

1 **Konzepte**

Ereignisraum

Die Menge $\Omega \neq \emptyset$ aller möglichen Ergebnisse des betrachteten Zufallsexperiments. Die Elemente $\omega \in \Omega$ heissen Elementarereignisse.

Potenzmenge

Die Potenzmenge von Ω , bezeichnet mit $\mathcal{P}(\Omega)$ oder 2^Ω ist die Menge aller Teilmengen von Ω . Ein Prinzipielles Ereignis ist eine Teilmenge $A \subseteq \Omega$, also eine Kollektion von Elementarereignissen. Die Klasse aller beobachtbaren Ereignisse ist \mathcal{F} .

σ -Algebra

Ein Mengensystem ist eine σ -Algebra falls

- (1) $\Omega \in \mathcal{F}$
- (2) Für jedes $A \in \mathcal{F}$ ist auch $A^c \in \mathcal{F}$
- (3) Für jede Folge $A_n, n \in \mathbb{N}$ mit $A_n \in \mathcal{F}$ für alle $n \in \mathbb{N}$ auch die Vereinigung $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$

Wahrscheinlichkeitsmass

Eine Abbildung $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$ mit folgenden Eigenschaften:

- (1) $\mathcal{P}[A] \geq 0$ für alle Ereignisse $A \in \mathcal{F}$
- (2) $\mathcal{P}[\Omega] = 1$
- (3) Für $A_i \in \mathcal{F}$ paarweise disjunkt gilt $\mathcal{P}[\bigcup_{i=1}^\infty A_i] = \sum_{i=1}^\infty \mathcal{P}[A_i]$

Es gelten weiter folgende Rechenregeln:

- $\mathcal{P}[A^c] = 1 - \mathcal{P}[A]$
- $\mathcal{P}[\emptyset] = 0$
- Für $A \subseteq B$ gilt $\mathcal{P}[A] \leq \mathcal{P}[B]$
- $\mathcal{P}[A \cup B] = \mathcal{P}[A] + \mathcal{P}[B] - \mathcal{P}[A \cap B]$

Diskrete Wahrscheinlichkeitsräume

Impliziert:

- Ω ist endlich oder abzählbar unendlich
- $\mathcal{F} = 2^\Omega$

Laplace Raum

Ist $\Omega = \{\omega_1, \dots, \omega_N\}$ endlich mit $|\Omega| = N$ und $\mathcal{F} = 2^\Omega$ sowie alle ω_i gleich wahrscheinlich mit $p_i = \frac{1}{n}$, so heisst Ω ein Laplace Raum und P die diskrete Gleichverteilung auf Ω . Dann ist für $A \subseteq \Omega$:

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

Bedingte Wahrscheinlichkeit

Seien A, B Ereignisse mit $P[A] > 0$. Die bedingte Wahrscheinlichkeit von B unter der Bedingung, dass A eintritt wird definiert durch:

$$\begin{aligned} P[B \mid A] &:= \frac{P[B \cap A]}{P[A]} \\ &= \frac{P[A \mid B] \cdot P[B]}{P[A]} \end{aligned}$$

Multiplikationsregel

Es gilt:

$$P[A \cap B] = P[A \mid B] \cdot P[B] = P[B \mid A] \cdot P[A]$$

Satz der totalen Wahrscheinlichkeit

Sei A_1, \dots, A_n eine Zerlegung von Ω in paarweise disjunkte Ereignisse, d.h. $\bigcup_{i=1}^n A_i = \Omega$ und $A_i \cap A_k = \emptyset \quad \forall i \neq k$. Dann gilt:

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

Beweis. Da $B \subseteq \Omega \implies B = B \cap \Omega = B \cap (\bigcup_{i=1}^n A_i) = \bigcup_{i=1}^n (B \cap A_i)$. Weiter sind alle Mengen der Art $(B \cap A_i)$ paarweise disjunkt, was bedeutet, dass $(B \cap A_i)$ eine disjunkte Zerlegung von B bilden. Damit folgt:

$$\begin{aligned} P[B] &= P\left[\bigcup_{i=1}^n (B \cap A_i)\right] \\ &= \sum_{i=1}^n P[B \cap A_i] = \sum_{i=1}^n P[B \mid A_i] \cdot P[A_i] \end{aligned}$$

Satz von Bayes

Sei A_1, \dots, A_n eine Zerlegung von Ω mit $P[A_i] > 0$ für $i \in \{1, \dots, n\}$. Sei B ein Ereignis mit $P[B > 0]$. Dann gilt für jedes k :

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

Beweis. Verwende Definition Bedingte Wahrscheinlichkeit, im Zähler Multiplikationsregel und im Nenner Satz der totalen Wahrscheinlichkeit.

Unabhängige Ereignisse (2)

Zwei ereignisse heissen (stochastisch) Unabhängig, falls

$$P[A \cap B] = P[A] \cdot P[B]$$

Ist $P[A] = 0$ oder $P[B] = 0$, so sind A, B immer unabhängig. Für $P[A] \neq 0$ gilt:

$$A, B \text{ unabhängig} \iff P[A \mid B] = P[A]$$

Unabhängige Ereignisse (∞)

Die Ereignisse A_1, \dots, A_n heissen (stochastisch) unabhängig, wenn für jede endliche Teilfamilie der Produktformel gilt, d.h. für $m \in \mathbb{N}$ und $\{k_1, \dots, k_m\} \subseteq \{1, \dots, n\}$:

$$P\left[\bigcap_{i=1}^n *A_{k_i}\right] = \prod_{i=1}^n P[A_{k_i}]$$

Diskrete Zufallsvariable

Eine reelwertige diskrete Zufallsvariable auf Ω ist eine Funktion $X : \Omega \mapsto \mathbb{R}$. Mit Ω ist natürlich auch $\mathcal{W}(X) = \{x_1, x_2, \dots\}$ endlich oder abzählbar.

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

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

- Die Gewichtsfunktion oder diskrete Dichte von X ist die Funktion $p_X : \mathcal{W}(X) \mapsto [0, 1]$, definiert durch:

$$p_X(x_k) := P[X = x_k] = P[\{\omega : X(\omega) = x_k\}]$$

Wobei gilt:

- $F_X(t) = P[X \leq t] = \sum_k \text{mit } x_k \leq t p_X(x_k)$
- Für jedes $x_k \in \mathcal{W}(X)$ gilt $0 \leq p_X(x_k) \leq 1$ und $\sum_{x_k \in \mathcal{W}(X)} p_X(x_k) = 1$
- $\mu_X(B) := P[X \in B] = \sum_{x_k \in B} p_X(x_k)$
- $\sum_{x_k \in \mathcal{W}(X)} p_X(x_k) = P[X \in \mathcal{W}(X)] = 1$

Indikatorfunktion

Für jede Teilmenge $A \subseteq \Omega$ ist die Indikatorfunktion I_A von A definiert durch:

$$I_A(\omega) := \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \in A^c \end{cases}$$

