Prof. V. Tassion
13. August 2023

WuS

Andrej Scheuer
ascheuer@student.ethz.ch

Mathematisches Framework

Wahrscheinlichkeitsraum

Grundraum: Menge Ω. $\omega \in \Omega$ heisst Elementarereigniss.

σ-Algebra Subset $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Erfüllt
E1 $\Omega \in \mathcal{F}$
E2 $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
E3 $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
E4 $\emptyset \in \mathcal{F}$
E5 $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
E6 $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$
E7 $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$

Wahrsch. Mass Sei Ω ein Grundraum und $\mathcal{F}$ eine σ-Algebra. Ein **Wahrsch. Mass** auf $(\Omega, \mathcal{F})$ ist eine Abbildung

$$\begin{aligned} \mathbb{P} : \mathcal{F} &\to [0,1] \\ A &\mapsto \mathbb{P}[A] \end{aligned}$$

welche die folgenden Bedingungen erfüllt:
E1 $\mathbb{P}[\Omega] = 1$
E2 (Zählbare Additivität) $\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A_i]$ falls $A = \bigcup_{i=1}^{\infty} A_i$ (disjunkt)
E3 $\mathbb{P}[\emptyset] = 0$
E4 (Additivität) $\mathbb{P}[A_1 \cup \ldots \cup A_k] = \mathbb{P}[A_1] + \ldots + \mathbb{P}[A_k]$ falls $A_i \cap A_j = \emptyset$ für $i \neq j$
E5 $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$
E6 $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$

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

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

Gilt auch für endliche Folgen.

Das Laplace-Modell ist definiert durch:
- $\mathcal{F} = \mathcal{P}(\Omega)$
- $\mathbb{P} : \mathcal{F} \to [0,1]$ definiert durch $\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \frac{|A|}{|\Omega|}$

Wichtige Ungleichungen

Monotonie Sei $A, B \in \mathcal{F}$ dann $A \subset B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$

Unoin Bound Seien $A_1, A_2, \ldots$ eine Folge von Ereignissen, dann:

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

Unoin Bound gilt auch für endliche Folgen.

Bedingte Wahrscheinlichkeit

Bedingte Wahrsch. Sei $(\Omega, \mathcal{F}, \mathbb{P})$. Seien A, B zwei Ereignisse mit $\mathbb{P}[B] > 0$. A gegeben B:

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

$\mathbb{P}[B|B] = 1$

Partition: $\Omega = A_1 \cup \ldots \cup A_n$ (paarweise disjunkt)

Totale Wahrsch. Sei $B_1, \ldots B_n$ eine Partition von Ω mit $\mathbb{P}[B_i] > 0$ für jedes $1 \leq i \leq n$:

$$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A| \ B_i]\mathbb{P}[B_i]$$

Bayes Sei $B_1, \ldots, B_n \in \mathcal{F}$ eine Partition von Ω mit $\mathbb{P}[B_i] > 0$ für jedes i. Für jedes Ereigniss A mit $\mathbb{P}[A] > 0$ gilt:

$$\forall i = 1, \ldots, n \quad \mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i]\mathbb{P}[B_i]}{\sum_{j=1}^{n} \mathbb{P}[A|B_j]\mathbb{P}[B_j]}$$

Unabhängigkeit

Unabhängigkeit zweier Ereignisse Sei $(\Omega, \mathcal{F}, \mathbb{P})$. Zwei Ereignisse sind **unabhängig**, falls

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

Sei $A, B \in \mathcal{F}$ zwei Ereignisse mit $\mathbb{P}[A], \mathbb{P}[B] > 0$. Dann sind die folgenden äquivalent: 1. $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$ 2. $\mathbb{P}[A|B] = \mathbb{P}[A]$ 3. $\mathbb{P}[B|A] = \mathbb{P}[B]$

Unabhängigkeit Eine Menge von Ereignisse $(A_i)_{i \in I}$ sind unabhängig falls:

$$\forall J \subset I \text{ endlich} \quad \mathbb{P}\left[\bigcup_{j \in J} A_j \right] = \prod_{j \in J} \mathbb{P}[A_j]$$

Falls A, B, C alle unabhängig sind gelten **alle** 4 Gleichungen:

$$\begin{aligned} \mathbb{P}[A \cap B] &= \mathbb{P}[A] \cdot \mathbb{P}[B] \\ \mathbb{P}[A \cap C] &= \mathbb{P}[A] \cdot \mathbb{P}[C] \\ \mathbb{P}[B \cap C] &= \mathbb{P}[B] \cdot \mathbb{P}[C] \\ \mathbb{P}[A \cap B \cap C] &= \mathbb{P}[A] \cdot \mathbb{P}[B] \cdot \mathbb{P}[C] \end{aligned}$$

Zufallsvariablen und Verteilungen

Zufallsvariablen

Zufallsvariable Sei $(\Omega, \mathcal{F}, \mathbb{P})$. Eine **Zufallsvariable** ist eine Abbildung $X : \Omega \to \mathbb{R}$ sodass für alle $a \in \mathbb{R}$

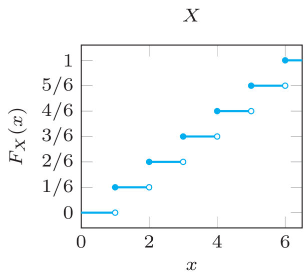
$$\{\omega ß \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$$

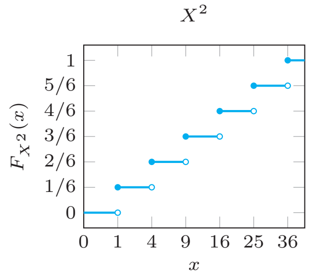
Verteilungsfunktion von X. $F_X : \mathbb{R} \to [0,1]$:

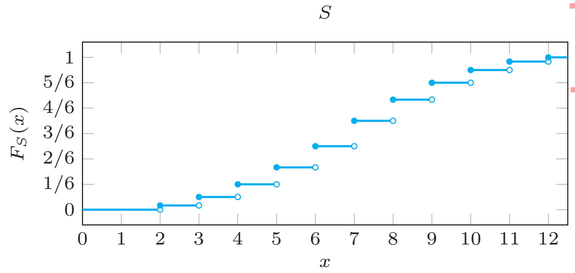
$$\forall a \in \mathbb{R} \quad F_X(a) = \mathbb{P}[X \leq a]$$

Zufallsexp: Werfe zwei Würfel. $\Omega = \{1, \ldots, 6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ und $\mathbb{P}[(\omega_1, \omega_2)] = 1/36, \forall \omega = (\omega_1, \omega_2) \in \Omega$.

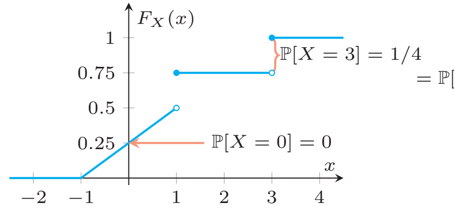
$$X : \begin{cases} \Omega & \to \mathbb{R} \\ (\omega_1, \omega_2) & \mapsto \omega_1 \end{cases} \quad X^2 : \begin{cases} \Omega & \to \mathbb{R} \\ (\omega_1, \omega_2) & \mapsto \omega_1^2 \end{cases} \quad S : \begin{cases} \Omega & \to \mathbb{R} \\ (\omega_1, \omega_2) & \mapsto \omega_1 + \omega_2 \end{cases}$$







$F_X(x)$ wobei X weder stetig noch diskret



$$\begin{aligned} &\mathbb{P}[0 \leq X \leq 1] \\ &= \mathbb{P}[0 \leq X < 1] + \mathbb{P}[X = 1] \\ &= 1/4 + 1/4 \\ &= 1/2 \end{aligned}$$

Prop. Seien $a < b$ reelle Zahlen

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

Eigenschaften Verteilungsfunktion Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Die Verteilungsfunktion $F = F_X : \mathbb{R} \to [0,1]$ von X erfüllt:
E1 F ist monoton wachsend.
E2 F ist rechtsseitig stetig[a].
E3 $\lim_{a \to -\infty} F(a) = 0$ und $\lim_{a \to \infty} F(a) = 1$

[a] $F(a) = \lim_{h \downarrow 0} F(a + h)$ für jedes $a \in \mathbb{R}$.

Unabhängigkeit

Unabhängig Seien $X_1, \ldots, X_n$ n Zufallsvariablen auf einem $(\Omega, \mathcal{F}, \mathbb{P})$. $X_1, \ldots X_n$ heissen **unabhängig** falls

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

Man kann zeigen, dass ZV $X_1, \ldots, X_n$ genau dann unabhängig sind, wenn:

$$\forall I_1 \subset \mathbb{R}, \ldots, I_n \subset R : \{X_1 \in \mathbb{R}\}, \ldots, \{X_n \in \mathbb{R}\} \text{ unabhängig}$$

gilt. Wobei I_k Intervalle sind.

