Wahrscheinlichkeit und Statistik

1 Grundlagen

1.1 Diskrete Wahrscheinlichkeitsräume

σ-Algebra Ein Mengensystem $\mathcal{F} \subset \mathcal{P}(\Omega)$ heisst σ-Algebra gdw.

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

Die Elemente der Algebra heissen Ereignisse.

Man kann ausserdem zeigen, dass für eine unendliche Folge $(A_i)_{i\geq 1}$ von Ereignissen gilt:

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$$

Wahrscheinlichkeitsmass Sei Ω ein Grundraum und sei $\mathcal{F}$ eine σ-Algebra. Eine Abbildung

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

heisst Wahrscheinlichkeitsmass auf $(\Omega, \mathcal{F})$, falls folgende Eigenschaften gelten:

- $\mathbb{P}[\Omega] = 1$
- (σ-Additivität) $\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ wenn $A = \bigcup_{i=1}^{\infty} A_i$ (disjunkte Vereinigung)

Auf der Potenzmenge $\mathcal{A}$ von Ω wird dann festgelegt:

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega), \quad A \in \mathcal{A}$$

Rechenregeln

- $\mathbb{P}[\emptyset] = 0$
- $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < ... < i_k \leq n} \mathbb{P}(A_{i_1} \cap ... \cap A_{i_k})$
- $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Für paarweise disjunkte Ereignisse $A_1, A_2, ..., A_k$ gilt:

$$\mathbb{P}[A_1 \cup \cdots \cup A_k] = \mathbb{P}[A_1] + \cdots + \mathbb{P}[A_k]$$

Für eine Folge $A_1, A_2, ...$ von (nicht notwendigerweise disjunkten) Ereignissen gilt:

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

1.2 Bedingte Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeit

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

Partition $\mathcal{B} = B_1, ..., B_n$ ist eine Partition von Ω gdw. B_i paarweise disjunkte Ereignisse sind und $\Omega = B_1 \cup \cdots \cup B_n$.

Totale Wahrscheinlichkeit Sei $B_1, ..., B_n$ eine Partition von Ω so dass $\mathbb{P}[B_i] > 0$ für jedes i gilt. Dann gilt für jedes $\mathcal{A} \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

Rechenregel

$$\mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)...\mathbb{P}(A_n|A_1 \cap ... \cap A_{n-1})$$

Satz von Bayes

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

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^{\complement}) \cdot (1 - \mathbb{P}(B))}$$

Im Allgemeinen gilt für eine Partition $\mathcal{B}$ von Ω und $\mathbb{P}(A) > 0$:

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}$$

1.3 Unabhängigkeit

Unabhängigkeit Sei $I = [1, ..., n]$. Eine Kollektion von Ereignissen $(A_i; i \in I)$ heisst (stochastisch) unabhängig wenn gilt:

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

Unabhängigkeit ist keine Eigenschaft der Ereignisse per se, sondern eine Eigenschaft in Bezug auf die Wahrscheinlichkeitsverteilung $\mathbb{P}$.

Unabhängigkeit von zwei Ereignissen Für zwei Ereignisse A, B mit $\mathbb{P}[A], \mathbb{P}[B] > 0$ sind folgende Aussagen Äquivalent:

- A, B sind unabhängig
- $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$
- $\mathbb{P}[A|B] = \mathbb{P}[A]$

2 Zufallsvariablen

Kriterium für eine Zufallsvariable X:

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

Indikatorfunktion Sei $A \in \mathcal{F}$. Dann ist die Indikatorfunktion $\mathbb{1}_A$ auf A:

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

Es gilt dann für eine Partition B

$$\mathbb{E}[\mathbb{1}_{B_i} \cdot X] = \mathbb{E}[\mathbb{P}[B_i] \cdot X] = \sum_{\omega \in B_i} \mathbb{P}[\{\omega\}] \cdot X(\omega)$$

$$\mathbb{E}[X|B](w) = \sum_{j \in I} \mathbb{1}_{B_j}(w) \cdot \mathbb{E}[X|B_j]$$

2.1 Verteilungsfunktion

Verteilungsfunktion Die Verteilungsfunktion F_X einer Z.V. X ist:

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

Eigenschaften

- F ist monoton wachsend
- F ist rechtsstetig: $\lim_{y \to b^+} F_X(y) = F_X(b)$
- $\lim_{a \to -\infty} F(a) = 0$
- $\lim_{a \to \infty} F(a) = 1$

Rechenregeln

- $\mathbb{P}[X \leq x] = 1 - \mathbb{P}[X > x]$
- $\mathbb{P}[X < x] = \mathbb{P}[X \leq x] - \mathbb{P}[X = x]$
- $\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a) \quad \forall a, b \in \mathbb{R}$
- $F_X(a) = \sum_{b \leq a} p_b$

2.2 Unabhängigkeit

Unabhängige Zufallsvariablen Die Zufallsvariablen $X_1, ..., X_n$ sind unabhängig falls $\forall x_1, x_2, ..., x_n \in \mathbb{R}$:

$$\mathbb{P}[X_1 \leq x_1, ..., X_n \leq x_n] = \mathbb{P}[X_1 \leq x_1] \ldots \mathbb{P}[X_n \leq x_n]$$

2.3 Diskrete Zufallsvariablen

Verteilung Sei X eine diskrete Zufallsvariable mit Werten in einer endlichen oder abzählbaren Menge $W \subset \mathbb{R}$. Die Zahlenfolge $(p(x))_{x \in W}$ definiert durch

$$\forall x \in W \quad p(x) := \mathbb{P}[X = x]$$

heisst Verteilung von X und erfüllt

$$\sum_{x \in W} p(x) = 1$$

Von Verteilung p zur Verteilungsfunktion F_X

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

Von Verteilungsfunktion F_X zur Verteilung p

$$W = \{\text{Position des Sprungs von } F_X\}$$
$$p(x) = \text{"Höhe des Sprungs" im Punkt } x \in W$$

2.4 Diskrete Verteilungen

Bernoulli Münzwurf mit Wahrscheinlichkeit p. $X \sim \mathrm{Ber}(p)$:

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

$$F(x) = \mathbb{P}[X \leq x] = \begin{cases} 0 & \text{falls } x < 0 \\ 1 - p & \text{falls } 0 \leq x < 1 \\ 1 & \text{falls } x \geq 1 \end{cases} \qquad \mathbb{E}[X] = p$$

$$\mathrm{Var}[X] = p \cdot (1 - p)$$

Binomial n mal Münzwurf, genau k sind Zahl. $X \sim \mathrm{Bin}(n, p)$:

