvariablen

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

Seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen:
$X \leq Y$ f.s. $\iff$ $\mathbb{P}(X \leq Y) = 1$

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

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

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

Die **Verteilungsfunktion** einer diskreten ZV X:

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

Die **Gewichtsfunktion** einer diskreten ZV X:

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

2.3 Diskrete Verteilungen

Bernoulli-Verteilung ($X \sim \text{Ber}(p)$):
$X(\Omega) = \{0, 1\}$ und die Gewichtsfunktion ist definiert durch

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

Binomialverteilung ($X \sim \text{Bin}(n, p)$):
Wiederholung von n unabhängigen Bernoulli-Experimenten mit gleichem Parameter p.

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

Geometrische Verteilung ($X \sim \text{Geo}(p)$):
Warten auf den 1-ten Erfolg.

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

Poisson-Verteilung ($X \sim \text{Poisson}(\lambda)$):
Grenzwert der Binomialverteilung für grosse n und kleine p.

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

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

2.4 Stetige Zufallsvariablen

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

$$F_X(a) = \int_{-\infty}^{a} f(x)\, dx \text{ für alle } a \in \mathbb{R}.$$

wobei $f : \mathbb{R} \to \mathbb{R}^+$ eine nicht-negative Funktion ist. f wird dann als **Dichte** von X benannt.
Wenn $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ messbar ist, ist die Zufallsvariable X **absolut stetig**.

Intuition: $f(x)\,dx$ ist die Wahrscheinlichkeit, dass $X \in [x, x + dx]$.

Von F_X zu f:
Sei X eine Zufallsvariable mit stetiger Verteilungsfunktion F_X und F_X stückweise $\mathcal{C}^1$, d.h. es gibt $x_0 = -\infty < \ldots < x_{n-1} < x_n = +\infty$, sodass F_X auf jedem Intervall (x_i, x_{i+1}) Element von $\mathcal{C}^1$ ist.
Dann ist X eine stetige Zufallsvariable und die Dichte f kann wie folgt konstruiert werden:

$$\forall x \in (x_i, x_{i+1}) \quad f(x) = F_X'(x).$$

Bedingte Dichte

Seien X, Y ZV auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit gemeinsamer Dichte $f_{X,Y}(x, y)$ und Randdichte $f_Y(y) \neq 0$. Dann ist die bedingte Dichte von X bedingt durch Y:

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

2.5 Stetige Verteilungen

Gleichverteilung ($X \sim \mathcal{U}([a, b])$):
Die Dichte ist auf dem Intervall $[a, b]$ gleich.

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

Exponentialverteilung ($T \sim \text{Exp}(\lambda)$):
Lebensdauer oder Wartezeit eines allg. Ereignisses (Stetiges Äquivalent zur Geometrischen Verteilung).

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

Normalverteilung ($X \sim \mathcal{N}(m, \sigma^2)$):
Häufig verwendete Verteilung. Undefiniert für $\sigma = 0$.

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

1. Seien $X_1, \ldots, X_n$ **unabhängige** normalverteilte ZV mit Parametern $(m_1, \sigma_1^2), \ldots, (m_n, \sigma_n^2)$, dann ist
$$Z = m_0 + \lambda_1 X_1 + \ldots + \lambda_n X_n$$
eine normalverteilte ZV mit Parametern $m = m_0 + \lambda_1 m_1 + \ldots + \lambda_n m_n$ und $\sigma^2 = \lambda_1^2 \sigma_1^2 + \ldots + \lambda_n^2 \sigma_n^2$.
2. Sei $Z \sim \mathcal{N}(0, 1)$ eine **standardnormalverteilte** Zufallsvariable. Dann gilt für $X \sim \mathcal{N}(m, \sigma^2)$
$$X = m + \sigma \cdot Z$$

3 Erwartungswert

Erwartungswert - Diskrete ZV

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

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

Wenn $X : \Omega \to \mathbb{N}_0$, kann man auch den Erwartungswert als

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$$

schreiben.

Erwartungswert - Stetige ZV

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

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

solange das Integral wohldefiniert ist.
Sei X eine stetige ZV mit $X \geq 0$ f.s., dann gilt:

$$\mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X > x)\, dx$$

3.1 Rechnen mit Erwartungswerten

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

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

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

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

Eine generellere Form wäre folgende Äquivalenz:
$X_1, X_2, ..., X_n$ unabhängig
$\iff$
Für jede $\phi_1 : \mathbb{R} \to \mathbb{R}, \dots, \phi_n : \mathbb{R} \to \mathbb{R}$ stückweise stetig, beschränkt gilt

$$\mathbb{E}(\phi_1(X_1) \cdots \phi_n(X_n)) = \mathbb{E}(\phi_1(X_1)) \cdots \mathbb{E}(\phi_n(X_n))$$

3.2 Ungleichungen

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

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

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

$$\mathbb{P}(X \ge c) \le \frac{\mathbb{E}(g(X))}{g(c)}$$

Einfache Version:
Sei X eine ZV mit $X \ge 0$ f.s., dann gilt für jedes $t > 0$:

$$\mathbb{P}(X \ge t) \le \frac{\mathbb{E}(X)}{t}$$

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

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

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

$$\phi(\mathbb{E}(X)) \le \mathbb{E}(\phi(X))$$

3.3 Varianz

Varianz

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

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

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

1. Sei X ein ZV, sodass $\mathbb{E}(X^2) < \infty$ und $a, b \in \mathbb{R}$:
$$\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X)$$
2. Seien $X_1, ..., X_n$ paarweise unabhängig. Dann gilt
$$\text{Var}(X_1 + \ldots + X_n) = \text{Var}(X_1) + \ldots + \text{Var}(X_n)$$

Kovarianz

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

$$\text{Cov}(X, Y) := \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

1. $\text{Cov}(X, X) = \text{Var}(X)$
2. X, Y unabhängig $\implies$ $\text{Cov}(X, Y) = 0$ (Die Umkehrung ist falsch!)
3. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

- Momenterzeugende Funktion: $M_X(t) := \mathbb{E}\left(e^{tX}\right)$.
- Die k-te Ableitung ist das k-te Moment von X: $m_k := \mathbb{E}(X^k)$.
- Das k-te zentrale Moment von X: $\mu_k := \mathbb{E}\left((X - \mu)^k\right)$.
- Für X stetig gilt deshalb $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$.
- Das k-te empirische Moment der Realisierung $(x_{1\ldots n})$: $\hat{m}_k(x_{1\ldots n}) := \frac{1}{n} \sum_{i=1}^{n} x_i^k$.