Gruppierung Seien $X_1, \ldots, X_n$ n unabhängige ZV und $\phi_1, \ldots, \phi_k$ beliebige Funktionen, dann sind

$$Y_1 = \phi_1(X_1, \ldots, X_{i_1}), \ldots, Y_k = \phi_k(X_{i_{k-1}}, \ldots, X_{i_k})$$

unabhängig

Unabhängig 2 Eine unendliche Folge von ZV $X_1, X_2, \ldots$ sind:
- **unabhängig**, falls $X_1, \ldots, X_n$ für jedes n unabhängig sind.
- **unabhängig und gleich verteilt** (i.i.d.), falls sie unabhängig sind und die selbe Verteilungsfunktion haben. $\forall i, j \ F_{X_i} = F_{X_j}$

Transformation Zufallsvariablen

Um ZV wie reelle Zahlen zu verwenden können wir sie mit $\phi : \mathbb{R} \to \mathbb{R}$ transformieren.

$$\phi(X) := \phi \circ X \qquad \begin{array}{ccccc} \Omega & \xrightarrow{X} & \mathbb{R} & \xrightarrow{\phi} & \mathbb{R} \\ \omega & \mapsto & X(\omega) & \mapsto & \phi(X(\omega)) \end{array}$$

Konstruktion ZV

$X_1, X_2, \dots$ sind i.i.d. Ber(1/2) falls $\forall x_1, \dots, x_n \in \{0,1\}$

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = \frac{1}{2^n}$$

Kolmogorov Es existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine unendliche Folge von ZV $X_1, X_2, \dots$ sodass es sich dabei um eine Folge von i.i.d. Bernoulli ZV mit Parameter 1/2 handelt.

Prop. Die Abbildung $Y : \Omega \to [0,1]$ definiert durch

$$Y(\omega) = \sum_{n=1}^{\infty} 2^{-n} X_n(\omega)$$

ist eine gleich verteilte ZV in $[0,1]$.

Verallgemeinerte Inverse Die generalisierte Inverse von F ist $F^{-1}]0,1[\to \mathbb{R}$ definiert durch $\forall \alpha \in]0,1[\quad F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$

Inv.transformations Sampling Sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Sei U eine gleich verteilte ZV in $[0,1]$. Dann hat die ZV $X = F^{-1}(U)$ die Verteilung $F_X = F$.

Sei $U \sim \mathcal{U}([0,1])$. Konstruiere aus U eine Ber(1/3)-verteilte ZV Z.

Die Vert.funk. F einer Ber(1/3) ZV und ihre verallg. Inverse $F^{-1} :]0,1[\to \mathbb{R}$ sind

$$F(a) = \begin{cases} 0, & \text{für } a < 0 \\ 2/3, & \text{für } 0 \leq a < 0 \\ 1, & \text{für } a \leq 1 \end{cases} \qquad F^{-1}(\alpha) = \begin{cases} 0, & \text{für } 0 < \alpha \leq 2/3 \\ 1, & \text{für } 2/3 < \alpha < 1 \end{cases}$$

Es Folgt, dass $Z := F^{-1}(U)$ Ber(1/3)-verteilt ist.

Sei $U \sim \mathcal{U}([0,1])$. Konstruiere aus U eine gleichverteilte ZV U' in $[-1,2]$.

Für $U' \sim \mathcal{U}([-1,2])$:

$$F(a) = \begin{cases} 0, & \text{für } a < -1 \\ \frac{a+1}{3}, & \text{für } -1 \leq a \leq 2 \\ 1, & \text{für} a \geq 2 \end{cases} \qquad F^{-1}(\alpha) = 3\alpha - 1$$

Es folgt: $U' := F^{-1}(U)$ mit $U' \sim \mathcal{U}([-1,2])$.

Seien $F_1, F_2, \dots$ eine Folge von Funktionen[a] $\mathbb{R} \to [0,1]$. Dann existiert ein $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Folge i.i.d. ZV $X_1, X_2, \dots$ auf diesem Wahrscheinlichkeitsraum, sodass

- jedes i X_i hat eine Verteilungsfunktion F_i (also $\mathbb{P}[X_i \leq x] = F_i(x)$).
- $X_1, X_2, \dots$ sind unabhängig.

[a]Funktionen müssen die Eigenschaften von Verteilungsfunktion erfüllen.

Diskrete und stetige Zufallsvariablen

(Un-)Stetigkeitspunkte von F

Wahrsch. eines gegeben Werts Sei $X : \omega \to \mathbb{R}$ eine ZV mit F. Dann gilt für jedes $a \in \mathbb{R}$: $\mathbb{P}[X = a] = F(a) - F(a-)$.
Wobei $F(a-) := \lim_{h \downarrow 0} F(a - h)$ (Linker Grenzwert).

Fast sichere Ereignisse

Fast sicher Sei $A \in \mathcal{F}$ ein Ereignis. A tritt **fast sicher** ein, wenn $\mathbb{P}[A] = 1$.

Diskrete Zufallsvariablen

Diskrete ZV Eine ZV $X : \Omega \to \mathbb{R}$ ist **diskret** falls eine endliche oder abzählbare Menge $W \subset \mathbb{R}$ existiert, sodass $X \in W$ fast sicher ist.

Falls Ω endlich oder abzählbar ist, ist jede ZV diskret.

Verteilung von X (Gewichtsfunktion) Sei X eine diskrete ZV mit Werten von einem endlichen/abzählbaren $W \subset \mathbb{R}$ als Argumenten. Die **Verteilung** von X ist die Folge $(p(x))_{x \in W}$ definiert durch $\forall x \in W$ $p(x) := \mathbb{P}[X = x]$.

Prop. Die Verteilung $(p(x))_{x \in W}$ einer diskreten ZV erfüllt $\sum_{x \in W} p(x) = 1$.

Sei X ein ZV

$$\forall \omega \in \Omega \quad X(\omega) := \begin{cases} -1 & \text{falls } \omega = 1,2,3 \\ 0 & \text{falls } \omega = 4 \\ 2 & \text{falls } \omega = 5,6 \end{cases} \qquad W = \{-1, 0, 2\}$$

Dann nimmt X in W fast sicher an und die Verteilung ist:

$$p(-1) = \frac{1}{2}, \quad , p(0) = \frac{1}{6}, \quad p(2) = \frac{1}{3}$$

Verteilung p vs Verteilungsfunktion F_X

Prop. Sei X eine diskrete ZV, die fast sicher Werte in W (endlich/abzählbar) annimmt.

$$\forall x \in \mathbb{R} \quad F_X(x) = \sum_{y \leq x, y \in W} p(y)$$

Beispiele diskreter ZV

Bernoulli Verteilung Sei $0 \leq p \leq 1$. Eine ZV X ist Bernoulli ZV mit Parameter p, falls sie Werte in $W = \{0,1\}$ nimmt und

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

$X \sim \text{Ber}(p)$

$$f_p(x) = \begin{cases} p^x (1-p)^{1-x} & x \in \{0,1\} \\ 0 & \text{sonst} \end{cases} \qquad F_p(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Binomial Verteilung Sei $0 \leq p \leq 1$ und $n \in \mathbb{N}$. Eine ZV x ist eine Binomial ZV mit Parametern n und p, falls sie Werte in $W = \{0, \dots, n\}$ nimmt und

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

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

Falls $X_1, \dots, X_n$ unabhängige Bernoulli ZV mit Parameter p sind. Dann ist $S_n := X_1 + \ldots + X_n$ eine Binomial ZV mit Parametern n und p.

Geometrische Verteilung Sei $0 \leq p \leq 1$. Eine ZV X ist eine Geometrische ZV mit Parameter p falls sie Werte in $W = \mathbb{N} \setminus \{0\}$ nimmt und

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

$X \sim \text{Geom}(p)$
Falls $p = 1$ und $k = 1$, verwenden wir die Konvention, dass $0^0 = 1$ und dementsprechend $\mathbb{P}[X = 1] = 1$.

Für eine ZV $T \in \mathbb{N}$ ist $T + 1 \sim \text{Geom}(p)$, woraus folgt: $\mathbb{E}[T + 1] = \mathbb{E}[T] + 1$.

Prop. Sei $X_1, X_2, \dots$ eine Folge von unendlich vielen unabhängigen Bernoulli ZV mit Parameter p. Dann ist $T := \min\{n \geq 1 : X_n = 1\}$ eine Geometrische ZV mit Parameter p.

Zwar kann T den Wert $+\infty$ annehmen, falls alle $X_i = 0$ sind, jedoch gilt $\mathbb{P}[T = \infty] = 0$.