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

$$\mathbb{E}[X] = n \cdot p$$

$$\mathrm{Var}[X] = n \cdot p \cdot (1 - p)$$

Negativ Binomial Anzahl Versuche die nötig sind, um r mal Zahl zu werfen. $X \sim \mathrm{NBin}(p, r)$:

$$\forall n \geq r \quad \mathbb{P}[X = n] = \binom{n-1}{r-1} p^r \cdot (1 - p)^{n-r}$$

$$\mathbb{E}[X] = \frac{r}{p}$$

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

Geometrisch Münze solange werfen bis Kopf zum ersten Mal erscheint. $X \sim \mathrm{Geom}(p)$:

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

$$F(n) = \mathbb{P}[X \leq n] = p \sum_{i=1}^{n} (1 - p)^{i-1} = 1 - (1 - p)^n$$

$$\mathbb{E}[X] = \frac{1}{p}$$

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

Gedächtnislosigkeit der geom. Verteilung Sei $T \sim \mathrm{Geom}(p)$ für $0 < p < 1$. Dann gilt:

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

Hypergeometrisch Wie binomial, aber ohne Zurücklegen. Aus N Elementen wird n mal gezogen, wobei M dieser Elemente einen Erfolg darstellen. $X \sim \mathrm{HG}(N, M, n)$:

$$\forall k \in \{0, ..., n\} \quad \mathbb{P}[X = k] = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}[X] = n \cdot \frac{M}{N}$$

$$\mathrm{Var}[X] = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N - n}{N - 1}$$

Poisson Approximiert die Binomialverteilung für grosse n und kleine p. Sei $\lambda > 0$. Dann gilt $X \sim \mathrm{Poisson}(\lambda)$:

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

$$F(n) = \sum_{k=0}^{n} \mathbb{P}[X = k] = p$$

$$\mathbb{E}[X] = \lambda, \qquad \mathbb{E}[X^2] = \lambda^2 + \lambda$$

$$\mathrm{Var}[X] = \lambda$$

2.5 Stetige Verteilungen

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

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

wobei $f : \mathbb{R} \to \mathbb{R}_+$ eine nicht-negative Funktion ist. Wir nennen dann f Dichte von X.

Es muss gelten:

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1$$

Gleichverteilung Eine stetige Zufallsvariable X heisst gleichverteilt ($X \sim \mathcal{U}([a, b])$) auf $[a, b]$ falls ihre Dichte gegeben ist durch

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

Eigenschaften:

- $\mathbb{P}[X \in [c, c + l]] = \frac{l}{b-a}$
- $F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$
- $\mathbb{E}[X] = \frac{a+b}{2}$
- $\mathrm{Var}[X] = \frac{1}{12} \cdot (b - a)^2$

Exponentialverteilung Das stetige Pendant zur diskreten geometrischen Verteilung. Modelliert häufig die Lebensdauer oder Wartezeit eines allgemeinen Ereignisses, z.B. Wartezeit bis zum ersten Kunden bei Neueröffnung eines Ladens.

Eine stetige Zufallsvariable T heisst exponentialverteilt mit Parameter $\lambda > 0$ ($T \sim \exp(\lambda)$) falls ihre Dichte gegeben ist durch

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

Eigenschaften:

- Wahrscheinlichkeit des Wartens ist exponentiell klein:
  $\forall t \geq 0 \quad \mathbb{P}[T > t] = e^{-\lambda t}.$.
- Gedächtnislosigkeit:
  $\forall t, s \geq 0 \quad \mathbb{P}[T > t + s \mid T > t] = \mathbb{P}[T > s]$
- $\mathbb{E}[T] = \frac{1}{\lambda}$
- $\mathrm{Var}[T] = \frac{1}{\lambda^2}$

Normalverteilung Eine stetige Zufallsvariable X heisst normal verteilt mit Parametern m und $\sigma^2 > 0$ ($X \sim \mathcal{N}(m, \sigma^2)$) falls ihre Dichte gegeben ist durch

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Tritt z.B. bei Messungen von physikalischen Grössen auf. Der reale Wert ist dann m, und σ stellt die Schwankungen / Ungenauigkeiten der Messung dar. Für kleine σ gilt die Messung als qualitativ hochwertig.

Eigenschaften:

- Seien $X_1, ..., X_n$ unabhängige normalverteile Zufallsvariablen mit Parametern $(m_1, \sigma_1^2), ..., (m_n, \sigma_n^2)$. Dann ist

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

eine normalverteile Zufallsvariable mit Parametern $m = m_0 + \lambda_1 m_1 + ... + \lambda_n m_n$ und $\sigma^2 = \lambda_1^2 \sigma_1^2 + ... + \lambda_n^2 \sigma_n^2$
- Standardnormalverteite Zufallsvariable: $X \sim \mathcal{N}(0,1)$. Für eine normalverteilte Zufallsvariable Z mit Parametern m und σ^2 gilt dann:

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

- Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = \frac{X}{n}$. Dann $Y \sim \mathcal{N}(\frac{\mu}{n}, \frac{\sigma^2}{n^2})$
- $\mathbb{E}[X] = m, \qquad \mathrm{Var}[X] = \sigma^2$

3 Erwartungswert

Allgemeiner Erwartungswert Sei $X : \Omega \to \mathbb{R}_+$ eine Zufallsvariable mit nicht-negativen Werten. Dann ist der Erwartungswert von X:

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

Für allgemeine Zufallsvariablen definieren wir X_+ und X_-, welche den positiven bzw. negativen Teil der Zufallsvariable darstellen und sonst 0 sind. Dann gilt $X = X_+ - X_-$ und $|X| = X_+ + X_-$.

Falls $\mathbb{E}[|X|] < \infty$, gilt:

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$$

Diskreter Erwartungswert Sei $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable mit Werten in W und $\phi : \mathbb{R} \to \mathbb{R}$ eine Abbildung. Dann gilt

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

Stetiger Erwartungswert Sei X eine stetige Zufallsvariable mit Dichte f. Dann gilt

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

Wenn $\phi : \mathbb{R} \to \mathbb{R}$ eine Abbildung, sodass $\phi(X)$ eine Zufallsvariable ist, dann gilt

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

Linearität des Erwartungswerts

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

Unabhängigkeit Falls X, Y unabhängige Zufallsvariablen, dann ist

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

Die andere Richtung gilt nicht. Wir brauchen eine stärkere Bedingung:

Seien $X_1, ..., X_n$ Zufallsvariablen. Die folgenden Aussagen sind äquivalent

1. $X_1, ..., X_n$ sind unabhängig
2. Für jede $\phi_1 : \mathbb{R} \to \mathbb{R}, ..., \phi_n : \mathbb{R} \to \mathbb{R}$ stückweise stetig, beschränkt gilt:

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

Extremwertformel Sei X eine Zufallsvarible, sodass $X \geq 0$ fast sicher. Dann gilt:

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

Satz Sei X eine Zufallsvariable. Sei $f : \mathbb{R} \to \mathbb{R}_+$ eine Abbildung, sodass $\int_{-\infty}^{\infty} f(x)dx = 1$. Dann sind folgende Aussagen äquivalent