3.4 Bedingter Erwartungswert

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.
Für ein beliebiges $B \in A, \mathbb{P}(B) > 0$ definieren wir den **bedingten Erwartungswert** X bedingt durch B als

$$\begin{aligned}\mathbb{E}(X \mid B) = \frac{\mathbb{E}(\mathbb{1}_B X)}{\mathbb{P}(B)} &= \sum_{x \in X(\Omega)} x \mathbb{P}(X = x \mid B) \\ &= \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\} \mid B)\end{aligned}$$

Bedingter Erwartungswert als Zufallsvariable
Wir betrachten eine Partition $\mathcal{B} = (B_i)_{i \in I}$ von Ω (B_i sind disjunkt und nichtleer, I abzählbar).
Dann definieren wir die **Zufallsvariable**

$$\mathbb{E}(X \mid \mathcal{B})(\omega) = \sum_{i \in I, \mathbb{P}(B_i > 0)} \mathbb{E}(X \mid B_i) \mathbb{1}_{B_i}(\omega)$$

1. **Intuition:** Die Information, die durch die Partition gegeben ist, ist dass eines der B_i eintreten wird. Bei der Realisierung durch das Eintreten des Elementarereignisses ω wird $\mathbb{E}(X \mid \mathcal{B})$ zu dem $\mathbb{E}(X \mid B_i)$ realisiert, bei welchem $\omega \in B_i$.
2. **Bemerkung:** Das $\mathcal{B}$ hat in der Vorlesung 2 verschiedene Bedeutungen. Es wird als Variable für sowohl die Borelsche σ-Algebra als auch die Partition von Ω verwendet.

4 Mehrere Zufallsvariablen

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

$$(x_1, \ldots, x_n) \mapsto F(x_1, \ldots, x_n) := \mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n)$$

4.1 Diskreter Fall - Gewichtsfunktion

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

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

Aus der gemeinsamen Gewichtsfunktion p bekommt man die gemeinsame Verteilungsfunktion mit

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

Seien $X_1, \ldots, X_n$ **diskrete** Zufallsvariablen in $(\Omega, \mathcal{F}, \mathbb{P})$, sodass $X_1 \in W_1, \ldots, X_n \in W_n$ f.s. für $W_1, \ldots, W_n \subset \mathbb{R}$ endlich oder abzählbar.
Für $\phi : \mathbb{R}^n \to \mathbb{R}$ beliebig, ist $Z = \phi(X_1, \ldots, X_n)$ eine diskrete Zufallsvariable mit $Z \in W = \phi(W_1 \times \ldots \times W_n)$ f.s. .
Die Gewichtsfunktion von Z ist gegeben durch $p_Z : W \to [0, 1]$:

$$p_Z(t) := \mathbb{P}(Z = t) = \sum_{\substack{x_1 \in W_1, \ldots, x_n \in W_n \\ \phi(x_1, \ldots, x_n) = t}} p(x_1, \ldots, x_n)$$

1. Mit dem vorherigen Satz können wir aus der gemeinsamen Verteilung die **Randverteilung** einer Zufallsvariablen extrahieren (wegsummieren). Wir verwenden dafür einfach die Funktion
$$\phi(x_1, \ldots, x_n) = x_i$$
2. Der Erwartungswert des Bildes der Funktion $\phi : \mathbb{R}^n \to \mathbb{R}$ ist
$$\mathbb{E}(\phi(X_1, \ldots, X_n)) = \sum_{x_1, \ldots, x_n} \phi(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$
3. Wir haben eine Äquivalenz:
$$\begin{aligned}&X_1, \ldots, X_n \text{unabhängig} \\ &\iff \\ &\forall x_1 \in W_1, \ldots, x_n \in W_n \\ &p(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1) \cdot \ldots \cdot \mathbb{P}(X_n = x_n)\end{aligned}$$

4.2 Stetiger Fall - Gemeinsame Dichte

Gemeinsame Dichte

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

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

für eine Funktion $f : \mathbb{R}^n \to [0, \infty)$, so heisst $f(x_1, \dots, x_n)$ die **gemeinsame Dichte** von $X_1, \dots X_n$.

1. $f(x_1, \dots, x_n) \geq 0$, und $= 0$ ausserhalb von $\mathcal{W}(X_1, \dots, X_n)$.
2. $$\mathbb{P}((X_1, \dots, X_n) \in A) = \int \cdots \int_{(x_1, \dots, x_n) \in A} f(x_1, \dots, x_n)\, dx_n \dots dx_1$$
   für $A \subseteq \mathbb{R}^n$ beliebig.
3. Haben X, Y die gemeinsame Verteilungsfunktion $F_{X,Y}$, so ist $F_X : \mathbb{R} \to [0, 1]$,
   $$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq \infty) = \lim_{y \to \infty} F_{X,Y}(x, y)$$
   die Verteilungsfunktion der *Randverteilung* von X. Analoges gilt für F_Y.
4. Falls X, Y eine gemeinsame Dichte $f(x, y)$ haben, so haben auch die Randverteilungen von X und Y Dichten $f_X : \mathbb{R} \to [0, \infty)$ und $F_Y : \mathbb{R} \to [0, \infty)$.
   $$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \text{ bzw. } f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$
   Die **Dichtefunktion** einer Randverteilung (Randdichte) entsteht aus der gemeinsamen Dichtefunktion durch "Wegintegrieren" der anderen Variable(n).

Wenn $X_1, \dots, X_n$ stetige ZV mit Dichten $f_1, \dots, f_n$, dann sind die folgenden Aussagen äquivalent:

- $X_1, \dots, X_n$ unabhängig
- $(X_1, \dots, X_n)$ ist stetig mit gemeinsamer Dichte
  $$f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$$
- Für alle $\phi_1 : \mathbb{R} \to \mathbb{R}, \dots, \phi_n : \mathbb{R} \to \mathbb{R}$ die stückweise stetig und beschränkt sind, gilt
  $$\mathbb{E}(\phi_1(X_1) \cdot \dots \cdot \phi_n(X_n)) = \mathbb{E}(\phi_1(X_1)) \cdot \dots \cdot \mathbb{E}(\phi_n(X_n))$$

4.3 Transformation von Zufallsvariablen

linearer Transformationssatz

Sei Z ein n-dimensionaler Zufallsvektor und $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ eine messbare Abbildung. Dann ist

$$H(\omega) = g(Z(\omega))$$

ein m-dimensionaler Zufallsvariable und ferner gilt

$$\mathbb{P}(H \in A) = \mathbb{P}(Z \in g^{-1}(A)).$$

Wenn g linear und umkehrbar (i.e. $g(x) = m + Bx$ mit $\det(B) \neq 0$) und unter Vorraussetzung, dass die Verteilung von Z absolut stetig ist, dann ist H auch absolut stetig und es gilt:

$$f_H(x) = \frac{1}{|\det(B)|} f_Z(B^{-1}(x - m)).$$

Beispielrechnung

$Z = (X, Y)$ 2-dim Zufallsvektor. Wir wollen die Dichte von $X + Y$ berechnen.
Man wäre versucht die Matrix B und den Vektor m wie folgt zu definieren:

$$B = \begin{pmatrix} 1 & 1 \end{pmatrix} \text{ und } m = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$\implies g((X, Y)) = \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} = X + Y$$

Dann wäre aber B (und somit g) nicht invertierbar! Deshalb wollen wir B so wählen, dass $g((X, Y)) = (X, X + Y)$:

$$\begin{pmatrix} X \\ X + Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$
$$\det(B) = 1 \neq 0 \implies B \text{ invertierbar}$$
$$B^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

Nach dem linearen Transformationssatz gilt

$$\begin{aligned} f_{X,X+Y}(x, z) &= \frac{1}{|\det(B)|} f_{X,Y}\left(B^{-1} \cdot \begin{pmatrix} x \\ z \end{pmatrix}\right) \\ &= 1 \cdot f_{X,Y}\left(\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix}\right) \\ &= f_{X,Y}(x, z - x) \end{aligned}$$

Aus der gemeinsamen Dichte $f_{X,X+Y}$ können wir die Dichte f_{X+Y} bestimmen.