Gedächtnislosigkeit Sei $T \sim \text{Geom}(p)$ für ein $0 < p < 1$. Dann

$$\forall n \geq 0, \forall k \geq 1 \quad \mathbb{P}[T \geq n + k | T > n] = \mathbb{P}[T \geq k]$$

Poisson Verteilung Sei $\lambda > 0$ eine positive reelle Zahl. Eine ZV X ist eine Poisson ZV mit Parameter λ, falls sie Werte in $W = \mathbb{N}$ nimmt und

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

$X \sim \text{Poisson}(\lambda)$.
Sind $X \sim \text{Poisson}(\lambda); Y \sim \text{Poisson}(\mu); \lambda, \mu > 0$ dann $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Sei $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$, dann $\forall k \in \mathbb{N}$ $\lim_{n \to \infty} \mathbb{P}[X_n = k] = \mathbb{P}[N = k]$, wobei N Poisson(λ).

Stetige Zufallsvariablen

Stetige ZV Eine ZV $X : \Omega \to \mathbb{R}$ ist stetig, falls ihre Verteilungsfunktion F_X als

$$F_X(a) = \int_{-\infty}^{a} f(x) \, \mathrm{d}x \quad \forall a \in \mathbb{R}$$

geschrieben werden kann. f ist eine nicht-negative Funktion $f : \mathbb{R} \to \mathbb{R}$, genannt **Dichte**.

$f(x) \, \mathrm{d}x$ stellt die Wahrscheinlichkeit dar, dass X einen Wert im (infinitessimalen) Intervall $[x, x + \mathrm{d}x]$ annimmt.

Prop. Die Dichte erfüllt:

$$\int_{-\infty}^{\infty} f(x) \, \mathrm{d}x = 1$$

Prop. Sei X eine ZV. Nehme an, dass die Verteilungsfunktion F_X stetig und stückweise $\mathcal{C}^1$ ist[a] sodass F_X auf jedem Intervall $]x_i, x_{i+1}[$ von $\mathcal{C}^1$ ist. Dann ist X eine stetige ZV und eine Dichte f kann durch

$$\forall x \in]x_i, x_{i+1}[\quad f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$$

definiert werden. (mit beliebigen Werten für $x_1, \dots, x_{n-1}$)

[a]Es existiert $x_0 = -\infty < x_1 < \ldots < x_{n-1} < x_n = +\infty$

Beispiele stetiger Zufallsvariablen

Gleichverteilt Eine stetige ZV X ist **gleich verteilt** in $[a, b]$, falls

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

ist. $X \sim \mathcal{U}([a,b])$.

Eigenschaften einer gleich verteilten ZV

- Die Wahrscheinlichkeit in einem Intervall $[c, c + \ell] \subset [a, b]$ zu landen, hängt nur von der Länge ℓ ab.

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

Exponentielle Verteilung (mit $\lambda > 0$) Eine stetige ZV T ist exponentiell mit Parameter λ, falls

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad F(x) = 1 - e^{-\lambda x}$$

ist. $T \sim \mathrm{Exp}(\lambda)$

Eigenschaften einer exponentiellen ZV

- Die Wartewahrscheinlichkeit ist exponentiell klein.

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

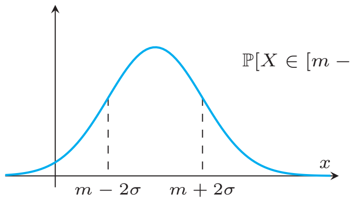
- Gedächtnislosigkeit $\forall t, s \geq 0$ $\mathbb{P}[T > t + s | T > t] = \mathbb{P}[T > s]$
- Sind $T_A \sim \mathrm{Exp}(\lambda), T_B \sim \mathrm{Exp}(\mu)$ dann ist die Dichte von $T_A + T_B$:

$$f_{T_A+T_B}(z) = \int_0^z f_{T_A}(x) \cdot f_{T_B}(z - x)\,\mathrm{d}x$$

Normalverteilung Eine stetige ZV X ist normal(-verteilt) mit Parametern m und $\sigma^2 > 0$, falls ihre Dichte gleich

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

ist. $X \sim \mathcal{N}\left(m, \sigma^2\right)$. Wobei m der Erwartungswert und σ^2 die Varianz sind. Falls $X \sim \mathcal{N}(0, 1) \Rightarrow X^2 \sim \mathcal{N}(1, 2)$.



$\mathbb{P}[X \in [m - 2\sigma, m + 2\sigma]] = 0.95$

Normalvert. $\Rightarrow$ Standardnormalvert. Sei $X \sim \mathcal{N}\left(m, \sigma^2\right)$. Um zur Standardnormalverteilung $Z \sim \mathcal{N}(0, 1)$ zu kommen rechnet man

$$Z = \frac{X - m}{\sigma}$$

Eigenschaften einer normalverteilten ZV

- Sind $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \ldots, X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ unabhängig, dann ist $X_1, \ldots, X_n \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.
- $X \sim \mathcal{N}(\mu, \sigma^2), a \in \mathbb{R} \Rightarrow X + a \sim \mathcal{N}(\mu + a, \sigma^2)$
- $X \sim \mathcal{N}(\mu, \sigma^2), a \in \mathbb{R} \setminus \{0\} \Rightarrow X/a \sim \mathcal{N}(\mu/a, \sigma^2/a^2)$
- Falls $X \sim \mathcal{N}(0, 1)$, dann ist $Z = m + \sigma \cdot X$ eine normalverteilten ZV mit Parametern m und σ^2.
- Falls X eine normalverteilte ZV mit m und σ^2 ist, dann ist das Wahrscheinlichkeitsmass hauptsächlich das Intervall $[m - 3\sigma, m + 3\sigma]$. $\mathbb{P}[|X - m| \geq 3\sigma] \leq 0.0027$

Erwartungswert

$X \sim$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
$\mathrm{Ber}(p)$	p	$p(1-p)$
$\mathrm{Bin}(n, p)$	np	$np(1-p)$
$\mathrm{Poisson}(\lambda)$	λ	λ
$\mathrm{Geom}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathcal{U}([a, b])$	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$
$\mathrm{Exp}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{N}\left(m, \sigma^2\right)$	m	σ^2

Erwartungswert generelle ZV

Erwartungswert Sei $X : \Omega \to \mathbb{R}_+$ eine ZV mit nicht-negativen Werten. Der Erwartungswert von X ist:

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

Der Erwartungswert kann endlich oder unendlich sein.

Prop. Sei X eine nicht-negative ZV. Dann ist $\mathbb{E}[X] \geq 0$, mit $\mathbb{E}[X] = 0$ gdw. $X = 0$ fast sicher.

Erwartungswert 2 Sei X eine ZV. Falls $\mathbb{E}[|X|] < \infty$, dann ist der Erwartungswert von X definiert durch $\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$.

$$X_+(\omega) = \begin{cases} X(\omega) & \text{falls } X(\omega) \geq 0 \\ 0 & \text{falls } X(\omega) < 0 \end{cases} \qquad X_-(\omega) = \begin{cases} -X(\omega) & \text{falls } X(\omega) \leq 0 \\ 0 & \text{falls } X(\omega) > 0 \end{cases}$$

Erwartungswert diskreter Zufallsvariablen

Prop. Sei $X : \Omega \to \mathbb{R}$ eine diskrete ZV mit Werten in W (endlich oder abzählbar), wobei W fast sicher ist.

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

gegeben, dass die Summe wohldefiniert ist.

Prop. Sei $X : \Omega \to \mathbb{R}$ eine diskrete ZV mit Werten in W (endlich oder abzählbar), wobei W fast sicher ist. Für jedes $\phi : \mathbb{R} \to \mathbb{R}$ gilt

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

gegeben, dass die Summe wohldefiniert ist.

Indikatorvariable

Sei $A \in \mathcal{F}$ ein Ereignis. Die **Indikatorfunktion** $\mathbb{1}_A$ von A ist definiert durch

$$\forall \omega \in \Omega \quad \mathbb{1}_A(\omega) = \begin{cases} 0 & \text{falls } \omega \notin A \\ 1 & \text{falls } \omega \in A \end{cases}$$

Dann ist $\mathbb{1}_A$ eine ZV.

$$\{\mathbb{1}_A \leq a\} = \begin{cases} \emptyset & \text{falls } a < 0 \\ A^c & \text{falls } 0 \leq a < 1 \\ \omega & \text{falls } a \geq 1 \end{cases}$$

$\emptyset$, A^c und Ω sind Elemente von $\mathcal{F}$. Weiterhin, wobei $X = 1_A$ gilt

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

Demnach ist $\mathbb{1}_A$ eine Bernoulli ZV mit Parameter $\mathbb{P}[A]$. $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$.

Erwartungswert stetiger Zufallsvariablen

Prop. Sei X eine stetige ZV mit Dichte f. Dann ist

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

gegeben, dass das Integral wohldefiniert ist.

Für die Normalverteilung: Falls $X \sim \mathcal{N}\left(m, \sigma^2\right)$, hat X die gleiche Verteilung wie $m + \sigma \cdot Y$, wobei Y die standard normalverteilte ZV ist.

$$\mathbb{E}[X] = \mathbb{E}[m + \sigma Y] = m + \sigma\mathbb{E}[X] = m + \sigma \cdot \int_{-\infty}^\infty x \cdot f_{0,1}(x)\,\mathrm{d}x = m + \sigma \cdot 0 = m$$

Prop. Sei X eine stetige ZV mit Dichte f. Sei $\phi : \mathbb{R} \to \mathbb{R}$ sodass $\phi(X)$ eine ZV ist. Dann folgt

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

gegeben, dass das Integral wohldefiniert ist.