1. X ist stetig mit Dichte f
2. Für jede Abbildung stückweise stetige, beschränkte Abbildung $\phi : \mathbb{R} \to \mathbb{R}$ gilt

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

3.1 Ungleichungen

Monotonie Seien X, Y zwei Zufallsvariablen, sodass $X \leq Y$ fast sicher. Falls beide Erwartungswerte wohldefiniert sind folgt dann

$$\mathbb{E}[X] \leq \mathbb{E}[Y] \text{ f.s.}$$

Markow Sei X eine nicht-negative Zufallsvariable. Für jedes $a > 0$ gilt dann

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

Chernoff Durch Anwenden der Markov-Ungleichung auf e^{tX}:

$$\mathbb{P}[X \geq a] = \mathbb{P}[e^{tX} \geq e^{ta}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

Jensen Sei X eine Zufallsvariable und $\phi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion. Falls $\mathbb{E}[\phi(X)]$ und $\mathbb{E}[X]$ wohldefiniert sind, gilt

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

Daraus folgt

- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$

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

$$\mathbb{P}[|X - m| \geq a] \leq \frac{\sigma_X^2}{a^2}, \quad \text{wobei } m = \mathbb{E}[X]$$

3.2 Varianz

Varianz und Standardabweichung Sei X eine Zufallsvariable, sodass $\mathbb{E}[X^2] < \infty$. Wir definieren die Varianz von X durch

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

Die Wurzel der Varianz, also σ_X nennen wir die Standardabweichung von X.

Eigenschaften der Varianz

1. Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$. Dann gilt

$$\sigma_X^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

2. Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$ und sei $\lambda \in \mathbb{R}$. Dann gilt
$$\sigma_{\lambda X}^2 = \lambda^2 \cdot \sigma_X^2 \quad \text{(i.e. Var}[\lambda X] = \lambda^2 \cdot \text{Var}[X])$$
3. Seien $X_1, ..., X_n$paarweise unabhängige Zufallsvariablen und $S = X_1 + \cdots + X_n$. Dann gilt
$$\sigma_S^2 = \sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2 \quad \text{(i.e. Var}[X+Y] = \text{Var}[X] + \text{Var}[Y])$$
4. Für eine Zufallsvariable X gilt
$$\text{Var}[\alpha \cdot X + \beta] = \alpha^2 \cdot \text{Var}[X], \quad \alpha, \beta \in \mathbb{R}$$
5. Für zwei beliebige Zufallsbariablen X, Y gilt
$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{Cov}[X, Y]$$

Kovarianz Seien X, Y zwei Zufallsvariablen mit endlichen zweiten Momenten $\mathbb{E}[X^2] < \infty$ und $\mathbb{E}[Y^2] < \infty$. Wir definieren die Kovarianz zwischen X und Y durch
$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$
Es gilt
$$X, Y \text{ unabhängig} \implies \text{Cov}(X, Y) = 0$$

4 Gemeinsame Verteilung

Gemeinsame diskrete Verteilung Seien $X_1, ..., X_n$ diskrete Zufallsvariablen und sei $W_i \subset \mathbb{R}$ endlich oder abzählbar, wobei $X_i \in W_i$ fast sicher. Die gemeinsame Verteilung von $(X_1, ..., X_n)$ ist eine Familie $p = (p(x_1, ..., x_n))_{x_1 \in W_1, ..., x_n \in W_n}$, wobei jedes Mitglied definiert ist durch
$$p(x_1, ..., x_n) = \mathbb{P}[X_1 = x_1, ..., X_n = x_n]$$

Satz Eine gemeinsame Verteilung von Zufallsvariablen $X_1, ..., X_n$ erfüllt
$$\sum_{x_1 \in W_1, ..., x_n \in W_n} p(x_1, ..., x_n) = 1$$
Ausserdem ist
$$\mathbb{P}[X = x] = \sum_y \mathbb{P}[X = x, Y = y]$$

Randverteilung Seien $X_1, ..., X_n$ diskrete Zufallsvariablen mit gemeinsamer Verteilung $p = (p(x_1, ..., x_n))_{x_1 \in W_1, ..., x_n \in W_n}$. für jedes i gilt
$$\forall z \in W_i \quad \mathbb{P}[X_i = z] = \sum_{x_j, j \neq i} p(x_1, ..., x_{i-1}, z, x_{i+1}, ..., x_n)$$

Erwartungswert des Bildes Seien $X_1, ..., X_n$ diskrete Zufallsvariablen mit gemeinsamer Verteilung $p = (p(x_1, ..., x_n))_{x_1 \in W_1, ..., x_n \in W_n}$. Sei $\phi : \mathbb{R}^n \to \mathbb{R}$, dann gilt
$$\mathbb{E}[\phi(X_1, ..., X_n)] = \sum_{x_1, ..., x_n} \phi(x_1, ..., x_n) p(x_1, ..., x_n)$$

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

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

Gemeinsame Dichte (stetig) Die Zufallsvariablen $X_1, ..., X_n : \Omega \to \mathbb{R}$ besitzen eine stetige gemeinsame Verteilung besitzen, falls eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}_+$ existiert, sodass
$$\mathbb{P}[X_1 \le a_1, ..., X_n \le a_n] = \int_{-\infty}^{a_1} \ldots \int_{-\infty}^{a_n} f(x_1, ..., x_n) dx_n \ldots dx_1$$
für jedes $a_1, ..., a_n \in \mathbb{R}$ gilt. Obige Abbildung f nennen wir gerade gemeinsame Dichte von $X_1, ..., X_n$.

Es gilt
$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, ..., x_n) dx_n \ldots dx_1 = 1$$
Intuitiv beschreibt $f(x, y) dx dy$ die Wahrscheinlichkeit, dass ein Zufallspunkt (X,Y) in einem Rechteck $[x, x + dx] \times [y, y + dy]$ liegt.

Erwartungswert unter Abbildungen Sei $\phi : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung. Falls $X_1, ..., X_n$ eine gemeinsame Dichte f besitzen, dann lässt sich der Erwartungswert der Zufallsvariablen $Z = \phi(X_1, ..., X_n)$ mittels
$$\mathbb{E}[\phi(X_1, ..., X_n)] = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \phi(x_1, ..., x_n) f(x_1, ..., x_n) dx_n \ldots dx_1$$
berechnen.