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_{X,X+Y}(x, z)\, dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx \end{aligned}$$

Falls X und Y unabhängig

$$= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x)\, dx$$

4.3.1 Charakterisierung der Dichte durch $\mathbb{E}$

Sei $\phi : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung und $X_1, \dots, X_n$ ZV mit gemeinsamer Dichte f. Dann lässt sich $\mathbb{E}(Z)$ für die Zufallsvariable $Z = \phi(X_1, \dots, X_n)$ mit

$$\mathbb{E}(Z) = \int \cdots \int_{\mathbb{R}^n} \phi(x_1, \dots, x_n) \cdot f(x_1, \dots, x_n)\, dx_n \dots dx_1$$

berechnen.
Dies reicht aber nicht, um die Dichte einer transformierten ZV zu berechnen. Mehrere Zufallsvariablen mit unterschiedlichen Dichten können den gleichen Erwartungswert haben.

Sei $f : \mathbb{R} \to \mathbb{R}_+$ eine Abblidung, sodass $\int_{-\infty}^{\infty} f(z)\, dz = 1$. Dann sind folgende Aussagen äquivalent

- Z ist stetig mit Dichte f
- Für jede stückweise stetige, beschränkte Abbildung $\psi : \mathbb{R} \to \mathbb{R}$ gilt
  $$\mathbb{E}(\psi(Z)) = \int_{-\infty}^{\infty} \psi(z) f(z)\, dz$$

Beispielrechnung

Wir können diese Erkenntnis nutzen, um die Dichte einer transformierten Zufallsvariable zu berechnen.
Seien X und Y zwei Zufallsvariablen mit gemeinsamer Dichtefunktion

$$f(x, y) = \begin{cases} \frac{1}{x^2 y^2} & \text{für } x \geq 1, y \geq 1 \\ 0 & \text{sonst.} \end{cases}$$

Bestimme die Dichtefunktion f_V der Zufallsvariable $V = XY$.
Sei $\psi : \mathbb{R} \to \mathbb{R}$ stückweise stetig und beschränkt. Wir definieren $\phi(x, y) = \psi(xy) = \psi(v)$ und berechnen

$$\begin{aligned} \mathbb{E}(\psi(V)) = \mathbb{E}(\phi(X, Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x, y) f(x, y)\, dx\, dy \\ &= \int_1^{\infty} \int_1^{\infty} \psi(xy) \frac{1}{x^2 y^2}\, dx\, dy \\ &\text{Substition } v = xy, dv = y\, dx \\ &= \int_1^{\infty} \int_y^{\infty} \psi(v) \frac{1}{v^2} \frac{dv}{y}\, dy \end{aligned}$$

$$\begin{aligned} A &= \{(v, y) \in \mathbb{R}^2 \mid 1 \leq y < \infty, y \leq v < \infty\} \\ &= \{(v, y) \in \mathbb{R}^2 \mid 1 \leq y \leq v, 1 \leq v < \infty\} \end{aligned}$$

Zeichnung hilft ;)

$$\begin{aligned} &= \int_1^{\infty} \int_1^{v} \psi(v) \frac{1}{v^2 y}\, dy\, dv \\ &= \int_1^{\infty} \psi(v) \frac{\ln(v)}{v^2}\, dv \\ &= \int_{-\infty}^{\infty} \psi(v) \cdot \frac{\ln(v)}{v^2} \mathbb{1}_{v \in [1, \infty)}\, dv \end{aligned}$$

$$\implies f_V(t) = \frac{\ln(v)}{v^2} \mathbb{1}_{v \in [1, \infty)}$$

genereller Transformationssatz

Sei Z ein n-dimensionaler Zufallsvektor mit Dichtefunktion $f_Z : \mathbb{R}^n \to \mathbb{R}_+$ und $\phi : \mathbb{R}^n \to \mathbb{R}^n$ stetig differenzierbar mit stetig differenzierbarer Umkehrabbildung ϕ^{-1}. Dann gilt für die Dichte f_U von $U = \phi(Z)$:

$$f_U(\vec{u}) = f_Z\left(\phi^{-1}(\vec{u})\right) \cdot |\det(J_{\phi^{-1}}(\vec{u}))|$$

Beweisidee. Für $A \subset \mathbb{R}^n$ gilt

$$\int_A f_U(\vec{u})\,d\vec{u} = \mathbb{P}(U \in A) = \mathbb{P}(Z \in \phi^{-1}(A)) = \int_{\phi^{-1}(A)} f_Z(\vec{z})\,d\vec{z}$$

Aus der mehrdimensionalen Integralrechnung folgt dann

$$\int_{\phi^{-1}(A)} f_Z(\vec{z})\,d\vec{z} = \int_A f_Z\left(\phi^{-1}(\vec{u})\right) \cdot |\det(J_{\phi^{-1}}(\vec{u}))|\,d\vec{u}$$

■

Beispielrechnung

Wir haben $Z = (X, Y)$, wobei X, Y unabhängig und exponentialverteilt mit $\lambda > 0$. Berechne die Dichtefunktion f_U von

$$U := \frac{X}{X+Y}$$

Wir definieren ϕ, so dass $(U, Y) = \phi(X, Y)$.

$$\phi(x,y) = \begin{pmatrix} \frac{x}{x+y} \\ y \end{pmatrix} \text{ und } \phi^{-1}(u,y) = \begin{pmatrix} \frac{uy}{1-u} \\ y \end{pmatrix}$$

Check: $\phi^{-1}\left(\frac{x}{x+y}, y\right) = \left(\frac{\frac{x}{x+y}y}{1-\frac{x}{x+y}}, y\right) = \left(\frac{xy}{x+y-x}, y\right) = (x, y)$.