Erwartungswert

Sei X eine diskrete Zufallsvariable mit Gewichtsfunktion $p_X(x)$, dann ist der Erwartungswert definiert durch:

$$E[X] := \sum_{x_k \in \mathcal{W}(X)} x_k \cdot p_X(x_k)$$

und hat folgende Eigenschaften:

- Linearität: $E[a \cdot X + b] = a \cdot E[X] + b$
- Monotonie: $X \leq Y \implies E[X] \leq E[Y]$
- Nimmt X nur Werte in \mathbb{N} an:

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

Erwartungswert von Funktionen

Sei X eine Diskrete Zufallsvariable mit Gewichtsfunktion $p_X(x)$ und $Y = g(X)$ für eine Funktion $Y : \mathbb{R} \mapsto \mathbb{R}$. Dann gilt:

E[Y] = E[g(X)] = \sum_{x_k \in \mathcal{W}(X)} g(x_k) \cdot p_X(x_k)

Varianz

Sei X eine diskrete Zufallsvariable. Ist $E[X^2] < \infty$, so heisst:

Var[X] := E[X - E[X]^2] = \sum_{x_k \in \mathcal{W}(X)} x_k - E[X]^2 \cdot p_X(x_k)

die Varianz von X . Es gilt weiter:

- $Var[X] = E[X^2] - E[X]^2$
- $Var[a \cdot X + b] = a^2 \cdot Var[X]$
- $Var[X - Y] = Var[X] + -1^2 \cdot Var[Y]$

Standardabweichung

\sigma(X) = \sqrt{Var[X]}

Gemeinsame Verteilung

Seien X_1, \dots, X_n beliebige Zufallsvariablen. Die Gemeinsame Verteilungsfunktion von X_1, \dots, X_n ist die Abbildung $F : \mathbb{R}^n \mapsto [0, 1]$, definiert durch:

(x_1, \dots, x_n) \mapsto F(x_1, \dots, x_n) := P[X_1 \le x_1, \dots, X_n \le x_n] = \sum_{y_1 \le x_1, \dots, y_n \le x_n} p(y_1, \dots, y_n)

Die Gemeinsame Gewichtsfunktion ist:

p(x_1, \dots, x_n) := P[X_1 = x_1, \dots, X_n = x_n]

Randverteilung

Haben X, Y die Gemeinsame Verteilungsfunktion F , so ist die Funktion $F_X : \mathbb{R} \mapsto [0, 1]$,

x \mapsto F_X(x) := P[X \le x, Y < \infty] = \lim_{y \rightarrow \infty} F(x, y)

Sind X, Y diskrete Zufallsvariablen mit $\mathcal{W}(Y) = \{y_1, y_2, \dots\}$ und gemeinsamer Gewichtsfunktion $p(x, y)$, so ist die Gewichtsfunktion der Randverteilung von X gegeben durch:

x \mapsto p_X(x) := \sum_{y_i \in \mathcal{W}(X)} P[X = x, Y = y_i]

Unabhängige Zufallsvariablen

Zufallsvariablen X_1, \dots, X_n heissen Unabhängig, falls gilt (äquivalent):

F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n) p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)

Unabhängige Ereignisse

Ereignisse A_1, \dots, A_n heissen Unabhängig, falls für beliebige Teilmengen $B_i \subseteq \mathcal{W}(X_i) \quad i = 1, \dots, n$ gilt (äquivalent):

P[X_1 \in B_1, \dots, X_n \in B_n] = \prod_{i=1}^n P[X_i \in B_i]

Funktionen von Zufallsvariablen

Seien X_1, \dots, X_n diskrete Unabhängige Zufallsvariablen und $f_i : \mathbb{R} \mapsto \mathbb{R}$ irgendwelche Funktionen. Sei weiter $Y_i := f_i(X_i)$. Dann sind die Zufallsvariablen Y_1, \dots, Y_n ebenfalls unabhängig.

Linearität des Erwartungswertes

Seien X_1, \dots, X_n diskrete Zufallsvariablen mit endlichen Erwartungswerten. $E[X_1], \dots, E[X_n]$. Sei $Y = a + \sum_{i=1}^n b_i \cdot X_i$ mit Konstanten a, b_1, \dots, b_n . Dann gilt:

E[Y] = a + \sum_{i=1}^n b_i \cdot E[X_i]

Kovarianz

Seien X, Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit endlichen Erwartungswerten. Dann ist die Kovarianz definiert als:

Cov(X, Y) := E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])]

Wobei $Cov(X, X) = Var[X]$.

Korrelation

Die Korrelation von X, Y ist definiert durch

\rho(X, Y) := \begin{cases} \frac{Cov(X, Y)}{\sigma(X) \cdot \sigma(Y)} & \text{falls } \sigma(X) \cdot \sigma(Y) > 0 \\ 0 & \text{sonst.} \end{cases}

und es gilt $|Cov(X, Y)| \leq \sigma(X) \cdot \sigma(Y)$ beziehungsweise $-1 \leq \rho(X, Y) \leq 1$.

Summenformel für Varianzen

Var\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n Var[X_i] + 2 \cdot \sum_{i < j} Cov(X_i, X_j)

ist aber $Cov(X, Y) = 0$ (X, Y paarweise unkorreliert), so wird die Summe linear.

Produkte von Zufallsvariablen

Seien X_1, \dots, X_n diskrete Zufallsvariablen mit endlichen Erwartungswerten. Falls X_1, \dots, X_n unabhängig sind, so ist:

E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n E[X_i]

Dann sind auch X_1, \dots, X_n paarweise unkorreliert und:

Var\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n Var[X_i]

da Unabhängig \implies paarweise Unabhängig \implies unkorreliert.

Bedingte Verteilung

Seien X, Y diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion $p(x, y)$. Die bedingte Gewichtsfunktion von X , gegeben dass $Y = y$, ist definiert durch:

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

für $p_Y(y) > 0$ und 0 sonst.

Kriterium für Unabhängigkeit

X, Y sind genau dann unabhängig, wenn für alle y mit $p_Y(y) > 0$ gilt:

p_{X|Y}(x|y) = p_X(x) \quad \forall x \in \mathcal{W}(X)

n tief k

\binom{n}{k} = \frac{n!}{k! \cdot (n - k)!}

Ableitung, Integration

Es gilt:

- **Summenregel** $(f(x) + g(x))' = f'(x) + g'(x)$
- **Produktregel** $(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$
- **Quotientenregel** $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}$ wenn $g(x) \neq 0$
- **Kettenregel** $(f(g(x)))' = f'(g(x)) \cdot g'(x)$
- **Partielle Integration** $\int_a^b f'(x) \cdot g(x) \, dx = [f(x) \cdot g(x)]_a^b - \int_a^b f(x) \cdot g'(x) \, dx$
- **Substitution** $\int_{\varphi(a)}^{\varphi(b)} f(x) \, dx = \int_a^b f(\varphi(t)) \cdot \varphi'(t) \, dt$
- $a + c, b + c \in I : \int_a^b f(t + c) \, dt = \int_{a+c}^{b+c} f(x) \, dx$
- **Logarithmus** $\int \frac{f'(t)}{f(t)} \, dt = \log(|f(x)|)$