Randverteilungen Falls X, Y eine gemeinsame Dichte $f_{X,Y}$ besitzt, dann gilt
$$\mathbb{P}[X \le a] = \mathbb{P}[X \in [-\infty, a], Y \in [-\infty, \infty]] = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx$$
Somit ist X stetig mit folgender Dichte
$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Unabhängigkeit stetiger Zufallsvariablen Seien $X_1, ..., X_n$ Zufallsvariablen mit Dichten $f_1, ..., f_n$. Dann sind folgende Aussagen äquivalent

1. $X_1, ..., X_n$ sind unabhängig
2. $X_1, ..., X_n$ sind insgesamt stetig mit gemeinsamer Dichte
$$f(x_1, ..., x_n) = f_1(x_1) \ldots f_n(x_n)$$

Falls $X_1, ..., X_n$ unabhängig sind, gilt:
$$\mathbb{P}[X_1 = x_1, ..., X_n = x_n] = \prod_{i \in [1, ..., n]} \mathbb{P}[X_i = x_i]$$

Bedingte Verteilungsfunktion Für zwei Z.V. ist die bedingte Verteilungsfunktion
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Bedingter Erwartungswert
$$\mathbb{E}[X|Y](y) = \int x f_{X|Y}(x|y)\, dx$$

4.1 Momente

***k*-tes Moment** Das k-te Moment einer Z.V. X wird definiert als
$$\mu_k = \mathbb{E}[(X - c)^k] = \int_{-\infty}^{\infty} (x - c)^k f(x)\, dx$$
Falls $c = 0$ ist es das k-te Raw Moment, wenn $c = \mathbb{E}[X]$ nennen wir es das k-te Central Moment.

Moment Generating Functions Für eine Z.V. X wird die MGF definiert als
$$M_X(t) = \mathbb{E}[e^{tX}], \quad t \in \mathbb{R}$$
Mit ableiten können wir Momente berechnen, da wir für $t = 0$ das k-te Moment $\mathbb{E}[X^k]$ erhalten:
$$\frac{d^k}{dt^k} M_X(t) = \mathbb{E}[\frac{d^k}{dt^k} e^{tX}] = \mathbb{E}[X^k e^{tX}]$$

Verteilung	MGF
Bernoulli	$(1-p)+pe^t$
Binomial	$(1-p+pe^t)^n$
Poisson	$e^{\lambda(e^t-1)}$
Geometrisch	$\frac{pe^t}{1-(1-p)e^t}$
Exponential	$\frac{\lambda}{\lambda-t}, \quad t<\lambda$
Normal	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$

5 Grenzwertsätze

Gesetz der grossen Zahlen Seien $X_1, X_2, ...$ unabhängige, identisch verteilte Z.V. mit $\mathbb{E}[|X_1|] < \infty$. Sei $m = \mathbb{E}[X_1]$, dann gilt

$$\lim_{n\to\infty} \frac{X_1 + ... + X_n}{n} = m$$

Zentraler Grenzwertsatz Sei $S_n = X_1 + \cdots + X_n$ für u.i.v. Z.V. X_i. Wenn der Erwartungswert $\mathbb{E}[X_1^2]$ wohldefiniert und endlich ist, und $m = \mathbb{E}[X_1]$ und $\sigma^2 = \mathrm{Var}[X_1]$, dann gilt für jedes a:

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

Der Satz besagt also, dass für grosse $n \in \mathbb{N}$ die Zufallsvariable

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

in Verteilung $Z \sim \mathcal{N}(0,1)$ ähnelt.

6 Schätzer

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

$$T = t(X_1, ..., X_n)$$

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

Einsetzen der Daten $x_1 = X_i(w), i \in \{1, ..., n\}$ liefert den Schätzwert $T(w) = t(X_1(w), ..., X_n(w))$ für den Parameter θ.

Erwartungstreu Ein Schätzer T heisst erwartungstreu ("unbiased") (für den Modellparameter), falls für alle $\theta \in \Theta$ gilt:

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

Interpretation: Im Durchschnitt (über alle möglichen Realisierungen w) schätzt T richtig, und zwar für jedes mögliche Modell $\mathbb{P}_\theta$.

Bias Sei $\theta \in \Theta$ und T ein Schätzer. Der Bias (erwarteter Schätzfehler) von T im Modell $\mathbb{P}_\theta$ ist definiert als

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

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

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

Prop. (MSE = Varianz von T + Bias)

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

Bemerkung MSE ist ein Kriterium für die Qualität eines Schätzers.

Maximum-Likelihood-Estimator (MLE)

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

Falls uiv.

$$L(x_1, ..., x_n) \coloneqq \begin{cases} \prod_{i=1}^n p(x_i; \theta) & \text{diskret} \\ \prod_{i=1}^n f(x_i; \theta) & \text{stetig} \end{cases}$$

Dann gilt für n Samples von Bernoulli verteilten Daten mit s mal 1 und $(n-s)$ mal 0:

$$T = \arg\max_\theta L(x_1, ..., x_n; \theta) = \arg\max_\theta \log(L(x_1, ..., x_n; \theta))$$

$$\implies \theta = \frac{s}{n} = \frac{1}{n}\sum_{i=1}^n x_1$$

Im Allg. finden wir θ indem wir L (also die gemeinsame Dichte) nach θ ableiten und das Minimum wählen.

Seien $X_1, ..., X_n$ uiv. $\sim \mathcal{N}(\mu, v)$ unter P_θ.

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

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi v}} \cdot e^{-\frac{(x_1-\mu)^2}{2v}}$$

$$T = \arg\max_\theta \log(L) = \sum_{i=1}^n \left(\log(\frac{1}{2\pi v}) - \frac{(x_i - \mu)^2}{2v}\right)$$

$$\implies \mu = \frac{1}{n}\sum_{i=1}^n x_i, \qquad v = \sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$$

Um einen Erwartungstreuen Schätzer zu erhalten skalieren wir mit $\frac{n}{n-1}$:

$$\mathbb{E}\left[\frac{n}{n-1}\cdot T\right] = \mathrm{Var}[X_1]$$

7 Konfidenzintervalle

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

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

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

Konfidenzintervall wenn Var **bekannt** Seien $X_1, ..., X_n$ u.i.v. normalverteilt mit m und $\sigma^2 = 1$. Der ML-Schätzer ist gegeben durch

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

Wir suchen nun für m Konfidenzintervalle der Form

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

wobei $C > 0$ von n unabhängig. Wir betrachten

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

wobei $Z = \sqrt{n}(T - m) = \frac{X_1 + ... + X_n - nm}{\sqrt{n}}$ standardnormalverteilt ist. Somit gilt:

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

Um ein Konfidenzintervall für m mit Niveau 95% zu erhalten wählen wir $c = 1.96$ (weil $2\Phi(1.96) - 1 \ge 0.95$):

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

χ^2-verteilt Eine stetige Z.V. X heisst χ^2-verteilt mit m Freiheitsgraden, falls ihre Dichte gegeben ist durch

$$f_X(y) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} \cdot y^{\frac{m}{2}-1} \cdot e^{-\frac{1}{2}y} \mathbb{1}_{y\ge 0}$$

Wir schreiben $X \sim \chi_m^2$.