Calculus

Linearität des Erwartungswert Seien $X, Y : \Omega \to \mathbb{R}$ zwei ZV, sei $\lambda \in \mathbb{R}$. Gegeben, dass der Erwartungswert wohldefiniert ist, gilt:

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

Wobei X und Y nicht unabhängig sein müssen.

Durch Induktion ergibt sich für jedes $n \in \mathbb{N}, n \geq 1$

$$\mathbb{E}[\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_n X_n] = \lambda_1 \mathbb{E}[X_1] + \lambda_2 \mathbb{E}[X_2] + \ldots + \lambda_n \mathbb{E}[X_n]$$

für beliebige X_i und λ_i.

$\mathbb{E}$ unabhängig Seien X, Y zwei unabhängige ZV, dann

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

Tailsum Formeln

Tailsum für nichtneg. ZV Sei X eine ZV, sodass $X \geq 0$ fast sicher.

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

Tailsum für diskrete ZV Sei X ein diskrete ZV, die Werte in $\mathbb{N} = \{0, 1, 2, \ldots\}$ nimmt. Dann

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

Charakterisierung via Erwartungswert

Dichte Sei X eine ZV. Sei $f : \mathbb{R} \to \mathbb{R}_+$ sodass $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$. Dann sind die folgenden Punkte äquivalent:

1. X ist stetig mit Dichte f.
2. Für jede messbare[a], beschränkte[b] Funktion $\phi : \mathbb{R} \to \mathbb{R}$

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

[a] Bedingung, die sicherstellt, dass $\phi(X)$ auch eine ZV ist.
[b] Für jedes $x \in \mathbb{R}$ existiert ein $C > 0$ sodass $|\phi(x)| \leq C$ ist.

Unabhängigkeit Seien X, Y zwei diskrete ZV. Die folgenden Punkte sind äquivalent:

1. X, Y sind unabhängig.
2. Für jedes messbare, beschränkte $\phi : \mathbb{R} \to \mathbb{R}$, $\psi : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[\phi(X) \cdot \psi(Y)] = \mathbb{E}[\phi(X)] \cdot \mathbb{E}[\psi(Y)]$$

Unabhängigkeit 2 Seien $X_1, \ldots, X_n$ n ZV. Dann sind die folgenden Punkte äquivalent:

1. $X_1, \ldots, X_n$ sind unabhängig.
2. Für jedes messbare, beschränkte $\phi_1 : \mathbb{R} \to \mathbb{R}, \ldots, \phi_n : \mathbb{R} \to \mathbb{R}$

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

Rezept Dichte

1. Fixiere $\phi : \mathbb{R} \to \mathbb{R}$ messbar, beschränkt (und unbekannt). $\phi(y) = \phi(h(x)) \overset{\sigma = \phi \circ h}{=} \sigma(x)$
2. $$\mathbb{E}[\phi(Y)] = \mathbb{E}[\sigma(X)] = \int_{-\infty}^{\infty} \sigma(x) \cdot f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \phi(h(x)) \cdot f_X(x)\,\mathrm{d}x = (*)$$
3. Variablenwechsel $z = h(x)$, $\frac{\mathrm{d}z}{\mathrm{d}x} = h'(x) \Leftrightarrow \mathrm{d}x = \frac{1}{h'(x)}\,\mathrm{d}z$
4. $$(*) = \int_{-\infty}^{\infty} \phi(z)\tilde{f}(z)\,\mathrm{d}z \Longrightarrow \tilde{f} = f_Y$$

$U \sim \mathcal{U}([0,1])$, $U' = a + (b-a)U = h(u)$

1. Fixiere ϕ (messbar, beschränkt). $\phi(U') = \phi(h(U)) = \sigma(U)$
2. $$f_U(x) = \mathbb{1}_{[0,1]}$$ $$\mathbb{E}[\phi(U')] = \mathbb{E}[\sigma(U)] = \int_{-\infty}^{\infty} \sigma(y) \cdot f_U(y)\,\mathrm{d}y = \int_0^1 \phi(h(y)) f_U(y)\,\mathrm{d}y$$
3. $z = h(y) = a + (b-a)y$, $\frac{\mathrm{d}z}{\mathrm{d}y} = b - a \Rightarrow \mathrm{d}y = \frac{1}{b-a}\,\mathrm{d}z$
4. $$\int_a^b \phi(z)\frac{1}{b-a}\,\mathrm{d}z = \int_{-\infty}^{\infty} \phi(z)\underbrace{\frac{1}{b-a}\mathbb{1}_{[a,b]}(z)}_{=\tilde{f}(z)}\,\mathrm{d}z$$

$\tilde{f}(z) = f_{U'}(z)$, $U' \sim \mathcal{U}[a,b]$

Seien X, Y zwei unabhängige ZV mit Dichte f_X bzw. f_Y. Bestimme die Dichte von $Z = \frac{X}{Y}$.

$$\mathbb{E}[\phi(Z)] = \mathbb{E}[\phi(X/Y)] = \iint_{-\infty}^{\infty} \phi(x/y) f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y =$$

$$\int_{-\infty}^{0}\left(\int_{-\infty}^{\infty} \phi(x/y) f_X(x)\,\mathrm{d}x\right) f_Y(y)\,\mathrm{d}y + \int_0^{\infty}\left(\int_{-\infty}^{\infty} \phi(x/y) f_X(x)\,\mathrm{d}x\right) f_Y(y)\,\mathrm{d}y$$

Integriere y einmal im negativen und einmal im positiven Bereich. Var.wechsel: $z = x/y, \mathrm{d}x = y\,\mathrm{d}z$.

$$\begin{aligned}\mathbb{E}[\phi(Z)] &= \int_{-\infty}^{0}\left(-\int_{-\infty}^{\infty} \phi(z) f_X(yz) y\,\mathrm{d}z\right) f_Y(y)\,\mathrm{d}y + \\ &\int_{-\infty}^{0}\left(\int_{-\infty}^{\infty} \phi(z) f_X(yz) y\,\mathrm{d}z\right) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \phi(z) f_x(yz)|y|\,\mathrm{d}z\right) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \phi(z)\left(\int_{-\infty}^{\infty} f_X(yz) f_Y(y)|y|\,\mathrm{d}y\right)\mathrm{d}z\end{aligned}$$

Z hat somit die Dichte: $\int_{-\infty}^{\infty} f_X(yz) f_Y(y)|y|\,\mathrm{d}y$.

Ungleichungen

Prop. Seien X, Y zwei ZV sodass $X \leq Y$. Dann $\mathbb{E}[X] \leq \mathbb{E}[Y]$. (Gegeben X, Y sind wohldefiniert)

Markov Ungleichung Sei X eine **nicht-negative** ZV. Dann gilt für jedes $a > 0$:

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

Jensens Ungleichung Sei X eine ZV. Sei $\phi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion. Falls $\mathbb{E}[\phi(X)]$ und $\mathbb{E}[X]$ wohldefiniert sind, dann

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

Cauchy-Schwarz Ungleichung Seien X, Y zwei ZV sodass $\mathbb{E}\left[X^2\right], \mathbb{E}\left[Y^2\right] < \infty$, dann gilt

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]}$$

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

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

Varianz

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

$$\sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \qquad \sigma_X^2 > 0$$

σ_X heisst die **Standardabweichung von X**.