We then have

$$\left|\det\left(J_{\phi^{-1}}(u,y)\right)\right| = \left|\det\begin{pmatrix} \frac{y}{1-u} + \frac{uy}{(1-u)^2} & 0 \\ \frac{u}{1-u} & 1 \end{pmatrix}\right| = \left|\frac{y(1-u)+uy}{(1-u)^2}\right| = \left|\frac{y}{(1-u)^2}\right|$$

Per genereller Transformationssatz gilt

$$f_{U,Y}(u,y) = f_{X,Y}\left(\frac{uy}{1-u}, y\right)\left|\frac{y}{(1-u)^2}\right| = \begin{cases} \lambda^2 e^{-\lambda\left(\frac{uy}{1-u}+y\right)}\left|\frac{y}{(1-u)^2}\right| & \text{if } \frac{uy}{1-u} \geq 0 \wedge y \geq 0 \\ 0 \cdot \left|\frac{y}{(1-u)^2}\right| & \text{sonst.} \end{cases}$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,Y}(u,y)\,dy = \int_0^{\infty} \frac{\lambda^2}{(1-u)^2} e^{-\frac{\lambda}{1-u}y} y \mathbb{1}_{u\in[0,1]}\,dy$$

per partielle Integration

$$= \mathbb{1}_{u\in[0,1]}$$

5 Konvergenz in Wahrscheinlichkeitsräumen

Unabhängigkeit einer Folge und iid./uiv.

Eine Folge von ZV $X_1, X_2, \ldots$ ist unabhängig, wenn $X_1, \ldots, X_n$ unabhängig ($\forall n \in \mathbb{N}$, nach der Definition in 2.1). Sie ist zudem **uiv./iid.**, falls $F_{X_i} = F_{X_j}, \forall i, j \in \mathbb{N}$.

In einem Wahrscheinlichkeitsraum können wir für eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ und einer ZV Z zwischen 3 Arten von Konvergenz unterscheiden:

1. **schwache Konvergenz / Konvergenz in Verteilung**
   Wir definieren $X_n \xrightarrow{d} Z$ (d for distribution) als
   $$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \lim_{n\to\infty} F_{X_n}(x) = F_Z(x) = \mathbb{P}(Z \leq x)$$
   für jede Stetigkeitsstelle $x \in \mathbb{R}$ von F_Z.
2. **Konvergenz in Wahrscheinlichkeit**
   Wir definieren $X_n \xrightarrow{\mathbb{P}} Z$ als
   $$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0$$
3. **Fast-sichere Konvergenz**
   Wir definieren $X_n \xrightarrow{\textbf{f.s.}} Z$ als
   $$\mathbb{P}(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = Z(\omega)\}) = 1$$

Wir haben dann auch

$$X_n \xrightarrow{\textbf{f.s.}} Z \implies X_n \xrightarrow{\mathbb{P}} Z \implies X_n \xrightarrow{d} Z$$

Die Umkehrung der Implikationen gilt nicht, wie folgende Beispiele zeigen:

1. $X_n \xrightarrow{d} Z \;\not\Longrightarrow\; X_n \xrightarrow{\mathbb{P}} Z$
   Sei $\Omega = \{0, 1\}$ und für alle $n \in \mathbb{N}$
   $$\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \frac{1}{2},\; X_n(\omega) = \begin{cases} 0 & \omega = 0 \\ 1 & \omega = 1 \end{cases}$$
   und
   $$\mathbb{P}(Z = 0) = \mathbb{P}(Z = 1) = \frac{1}{2},\; Z(\omega) = \begin{cases} 1 & \omega = 0 \\ 0 & \omega = 1 \end{cases}$$
   Aus $F_{X_n} = F_Z$ folgt direkt $X_n \xrightarrow{d} Z$.
   Da aber $|X_n(\omega) - Z(\omega)| = 1, \forall \omega \in \Omega$ und demzufolge $\lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \epsilon) \not\to 0$ für $n \to \infty$, gilt
   $$X_n \stackrel{\mathbb{P}}{\not\to} Z$$
2. $X_n \xrightarrow{\mathbb{P}} Z \;\not\Longrightarrow\; X_n \xrightarrow{\textbf{f.s.}} Z$
   Wir betrachten den Wahrscheinlichkeitsraum $([0,1], \mathcal{B}, \mathbb{P})$. Für ein beliebiges $n \in \mathbb{N}$ sei $k = \lfloor \log_2(n) \rfloor$ und $j \in \{0, \ldots, 2^k - 1\}$, sodass $n = 2^k + j$.
   Dann definieren wir
   $$X_n(\omega) = \mathbb{1}_{\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]}(\omega).$$
   und
   $$Z(\omega) = 0 \quad \forall \omega \in \Omega = [0,1]$$
   Zur Visualisierung würde die Folge so aussehen
   $$X_1 = \mathbb{1}_{[0,1]}, X_2 = \mathbb{1}_{[0,\frac{1}{2}]}, X_3 = \mathbb{1}_{[\frac{1}{2},1]}, X_4 = \mathbb{1}_{[0,\frac{1}{4}]} \text{ etc.}$$
   Wir hätten dann
   $$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0 \implies X_n \xrightarrow{\mathbb{P}} Z$$
   Aber für jedes $\omega \in [0,1]$ finden wir unendlich viele X_n mit $X_n(\omega) = 1$ und deshalb
   $$\mathbb{P}(\{\omega \in [0,1] \mid \lim_{n\to\infty} X_n(\omega) = Z(\omega)\}) = 0 \implies X_n \stackrel{\textbf{f.s.}}{\not\to} Z$$

5.1 Gesetz der grossen Zahlen

starkes Gesetz der grossen Zahlen

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

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

gilt dann

$$\overline{X}_n \longrightarrow \mu \textbf{ f.s.}$$

Dies ist eine fast-sichere Konvergenz.

schwaches Gesetz der grossen Zahlen

Sei $X_1, X_2, \ldots$ eine Folge von paarweise unkorrelierten Zufallsvariablen, die alle den gleichen Erwartungswert $\mathbb{E}(X_i) = \mu$ und die gleiche Varianz $\mathrm{Var}(X_i) = \sigma^2$ haben. Sei

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

Dann konvergiert $\overline{X}_n$ für $n \to \infty$ in Wahrscheinlichkeit gegen $\mu = \mathbb{E}(X_i)$, d.h.