Falls $m = 2$, so ist X exponential-verteilt:

$$F_X(y) = \frac{1}{2} e^{-\frac{1}{2}y} \mathbb{1}_{y \ge 0}$$

$$\implies X \sim \exp(\frac{1}{2})$$

Gamma-Funktion Γ ist die Gamma-Funktion mit

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

wobei für natürliche Zahlen n gilt

$$\Gamma(n) = (n-1)!$$

Satz χ^2-verteilt Für Z.V. $X_1, ..., X_n$ u.i.v. mit $X_1 \sim \mathcal{N}(0,1)$, so ist die Summe

$$Y := \sum_{i=1}^m X_i^2 \sim \chi_m^2$$

***t*-verteilt** Eine stetige Z.V. X heisst t-verteilt mit m Freiheitsgraden, falls ihre Dichte gegeben ist durch

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

Wir schreiben $X \sim t_m$.

Bemerkungen:

- Für $m \to \infty$ erhält man asymptotisch eine $\mathcal{N}(0,1)$-Verteilung.
- symmetrisch um 0

Satz *t*-verteilt Für X, Y unabhängig mit $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi_m^2$, ist der Quotient

$$Z := \frac{X}{\sqrt{\frac{Y}{m}}}$$

t-verteilt mit m Freiheitsgraden, also $Z \sim t_m$.

Strichprobenmittel und -varianz Für ZVen $X_1, ..., X_N$ sind Strichprobenmittel und Stichprobenvarianz definiert durch

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

Für $X_1, ..., X_n$ u.i.v. $\sim \mathcal{N}(\mu, \sigma^2)$ sind $\overline{X}_n$ und S^2 unabhängig.

Normalverteilung mit μ und σ^2 unbekannt Wir verwenden Stichprobenmittel und -varianz und setzen je ein Intervall um diese Schätzer. Zuerst für μ. Wir wollen erreichen, dass gilt

$$1 - \alpha \leq \mathbb{P}_\theta\left[[\overline{X}_n - \ldots, \overline{X}_n + \ldots] \ni \mu\right] = \mathbb{P}_\theta\left[|\overline{X}_n - \mu| \leq \ldots\right]$$

Für jedes $\theta \in \Theta$ ist

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

also wollen wir

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

Um ein möglichst kurzes Intervall zu erhalten, erfüllen wir diese Bedingung mit Gleichheit und brauchen dann gerade

$$\frac{\ldots}{S/\sqrt{n}} = t_{n-1, 1-\frac{\alpha}{2}}$$

Also erhalten wir als Konfidenzintervall für μ zum Niveau $1 - \alpha$

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

Um ein Konfidenzintervall für σ^2 zu konstruieren benutzen wir die Tatsache, dass

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

Also ist mit der Notation $\chi_{m,\gamma}^2$ für das γ-Quantil einer χ_m^2-Verteilung

$$\begin{aligned} 1 - \alpha &= \mathbb{P}_\theta\left[\chi_{n-1,\frac{\alpha}{2}}^2 \leq \frac{1}{\sigma^2}(n-1)S^2 \leq \chi_{n-1,1-\frac{\alpha}{2}}^2\right] \\ &= \mathbb{P}_\theta\left[\frac{(n-1)S^2}{\chi_{n-1,1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1,\frac{\alpha}{2}}^2}\right] \end{aligned}$$

und das Konfidenzintervall für σ^2 zum Niveau $1 - \alpha$ wird

$$C(X_1, ..., X_n) = \left[\frac{(n-1)S^2}{\chi_{n-1,1-\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{n-1,\frac{\alpha}{2}}^2}\right]$$

7.1 Appriximative Konfidenzintervalle

Oft ist ein Schätzer T eine Funktion einer Summe $\sum_{i=1}^n Y_i$ wobei die Y_i im Modell P_θ i.i.d. sind. Nach dem zentralen Grenzwertsatz ist für grosse n

$$\sum_{i=1}^n Y_i \text{ approx. normalverteilt unter } \mathbb{P}_\theta$$

mit Parametern $\mu = n\mathbb{E}_\theta[Y_i]$ und $\sigma^2 = n\mathrm{Var}_\theta[Y_i]$.

8 Tests

Null- und Alternativhypothese Wir wollen eine Entscheidung zwischen zwei Modellklassen treffen, der Nullhypothese $\Theta_0 \subseteq \Theta$ und der Alternativhypothese $\Theta_A \subseteq \Theta$ wobei $\Theta_0 \cap \Theta_A = \emptyset$.

Explizit formuliert ist also die Nullhypothese H_0: "der wahre (aber unbekannte) Parameter θ liegt in der Menge Θ_0.

Eine starke Aussage kann erreicht werden, wenn H_0 verworfen wird.

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

- T (Teststatistik) eine Zufallsvariable der Form $T = t(X_1, ..., X_n)$ ist
- $K \subseteq \mathbb{R}$ (kritischer oder Verwerfungsbereich) eine (deterministische) Teilmenge von $\mathbb{R}$ ist

Bei einem Test wird die Nullhypothese H_0 entweder verworfen (falls $T(\omega) \in K$) oder angenommen (falls $T(\omega) \notin K$).

Fehler 1. Art H_0 verworfen obwohl richtig

$$\mathbb{P}_\Theta[T \in K] \text{ für } \theta \in \Theta_0$$

Fehler 2. Art H_0 akzeptiert obwohl falsch

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

Signifikanzniveau Ein kleines Signifikanzniveau bedeutet eine geringe Chance auf Fehler 1. Art.

Sei $\alpha \in (0,1)$. Ein Test (T, K) besitzt Signifikanzniveau α, falls

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

Macht Eine grosse Macht bedeutet eine geringe Wahrscheinlichkeit für Fehler 2. Art.

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

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

Likelihood Quotient Seien $\theta_0 \neq \theta_A$ zwei fixierte Zahlen, und $X_1, ..., X_n$ entweder diskret oder gemeinsamstetig unter $\mathbb{P}_{\theta_0}$ und unter $\mathbb{P}_{\theta_A}$. Somit ist nach Annahme auch die Likelihood-Funktion $L(x_1, ..., x_n; \theta)$ für θ_0 und θ_A wohldefiniert.

Für jedes $x_1, ..., x_n$ definieren wir den Likelihood-Quotienten durch

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

Falls $L(x_1, ..., x_n; \theta_0) = 0$, setzen wir $R(x_1, ..., x_n) = +\infty$.

Intuitiv gilt für einen grosse Quotienten, dass die Beobachtungen $x_1, ..., x_n$ als Resultate im Modell $\mathbb{P}_{\theta_A}$ deutlich wahrscheinlicher sind als im Modell $\mathbb{P}_{\theta_0}$. Die Daten widersprechen $\mathbb{P}_{\theta_0}$ im Vergleich zu $\mathbb{P}_{\theta_A}$

LQ-Test Sei $c \geq 0$ der Likelihood-Quotient-Test (LQ-Test) mit Param c ist definiert durch

- $T = R(x_1, ..., x_n)$
- $K = (c, \infty]$

Lemma Neyman-Pearson Sei $c \geq 0$. Sei (T, K) ein LQ-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

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

Intuitiv ist jeder LQ-Test in dem Sinn optimal, dass jeder andere Test mit kleinerem Signifikanzniveau auch eine kleinere Macht hat.