1. Sei X eine ZV mit $\mathbb{E}\left[X^2\right] < \infty$. Dann $\sigma_X^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.
2. Seien $X_1, \ldots, X_n$ n (paarweise) unabhängige ZV und $S = X_1 + \ldots + X_n$. Dann $\sigma_S^2 = \sigma_{X_1}^2 + \ldots + \sigma_{X_n}^2$
3. $\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X]$

Kovarianz

Kovarianz Seien X, Y zwei ZV. Sei $\mathbb{E}\left[X^2\right] < \infty$ und $\mathbb{E}\left[Y^2\right] < \infty$ (endliches zweites Moment). Die **Kovarianz zwischen X und Y** ist definiert durch

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

- X, Y unabhängig $\Longrightarrow \mathrm{Cov}(X, Y) = 0$
- X, Y unabhängig $\Longleftrightarrow$ $\forall \phi \psi$ stückweise stetig, beschränkt. $\mathrm{Cov}(\phi(X), \psi(Y)) = 0$

Gemeinsame Verteilung

Diskrete gemeinsame Verteilungen

Gemeinsame Verteilung Seien $X_1, \ldots X_n$ n diskrete ZV mit $X_i \in W_i$ fast sicher, für beliebige W_i endlich/abzählbar. Die gemeinsame Verteilung von $(X_1, \ldots X_n)$ ist die Menge $p = \big(p(x_1, \ldots, x_n)\big)_{x_1 \in W_1, \ldots, x_n \in W_n}$ ist definiert durch

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

Seien $X \sim \mathrm{Ber}(1/2)$ und $Y \sim \mathrm{Ber}(1/2)$ unabhängig. Die Verteilung von (X, Y) ist gegeben durch

$$\forall x, y \in \{0, 1\}\ p(x, y) = \frac{1}{4}$$

Die Verteilung von (X, X) ist gleich

$$\forall x, y \in \{0, 1\}\ p(x, y) = \begin{cases} \frac{1}{2} & x = y \\ 0 & x \neq y \end{cases}$$

Die Verteilung von $(X, X + Y)$ ist

x	y	$p(x, y)$
0	0	1/4
0	1	1/4
0	2	0
1	0	1/4
1	1	1/4
1	2	0

Prop. Die gemeinsame Verteilung für beliebige ZV $X_1, \ldots, X_n$ erfüllt

$$\sum_{x_1 \in W_1, \ldots, x_n \in W_n} p(x_1, \ldots, x_n) = 1$$

Prop. Sei $n \geq 1$ und $\phi : \mathbb{R}^n \to \mathbb{R}$ eine beliebige Funktion. Seien $X_1, \ldots, X_n$ n diskrete ZV auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in endlichen/abzählbaren Mengen $W_1, \ldots, W_n$ fast sicher. Dann ist $Z = \phi(X_1, \ldots, X_n)$ eine diskrete ZV mit Werten in $W = \phi(W_1 \times \ldots \times W_n)$ fast sicher und eine Verteilung

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

Seien $X \sim \mathrm{Ber}(1/2)$ und $Y \sim \mathrm{Ber}(1/2)$ unabhängig und $X := X + Y$. Durch Anwendung auf $\phi(x, y) = x + y$ ergibt sich:

$$\mathbb{P}[Z = 0] = \sum_{\substack{x, y \in \{0,1\} \\ x + y = 0}} \mathbb{P}[X = x, Y = y] = \mathbb{P}[X = 0, Y = 0] = \frac{1}{4}$$

$$\mathbb{P}[Z = 1] = \sum_{\substack{x, y \in \{0,1\} \\ x + y = 1}} \mathbb{P}[X = x, Y = y] = \mathbb{P}[X = 0, Y = 1] + \mathbb{P}[X = 1, Y = 0] = \frac{1}{2}$$

$$\mathbb{P}[Z = 2] = \sum_{\substack{x, y \in \{0,1\} \\ x + y = 2}} \mathbb{P}[X = x, Y = y] = \mathbb{P}[X = 1, Y = 1] = \frac{1}{4}$$

Randverteilung Seien $X_1, \ldots, X_n$ n diskrete ZV mit gemeinsame Verteilung $p = \big(p(x_1, \ldots, x_n)\big)_{x_1 \in W_1, \ldots, x_n \in W_n}$. Für jedes i gilt

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

Erwartungswert der Abbildung Seien $X_1, \ldots, X_n$ n diskrete ZV mit gemeinsame Verteilung $p = \big(p(x_1, \ldots, x_n)\big)_{x_1 \in W_1, \ldots, x_n \in W_n}$. Sei $\phi : \mathbb{R}^n \to \mathbb{R}$, dann

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

wenn die Summe wohldefiniert ist.

Unabhängigkeit Seien $X_1, \ldots, X_n$ n diskrete ZV mit gemeinsame Verteilung $p = \big(p(x_1, \ldots, x_n)\big)_{x_1 \in W_1, \ldots, x_n \in W_n}$. Die folgenden Aussagen sind äquivalent:

- $X_1, \ldots, X_n$ sind unabhängig.
- $p(x_1, \ldots, x_n) = \mathbb{P}[X_1 = x_1] \cdot \ldots \cdot \mathbb{P}[X_n = x_n]$ für jedes $x_1 \in W_1, \ldots, x_n \in W_n$

Stetige gemeinsame Verteilung

Stetige gemeinsame Verteilung Sei $n \geq 1$ für ZVs $X_1, \dots, X_n : \Omega \to \mathbb{R}$ haben eine **gemeinsame, stetige Verteilung** falls eine Funktion f $f : \mathbb{R}^n \to \mathbb{R}_+$ existiert, sodass

$$\mathbb{P}[X_1 \leq a_1, \dots, X_n \leq a_n] = \int_{-\infty}^{a} \dots \int_{-\infty}^{b} f(x_1, \dots, x_n)\, \mathrm{d}x_1 \dots \mathrm{d}x_n$$

für jedes $a_1, \dots, a_n \in \mathbb{R}$. Eine solche Funktion f heisst **gemeinsame Dichte** von $(X_1, \dots, X_n)$.

Prop. Sei f eine gemeinsame Dichte von $(X_1, \dots, X_n)$. Dann gilt

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots x_n)\, \mathrm{d}x_n \dots \mathrm{d}x_1 = 1$$

Erwartungswert der Abbildung Sei $\phi : \mathbb{R}^2 \to \mathbb{R}$. Falls $X_1, \dots, X_n$ eine gemeinsame Dichte haben, dann kann der der Erwartungswert der ZV $Z = \phi(X_1, \dots, X_n)$ mit

$$\mathbb{E}[\phi(X_1, \dots, X_n)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \phi(x_1, \dots x_n) \cdot f(x_1, \dots, x_n)\, \mathrm{d}x_n \dots \mathrm{d}x_1$$

berechnet werden. (Integral muss Wohldefiniert sein.)

Prop. Seien $X_1, \dots, X_n$ n ZV mit einer gemeinsamen Dichte $f = f_{X_1, \dots, X_n}$. Dann für jedes i, ist X_i eine stetige ZV mit Dichte f_i, definiert durch

$$f_i(z) = \int_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \mathbb{R}^{n-1}} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)\, \mathrm{d}x_1 \dots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1}\, \mathrm{d}x_n$$

für jedes $z \in \mathbb{R}$

Falls X, Y eine gemeinsame Dichte haben, gilt

$$\mathbb{P}[X \leq a] = \mathbb{P}\Big[X \in [-\infty, a], Y \in [-\infty, \infty]\Big] = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y \right) \mathrm{d}x$$

Falls X, Y eine gemeinsame Dichte $f_{X,Y}$ haben, gilt für X resp. Y

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

Unabhängigkeit stetiger ZV Seien $X_1, \dots X_n$ n stetige ZV mit den jeweiligen Dichten $f_1, \dots, f_n$. Die folgenden Aussagen sind äquivalent:
1. $X_1, \dots, X_n$ sind unabhängig.
2. $X_1, \dots, X_n$ sind gemeinsam stetig mit gemeinsamer Dichte $f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$

Zwei unabhängige, stetige ZV sind automatisch gemeinsam stetig.

Asymptotische Resultate

In diesem Kapitel ist $(\Omega, \mathcal{F}, \mathbb{P})$ und eine unendliche Folge von i.i.d. ZV $X_1, X_2, \dots$ fixiert. Also gegeben ein ZV $X_i : \Omega \to \mathbb{R}$ sodass

$$\begin{matrix} \forall i_1 < \dots < i_k \\ \forall x_1, \dots, x_k \in \mathbb{R} \end{matrix} \quad \mathbb{P}[X_{i_1} \leq x_1, \dots, X_{x_k} \leq x_k] = F(x_1) \cdot \dots \cdot F(x_k)$$

Der **Empirische Durchschnitt** ist definiert durch

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$$

Gesetz der grossen Zahlen

Sei $\mathbb{E}[|X_1|] < \infty$. Falls wir $m = \mathbb{E}[X_1]$ definieren folgt

$$\lim_{n \to \infty} \frac{X_1 + \dots + X_n}{n} = m \quad \text{fast sicher}$$

Da die ZV i.i.d. macht es keinen Unterschied, dass das Theorem nur durch X_1 definiert ist.

- Falls $X_1, X_2, \dots$ i.i.d. Ber(p)

$$\lim_{n \to \infty} \frac{X_1 + \dots + X_n}{n} = p \quad \text{fast sicher}$$

- Falls $T_1, T_2, \dots$ i.i.d. Exp(λ)

$$\lim_{n \to \infty} \frac{T_1 + \dots + T_n}{n} = \lambda \quad \text{fast sicher}$$

Sei $X_1, X_2, \dots$ eine Folge von i.i.d. ZV mit $\mathbb{E}[|X_1|] < \infty$. Dann

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}[X_1]$$

fast sicher.