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > \varepsilon) = 0 \text{ i.e. } \overline{X}_n \xrightarrow{\mathbb{P}} \mu$$

Bemerkung:
Zur Erinnerung:

$$X_i, X_j \text{ unkorreliert} \iff \mathrm{Cov}(X_i, X_j) = 0$$

Wir haben auch

$$X_i, X_j \text{ unabhängig} \implies X_i, X_j \text{ unkorreliert}$$

5.2 Zentraler Grenzwertsatz

Zentraler Grenzwertsatz

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

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

also

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

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

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

beziehungsweise

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

Beispielrechnung

Seien $(X_i)_{i\ge1}, (Y_i)_{i\ge1}$ und $(Z_i)_{i\ge1}$ Folgen von iid. ZV mit

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

und analog für Y_1 und Z_1. Wir definieren

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

Die Folge $\left((S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\right)_{n\ge1}$ wird zufällige Irrfahrt in $\mathbb{Z}^3$ genannt. Sei $\alpha > \frac{1}{2}$. Zeige, dass

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

wobei $\|(x,y,z)\|_2 := \sqrt{x^2+y^2+z^2}$ die euklidische Norm ist.
Schritt 1: Für alle $\alpha > 1/2$ zeigen wir $\mathbb{P}(|S_n^{(x)}| \le n^\alpha) \overset{n\to\infty}{\longrightarrow} 1$.
Da $\mathbb{E}(X_i) = 0$ und $\text{Var}(X_i) = 1$ folgt für $a \in \mathbb{R}$ beliebig per ZGS

$$\mathbb{P}\left(S_n^{(x)} \le a\sqrt{n}\right) = \mathbb{P}\left(\frac{S_n^{(x)}}{\sqrt{n}} \le a\right) \overset{n\to\infty}{\longrightarrow} \Phi(a)$$

und somit auch

$$\mathbb{P}\left(|S_n^{(x)}| \le a\sqrt{n}\right) = \mathbb{P}\left(S_n^{(x)} \le a\sqrt{n}\right) - \mathbb{P}\left(S_n^{(x)} \le -a\sqrt{n}\right) \overset{n\to\infty}{\longrightarrow} \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$

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

$$\mathbb{P}\left(|S_n^{(x)}| \le n^\alpha\right) = \mathbb{P}\left(|S_n^{(x)}| \le n^\beta\sqrt{n}\right) \longrightarrow \lim_{n\to\infty}(2\Phi(n^\beta) - 1) = 1$$

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

Schritt 2: $\forall \alpha > 1/2, \mathbb{P}\left(\left\|\left(S_n^{(x)}, S_n^{(y)}, S_n^{(z)}\right)\right\|_2 \le n^\alpha\right) \overset{n\to\infty}{\longrightarrow} 1$.
Sei $\alpha' \in (1/2, \alpha)$. Dann folgt

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

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

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

6 Schätzer

Wir treffen folgende Annahmen:

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

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

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

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

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

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

Sei $\theta \in \Theta$ und T ein Schätzer. Der **Bias** (erwartete 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:

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

6.1 Maximum-Likelihood-Methode

6.1.1 Likelihood-Funktion, ML-Schätzer

Die Likelihood-Funktion ist definiert als

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

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

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

6.1.2 Anwendung der Methode

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

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

Beispielrechnung mit Randstelle

Wir betrachten den Parameterraum $\Theta = \mathbb{R}_+ \times R_+$ mit $\theta = (\theta_1, \theta_2)$ und die Modellfamilie $(\mathbb{P}_\theta)_{\theta\in\Theta}$, wobei die ZV $X_1, \dots, X_n$ iid. mit
$f_{\theta_1,\theta_2}(x) = \begin{cases} \theta_2 e^{\theta_1\theta_2 - \theta_2 x} & \text{falls } x \ge \theta_1, \\ 0 & \text{sonst.} \end{cases}$
Bestimme den ML-Schätzer für $\theta = (\theta_1, \theta_2)$:
Die Likelihood-Funktion ist

$$\begin{aligned}L(x_1, \dots, x_n; \theta_1, \theta_2) &= \prod_{i=1}^n f_{\theta_1,\theta_2}(x_i) = \prod_{i=1}^n \theta_2 e^{\theta_1\theta_2 - \theta_2 x_i} \mathbb{1}_{x_i\in[\theta_1,\infty)}\\ &= \theta_2^n \exp\left(n\theta_1\theta_2 - \theta_2\sum_{i=1}^n x_i\right)\prod_{i=1}^n \mathbb{1}_{x_i\in[\theta_1,\infty)}\\ &= \theta_2^n \exp\left(n\theta_1\theta_2 - \theta_2\sum_{i=1}^n x_i\right)\mathbb{1}_{\min(x_i)\ge\theta_1}\end{aligned}$$

Wenn wir davon jetzt die Log-Likelihood Funktion nehmen würden, und diese ableiten, kommen wir auf etwas undefiniertes. Das liegt daran, dass sobald $\theta_1 > \min(x_i)$ gibt es einen Sprung zu 0.
Da $\theta_2 > 0$ folgt

$$L(x_1, ..., x_n; \theta_1, \theta_2) > 0 \iff \forall i \in \{1, \dots, n\} : x_i \ge \theta_1$$

Um $L(x_1, ..., x_n; \theta_1, \theta_2)$ zu maximieren, schränken wir den Ursprungsraum mit $\theta_1 \le \min_{1\le i\le n}(x_i)$ ein und bestimmen die Log-Likelihood Funktion als

$$f(\theta_1, \theta_2) = \log\left(L(x_1, ..., x_n; \theta_1, \theta_2)\right) = n\log(\theta_2) + n\theta_1\theta_2 - \theta_2\sum_{i=1}^n x_i$$

Da $\theta_2 > 0$ ist (unter der Einschränkung) die Log-Likelihood Funktion für $\theta_1 = \min_{1\le i\le n}(x_i)$ maximal (unabhängig von θ_2). Somit können wir θ_1 so fixieren und $\log(L)$ separat nach θ_2 maximieren.

$$\begin{aligned}\frac{\delta f}{\delta\theta_2} = \frac{n}{\theta_2} + n\theta_1 - \sum_{i=1}^n x_i &= 0\\ n\theta_1 - \sum_{i=1}^n x_i &= -\frac{n}{\theta_2}\\ \theta_2 &= \frac{n}{\sum_{i=1}^n x_i - n\theta_1}\end{aligned}$$

Überprüfen des kritischen Punktes:

$$\frac{\delta^2}{\delta^2\theta_2^2} f\left(\theta_1, \frac{n}{\sum_{i=1}^n x_i - n\theta_1}\right) = -\frac{n}{\frac{n}{\left(\sum_{i=1}^n x_i - n\theta_1\right)^2}}$$

$$= -\left(\sum_{i=1}^n x_i - n\theta_1\right)^2 < 0$$

Daraus erhalten wir die Maximimum-Likelihood-Schätzer für θ_1 und θ_2:

$$T_1 = \min_{1\le i\le n} X_i \text{ und } T_2 = \frac{n}{\sum_{i=1}^n X_i - nT_1}$$

6.2 Momentenmethode /-schätzer:

1. Sei $X_1, ..., X_n$ iid. eine Stichprobe.
2. Sei Θ ein m-dimensionaler Parameterraum.
3. Stelle für $\theta = (\theta_1, \ldots \theta_m)$ ein Gleichungssystem auf, in dem das k-te empirische Moment dem k-ten Moment gleichgesetzt wird: $\hat{m}_k(x_1, \ldots x_n) = g_k(\theta_1, \ldots \theta_m), k \in \{1, \ldots m\}$.
4. Der Vektor $\hat{\theta}(X_1, \ldots X_m)$ heisst Momentenschätzer des Parameters θ.