Substitution Beispiel

$$\int \cos(x^2) 2x \, dx$$

$$\int \cos(u) du$$

$$u = x^2$$

$$\frac{du}{dx} = \frac{dx^2}{dx} = 2x$$

2 Diskrete Verteilungen

Diskrete Gleichverteilung

Die diskrete Gleichverteilung auf einer endlichen Menge $\mathcal{W} = \{x_1, \dots, x_n\}$ gehart zu einer Zufallsvariablen X mit Wertebereich \mathcal{W} und Gewichtsfunktion:

$$p_X(x_k) = P[X = x_k] = \frac{1}{N}$$

$$k \in \{1, \dots, N\}$$

Unabhangige 0-1-Experimente

Es sei $A_i := \{\text{Erfolg beim } i\text{-ten Experiment}\}$ und:

- Die A_i sind unabhangig
- $P[A_i] = p$ fur alle i
-

$$Y_i(\omega) = \begin{cases} 1 & \omega \in A_i \\ 0 & \omega \notin A_i \end{cases}$$

Bernoulli-Verteilung

Ein einziges 0-1-Experiment mit $\mathcal{W}(X) = \{0, 1\}$. Die Gewichtsfunktion ist gegeben durch $p_X(1) = p$, sowie $p_X(0) = 1 - p$. Man schreibt kurz $X \sim Be(p)$. Es gilt:

$$E[X] = 1 \cdot P[X = 1] + 0 \cdot P[X = 0] = p$$

$$Var[X] = E[X^2] - E[X]^2 = p \cdot (1 - p)$$

Binomialverteilung

Beschreibt die Anzahl der Erfolge bei n unabhangigen 0-1-Experimenten mit Erfolgsparameter p . Also ist die Zufallsvariable respektive Gewichtsfunktion:

$$X = \sum_{i=1}^n I_{A_i} = \sum_{i=1}^n Y_i$$

$$p_X(k) = P[X = k] = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

und man schreibt kurz $X \sim Bin(n, p)$. Es gilt weiter:

$$E[X] = \sum_{i=1}^n E[Y_i] = n \cdot p$$

$$Var[X] = \sum_{i=1}^n Var[Y_i] = n \cdot p \cdot (1 - p)$$

Geometrische Verteilung

Bei einer unendlichen Folge von unabhangigen 0-1-Experimenten mit Erfolgsparameter p sein X die Wartezeit zum ersten Erfolg:

$$X = \inf\{i \in \mathbb{N} : A_i \text{ tritt ein}\}$$

$$p_X(k) = P[X = k] = p \cdot (1 - p)^{k-1}$$

wir schreiben $X \sim Geom(p)$ und es gilt:

$$E[X] = \sum_{i=0}^{\infty} (1 - p)^i = \frac{1}{1 - (1 - p)} = \frac{1}{p}$$

$$E[X \cdot (X - 1)] = \frac{2(1 - p)}{p^2}$$

$$Var[X] = \frac{1 - p}{p^2}$$

Negativbinomiale Verteilung

Bei einer unendlichen Folge von unabhangigen 0-1-Experimenten mit Erfolgsparameter p sein X die Wartezeit zum r -ten Erfolg ($r \in \mathbb{N}$):

$$X = \inf\{k \in \mathbb{N} : \sum_{i=1}^k I_{A_i} = r\}$$

$$p_X(k) = P[X = k] = \binom{k-1}{r-1} \cdot p^r \cdot (1 - p)^{k-r}$$

wir schreiben $X \sim NB(r, p)$ und es gilt:

$$E[X] = \sum_{i=1}^r E[X_i] = \frac{r}{p}$$

$$Var[X] = \sum_{i=1}^r Var[X_i] = \frac{r \cdot (1 - p)}{p^2}$$

Hypergeometrische Verteilung

In einer Urne seien n Gegenstande, davon r vom Typ 1 und $n - r$ vom Typ 2. Man zieht ohne zurucklegen m der Gegenstande. Die Zufallsvariable X beschreibt die Anzahl der Gegenstande vom Typ 1 in der Stichprobe. Der Wertebereich von X ist $\mathcal{W}(X) = \{0, 1, \dots, \min(m, r)\}$ und:

$$p_X(k) = \frac{\binom{r}{k} \cdot \binom{n-r}{m-k}}{\binom{n}{m}}$$

fur $k \in \mathcal{W}(X)$

$$E[X] = \sum_{i=1}^n i \cdot p_X(i) = m \cdot \frac{r}{n}$$

(Nicht im Skript)

$$Var[X] = m \cdot \frac{r}{n} \left(1 - \frac{r}{n}\right) \cdot \frac{n - m}{n - 1}$$

(Nicht im Skript)

Poisson Verteilung

Die Poisson Verteilung mit Parameter $\lambda \in (0, \infty)$ ist eine Verteilung auf der Menge $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ mit Gewichtsfunktion:

$$p_X(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

fur $k = 0, 1, 2, \dots$

$$E[X] = \sum_{i=1}^n i \cdot \frac{\lambda^i}{i!} e^{-\lambda} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$E[X^2] = \lambda^2 + \lambda$$

$$Var[X] = \lambda$$

Ist eine Zufallsvariable X Poisson verteilt mit Parameter λ schreiben wir $X \sim P(\lambda)$.

3 Zufallsvariablen

Zufallsvariable

Sein (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum. Also Ω ein Grundraum, $\mathcal{F} \subseteq 2^\Omega$ die beobachtbaren Ereignisse und P ein Wahrscheinlichkeitsmass auf \mathcal{F} . Eine (reelwertige) Zufallsvariable auf Ω ist eine messbare Funktion $X : \Omega \mapsto \mathbb{R}$. Das bedeutet, dass die Menge $\{X \leq t\} = \{\omega : X(\omega) \leq t\}$ fur jedes t ein beobachtbares Ereigniss sein muss.

Verteilungsfunktion

Die Verteilungsfunktion von X ist die Abbildung $F_X : \mathbb{R} \mapsto [0, 1]$:

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

und hat die Eigenschaften:

- F_X ist wachsend und rechtsstetig. Das bedeutet, dass $F_X(s) \leq F_X(t)$ fur $s \leq t$ gilt und $F_X(u) \rightarrow F_X(t)$ fur $u \rightarrow t$ mit $u > t$
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$ und $\lim_{t \rightarrow +\infty} F_X(t) = 1$

Dichtefunktion

Das Analogon der Gewichtsfunktion im Diskreten Fall. Eine Zufallsvariable X mit Verteilungsfunktion $F_X(t) = P[X \leq t]$ heisst (absolut) stetig mit Dichte (funktion) $f_X : \mathbb{R} \mapsto [0, \infty)$, falls gilt:

$$F_X(t) = \int_{-\infty}^t f_X(s) \, ds$$

fur alle $t \in \mathbb{R}$

und hat die Eigenschaften:

- $f_X \geq 0$ und $f_X = 0$ ausserhalb von $\mathcal{W}(X)$.
- $\int_{-\infty}^{\infty} f_X(s) \, ds = 1$; das folgt aus $\lim_{t \rightarrow +\infty} F_X(t) = 1$