Anwendung: Monte-Carlo Integration

Das Gesetz der grossen Zahlen kann verwendet werden, um komplizierte Integrale zu approximieren.
Sei $g : [0, 1] \to \mathbb{R}$ sodass $\int_0^1 |g(x)|\, \mathrm{d}x < \infty$. Das Ziel ist es $I = \int_0^1 g(x)\, \mathrm{d}x$ zu berechnen.
Um I zu approximieren, interpretiere es als Erwartungswert. Sei U eine normalverteilte ZV in $[0, 1]$. Dann ist

$$\mathbb{E}[g(U)] = \int_0^1 g(x)\, \mathrm{d}x = I$$

Sei $X_1, X_2, \dots$ i.i.d. sodass $\forall n$ $X_n = g(U_n)$ wobei $U_1, U2, \dots$ eine Folge von i.i.d. normalverteilten ZV in $[0, 1]$ ist. Dann ist

$$\mathbb{E}[|X_1|] = \int_0^1 |g(x)|\, \mathrm{d}x < \infty$$

und $\mathbb{E}[X_1] = I$. Mit dem Gesetz der grossen Zahlen folgt

$$\lim_{n \to \infty} \frac{g(U_1) + \dots + g(U_n)}{n} = I$$

Konvergenz der Verteilung

Zwei ZV X und Y haben ähnliche wahrscheinlichkeitstechnische Eigenschaften wenn ihre Verteilungsfunktionen F_X und F_Y nahe beieinander sind.

Konvergenz der Verteilung Seien $(X_n)_{n \in \mathbb{N}}$ und X ZVs.

$$X_n \overset{\text{Approx}}{\approx} X \text{ wenn } n \to \infty$$

falls für jedes $x \in \mathbb{R}$

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

Grenzwertsätze

Wie weit ist $\frac{X_1 + \dots + X_n}{n}$ typischerweise von m entfernt?

Fluktuation von Normalverteilten ZV

Sei $Z = \frac{X_1 + \dots + X_n}{n} - m$ eine ZV. $Z \sim \mathcal{N}\left(\overline{m} = 0, \overline{\sigma}^2 = \frac{1}{n}\sigma^2\right)$. Die Standardabweichung $\overline{\sigma} = \frac{1}{\sqrt{n}}\sigma$ repräsentiert die typischen Abweichungen von Z.

Grob gesagt ist die Abweichung zwischen $\frac{X_1 + \dots + X_n}{n}$ und m von Ordnung $\frac{\sigma}{\sqrt{n}}$.
Um Abweichungen der Ordnung 1 zu bekommen, skalieren wir Z um einen Faktor $\frac{\sqrt{n}}{\sigma}$ (immer noch normalverteilt) $S_n = \sum_{i=1}^{n} X_i$

$$\frac{\sqrt{n}}{\sigma} Z = \frac{X_1 + \dots + X_n - n \cdot m}{\sqrt{\sigma^2 n}} = \frac{S_n - n \cdot m}{\sqrt{\sigma^2 n}}$$

Zentraler Grenzwertsatz Sei $\mathbb{E}[X_1^2]$ wohldefiniert und endlich. Für $m = \mathbb{E}[X_1]$ und $\sigma^2 = \mathrm{Var}(X_1)$ folgt

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

für jedes $a \in \mathbb{R}$ $\Phi(-a) = 1 - \Phi(a)$.

Für grosse n sieht die Verteilung von

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

wie $\mathcal{N}(0, 1)$ aus. Mit $Z_n \approx Z$ für $n \to \infty$ ist $Z \sim \mathcal{N}(0, 1)$.

Statistische Grundideen

- **Stichproben**: Gesamtheit der Beobachtungen oder ZV.
- Stichprobenumfang: Anzahl Stichproben.
- **Parameter θ** beschreibt Modell für Stichprobe.
- Man betrachtet eine **Familie von Wahrsch.räumen** mit festem Grundraum $(\Omega, \mathcal{F})$ und für jedes θ aus dem **Parameterraum** Θ ein Wahrscheinlichkeitsmass $\mathbb{P}_\theta$ um $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ zu bilden

Schätzer

Setup: • Parameterraum $\Theta \subset \mathbb{R}$ • Grundraum Ω • Sigma-Algebra $\mathcal{F}$ • $(\mathbb{P}_\theta)_{\theta \in \Theta}$ Familie von Wahrscheinlichkeitsmasse auf $(\Omega, \mathcal{F})$ • $X_1, \dots, X_n$ ZV auf $(\Omega, \mathcal{F})$

Grundbegriffe

Schätzer Zufallsvariable $T : \Omega \to \mathbb{R}$ der Form

$$T = t(X_1, \dots, X_n)$$

wobei $t : \mathbb{R}^n \to \mathbb{R}$.

Bias

Erwartungstreu Ein Schätzer ist **erwartungstreu** für θ, falls für alle $\theta \in \Theta$ gilt: $\mathbb{E}_\theta[T] = \theta$.

T schätzt also richtig, unabhängig davon welches Modell $\mathbb{P}_\theta$ zugrunde liegt.

Bias 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 („mean squared error“, **MSE**) von T im Modell $\mathbb{P}_\theta$ ist definiert als

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

Maximum-Likelihood (ML-Methode)

Likelihood-Funktion

$$L(x_1,\dots,x_n;\theta) := \begin{cases} p_{\vec{X}}(x_1,\dots,x_n;\theta) & \text{im diskreten Fall} \\ f_{\vec{X}}(x_1,\dots,x_n;\theta) & \text{im stetigen Fall} \end{cases}$$

wobei $p_{\vec{X}}(x_1,\dots,x_n;\theta) = \mathbb{P}_\theta[X_1 = x_1,\dots,X_n = x_n]$
Die Funktion $\log L(x_1,\dots,x_n;\theta)$ heisst **log-Likelihood-Funktion**.

Falls die X_i unter $\mathbb{P}_\theta$ i.i.d. sind gilt:

$$p_{\vec{X}}(x_1,\dots,x_n;\theta) = \prod_{i=1}^n p_X(x_i;\theta) \qquad \sum_{i=1}^n \log\Big(p_X(x_i;\theta)\Big)$$

$$f_{\vec{X}}(x_1,\dots,x_n;\theta) = \prod_{i=1}^n f_X(x_i;\theta) \qquad \sum_{i=1}^n \log\Big(f_X(x_i;\theta)\Big)$$

ML-Schätzer Für jedes $x_1,\dots,x_n$ sei $t_{\mathrm{ML}}(x_1,\dots,x_n) \in \mathbb{R}$ der Wert, der $\theta \mapsto L(x_1,\dots,x_n;\theta)$ als Funktion maximiert. D.h.

$$L(x_1,\dots,x_n;t_{\mathrm{ML}}(x_1,\dots,x_n)) = \max_{\theta\in\Theta} L(x_1,\dots,x_n)$$

Ein **Maximum-Likelihood-Schätzer (ML-Schätzer)** T_{ML} für θ wird definiert durch

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

Meistens sind $X_1,\dots,X_n$ i.i.d. und dann ist es einfacher $\log L$ zu maximieren, da es sich dann um eine Summe handelt und man dann nur die Nullstellen der Ableitung nach θ suchen muss.

Rezept ML-Schätzer
1. Bestimme log-Likelihood-Funktion.
2. Setze Ableitung $\frac{\mathrm{d}}{\mathrm{d}\theta}$ gleich 0.
3. Löse nach θ auf.
4. T_{ML} ist (fast) gleich θ

Sei $\Theta = [0,1]$ und $X_1,\dots,X_n$ unter $\mathbb{P}_\theta$ i.i.d. mit $X_1 \sim \mathrm{Geom}(\theta)$. Bestimme T_{ML}.

Die log-Likelihood-Funktion ist

$$n\cdot\log(\theta) + (x_1,\dots,x_n - n)\cdot\log(1-\theta)$$

Die Ableitung gleich 0 gesetzt ist

$$\frac{n}{\theta} - \frac{x_1,\dots,x_n - n}{1-\theta} = 0$$

Löse nach θ auf

$$\theta = \frac{n}{x_1,\dots,x_n}$$

Konstruiere T_{ML}

$$T_{\mathrm{ML}} = \frac{n}{X_1,\dots,X_n}$$

Verschiedene T_{ML}

$\mathrm{Ber}(\theta)$	$\frac{1}{n}\sum_{i=1}^n X_i$
$\mathrm{Bin}(k,\theta)$	$\frac{1}{k}\overline{X}_n = \frac{1}{nk}\sum_{i=1}^n X_i$
$\mathrm{Poisson}(\theta)$	$\frac{1}{n}\sum_{i=1}^n X_i$
$\mathrm{Geom}(\theta)$	$n/\sum_{i=1}^n X_i$
$\mathcal{U}([a,b])$	$\max\{X_1,\dots,X_n\}$
$\mathrm{Exp}(\theta)$	$\frac{1}{\overline{X}_n} = n/\sum_{i=1}^n X_i$
$\mathcal{N}(\mu,\sigma^2)$	$\hat{\mu} = \overline{X}_n \quad \hat{\sigma}^2 = S^2$

Modelle mit mehreren Parametern

Empirisches Moment Für $k \in \{1,\dots,m\}$ sei das k-te Moment empirische Moment oder Stichprobenmoment $\hat{m}_k$ der Realisierung $(x_1,\dots,x_n)$:

$$\hat{m}_k(x_1,\dots,x_n) := \frac{1}{n}\sum_{i=1}^n x_i^k$$

Normalverteile Stichprobe

Sei $X_1,\dots X_n$ i.i.d. $\mathcal{N}\left(\mu,\sigma^2\right)$-verteilt mit unbekannten Parameter $\theta = \left(\mu,\sigma^2\right)$. Damit berechnen wir mit den Ableitungen der $\log L$

$$T_1 = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n$$

$$T_2 = \frac{1}{n}\sum_{i=1}^n (X_1 - \overline{X}_n)^2$$

möchten wir aber noch, dass der Schätzer erwartungstreu wird, so wählen wir für $T_2 = S^2$:

Empirische Stichprobenvarianz

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

Der zentrale Grenzwertsatz liefert einen allgemeinen approximative Zugang.