7 Tests

Die **Nullhypothese** H_0 und die **Alternativhypothese** H_A sind zwei Teilmengen $\Theta_0 \subseteq \Theta, \Theta_A \subseteq \Theta$ wobei $\Theta_0 \cap \Theta_A = \varnothing$. Falls keine explizite Alternativhypothese spezifiziert ist, so hat man $\Theta_A = \Theta \setminus \Theta_0$.
Eine Hypothese heisst *einfach*, falls die Teilmenge aus einem einzelnen Wert besteht; sonst *zusammengesetzt*.

Definition Test

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

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

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

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

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

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

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

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

7.1 Signifikanzniveau und Macht

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

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

Es ist meist unser primäres Ziel, die Fehler 1. Art zu minimieren. Das sekundäre Ziel ist, Fehler 2. Art zu vermeiden. Hierfür definieren wir die Macht eines Tests als Funktion:

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

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

7.2 Konstruktion von Tests

Wir nehmen an, dass $X_1, \ldots, X_n$ diskret oder gemeinsam stetig unter $\mathbb{P}_{\theta_0}$ und $\mathbb{P}_{\theta_A}$ sind, wobei $\Theta_0 \cap \Theta_A = \emptyset$ einfach sind ($\theta_0 \in \Theta_0 \wedge \theta_A \in \Theta_A$).
Der Likelihood-Quotient ist somit wohldefiniert:

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

(Falls $L(x_1, \ldots, x_n; \theta_0) = 0$ setzen wir $R(x_1, \ldots, x_n) = +\infty$.)
Für zusammengesetzte Θ_0 und Θ_A können wir den verallg. Likelihood-Quotient definieren:

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

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

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

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

Neyman-Pearson-Lemma
Der LQ-Test ist optimal, da jeder andere Test mit kleinerem (oder gleichem) Signifikanzniveau auch eine kleinere (oder gleiche) Macht hat.

7.3 Häufige Fälle

Normalverteilt - μ unbekannt, σ^2 bekannt (z-Test)

Erwartungstreuer Schätzer: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$
Verteilung unter $\mathbb{P}_\theta$: $\frac{\overline{X}_n - \theta_0}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0,1)$

1. Modell $X_1, \ldots, X_n \sim \mathcal{N}(\theta, \sigma^2)$ iid. unter $\mathbb{P}_\theta$
2. Hypothesen $H_0 : \theta = \theta_0$ und $H_A : \theta > \theta_0$, $H_A : \theta < \theta_0$ (einseitig) oder $H_A : \theta \neq \theta_0$ (zweiseitig)
3. Test $T = \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)$ unter $\mathbb{P}_{\theta_0}$
4. Verwerfungsbereich $K_> = (c_>, \infty)$, $K_< = (-\infty, -c_<)$ oder $K_{\neq} = (-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$
5. **Fall 1** $\alpha = \mathbb{P}_{\theta_0}(T \in K_>) = \mathbb{P}_{\theta_0}(T > c_>)$
Fall 2 $\alpha = \mathbb{P}_{\theta_0}(T \in K_<) = \mathbb{P}_{\theta_0}(T < -c_<) = 1 - \mathbb{P}_{\theta_0}(T \le c_<)$
Fall 3 $\alpha = \mathbb{P}_{\theta_0}(T \in K_{\neq}) = \mathbb{P}_{\theta_0}(T < -c_{\neq}) + \mathbb{P}_{\theta_0}(T > c_{\neq}) = \mathbb{P}_{\theta_0}(T < -c_{\neq}) + 1 - \mathbb{P}_{\theta_0}(T \le c_{\neq})$

Normalverteilt - μ, σ^2 unbekannt (t-Test)

Wir definieren $\vec{\theta} = (\mu, \sigma^2)$ und den Varianz-Schätzer $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

1. Modell $X_1, \ldots, X_n \sim \mathcal{N}(\theta, \sigma^2)$ iid. unter $\mathbb{P}_{\vec{\theta}}$
2. Hypothesen: $\Theta_0 = \{\mu_0\} \times (0, \infty)$, für die Alternativhypothese gibt es wieder die drei Fälle mit $\mu_A > \mu_0$, $\mu_A < \mu_0$ und $\mu_A \neq \mu_0$.
3. Teststatistik $T = \frac{\overline{X}_n - \mu_0}{\sqrt{S^2/n}} \sim t_{n-1}$
4. Verewerfungsbereich: $K_>$, $K_<$ oder $K_{\neq}$
5. Für Signifikanzniveau α, können wir die kritischen Werte als $c_> = t_{n-1,1-\alpha}, c_< = t_{n-1,1-\alpha}$ und $c_{\neq} = t_{n-1,1-\frac{\alpha}{2}}$ wählen. Hierbei bezeichnen wir mit $t_{m,\gamma}$, das γ-Quantil einer t_m-Verteilung (i.e. derjenige Wert $z = t_{m,\gamma}$, so dass für $X \sim t_m$ $\mathbb{P}(X \le z) = \gamma$ gilt).

Gepaarter Zweistichprobentest: μ_X, μ_Y, gleiche Varianz σ^2

Sei $X_1, ..., X_n$ iid. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_n$ iid. $\sim \mathcal{N}(\mu_Y, \sigma^2)$, wobei X_i, Y_i unabhängig.
Dann ist für $Z_i := X_i - Y_i$ die ZV $Z_1, \ldots, Z_n$ iid. $\sim \mathcal{N}(\mu_X - \mu_Y, 2\sigma^2)$.
Für bekanntes σ können wir auf den Z_i dann den z-Test ausführen, wenn unbekannt dann der t-Test.

Ungepaarter Zweistichprobentest: μ_X, μ_Y, gleiche Varianz σ^2

Sei $X_1, ..., X_n$ iid. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_m$ iid. $\sim \mathcal{N}(\mu_Y, \sigma^2)$, wobei $m \neq n$, $X_1, ..., X_n$ und $Y_1, ..., Y_m$ unabhängig.

1. σ^2 **bekannt.** Dann haben wir folgende Teststatistik

$$T := \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0, 1) \text{ unter jedem } \mathbb{P}_\theta$$

Mit dem können wir dann den z-Test ausführen.

2. σ^2 **unbekannt.** Empirische Varianzen der einzelnen beiden Datensätzen

$$S_X^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \text{ und } S_Y^2 := \frac{1}{m-1}\sum_{j=1}^m (Y_j - \overline{Y}_m)^2$$

kombinieren wir zu einer empirischen Varianz

$$S^2 := \frac{1}{m+n-2}((n-1)S_X^2 + (m-1)S_Y^2)$$

Daraus die Teststatistik

$$T := \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{m+n-2} \text{ unter jedem } \mathbb{P}_\theta$$

In diesem Fall machen wir damit ein t-Test.