Gleichverteilung

Die Gleichverteilung auf dem Intervall $[a, b]$ ist ein Modell für die Zufällige Wahl eines Punktes in $[a, b]$. Die zugehörige Zufallsvariable X hat den Wertebereich $\mathcal{W}(X) = [a, b]$, sowie

$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq t \leq b \\ 0 & \text{sonst.} \end{cases}$$
$$F_X(t) = \begin{cases} 0 & \text{für } t < a \\ \frac{t-a}{b-a} & \text{für } a \leq t \leq b \\ 1 & \text{für } t > b. \end{cases}$$

wir schreiben kurz $X \sim U(a, b)$.

$$E[X] = \frac{a+b}{2} \qquad Var[X] = \frac{(b-a+1)^2-1}{12}$$

Exponentialverteilung

Die Exponentialverteilung mit Parameter $\lambda > 0$ ist das stetige Analogon der Geometrischen Verteilung. Die zugehörige Zufallsvariable X hat $\mathcal{W}(X) = [0, \infty)$, Dichte und Verteilungsfunktion:

$$f_X(t) = \begin{cases} \lambda \cdot e^{-\lambda t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$
$$F_X(t) = \int_{-\infty}^t f_X(s) \, ds = \begin{cases} 1 - e^{-\lambda t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

wir schreiben kurz $X \sim Exp(\lambda)$. Weiter ist die Funktion Gedächtnislos, dh. $P[X > t + s \mid X > s] = P[X > t]$.

$$E[X] = \frac{1}{\lambda} \qquad Var[X] = \frac{1}{\lambda^2}$$

Normalverteilung

Die Normalverteilung hat zwei Parameter: $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Die zugehörige Zufallsvariable X hat den Wertebereich $\mathcal{W}(X) = \mathbb{R}$ und die Dichtefunktion:

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \qquad \text{für } t \in \mathbb{R}$$

welche symmetrisch um μ ist. Wir schreiben kurz: $X \sim \mathcal{N}(\mu, \sigma^2)$.

Standard Normalverteilung

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

$$\Phi(t) = \int_{-\infty}^t \varphi(s) \, ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}s^2} \, ds$$

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

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

deshalb genügt es Φ zu tabellieren.

$$\Phi(-z) = 1 - \Phi(z)$$

Normalapproximation

Wenn $S_n \sim Bin(n, p)$ dann

$$S_n \sim_{approx} N(np, np(1-p))$$

Erwartungswert

Ist X stetig mit Dichte $f_X(x)$, so ist der Erwartungswert:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx$$

sofern das Integral absolut konvergiert. Ist das Integral nicht absolut konvergent, so existiert der Erwartungswert nicht.

Erwartungswert einer Funktion

Sei X eine Zufallsvariable und $Y = g(X)$ eine weitere Zufallsvariable. Ist X stetig mit Dichte f_X , so ist

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, dx$$

Momente & Absolute Momente

Sei X eine Zufallsvariable und $p \in \mathbb{R}^+$. Wir definieren:

- p -te absolute Moment von X : $M_p := E[|X|^p]$
- falls $M_n < \infty$ für ein n , dann ist das n -te (rohe) Moment von X durch $m_n := E[X^n]$ definiert.
- Das n -te zentralisierte Moment von X ist durch $\mu_n := E[(X - E[X])^n]$ definiert.

Es gilt weiter, dass $M_n < \infty$ für $n \in \mathbb{N} \implies |m_n| \leq M_n$.

$$M_p = \int_{-\infty}^{\infty} |x|^p f_X(x) \, dx$$

$$m_n = \int_{-\infty}^{\infty} x^n f_X(x) \, dx$$

$$p \leq q \wedge M_q < \infty \implies M_p < \infty$$

Gemeinsame Verteilung/Dichte

Die Gemeinsame Verteilungsfunktion von Zufallsvariablen X_1, \dots, X_n ist die Abbildung $F : \mathbb{R}^n \mapsto [0, 1]$ mit:

$$F(x_1, \dots, x_n) := P[X_1 \leq x_1, \dots, X_n \leq x_n] \\ = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) \, dt_n \dots t_1$$

dann heisst $f(x_1, \dots, x_n)$ die gemeinsame Dichte, welche folgende Eigenschaften hat:

- $f(x_1, \dots, x_n) \geq 0$ und $= 0$ ausserhalb von $\mathcal{W}(X_1, \dots, X_n)$.
- $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) \, dt_n \dots t_1 = 1$
- $P[(X_1, \dots, X_n) \in A] = \int_{(x_1, \dots, x_n) \in A} f(t_1, \dots, t_n) \, dt_n \dots t_1$ für $A \subseteq \mathbb{R}^n$

Randverteilung

Haben X, Y die gemeinsame Verteilungsfunktion F , dann ist:

$$F_X(x) = P[X \leq x] = P[X \leq x, Y < \infty] = \lim_{y \rightarrow \infty} F(x, y)$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy$$

Unabhängigkeit

Die Zufallsvariablen X_1, \dots, X_n heissen unabhängig, falls gilt (äquivalent):

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \\ f(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n)$$

für alle x_1, \dots, x_n .

Bedingte Verteilungen

Es gilt:

$$f_{X_1 \mid X_2}(x_1 \mid x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)} \\ P[Y > t \mid Y < a] = \frac{P[t < Y < a]}{P[Y < a]} \\ E[X_1 \mid X_2] = \int x_1 \cdot f_{x_1 \mid x_2}(x_1 \mid x_2) \, dx_1$$

Summen von Zufallsvariablen

Sei $Z = X + Y$ eine Zufallsvariable mit:

$$F_Z(z) = P[Z \leq z] = P[X + Y \leq z] \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) \, dy \, dx \\ f_Z(z) = \int_{-\infty}^{\infty} f(z-y, y) \, dy$$

Transformationen

Sei X eine Zufallsvariable mit Verteilung und Dichte. Sei $g : \mathbb{R} \mapsto \mathbb{R}$ eine messbare Funktion. Betrachte nun $Y = g(X)$, wir suchen Verteilung und Dichte von Y :

$$F_Y(t) = P[Y \leq t] = P[g(X) \leq t] = \int_{A_g} f_X(s) \, ds \\ A_g := \{s \in \mathbb{R} \mid g(s) \leq t\}$$

Wobei man die Dichte durch ableiten der Verteilung erhält.