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

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

Ein Beispiel ist der rechtsseitige Test $K_t = (t, \infty)$.

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 Zufallsvariable

$$p\text{-Wert} = G(T)$$

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

Anmerkungen:

- Der p-Wert hängt direkt von den anfänglichen Beobachtungen $X_1, ..., X_n$ ab.
- Für einen p-Wert mit Wert p gilt, dass alle Tests mit Signifikanzniveau $\alpha > p$ die Nullhypothese H_0 verwerfen würden und alle Tests mit Signifikanzniveau $\alpha \leq p$ die Nullhypothese H_0 nicht verwerfen würden.
- Der p-Wert ist nur von der Nullhypothese (und nicht der Alternativhypothese) abhängig.

Intuition: p-Wert klein $\implies$ H_0 wird wahrscheinlich verworfen. Aliter: Der p-Wert beschreibt die Wahrscheinlichkeit der beobachteten Daten, angenommen dass H_0 wahr ist. Ein kleiner Wert zeigt, dass die beobachteten Daten eher unwahrscheinlich sind.

9 Rezepte

Erwartungswert einer Summe

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

Unabhängigkeit prüfen

- Prüfe $I \subseteq \{1, ..., n\} : \mathbb{P}[\bigcap_{i \in I} X_i] = \prod_{i \in I} \mathbb{P}[X_i]$
- Prüfe die Randdichte $f(x_1, ..., x_n) = f(x_1) \cdot ... \cdot f(x_n)$
- Prüfe Verteilungsfunktion $F(x_1, ..., x_n) = F(x_1) \cdot ... \cdot F(x_n)$

Standard-Normalverteilt Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann ist

$$\phi(X) := \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Standardnormalverteilungstabelle Für $Z \sim \mathcal{N}(0, 1)$ gilt:

$$\Phi(c) = \mathbb{P}[Z \leq c] = \mathbb{P}[Z \geq -c] = 1 - \mathbb{P}[Z \leq -c] = 1 - \Phi(-c)$$

Beispiel: Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, $a \in \mathbb{R}$. Finde b in

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

1. X normalisieren. $\mathbb{P}[X \leq b] = \mathbb{P}[\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}] = a$
2. $\frac{b-\mu}{\sigma} = \Phi^{-1}(a)$
3. $b = \Phi^{-1}(a) \cdot \sigma + \mu$

MLE-Schätzer finden

1. (Gemeinsame) Dichtefunktion finden
2. Log-Likelihood Funktion definieren

$$\log L(x_1, ..., x_n; \theta) = \log f(x_1, ..., x_n; \theta)$$

$$\text{falls u.i.v.: } = \prod_{i=1}^{n} f(x_i; \theta)$$

3. Ableiten nach θ und Nullsetzen $\implies \theta^*$
4. Prüfen ob zweite Ableitung an der Stelle θ^* strikt kleiner als 0

9.1 Tests

z-Test Nomalverteilt, μ unbekannt, σ^2 bekannt.

1. Modell: $X_1, ..., X_n \sim \mathcal{N}(\theta, \sigma^2)$ uiv. unter $\mathbb{P}_\theta$
2. Hypothesen: $H_0 : \theta = \theta_0$ und $H_A : \theta\, [op]\, \theta_0$, $op \in \{>, <, \neq\}$
3. Teststatistik: $T = \frac{\overline{X}_n - \theta_0}{\sqrt{\sigma^2/n}}$
4. Verteilung Statistik: $T \sim \mathcal{N}(0, 1)$ unter $\mathbb{P}_{\theta_0}$
5. Verwerfungsbereich:
   a) $K_> = (c_>, \infty)$.
   $\mathbb{P}_{\theta_0}[T \in K_>] = \mathbb{P}_{\theta_0}[T > c_>] = 1 - \mathbb{P}_{\theta_0}[T \leq c_>] = \alpha \implies c_> = \Phi^{-1}(1 - \alpha) = -\Phi^{-1}(\alpha)$
   b) $K_< = (-\infty, c_<)$.
   $\mathbb{P}_{\theta_0}[T \in K_<] = \mathbb{P}_{\theta_0}[T < c_<] = 1 - \mathbb{P}_{\theta_0}[T \leq c_<] = \alpha \implies c_< = \Phi^{-1}(\alpha) = -\Phi^{-1}(1 - \alpha)$
   c) $K_{\neq} = (-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$.
   $\mathbb{P}_{\theta_0}[T \in K_{\neq}] = \mathbb{P}_{\theta_0}[T < -c_{\neq}] + \mathbb{P}_{\theta_0}[T > c_{\neq}] = 2 \cdot \mathbb{P}_{\theta_0}[T \leq c_{\neq}] = \alpha \implies c_{\neq} = \Phi^{-1}(1 - \frac{\alpha}{2}) = -\Phi^{-1}(\frac{\alpha}{2})$

t-Test Normalverteilt, μ unbekannt, σ^2 unbekannt.

1. Modell: $X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ uiv. unter $\mathbb{P}_\theta$
2. Hypothesen: $H_0 : \theta = \vec{\theta_0} = \{\mu_0\} \times (0, \infty)$ und H_A ähnlich wie oben
3. Teststatistik: $T = \frac{\overline{X}_n - \theta_0}{\sqrt{S^2/n}}$
4. Verteilung Statistik: $T \sim t_{n-1}$ unter $\mathbb{P}_{\theta_0}$
5. Verwerfungsbereich: wie oben, wobei $c_> = t_{n-1,1-\alpha}$, $c_< = t_{n-1,\alpha} = -t_{n-1,1-\alpha}$, $c_{\neq} = t_{n-1,1-\frac{\alpha}{2}}$

Gepaarter Zweistichprobentest Normalverteilt, μ_x, μ_y unbekannt, σ^2 gleich und bekannt.

1. Modell: $X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ uiv. unter $\mathbb{P}_\theta$ und $Y_1, ..., Y_n \sim \mathcal{N}(\mu, \sigma^2)$ uiv. unter $\mathbb{P}_\theta$. Sei $Z_i = X_i - Y_i \sim \mathcal{N}(\mu_z = \mu_x - \mu_y, 2\sigma^2)$
2. Hypothesen: $H_0 : \mu_z = 0$ und H_A ähnlich wie beim z-Test
3. Teststatistik: $T = \frac{(\overline{X}_n - \overline{Y}_n) - (\mu_x - \mu_y)}{\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}}$
4. Verteilung Statistik: $T \sim \mathcal{N}(0, 1)$ unter $\mathbb{P}_{\theta_0}$
5. Verwerfungsbereich: wie beim z-Test

Gepaarter Zweistichprobentest Normalverteilt, μ_x, μ_y unbekannt, σ^2 gleich und unbekannt.