7.4 p-Wert

Sei $T = t(X_1, \ldots, X_n)$ eine Teststatistik und $(T, K_t)_{t\geq 0}$ eine Familie von Tests.

Geordnete Teststatistik

Eine Familie von Tests heisst geordnet bzgl. T falls $K_t \subset \mathbb{R}$ und $s \leq t \implies K_t \subseteq K_s$. Beispiele:

- $K_t = (t, \infty)$ (rechtsseitiger Test)
- $K_t = (-\infty, -t)$ (linksseitiger Test)
- $K_t = (-\infty, -t) \cup (t, \infty)$ (beidseitiger Test)

Definition p-Wert

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

$$G : \Omega \mapsto [0,1], \quad G(\omega) = \mathbb{P}_{\theta_0}[T \in K_{t(X_1(\omega),\ldots,X_n(\omega))}]$$

Intuitiv: Wenn wir den Verwerfungsbereich mit dem realisierten Wert der Teststatistik bestimmen würden; was wäre das Signifikanzniveau (i.e. Fehler 1. Art)?
Der p-Wert hat folgende Eigenschaften:

1. Sei T stetig und $K_t = (t, \infty)$. Dann 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.

Insgesamt gilt also:

kleiner p-Wert $\implies$ H_0 wird wahrscheinlich verworfen

8 Konfidenzintervalle

Definition Konfidenzintervall

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

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

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

Wenn wir einen Schätzer $T = T_{ML} \sim \mathcal{N}(\theta, \mathrm{Var}(X))$ haben, suchen wir ein Konfidenzintervall der Form

$$I = \left[T - c\sqrt{\mathrm{Var}(X)}, T + c\sqrt{\mathrm{Var}(X)}\right]$$

Hierbei gilt:

$$\begin{aligned}
&\mathbb{P}_\theta\left(T - c\sqrt{\mathrm{Var}(X)} \leq \theta \leq T + c\sqrt{\mathrm{Var}(X)}\right) \\
&= \mathbb{P}_\theta(-c \leq Z \leq c) \\
&= \mathbb{P}_\theta(Z \leq c) - \mathbb{P}_\theta(Z < -c) \\
&= \mathbb{P}_\theta(Z \leq c) - (1 - \mathbb{P}_\theta(Z \leq c)) \\
&= 2\Phi(c) - 1
\end{aligned}$$

wobei $Z = \frac{T-\theta}{\sqrt{\mathrm{Var}(X)}} \sim \mathcal{N}(0,1)$ ist.

8.1 Approximatives Konfidenzintervall

Wir können den zentralen Grenzwertsatz benutzen, um eine standardnormalverteilte ZV zu erhalten, und damit die Konfidenzintervalle zu bestimmen.

9 Aufgaben

9.1 Multiple Choice

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

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

Sei Y eine stetige Zufallsvariable. Für alle $s, t \in \mathbb{R}^+$:
$\exists \lambda > 0.\ Y \sim Exp(\lambda) \iff \mathbb{P}(Y > s) = \mathbb{P}(Y > s + t \mid Y > t)$

- ✓ wahr.
- ☐ falsch.

9.2 Aufgaben Wahrscheinlichkeit

Dichte von $\max(X_1, X_2)$

Seien $X_1, X_2 \sim \mathcal{U}[0,1]$ unabhängige ZV und sei $X = \max(X_1, X_2)$. Berechne die Dichtefunktion von X und $\mathbb{P}[X_1 \leq x \mid X \geq y]$.

$$\begin{aligned}
F_X(t) &= \mathbb{P}[\max(X_1, X_2) \leq t] \\
&= \mathbb{P}[X_1 \leq t] \cdot \mathbb{P}[X_2 \leq t] = F_{X_1}(t) \cdot F_{X_2}(t) \\
f_X(t) &= \frac{d}{dt} F_{X_1}(t) \cdot F_{X_2}(t) = \frac{d}{dt} t^2 \cdot \mathbb{I}_{0\leq t\leq 1} = 2t \cdot \mathbb{I}_{0 \leq t \leq 1}
\end{aligned}$$

Für die Wahrscheinlichkeit brauchen wir eine Fallunterscheidung:

1. $x < 0$ oder $1 < x$:
$$\mathbb{P}(X_1 \leq x \mid X \geq y) = 0 \text{ bzw. } 1$$
2. $0 \leq x \leq y \leq 1$:
$$\frac{\mathbb{P}(X_1 \leq x, X_2 \geq y)}{\mathbb{P}(X \geq y)} = \frac{x(1-y)}{1-y^2}$$
3. $0 \leq y \leq x \leq 1$:
$$\frac{\mathbb{P}(X_1 < y, X_2 \geq y) + \mathbb{P}(y \leq X_1 \leq x)}{\mathbb{P}(X \geq y)} = \frac{x - y^2}{1-y^2}$$

Gemeinsame Dichte

Bestimme die gemeinsame Dichte von $P \sim \mathcal{U}[0,1]$ und $H \sim \mathcal{U}[0,P]$.
Wir wissen:

$$f_P(p) = \mathbb{I}_{p\in[0,1]} \quad f_{H|P}(h \mid p) = \frac{1}{p} \cdot \mathbb{I}_{h\in[0,p]}$$

Somit ist:

$$f_{P,H}(p,h) = f_P(p) \cdot f_{H|P}(h \mid p) = \frac{1}{p} \cdot \mathbb{I}_{0\leq h\leq p\leq 1}$$

Maximum und Minimum gleichverteilter ZVen

Seien U_1, U_2, U_3 unabhängige, $\mathcal{U}([0,1])$-verteilte Zufallsvariablen.
Wir betrachten die stetigen ZV $L := \min(U_1, U_2, U_3)$ und $M := \max(U_1, U_2, U_3)$.
Zeige für beliebige $\phi, \psi : \mathbb{R} \to \mathbb{R}$ stückweise stetig und beschränkt, dass

$$\mathbb{E}(\phi(M) \cdot \psi(L)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(m) \cdot \psi(l) \cdot 6(m-l) \mathbb{1}_{0\leq l\leq m\leq 1}\, dl\, dm$$

Wegen Unabhängigkeit ist die gemeinsame Dichte durch $f(u_1, u_2, u_3) = \mathbb{1}_{u_1\in[0,1]} \mathbb{1}_{u_2\in[0,1]} \mathbb{1}_{u_3\in[0,1]}$ bestimmt.

$$\begin{aligned}
&\mathbb{E}(\phi(M)\psi(L)) \\
&= \int_{\mathbb{R}^3} \phi(\max(u_1,u_2,u_3))\psi(\min(u_1,u_2,u_3)) f(u_1,u_2,u_3)\, du_1\, du_2\, du_3 \\
&= \int_0^1 \int_0^1 \int_0^1 \phi(\max(u_1,u_2,u_3))\psi(\min(u_1,u_2,u_3))\, du_1\, du_2\, du_3
\end{aligned}$$