Anwendung von Transformationen

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

$$\begin{aligned} F_Y(t) &= P[Y \leq t] = P[F^{-1}(X) \leq t] \\ &= P[X \leq F(t)] = F(t) \end{aligned}$$

Markov Ungleichung

Sei X eine Zufallsvariable und ferner $g : \mathcal{W}(X) \mapsto [0, \infty)$ eine wachsende Funktion. Für jedes $c \in \mathbb{R}$ mit $g(c) > 0$ gilt dann:

$$P[X \geq c] \leq \frac{E[g(X)]}{g(c)}$$

Chebyshev-Ungleichung

Sei Y eine Zufallsvariable mit endlicher Varianz. Für jedes $b > 0$ gilt dann:

$$P[|Y - E[Y]| \geq b] \leq \frac{Var[Y]}{b^2}$$

Schwaches Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, die alle den gleichen Erwartungswert $E[X_i] = \mu$ und die gleiche Varianz $Var[X_i] = \sigma^2$ 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 $n \rightarrow \infty$ in Wahrscheinlichkeit/stochastisch gegen $\mu = E[X_i]$, d.h.:

$$P\left[|\overline{X}_n - \mu| > \varepsilon\right] \xrightarrow{n \rightarrow \infty} 0 \qquad \text{für jedes } \varepsilon > 0$$

(Statt unabhängig genügt auch $Cov(X_i, X_k) = 0$ für $i \neq k$)

Starkes Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, die alle dieselbe Verteilung haben, und ihr Erwartungswert $\mu = E[X_i]$ sei endlich. Für:

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

gilt dann

$$\overline{X}_n \xrightarrow{n \rightarrow \infty} \mu \qquad \text{P-fast sicher}$$

d.h.:

$$P\left[\left\{\omega \in \Omega : \overline{X}_n(\omega) \xrightarrow{n \rightarrow \infty} \mu\right\}\right] = 1$$

i.i.d. / u.i.v.

Independent identically distributed

Zentraler Grenzwertsatz

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

$$\lim_{n \rightarrow \infty} P\left[\frac{S_n - n \cdot \mu}{\sigma \sqrt{n}} \leq x\right] = \Phi(x) \qquad \text{für alle } x \in \mathbb{R}$$

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

4 Tabellen

4.1 Ableitungen

F(x)	f(x)	f'(x)
$(x-1)e^x$	xe^x	$(x+1)e^x$
$\frac{x^{-a+1}}{-a+1}$	$\frac{1}{x^a}$	$\frac{a}{x^{a+1}}$
$\frac{x^{a+1}}{a+1}$	$x^a \ (a \neq -1)$	$a \cdot x^{a-1}$
$\frac{1}{k \ln(a)} a^{kx}$	a^{kx}	$ka^{kx} \ln(a)$
$\ln x $	$\frac{1}{x}$	$-\frac{1}{x^2}$
$\frac{2}{3}x^{3/2}$	\sqrt{x}	$\frac{1}{2\sqrt{x}}$
$-\cos(x)$	$\sin(x)$	$\cos(x)$
	$\frac{\sin(x)^2}{2}$	$\sin(x) \cos(x)$
$\sin(x)$	$\cos(x)$	$-\sin(x)$
$\frac{1}{2}(x - \frac{1}{2} \sin(2x))$	$\sin^2(x)$	$2 \sin(x) \cos(x)$
$\tan(x) - x$	$\tan(x)^2$	$2 \sec(x)^2 \tan(x)$
$-\cot(x) - x$	$\cot(x)^2$	$-2 \cot(x) \csc(x)^2$
$\frac{1}{2}(x + \frac{1}{2} \sin(2x))$	$\cos^2(x)$	$-2 \sin(x) \cos(x)$
$-\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 $	$\frac{1}{\ln(a)x}$

F(x)	f(x)
$\arcsin(x)/\arccos(x)$	$\frac{1/-1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$x \arcsin(x) + \sqrt{1-x^2}$	$\arcsin(x)$
$x \arccos(x) - \sqrt{1-x^2}$	$\arccos(x)$
$x \arctan(x) - \frac{1}{2} \ln(1+x^2)$	$\arctan(x)$
$\ln(\cosh(x))$	$\tanh(x)$
$x^x \ (x > 0)$	$x^x \cdot (1 + \ln x)$
$f(x)^{g(x)}$	$e^{g(x) \ln(f(x))}$
$f(x) = \cos(\alpha)$	$f(x)^n = \sin(x + n \frac{\pi}{2})$
$f(x) = \frac{1}{ax+b}$	$f(x)^n = (-1)^n * a^n * n! * (ax+b)^{-n+1}$
$-\ln(\cos(x))$	$\tan(x)$
$\ln(\sin(x))$	$\cot(x)$
$\ln(\tan(\frac{x}{2}))$	$\frac{1}{\sin(x)}$
$\ln \tan(\frac{x}{2} + \frac{\pi}{4})$	$\frac{1}{\cos(x)}$
f(x)	F(x)
$\int f'(x)f(x) \, dx$	$\frac{1}{2}(f(x))^2$
$\int \frac{f'(x)}{f(x)} \, dx$	$\ln f(x) $
$\int_{-\infty}^{\infty} e^{-x^2} \, dx$	$\sqrt{\pi}$
$\int (ax+b)^n \, dx$	$\frac{1}{a(n+1)}(ax+b)^{n+1}$
$\int x(ax+b)^n \, dx$	$\frac{(ax+b)^{n+2}}{(n+2)a^2} - \frac{b(ax+b)^{n+1}}{(n+1)a^2}$
$\int (ax^p+b)^n x^{p-1} \, dx$	$\frac{(ax^p+b)^{n+1}}{ap(n+1)}$
$\int (ax^p+b)^{-1} x^{p-1} \, dx$	$\frac{1}{ap} \ln ax^p+b $
$\int \frac{ax+b}{cx+d} \, dx$	$\frac{ax}{c} - \frac{ad-bc}{c^2} \ln cx+d $
$\int \frac{1}{x^2+a^2} \, dx$	$\frac{1}{a} \arctan \frac{x}{a}$
$\int \frac{1}{x^2-a^2} \, dx$	$\frac{1}{2a} \ln \left \frac{x-a}{x+a} \right $
$\int \sqrt{a^2+x^2} \, dx$	$\frac{x}{2} f(x) + \frac{a^2}{2} \ln(x+f(x))$

Momenterzeugende Funktion

Die Momenterzeugende Funktion einer Zufallsvariable X ist:

M_X(t) := E[e^{t \cdot X}] \qquad \text{f\u00fcr } t \in \mathbb{R}

Grosse Summenabweichung

Seien X_1, \dots, X_n i.i.d. Zufallsvariablen, f\u00fcr welche die Momenterzeugende Funktion $M_X(t)$ f\u00fcr alle $t \in \mathbb{R}$ endlich ist. F\u00fcr jedes $b \in \mathbb{R}$ gilt dann:

P[S_n \geq b] \leq \exp\left(\inf_{t \in \mathbb{R}} (n \cdot \log M_X(t) - t \cdot b)\right)

Chernoff Schranken

Seien X_1, \dots, X_n unabh\u00e4ngig mit $X_i \sim Be(p)$ und $S_n = \sum_{i=1}^n X_i$. Sei $\mu_n := E[S_n] = \sum_{i=1}^n p_i$ und $\delta > 0$. Dann gilt:

P[S_n \geq (1 + \delta) \cdot \mu_n] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^{\mu_n}

5 Sch\u00e4tzer

Sch\u00e4tzer

Wir suchen ein Modell f\u00fcr eine Stichprobe X_1, \dots, X_n und haben einen Parameterraum $\vartheta \subseteq \Theta$ und f\u00fcr jedes ϑ einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P_\vartheta)$. Wir m\u00f6chten nun die Parameter $\vartheta_1, \dots, \vartheta_m$ bestimmen. Ein Sch\u00e4tzer T_j f\u00fcr einen Parameter ϑ_j ist eine Zufallsvariable der Form $T_j := t_j(X_1, \dots, X_n)$ f\u00fcr eine Sch\u00e4tzfunktion t_j .