1.6 Seien $X_1,\dots X_n$ i.i.d. $\sim\mathcal{N}\left(\mu,\sigma^2\right)$. Dann gilt:
1. $\overline{X}_n$ ist normalverteilt gemäss $\mathcal{N}\left(\mu,\frac{1}{n}\sigma^2\right)$, und dann gilt $\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
2. $\frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$ ist χ^2-verteilt mit $n-1$ Freiheitsgraden.
3. $\overline{X}_n$ und S^2 sind unabhängig.
4. Der Quotient

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

ist t-verteilt mit $n-1$ Freiheitsgraden.

Konfidenzintervalle

Setup: • Parameterraum $\Theta \subset \mathbb{R}$ • Grundraum Ω • Sigma-Algebra $\mathcal{F}$ • $(\mathbb{P}_\theta)_{\theta\in\Theta}$ Familie von Wahrscheinlichkeitsmasse auf $(\Omega,\mathcal{F})$ • $X_1,\dots,X_n$ ZV auf $(\Omega,\mathcal{F})$

Definition

Um zu beschreiben wie weit T_{ML} vom wahren Wert p ist, führen wir Konfidenzintervalle ein.

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

$$\forall\theta\in\Theta \qquad \mathbb{P}_\theta[A \le \theta \le B] \ge 1-\alpha$$

wobei A,B ZV der Form $A = a(X_1,\dots,X_n)$, $B = b(X_1,\dots,X_n)$ mittels $a,b:\mathbb{R}^n\to\mathbb{R}$ sind. Für Niveau 95%: $c_{1-\alpha/2} = c_{0.975} = 1.96$

θ ist deterministisch und nicht zufällig. Die stochastischen Elemente sind gerade die Schranken $A = a(X_1,\dots,X_n)$ und $B = b(X_1,\dots,X_n)$.

Normal mit σ und m unbekannt

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

Seien $X_1,\dots,X_n$ i.i.d. $\sim\mathcal{N}\left(\mu,\sigma^2\right)$. Dann sind $\overline{X}_n$ und S^2 unabhängig.

Konstruktion von Konfidenzintervallen

Fall 1: σ^2 bekannt. Seien $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu,\sigma^2\right)$, $\theta = \mu \in \mathbb{R}$.

$$\Longrightarrow I = \left[\overline{X}_n - c_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}, \overline{X}_n + c_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right]$$

für $c_{1-\frac{\alpha}{2}}$ sodass $\Phi(c_{1-\frac{\alpha}{2}}) \ge 1-\frac{\alpha}{2}$
I ist ein $(1-\alpha)$-Konfidenzintervall für μ.

Fall 2: Seien $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu,\sigma^2\right)$, $\theta = \left(\mu,\sigma^2\right)$.

$$\Longrightarrow I = \left[\overline{X}_n - c\sqrt{\frac{S^2}{n}}, \overline{X}_n + c\sqrt{\frac{S^2}{n}}\right]$$

c wird so gewählt, dass $F_{n-1}(c) \ge 1-\frac{\alpha}{2}$
Wobei F_{n-1} Verteilungsfunktion von t_{n-1} ist.

Fall 3: Seien $X_1,\dots,X_n$ i.i.d. mit $(\mathbb{P}_\theta)_{\theta\in\Theta}$.
Aus dem zentralen GWS. folgt

$$\underbrace{\frac{X_1+\ldots+X_n - n\mu}{\sqrt{\sigma^2 n}}}_{=Z} \overset{\text{approx.}}{\approx} \mathcal{N}(0,1)$$

$$\Rightarrow \mathbb{P}[-c \le Z \le c] \approx 2\Phi(c) - 1 \ge 1-\alpha \Leftrightarrow \Phi(c) \ge 1-\frac{\alpha}{2}$$

$$\Rightarrow \mathbb{P}\left[-c \le \frac{X_1+\ldots+X_n - \mathbb{E}[X_1]\cdot n}{\sqrt{\mathrm{Var}(X_1)\cdot n}} \le c\right]$$

$\mathbb{P}[-c \le Z \le c]$ so umformen, dass sich $\mathbb{P}[A \le \theta \le B]$ ergibt, dann ist K.I. $[A,B]$.

Konfidenzintervall eines normalen Modells mit Varianz 1 und unbk. Mittelwert

Seien $X_1,\dots,X_n$ i.i.d. und $\sim\mathcal{N}(m,1)$. Betrachte also ein stochastisches Modell mit bekannter Varianz $\left(\sigma^2 = 1\right)$ aber unbekanntem Mittelwert m.
Der Maximum-Likelihood-Schätzer ist gegeben durch

$$T = T_{\mathrm{ML}} = \frac{X_1+\ldots+X_n}{n}$$

Wir suchen nun für m Konfidenzintervalle der Form

$$I = \left[T - \frac{c}{\sqrt{n}}, T + \frac{c}{\sqrt{n}}\right]$$

wobei $c > 0$ eine von n unabhängige Konstante ist. Betrachte zuerst

$$\mathbb{P}_\theta\left[T - \frac{c}{\sqrt{n}} \leq m \leq T + \frac{c}{\sqrt{n}}\right] = \mathbb{P}_\theta[-c \leq Z \leq c]$$

wobei $Z = \sqrt{n}(t - m) = \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}}$ eine standardnormalverteilte ZV ist. Somit ist

$$\mathbb{P}_\theta[-c \leq Z \leq c] = \mathbb{P}_\theta[Z \leq c] - \mathbb{P}_\theta[X < -c] = 2\Phi(c) - 1$$

Mittels Tabelle: $2\Phi(1.96) - 1 \geq 0.95$. Somit folgt für $c = 1.96$

$$\mathbb{P}_\theta\left[T - \frac{1.96}{\sqrt{n}} \leq m \leq T + \frac{1.96}{\sqrt{n}}\right] \geq \frac{95}{100}$$

$$I = \left[T - \frac{1.96}{\sqrt{n}}, T + \frac{1.96}{\sqrt{n}}\right]$$

Was nach Definition ein Konfidenzintervall für m mit Niveau 95% ist.

Verteilungsaussagen

χ^2-Verteilung Die χ^2-Verteilung mit m Freiheitsgraden gehört zu einer stetigen ZV Y mit Dichtefunktion

$$f_Y(y) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} e^{-\frac{1}{2}y} \text{ für } y \geq 0$$

Die Gamma-Funktion ist

$$\Gamma(v) := \begin{cases} (n-1)! & \text{falls } v = n \in \mathbb{N} \\ \int_0^\infty t^{v-1} e^{-t}\, dt & \text{sonst} \end{cases}$$

Die χ^2 Verteilung mit m Freiheitsgraden ist ein Spezialfall einer $\mathcal{N}(\alpha, \lambda)$-Verteilung mit $\alpha = \frac{m}{2}$ und $\lambda = \frac{1}{2}$.
Für $m = 2$ ergibt das eine Exponetialverteilung mit Parameter $\frac{1}{2}$.

Sind die ZV $X_1, \ldots, X_m$ i.i.d. $\sim \mathcal{N}(0,1)$, so ist die Summe $Y := \sum_{i=1}^m X_i^2 \sim \chi^2_m$. $Z := \frac{1}{m}\sum_{i=1}^m X_i^2 \sim \chi^2_1$.

t-Verteilung Die t-Verteilung mit m Freiheitsgraden gehört zu einer stetigen ZV z mit Dichtefunktion

$$f_Z(z) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)}\left(1 + \frac{z^2}{m}\right)^{-\frac{m+1}{2}} \quad \text{für } Z \in \mathbb{R}$$

- $m = 1$: Cauchy-Verteilung
- $m \to \infty$: $\mathcal{N}(0,1)$

Die t-Verteilung ist symmetrisch um 0 und geht langsamer gegen 0 als die Normalverteilung.

Sind $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi^2_m$ unabhängig ist der Quotient $Z := \frac{X}{\sqrt{(1/m)Y}}$ t-Verteilt mit m Freiheitsgraden.