1. Modell: $X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ uiv. unter $\mathbb{P}_\theta$ und $Y_1, ..., Y_n \sim \mathcal{N}(\mu, \sigma^2)$ uiv. unter $\mathbb{P}_\theta$. Sei $Z_i = X_i - Y_i \sim \mathcal{N}(\mu_z = \mu_x - \mu_y, 2\sigma^2)$
2. Hypothesen: $H_0 : \mu_z = 0$ und H_A ähnlich wie beim z-Test
3. Teststatistik: $T = \frac{(\overline{X}_n - \overline{Y}_n) - (\mu_x - \mu_y)}{\sqrt{S^2(\frac{1}{n} + \frac{1}{m})}}$
4. Verteilung Statistik: $T \sim t_{n-1}$ unter $\mathbb{P}_{\theta_0}$
5. Verwerfungsbereich: wie beim t-Test

Bernoulliverteilte Testergebnisse

1. Modell: $X_1, ..., X_n \sim \text{Ber}(p)$ uiv. unter $\mathbb{P}_p$, p unbekannt
2. Hypothesen: $H_0 : p = p_0$, $H_A : p \neq p_0$ (ähnlich wie z-Test)
3. Teststatistik: $T = \sum_{i=1}^{n} X_i$
4. Verteilung Statistik: $T \sim \text{Bin}(n, p_0)$ unter $\mathbb{P}_{p_0}$
5. Verwerfungsbereich: $\mathbb{P}_{p_0}[T \in K] = \mathbb{P}_{p_0}[T > k] \leq 2.5\% \iff \mathbb{P}_{p_0}[T \leq k] \geq 97.5\%$

p-Wert berechnen Gegeben Daten $x_1, ..., x_n$, ein Schätzer T und eine einfache Nullhypothese $H_0 : \theta = \theta_0$. Wir definieren Test (T, K) mit $K_t = (t, \infty)$ für rechtsseitigen Test, $K_t = (-\infty, -t)$ für linksseitig oder $K_t = (-\infty, -t) \cup (t, \infty)$ für beidseitig wobei $t = T(x_1, ..., x_n)$. Der p-Wert ist definiert als:

$$\mathbb{P}_{\theta_0}[T \in K_t]$$

Macht berechnen Für den Rechtsseitigen Test gilt z.B.:

$$\mathbb{P}_{\theta_A}[T > c_>] = \mathbb{P}_{\theta_A}\left[\frac{\overline{X}_n - \mu_0}{\sqrt{\sigma^2/n}} > c_>\right]$$

$$= \mathbb{P}_{\theta_A}\left[\frac{\overline{X}_n - \mu_0}{\sqrt{\sigma^2/n}} + \frac{\mu_0 - \mu_A}{\sqrt{\sigma^2/n}} > c_> + \frac{\mu_0 - \mu_A}{\sqrt{\sigma^2/n}}\right]$$

10 Aufgaben

Konfidenzintervall für $\mathcal{N}(\mu, \sigma^2)$ Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit σ bekannt.

a) Formel für 95%-Konfidenzintervall für μ nach n Stichproben: Sei $\overline{X}$ ein Schätzer für X. Wir wissen, dass

$$\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Also gilt

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right)$$

wobei $z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$. Durch Umformen erhalten wir

$$\mathbb{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

Also hat das Vertrauensintervall die Form

$$\overline{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

b) Berechne das realisierte Konfidenzintervall für beobachteten Stichprobenmittelwert $\overline{x}$:

$\overline{x}$ für $\overline{X}$ einsetzen, und $z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2} = 1.96$ für $\alpha = 0.05$.

c) Berechne Anzahl Stichproben, um Konfidenzintervall kleiner als m zu erhalten:

Folgendes nach n auflösen:

$$2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq m$$

Approx. Konfidenzintervall für Binom. Sei N Anz. Fische, zuerst 500 gefangen und markiert und anschliessend 200 gefangen und Anz. $X \sim \text{Bin}(n = 200, \theta)$ markierte Fische gezählt. θ ist die W'keit, dass ein gezogener Fisch markiert ist. Wir haben

$$\theta = \frac{\text{Anz. markierte Fische}}{\text{Totale Anz. Fische}} = \frac{500}{N}$$

.

Wir schätzen θ durch $T = \frac{X}{n}$, bei einer Beobachtung von $\overline{X} = 40$ ist dann $T^{(\theta)}(\omega) = X(\omega)/n = 40/200 = 1/5$. Durch Umformen erhalten wir $T^{(N)} = \frac{500}{T^{(\theta)}} = \frac{500n}{X}$. Realisiert gibt das

$$T^{(N)}(\omega) = \frac{500}{T^{(\theta)}(\omega)} = 2500$$

Bestimme nun approx. Konfidenzintervall für θ und N, je mit Level 95%:

Da $X \sim \text{Bin}(n, \theta)$ mit $\mathbb{E}[X] = n\theta$ und $\text{Var}[X] = n\theta(1-\theta)$, ist

$$Y = \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \sim \mathcal{N}(0, 1)$$

für $n \to \infty$ nach dem Zentralen Grenzwertsatz. Für $\theta \in [0, 1]$ gilt dann:

$$\mathbb{P}_\theta\left[-1.96 \leq \frac{X - n\theta}{\sqrt{n/4}} \leq 1.96\right]$$

$$\geq \mathbb{P}_\theta\left[-1.96 \leq \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \leq 1.96\right] \geq 0.95$$

weil für alle $\theta \in [0, 1]$ gilt $\theta(1 - \theta) \leq 1/4$.

Anschliessend auflösen, bis nur noch θ in der Mitte steht und den realisierten Wert einsetzen.

Verwenden, dass $N = \frac{500}{\theta}$ und Konfidenzintervall-Grenzen anpassen.

Normalverteilung bei fixem Mittelwert $X_1, ..., X_n \sim \mathcal{N}(0, \sigma^2)$ u.i.v. unter $\mathbb{P}_\theta$, wobei $\theta = \sigma^2 > 0$. $T_{ML} = \frac{1}{n} \sum i = 1nX_i^2$ ist ein erwartungstreuer Schätzer.

Da $\frac{X_i}{\sigma} \sim \mathcal{N}(0,\ 1)$ ist

$$Z = \frac{n}{\sigma^2} T_{ML} = \sum i = 1n\left(\frac{X_i}{\sigma}\right)^2 \sim \chi_n^2 \textit{ unter } \mathbb{P}_\theta$$

Wir suchen Konfidenzintervall

$$[A,\ B] = \left[\frac{nT_{ML}}{c_1},\ \frac{nT_{ML}}{c_2}\right]$$

wobei $c_1,\ c_2 > 0$.