Sch\u00e4tzwert

Ein Sch\u00e4tzwert ist das Ergebnis einer konkreten Berechnung, eine Zahl. Sie entsteht durch das Einsetzen konkreter Daten in einen Sch\u00e4tzer: $T_j(\omega) = t_j(x_1, \dots, x_n)$ und liefert damit einen Wert f\u00fcr genau einen Parameter.

Eigenschaften von Sch\u00e4tzern

Sei T ein Sch\u00e4tzer.

- T ist erwartungstreu, falls $E_\vartheta[T] = \vartheta$ gilt. T sch\u00e4tzt im Mittel also richtig.
- Bias := $E_\vartheta[T] - \vartheta$. Ein erwartungstreuer Sch\u00e4tzer hat also keinen Bias.
- Mittlere Quadratische Sch\u00e4tzfehler $MSE_\vartheta[T] := E_\vartheta[(T - \vartheta)^2]$.
- Eine Folge $T^{(n)}$ von Sch\u00e4tzern heisst konsistent f\u00fcr ϑ , falls $T^{(n)}$ f\u00fcr $n \rightarrow \infty$ in P_ϑ -Wahrscheinlichkeit gegen ϑ konvergiert, d.h. f\u00fcr jedes $\vartheta \in \Theta$ gilt:

\lim_{n \rightarrow \infty} P_\vartheta \left[\left| T^{(n)} - \vartheta \right| > \varepsilon \right] = 0

Maximum-Likelihood Methode

(Analog im diskreten Fall.) In einem Modell P_ϑ sind die Zufallsvariablen X_1, \dots, X_n stetig mit einer gemeinsamen Dichtefunktion $f(x_1, \dots, x_n, \vartheta)$. Oft sind die X_i i.i.d. und man erh\u00e4lt:

f(x_1, \dots, x_n, \vartheta) = P[X_1 = x_1, \dots, X_n = x_n] = \prod_{i=1}^n f_X(x_i, \vartheta)

Wir nehmen nun an, dass die Daten die wir erhalten haben sehr Wahrscheinlich sind und versuchen nun folgende Likelihood funktion zu Maximieren durch Anpassungen an ϑ :

L(x_1, \dots, x_n; \vartheta) := f(x_1, \dots, x_n; \vartheta) \log L(x_1, \dots, x_n; \vartheta) := \log f(x_1, \dots, x_n; \vartheta)

letzteres kann bei Produkt zu Summe umwandlung hilfreich sein.

Sei $\Theta = [0, 1]$. Wir betrachten die Modellfamilie $P_{\theta \in \Theta}$, wobei X_1, \dots, X_n unter \mathbb{P}_θ unabh\u00e4ngig und identisch verteilt sind mit $X_1 \sim \text{Geom}(\theta)$. Was ist die Likelihood-Funktion $L(x_1, \dots, x_n; \theta)$ f\u00fcr $x_1, \dots, x_n \in \{1, 2, \dots\}$?

L(x_1, \dots, x_n; \theta) = (P_\theta)[X_1 = x_1, \dots, X_n = x_n] = \prod_{i=1}^n \mathbb{P}_\theta[X_i = x_i] = \theta^n \cdot (1 - \theta)^{x_1 + \dots + x_n - n}

Was ist der Maximum-Likelihood-Sch\u00e4tzer T_{ML} f\u00fcr θ ?

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

Wir setzen nun die Ableitung der log-Likelihood-Funktion nach θ gleich 0 und erhalten:

\frac{n}{\theta} - \frac{x_1 + \dots + x_n - n}{1 - \theta} = 0 \iff n - n\theta = (x_1 + \dots + x_n) \cdot \theta - n\theta \iff \theta = \frac{n}{x_1 + \dots + x_n} = \frac{n}{X_1 + \dots + X_n}

Empirisches Moment

F\u00fcr $k \in \{1, \dots, m\}$ 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

Momentenmethode

Der Momentenmethode liegt zugrunde, dass die Momente einer Zufallsvariable bzw. einer Wahrscheinlichkeitsverteilung durch Stichprobenmomente gesch\u00e4tzt werden k\u00f6nnen. Sei X_1, \dots, X_n eine Stichprobe und Θ der Parameterraum. F\u00fcr jeden Parameter $\vartheta = (\vartheta_1, \dots, \vartheta_m) \in \Theta$ sei X_1, \dots, X_n i.i.d. unter dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P_\vartheta)$. Methode:

1. F\u00fcr gegebene Realisierungen x_1, \dots, x_n bestimme f\u00fcr jedes $k \in \{1, \dots, m\}$ das k -te empirische Moment
 2. Stelle ein Gleichungssystem f\u00fcr die Unbekannten Parameter $\vartheta_1, \dots, \vartheta_m$ auf, in dem das k -te empirische Moment dem k -ten Moment gleichgesetzt wird, also:
- \hat{m}_k(x_1, \dots, x_n) = g_k(\vartheta_1, \dots, \vartheta_m) \qquad k \in \{1, \dots, m\}
3. Existiert eine Eindeutige L\u00f6sung so wird das unsere Sch\u00e4tzung f\u00fcr ϑ .

Momentensch\u00e4tzer

Der Vektor $\hat{\vartheta}(X_1, \dots, X_m)$ heisst Momentensch\u00e4tzer des Parameters ϑ .

Beispiel: Normalverteilte Stichprobe

Sei X_1, \dots, X_n i.i.d. $\mathcal{N}(\mu, \sigma^2)$ -verteilt mit unbekannten Parametern $\vartheta = (\mu, \sigma^2)$. Damit berechnen wir mit der log max likelihood funktion Ableitungen setzen diese zu 0 und bekommen:

T_1 = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}_n T_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}_n^2

m\u00f6chten wir aber noch, dass der Sch\u00e4tzer erwartungstreu wird, so w\u00e4hlen wir f\u00fcr $T_2 = S^2$:

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

Normalverteilte Stichproben

Seien X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

- $\overline{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ und $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$.
- $\frac{n-1}{\sigma^2} S^2 = \left(\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \overline{X}_n)^2\right) \sim \chi_{n-1}^2$
- \overline{X}_n und S^2 sind unabh\u00e4ngig
- $\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\frac{S/\sigma}{\sigma/\sqrt{n}}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1} \frac{n-1}{\sigma^2} S^2}} \sim t_{n-1}$

6 Statistik

χ²-Verteilung

Die χ²- Verteilung mit n Freiheitsgraden (bez. χ²_n) gehört zu einer stetigen Zufallsvariable Y mit Dichtefunktion:

f_Y(y) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{1}{2}y}

wobei dies ein Spezialfall der Ga(α, λ) Verteilung ist mit α = \frac{n}{2} und λ = \frac{1}{2}. Sind die Zufallsvariablen X_1, \dots, X_n i.i.d. \sim \mathcal{N}(0, 1), so ist die Summe Y := \sum_{i=1}^n X_i^2 \sim \chi_n^2.

t-Verteilung

Die t-Verteilung mit n Freiheitsgraden gehört zu einer stetigen Zufallsvariable Z mit Dichtefunktion

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

für n = 1 ist das eine Cauchy Verteilung und für n \to \infty erhält man \mathcal{N}(0, 1). Sind X, Y unabhängig und X \sim \mathcal{N}(0, 1) und Y \sim \chi_n^2, so ist der Quotient:

Z := \frac{X}{\sqrt{\frac{1}{n}Y}} \sim t_n, \text{ also } t\text{-Verteilt mit } n \text{ Freiheitsgraden}

Hypothesen

Es gibt:

- Hypothese H_0 : \vartheta \in \Theta_0
- Alternative H_A : \vartheta \in \Theta_A

Man verwirft die Hypothese genau dann, wenn der realisierte Wert im Verwerfungsbereich K liegt.