Wir zerteilen jetzt dieses Integral in 6 Cases mit Indikatorfunktionen (die einzelnen Integrale summiert ergeben das gesuchte Integral). Beispielrechnung mit $\mathbb{1}_{u_1\leq u_2\leq u_3}$ (andere Fälle analog).

$$\begin{aligned}
&= \int_0^1 \int_0^1 \int_0^1 \phi(\max(u_1,u_2,u_3))\psi(\min(u_1,u_2,u_3)) \mathbb{1}_{u_1\leq u_2\leq u_3}\, du_1\, du_2\, du_3 \\
&= \int_0^1 \int_0^1 \int_0^1 \phi(u_3)\psi(u_1) \mathbb{1}_{u_1\leq u_2\leq u_3}\, du_1\, du_2\, du_3 \\
&= \int_0^1 \phi(u_3) \left(\int_0^{u_3} \psi(u_1) \left(\int_{u_1}^{u_3} du_2 \right) du_1 \right) du_3 \\
&= \int_0^1 \int_0^1 \phi(u_3)\psi(u_1)(u_3 - u_1) \mathbb{1}_{u_1\leq u_3}\, du_1\, du_3
\end{aligned}$$

Die anderen 5 Fälle sind analog und deshalb folgt

$$\mathbb{E}(\phi(M) \cdot \psi(L)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(m) \cdot \psi(l) \cdot 6(m-l) \mathbb{1}_{0\leq l\leq m\leq 1}\, dl\, dm$$

10 Tabellen & Diverses

10.1 Grenzwerte

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

Partielle Integration

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

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

Substitution

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

- $g'(x)$ muss sich herauskürzen, sonst nutzlos.
- Grenzen substituieren nicht vergessen.
- Alternativ: unbestimmtes Integral berechnet werden und dann u wieder durch x substituieren.
- Man kann auch das Theorem in die andere Richtung anwenden:
$$\int_a^b f(u)\,du = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(x))g'(x)\,dx$$
- Sei $\underline{X}, Y$ kompakt, $f : Y \subset \mathbb{R}^n \to \mathbb{R}$ stetig.
Sei $\gamma : \underline{X} \to Y$ mit $\underline{X} = \underline{X}_0 \cup B, Y = Y_0 \cup C$ (B, C Rand von $\underline{X}, Y$).
Wenn $\gamma : \underline{X}_0 \to Y_0$ bijektiv und C^1 mit $\det(J_\gamma(x)) \neq 0, \forall x \in \underline{X}_0$, dann gilt
$$\int_Y f(y)\,dy = \int_{\underline{X}} f(\gamma(x))|\det(J_\gamma(x))|\,dx$$

10.2 Ableitungen

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

10.3 Weitere Ableitungen

$\mathbf{F(x)}$	$\mathbf{f(x)}$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$\frac{-1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$x^x\ (x>0)$	$x^x \cdot (1+\ln x)$

Gamma-Verteilung

Die Gamma-Verteilung ist eine stetige Verteilung mit der Dichtefunktion

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

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

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

1. $\sum_{i=1}^n X_i^2 \sim \chi_n^2$
2. $\frac{1}{n}\sum_{i=1}^n X_i \sim \chi_1^2$
3. $X_1^2 + X_2^2 \sim Exp\left(\frac{1}{2}\right)$
4. Sei $Y \sim \chi_m^2$ unabhängig von $X \sim \mathcal{N}(0,1)$. Dann gilt
$$\frac{X}{\sqrt{\frac{1}{m}Y}} \sim t_m$$
5. Es gilt $\lim_{m\to\infty} t_m \sim \mathcal{N}(0,1)$ verteilt, für endliche m is t_m langschwänziger als $\mathcal{N}(0,1)$.

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

1. $\frac{n-1}{\sigma^2}S^2 \sim \chi_{n-1}^2$
2. $\overline{X}_n$ und S^2 sind unabhängig.
3. $$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}$$

10.4 MLE Schätzer

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

10.5 Verteilungen

Verteilung	Parameter	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	$p_X(t)/f_X(t)$	$F_X(t)$
Gleichverteilung	n: Anzahl Ereignisse x_i: Ereignisse	$\frac{1}{n}\sum_{i=1}^n x_i$	$\frac{1}{n}\sum_{i=1}^n x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^n x_i\right)^2$	$\frac{1}{n}$	$\frac{\lvert\{k: x_k \le t\}\rvert}{n}$
Bernoulli	p : ErfolgsWK	p	$p\cdot(1-p)$	$p^t(1-p)^{1-t}$	$1-p$ für $0 \le t < 1$
Binomial	n: Anzahl Versuche p : ErfolgsWK	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	$\frac{\lambda^t}{t!}e^{-\lambda}$	$e^{-\lambda}\sum_{k=0}^t \frac{\lambda^k}{k!}$
Gleichverteilung	$[a,b]$: Intervall	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$	$\begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst} \end{cases}$	$\begin{cases} 0 & x \le a \\ \frac{t-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$
Exponentialverteilung	$\lambda : \frac{1}{\mathbb{E}[X]}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\begin{cases} \lambda e^{-\lambda t} & t \ge 0 \\ 0 & t < 0 \end{cases}$	$\begin{cases} 1-e^{-\lambda t} & t > 0 \\ 0 & t \le 0 \end{cases}$
Normalverteilung	σ^2: Varianz $\mu : \mathbb{E}[X]$	μ	σ^2	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^t e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\mathrm{d}y$
χ^2-Verteilung	n: Freiheitsgrad	n	$2n$	$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}t^{\frac{n}{2}-1}e^{-\frac{t}{2}}$ für $t > 0$	$P\left(\frac{n}{2}, \frac{t}{2}\right)$
t-Verteilung	n: Freiheitsgrad	$\begin{cases} 0 & n > 1 \\ \text{undef.} & \text{sonst} \end{cases}$	$\begin{cases} \frac{n}{n-2} & n > 2 \\ \infty & 1 < n \le 2 \\ \text{undef.} & \text{sonst} \end{cases}$	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$	I'd rather not

Gamma-Funktion

$$\Gamma(v) := \int_0^\infty t^{v-1}e^{-t}\,dt, v \ge 0.$$

Es gilt $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$.

Binomischer Lehrsatz

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k}y^k$$

Cauchy Produkt

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

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

11 Quellen

Dieses Cheatsheet wurde von vorherigen (Julian Steinmann, Danny Camenisch) inspiriert (vor allem Kapitel 6-10). Definitionen und Aufgaben stammen aus den Slides von Prof. Teichmann, dem Skript (M. Schweizer, 2021) und den Übungsserien vom Frühlingssemester 2023.