Sei $X_0, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(0,1)$, dann

$$t_n \sim \frac{X_0}{\sqrt{\frac{X_1^2 + \ldots + X_n^2}{n}}}$$

$$t_{n-1} \sim \frac{X_0 + X_1}{\sqrt{\frac{2X_2^2, \ldots, 2X_n^2}{n-1}}}$$

Letzteres ergibt sich aus: $Y = \frac{X_0 + X_1}{\sqrt{2}}$ und $Z = X_2^2, \ldots, X_n^2$. Nun ist $Y \sim \mathcal{N}(0,1)$, $Z \sim \chi^2_{n-1}$ und Y, Z unabhängig. Daraus folgt

$$t_{n-1} \sim \frac{Y}{\sqrt{Z/n-1}}$$

Approximative Konfidenzintervalle

Gegeben ein $\mathcal{N}(0,1)$ verteilter Schätzer, berechne approximatives K.I. wie folgt:

1. Berechne Konfidenzintervall wie normal.
2. Forme um, bis man Schätzer einsetzen kann.

Tests

Null- und Alternativhypothese

Es existiert bereits eine Vermutung, wo in Θ der richtige, unbekannte Parameter θ liegt. Grundproblem: Entscheidung zwischen zwei konkurrierenden Modellklassen zu treffen.

$$\begin{array}{lll} \text{Nullhypothese} & H_0 : \theta \in \Theta_0 & \Theta_0 \subseteq \Theta \\ \text{Alternativhypothese} & H_A : \theta \in \Theta_A & \Theta_A \subseteq \Theta \end{array}$$

$$\text{wobei } \Theta_0 \cap \Theta_A = \emptyset.$$

Falls keine explizite Alternative gegeben ist: $\Theta_A = \Theta_0^c = \Theta \setminus \Theta_0$.
Hypothesen heissen **einfach**, wenn sie aus einem einzelnem Wert, θ_0 bzw. θ_A, bestehen; sonst **zusammengesetzt**.

Test und Entscheidung

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

- T eine ZV der Form $T = t(X_1, \ldots, X_n)$ ist, und
- $K \subseteq \mathbb{R}$ eine (deterministische) Teilmenge von $\mathbb{R}$ ist.

Die ZV $T = t(X_1, \ldots, X_n)$ heisst dann **Teststatistik**, und K heisst **kritischer Bereich** oder **Verwerfungbereich**.

Entscheidungsregel

- Die Hypothese H_0 wird verworfen, falls[a] $T(\omega) \in K$.
- Die Hypothese H_0 wird nicht verworfen, falls $T(\omega) \notin K$.

[a] Wobei $T(\omega)$ die Teststatistik $T(\omega) = t(X_1(\omega), \ldots, X_n(\omega))$ ist.

Die Entscheidung des Test hängt via $T(\omega)$ von ω ab. Weil T eine ZV ist, ist die Menge $\{T \in K\}$ ein Ereignis, und wir können $\mathbb{P}_\theta[T \in K]$ in jedem Modell $\mathbb{P}_\theta$ betrachten. Die Entscheidung bei einem Test kann falsch herauskommen:

1. **Fehler 1. Art**: Nullhypothese wird zu Unrecht verworfen (d.h. obwohl sie richtig ist). Passiert für $\theta \in \Theta_0$ und $T \in K$. Deshalb heisst $\mathbb{P}_\theta[T \in K]$ für $\theta \in \Theta_0$ die Wahrsch. für einen Fehler 1. Art.
2. **Fehler 2. Art**: Nullhypothese wird zu Unrecht nicht verworfen (d.h. man akzeptiert sie, obwohl sie falsch ist). Passiert für $\theta \in \Theta_A$ und $T \notin K$. Deshalb heisst $\mathbb{P}_\theta[T \notin K] = 1 - \mathbb{P}_\theta[T \in K]$ für $\theta \in \Theta_A$ die Wahrsch. für einen Fehler 2. Art.

Signifikanzniveau und Macht

Signifikanzniveau Sei $\alpha \in]0,1[$. Ein Test (T, R) besitzt **Signifikanzniveau** α, falls

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

Macht Die **Macht** eines Tests (T, K) wird definiert als folgende Funktion

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

- Um einen Fehler 1. Art zu minimieren, fixieren wir ein Parameter α, und designen einen Test zum Signifikanzniveau α.
- Um einen Fehler 2. Art zu vermeiden, suche den Test mit grösster Macht (nachdem Signifikanzniveau α gefunden wurde). D.h. minimiere die Grösse $1 - \beta(\theta) = \mathbb{P}_\theta[T \notin K]$ für $\theta \in \Theta_A$.

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

Die Entscheidung ist eine **Interpretation** der Übereinstimmung zwischen Daten und Modell, und nie ein Beweis.

Bemerkung

Wenn die Teststatistik T diskret ist, kann ein vorgegebene Niveau α in der Regel nicht genau eingehalten werden, d.h. es ist unmöglich, einen kritischen Bereich K mit $\mathbb{P}_{\theta_0}[T \in K] = \alpha$ zu finden. Ein Ausweg bietet ein sogenannter **randomisierter Test**. Man wählt $\gamma \in [0,1]$ so, dass $\gamma\mathbb{P}_{\theta_0}[T > c] + (1-\gamma)\mathbb{P}_{\theta_0}[T > c+1] = \alpha$ gilt und entscheidet dann wie folgt: Ist $T > c$, so verwirft man H_0 mit Wahrsch. γ, d.h. H_0 wird abgelehnt, falls erstens $T > c$ gilt und zweitens eine unabhängige $\mathcal{U}(0,1)$-verteilte ZV einen Wert $\leq \gamma$ realisiert.

Konstruktion von Tests

Annahmen: • $\theta_0 \neq \theta_A$ zwei fixierte Zahlen. • Null- und Alternativhypothesen sind von einfacher Form. • ZV $X_1, \ldots, X_n$ sind entweder diskret oder gemeinsam stetig unter $\mathbb{P}_{\theta_0}$ und $\mathbb{P}_{\theta_A}$ (somit ist Likelihood-Funktion für $\theta = \theta_0$ und $\theta = \theta_A$ wohldefiniert).

Likelihood-Quotient Für jedes $x_1, \ldots, x_n$, definieren wir den Likelihood-Quotienten durch

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

Als Konvention setzen wie $R(x_1, \ldots, x_n) = +\infty$, falls $L(x_1, \ldots, x_n; \theta_0) = 0$.

Likelihood-Quotienten-Test Sei $c \geq 0$. Der Likelihood-Quotienten-Test mit Parameter c ist ein Test (T, K), wobei Teststatistik und Verwerfungbereich gegeben sind durch

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

Neymann-Pearson-Lemma Sei $c \geq 0$. Sei (T, K) ein Likelihood-Quotienten-Test mit Parameter c und Signifikanzniveau $\alpha^* := \mathbb{P}_{\theta_0}[T > c]$. Ist (T', K') ein anderer Test mit Signifikanzniveau $\alpha \leq \alpha^*$, so gilt

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

Verallgemeinerter Likelihood-Quotient

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

oder auch

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

und wähle als Teststatistik $T_0 := R(X_1, \ldots, X_n)$ bzw. $\tilde{T} := \tilde{R}(X_1, \ldots, x_n; \theta)$ mit kritischem Bereich $K_0 :=]c_0, \infty[$.

Der p-Wert

Sei $X_1, \ldots, X_n$ ein Stichprobe vom Umfang n. Wir wollen $H_0 : \theta = \theta_0$ gegen $H_A : \theta \in \Theta_A$ testen.

Geordnete Testsammlung Sei T eine Teststatistik. Eine Familie von Tests $(T, (K_t)_{t \geq 0})$ heiss **geordnet bzgl. T** falls $K_i \subset T$ und

$$s \leq t \Longrightarrow K_s \supset K_t$$

gilt.

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

$$\text{p-Wert} = G(T) = P_{\mu_0}[|T| \geq a] \qquad a := \text{beobachteter Wert}$$

wobei $G : \mathbb{R}_+ \to [0,1]$ mittels $G(t) = \mathbb{P}_{\theta_0}[T \in K_t]$ definiert ist.

- Der p-Wert hängt direkt von den anfänglichen Beobachtungen $X_1, \dots, X_n$ ab.
- p-Wert liegt stets in $[0,1]$. Falls T stetig und $K_t =]t, \infty[$ dann ist der p-Wert unter $\mathbb{P}_{\theta_0}$ auf $[0,1]$ gleichverteilt.
- p-Wert liefert Information, welche Tests in $(T, K_t)_{t \geq 0}$ die Nullhypothese H_0 ablehnen.
  Für einen p-Wert mit Wert p gilt. dass alle Tests mit Signifikanzniveau $\alpha > p$ H_0 verwerfen und alle Tests mit $\alpha \leq p$ H_0 nicht verwerfen.
- Der p-Wert ist nur von H_0 abhängig.
- p-Wert klein $\implies H_0$ wird wahrscheinlich verworfen.