Fehler

Es gibt folgende Fehler:

- 1. Art: Hypothese wird zu unrecht abgelehnt. P_{\vartheta}[T \in K] für \vartheta \in \Theta_0
- 2. Art: Hypothese wird zu unrecht nicht verworfen. P_{\vartheta}[T \notin K] für \vartheta \in \Theta_A

Meist kann man nicht beides minimieren, also geht man wie folgt vor:

1. Man wählt ein **Signifikanzniveau** \alpha \in (0, 1) und kontrolliert die Wahrscheinlichkeit eines Fehlers 1. Art durch

\sup_{\vartheta \in \Theta_0} P_{\vartheta}[T \in K] \leq \alpha

2. Man versucht die Wahrscheinlichkeit für einen Fehler zweiter Art P_{\vartheta}[T \notin K] für \vartheta \in \Theta_A zu minimieren. Dazu maximiert man die **Macht des Tests**:

\beta : \Theta_A \mapsto [0, 1] \quad \vartheta \mapsto \beta(\vartheta) := P_{\vartheta}[T \in K]

Somit ist es schwieriger eine Hypothese zu verwerfen als zu behalten. In einem Test **verwendet man deshalb immer als Hypothese die Negation der eigentlich gewünschten Aussage.**

Likelihood Quotient

Sei L(x_1, \dots, x_n; \vartheta) die Likelihood Funktion und \vartheta_0 \in \Theta_0 sowie \vartheta_A \in \Theta_A. Dann definieren wir:

R(x_1, \dots, x_n; \vartheta_0, \vartheta_A) = \frac{L(x_1, \dots, x_n; \vartheta_A)}{L(x_1, \dots, x_n; \vartheta_0)}

je grösser der Quotient, desto wahrscheinlicher die Alternative. Es gibt auch:

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

\tilde{R}(x_1, \dots, x_n) = \frac{\sup_{\vartheta \in \Theta_A \cup \Theta_0} L(x_1, \dots, x_n; \vartheta)}{\sup_{\vartheta \in \Theta_0} L(x_1, \dots, x_n; \vartheta)}

Wähle Konstante c_0 für K_0 = (c_0, \infty) mithilfe von Signifikanzniveau.

Neyman-Pearson Lemma

Sei \Theta_0 = \{\vartheta_0\} und \Theta_A = \{\vartheta_A\}. Wie oben sei T = R(X_1, \dots, X_n; \vartheta_0, \vartheta_A) und K := (c, \infty), sowie \alpha^* := P_{\vartheta_0}[T \in K] = P_{\vartheta_0}[T > c]. Der Likelihood Quotienten Test mit Teststatistik T und kritischem Bereich K ist dann in folgendem Sinn optimal: Jeder andere Test mit Signifikanzniveau \alpha \leq \alpha^* hat eine kleinere Macht (bez. Grössere WS Fehler 2. Art).

p-Wert

(Nach Wikipedia) Der p-Wert ist die Wahrscheinlichkeit ein mindestens so extremes Testergebnis zu erhalten, wenn die Nullhypothese gelten würde:

p(x) = P[X \leq x \mid H_0] \text{ oder } P[X \geq x \mid H_0]

p(x) = 2 \cdot \min\{P[X \leq x \mid H_0], P[X \geq x \mid H_0]\}

z-Test

Normalverteilung, Test für Erwartungswert bei bekannter Varianz. Hier sind X_1, \dots, X_n i.i.d. \mathcal{N}(\vartheta, \sigma^2). Wir möchten die Hypothese H_0 : \vartheta = \vartheta_0 testen. Mögliche Alternativen H_A sind \vartheta > \vartheta_0, \vartheta < \vartheta_0 (einseitig) oder \vartheta \neq \vartheta_0 (zweiseitig). Die Teststatistik ist:

T = \frac{\bar{X} - \vartheta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{unter } P_{\vartheta_0}

Und die Verwerfungsbereiche:

\vartheta < \vartheta_0 \quad (-\infty, z_{\alpha})

\vartheta > \vartheta_0 \quad (z_{1-\alpha}, \infty)

\vartheta \neq \vartheta_0 \quad (-\infty, z_{\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, \infty)

Wobei die z Werte in der Tabelle nachgeschaut werden können und es gilt z_{\alpha} = -z_{1-\alpha}.

Beispiel Fehler 2. Art Berechnen

Nehme an: einseitiger z-Test, T = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}, \mu_0 = 70.

H_0 : \mu = \mu_0 \quad H_A : \mu < \mu_0

Kritischer Bereich mit 5% niveau: K = (-\infty, -1.645). Wir nehmen an, dass T = \frac{\bar{X}_n - \mu_A}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1):

P_{\mu_A}[T \notin K] = P_{\mu_A}[T > -1.645]

= P_{\mu_A}\left[\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > -1.645\right]

= P_{\mu_A}\left[\frac{\bar{X}_n - \mu_A}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} - 1.645\right]

= 1 - P_{\mu_A}\left[\frac{\bar{X}_n - \mu_A}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} - 1.645\right]

= 1 - \Phi\left(\frac{\mu_0 - \mu_A}{\sigma - \sqrt{n}} - 1.645\right) \quad \text{Weil } \sim \mathcal{N}(0, 1)

t-Test

Normalverteilung, Test für Erwartungswert bei unbekannter Varianz. Hier sind X_1, \dots, X_n i.i.d. \sim \mathcal{N}(\mu, \sigma^2) unter P_{\vec{\vartheta}}, wobei \vec{\vartheta} = (\mu, \sigma^2). Wir wollen die Hypothese \mu = \mu_0 testen. Die Teststatistik ist:

T := \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{unter } P_{\vartheta_0}

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

Und die Verwerfungsbereiche:

c_< = t_{n-1, \alpha} \quad (-\infty, c_<) \quad \mu < \mu_0

c_> = t_{n-1, 1-\alpha} \quad (c_>, \infty) \quad \mu > \mu_0

c_{\neq} = t_{n-1, 1-\frac{\alpha}{2}} \quad (-\infty, c_<) \cup (c_>, \infty) \quad \mu \neq \mu_0

Wobei gilt t_{m, \alpha} = -t_{m, 1-\alpha}.

Gepaarter Zweiprobe-Test

Hier sind X_1, \dots, X_n i.i.d. \sim \mathcal{N}(\mu_X, \sigma^2) und Y_1, \dots, Y_n i.i.d. \sim \mathcal{N}(\mu_Y, \sigma^2) unter P_{\vartheta}. Insbesondere ist m = n und die Varianz beider Stichproben dieselbe. Differenzen Z_i := X_i - Y_i sind unter P_{\vartheta} i.i.d. \mathcal{N}(\mu_X - \mu_Y, 2\sigma^2). Dann analog z und t-Test. (Setzt natürliche Paarung von Daten voraus!)

Ungepaarter Zweiprobe

Hier sind unter P_ϑ die Zufallsvariablen X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und Y_1, \dots, Y_m i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$, wobei die Varianz in beiden Fallen dieselbe ist.

- Bei bekannter Varianz:

$$H_0 : \mu_X - \mu_Y = \mu_0 \quad (z.B. \mu_0 = 0)$$
$$T = \frac{\overline{X}_n - \overline{Y}_m - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0, 1)$$

Die kritischen Werte fur den Verwerfungsbereich sind wie oben geeignete Quantile der $\mathcal{N}(0, 1)$ -Verteilung, je nach Alternative. Das ist der ungepaarte Zweistichproben- z -Test.

- Bei unbekannter Varianz:

$$S_X^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$$
$$S_Y^2 := \frac{1}{m-1} \sum_{j=1}^m (Y_j - \overline{Y}_m)^2$$
$$S^2 := \frac{1}{m+n-2} ((n-1) \cdot S_X^2 + (m-1) \cdot S_Y^2)$$
$$T = \frac{\overline{X}_n - \overline{Y}_m - (\mu_X - \mu_Y)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

unter jedem P_ϑ . Dieser Test heisst ungepaarter Zweistichproben- t -Test.

Konfidenzbereich

Ein Konfidenzbereich fur ϑ zu Daten x_1, \dots, x_n ist eine Menge $C(x_1, \dots, x_n) \subseteq \Theta$. Damit ist $C(X_1, \dots, X_n)$ eine zufallige Teilmenge von Θ . Dieses C heisst Konfidenzbereich zum Niveau $1 - \alpha$, falls fur alle $\vartheta \in \Theta$ gilt:

$$P_\vartheta[\vartheta \in C(X_1, \dots, X_n)] \geq 1 - \alpha$$

Beispiel Konfidenzbereich

Machen wir den Ansatz:

$$C(X_1, \dots, X_n) = [\overline{X}_n - \dots, \overline{X}_n + \dots]$$

so wollen wir erreichen, dass gilt:

$$1 - \alpha \leq P_\vartheta[\vartheta \in C(X_1, \dots, X_n)]$$
$$= P_\vartheta [\mu \in [\overline{X}_n - \dots, \overline{X}_n + \dots]] = P_\vartheta [|\overline{X}_n - \mu| \leq \dots]$$

Nach Satz 7.1 ist fur jedes $\vartheta \in \Theta$:

$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1} \quad \text{unter } P_\vartheta$$
$$1 - \alpha \leq P_\vartheta \left[\left| \frac{\overline{X}_n - \mu}{S/\sqrt{n}} \right| \leq \frac{\dots}{S/\sqrt{n}} \right]$$

also erhalten wir das Konfidenzintervall fur μ zum Niveau $1 - \alpha$:

$$C(X_1, \dots, X_n) = \left[\overline{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

7 Diskrete Verteilungen

Verteilung	Parameter	$\mathbb{E}[X]$	$\text{Var}(X)$	$p_X(t)$	$F_X(t)$
Gleichverteilung	n : Anzahl Ereignisse x_i : Ereignisse	$\frac{1}{n} \sum_{i=1}^n x_i$	$\frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n^2} \left(\sum_{i=1}^n x_i \right)^2$	$\frac{1}{n}$	$\frac{\lfloor (k+0.5)/n \rfloor}{n}$
Bernoulli	p : ErfolgsWK	p	$p \cdot (1-p)$	$p^t (1-p)^{1-t}$	$1-p$ fur $0 \leq t < 1$
Binomial	p : ErfolgsWK n : Anzahl Versuche	np	$np(1-p)$	$\binom{n}{t} p^t (1-p)^{n-t}$	$\sum_{k=0}^t \binom{n}{k} p^k (1-p)^{n-k}$
Geometrisch	p : ErfolgsWK t : Anzahl Versuche	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$p(1-p)^{t-1}$	$1 - (1-p)^t$
Negativ Binomial	$r > 0$: Erfolge bis Abbruch p : ErfolgsWK t : Misserfolge	$\frac{r}{1-p}$	$\frac{r}{(1-p)^2}$	$\binom{t+r-1}{r-1} \cdot (1-p)^r p^t$	$F_{\text{Binomial}}(t; n=t+r, p)$
Hypergeometrisch	N : Anzahl aller Elemente $M \leq N$: Anzahl moglicher Erfolge $n \leq N$: Anzahl Elemente in der Stichprobe	$n \frac{M}{N}$	$n \frac{M}{N} \left(1 - \frac{M}{N} \right) \frac{N-n}{N-1}$	$\frac{\binom{M}{t} \binom{N-M}{n-t}}{\binom{N}{n}}$	$\sum_{k=0}^t \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$
Poisson	λ : Erwartungswert und Varianz	λ	λ	$\frac{\lambda^t}{t!} e^{-\lambda}$	p

8 Stetige Verteilungen

Verteilung	Parameter	$\mathbb{E}[X]$	$\text{Var}(X)$	$f_X(t)$	$F_X(t)$
Gleichverteilung	$[a, b]$: Intervall	$\frac{a+b}{2}$	$\frac{1}{12} (b-a)^2$	$\frac{1}{b-a}$	$\frac{t-a}{b-a}$
Exponentialverteilung	λ : $\frac{1}{\text{MW}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$	$\begin{cases} 1 - e^{-\lambda t}, & t > 0 \\ 0, & t \leq 0 \end{cases}$
Normalverteilung	σ^2 : Varianz μ : $\mathbb{E}[X]$	μ	σ^2	$\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} = -\infty < t < \infty$	$\frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2} dy$
χ^2 -Verteilung	n : Freiheitsgrad	n	$2n$	$\frac{1}{2^n \Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}, \quad t > 0$	$\text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$
Gamma-Verteilung	α, λ	$\frac{n}{\alpha}$	$\frac{n}{\alpha^2}$	$\frac{\lambda^\alpha \mu^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}, \quad t > 0 \quad \alpha, \lambda > 0$	$\frac{1}{\Gamma(\alpha)} \gamma(\alpha, \lambda t)$
t-Verteilung	n : Freiheitsgrad	0 fur $n > 1$ sonst undef.	$\begin{cases} \frac{n-2}{n-4} & n > 2 \\ \infty & 1 < n \leq 2 \\ \text{undef.} & \text{sonst} \end{cases}$	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n} \right)^{-\frac{n+1}{2}}$	cdf