Für $\theta \in \Theta$ beliebig

$$\mathbb{P}_\theta\left[\frac{nT_{ML}}{c_1} \leq \sigma^2 \leq \frac{nT_{ML}}{c_2}\right] = \mathbb{P}_\theta[c_2 \leq Z \leq c_1]$$

11 Anhang

Reihen

$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ (Geom.)	$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$ (Teleskop)
$\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$ (Exp.)	$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$
$\sum_{k=1}^{\infty} n \cdot q^n = \frac{q}{(1-q)^2}$	$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$
$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$	$\sum_{k=1}^{n} z^k = \frac{1-z^{n+1}}{1-z}$

11.1 Identitäten

p-q-Formel $x^2 + p \cdot x + q = 0$ $\left(p = \frac{b}{a},\ q = \frac{c}{a}\right)$

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$$

Mitternachtsformel: $x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Normalform Kreisgleichung: $r^2 = (x - m_1)^2 + (y - m_2)^2$

mit Mittelpunkt (m_1, m_2)

Wurzeln:

$\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{a \cdot b}$

$(\sqrt[n]{a})^m = \sqrt[n]{a^m}$

$\sqrt[m]{\sqrt[n]{a}} = \sqrt[m \cdot n]{a}$

Logarithmus:

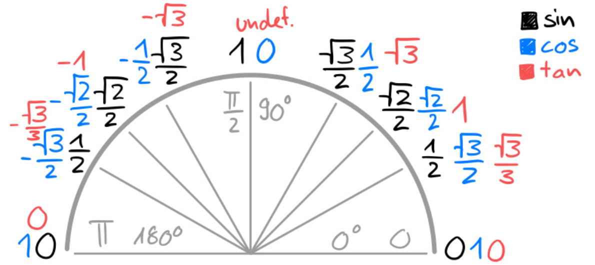
$\log_a(u \cdot v) = \log_a(u) + \log_a(v)$

$\log_a(\frac{u}{v}) = \log_a(u) - \log_a(v)$

$\log_a(u^v) = v \cdot \log_a(u)$

$\log_a(\sqrt[n]{u}) = \frac{1}{n} \cdot \log_a(u)$

$\log_b(u) = \frac{\log_a(u)}{\log_a(b)}$



Binomialkoeffizient: $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

Stirling's formula: $n! \approx \frac{\sqrt{2\pi n} \cdot n^n}{e^n}$

Inverse $f^{-1}(y)$ berechnen:

$f(x) = y = \ln(x - 17) + 2 \Longleftrightarrow f^{-1}(y) = x = e^{y-2} + 17$

- $e^{ix} = \cos(x) + i\sin(x)$
- $\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}$
- $\cos(x) = \frac{e^{ix} + e^{-ix}}{2}$
- $\sinh(x) = \frac{e^x - e^{-x}}{2} = \frac{e^{2x} - 1}{2e^x}$
- $\cosh(x) = \frac{e^x + e^{-x}}{2} = \frac{e^{2x} + 1}{2e^x}$
- $\int\limits_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$
- $\sin^2(x) + \cos^2(x) = 1$
- $\tan(x) = \frac{\sin(x)}{\cos(x)}$
- $\sin(x \pm y) = \sin(x) \cdot \cos(y) \pm \cos(x) \cdot \sin(y)$
- $\cos(x \pm y) = \cos(x) \cdot \cos(y) \mp \sin(x) \cdot \sin(y)$
- $\sin(2x) = 2\sin(x)\cos(x)$
- $\cos(2x) = \cos^2(x) - \sin^2(x)$

11.2 Integration

Trick 1: Linearität

Stellt meist einen ersten Schritt dar, um das Problem auf kleinere Probleme zu reduzieren. Wir nutzen dabei aus, dass

$$\int_a^b \alpha f(x) + \beta g(x)\ dx = \alpha \int_a^b f(x)\ dx + \beta \int_a^b g(x)\ dx$$

Beispiel:

$$\int_{-1}^{1} \frac{(6x^2+3)\sin(x^3)}{2}\ dx = \int_{-1}^{1} 3x^2 \sin(x^3)\ dx + \frac{3}{2}\int_{-1}^{1} \sin(x^3)\ dx$$

Trick 2: Substitution

$$f(g(x)) + c = \int g'(x) f'(g(x))\ dx$$

$$\int f(y)\ dy = \int g'(x) f(g(x))\ dx$$

$$\int_{g(a)}^{g(b)} f(y)\ dy = \int_a^b g'(x) f(g(x))\ dx$$

Wichtig: Die Grenzen müssen ebenfalls substituiert werden!

Beispiel: berechne $\int_0^2 x \cdot e^{x^2} dx$. Wir substituieren $u = x^2$. es gilt $\frac{du}{dx} = 2x \iff dx = \frac{du}{2x}$.

$$\int_0^2 x \cdot e^{x^2} dx = \int_0^4 x \cdot e^u \frac{du}{2x} = \int_0^4 \frac{1}{2} e^u du$$
$$= \frac{1}{2} e^u \Big|_0^4 = \frac{1}{2}(e^4 - 1)$$

Beispiel: ($u = 1 + x^3$, $\frac{du}{dx} = 3x^2$, $dx = \frac{du}{3x^2}$)

$$\int_0^2 x^2 \log(1+x^3)\ dx = \int_1^9 x^2 \log u\ \frac{du}{3x^2} = \frac{1}{3}\int_1^9 \log u\ du$$
$$= \frac{1}{3}\left[x(\log x - 1)\right]_1^9 = ... = 3\log 9 - \frac{8}{3}$$

Beispiel: ($u = x^2$, $\frac{du}{dx} = 2x$, $dx = \frac{du}{2x}$)

$$\int_0^1 2x^3 x^{x^2}\ dx = \int_0^1 2xue^u\ \frac{du}{2x} = \int_0^1 ue^u\ du$$
$$\overset{\text{PI}}{=} \left[ue^u\right]_0^1 - \int_0^1 e^u\ du = \left[ue^u\right]_0^1 - \left[e^u\right]_0^1 = e - (e-1) = 1$$

Trick 3: Partielle Integration

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

$$\int_a^b f'(x) g(x)\ dx = f(b)g(b) - f(a)g(a) - \int_a^b f(x) g'(x)\ dx$$

Beispiel:

$$\int \log x\ dx = \int 1 \cdot \log x\ dx = x \log x - \int x \cdot \frac{1}{x}\ dx$$
$$= x \log x - x + c$$

Trick 4: ungerade Funktionen

Sei $f(x) : [-a, a] \to \mathbb{R}$ stetig und ungerade, d.h. $f(-x) = -f(x)$. Dann gilt:

$$\int_{-a}^{a} f(x)\ dx = 0$$

Trick 5: Partialbruchzerlegung

$$\frac{dx+f}{ax^2+bx+c} = \frac{A}{Bx+C} + \frac{D}{Ex+F}$$

Bei doppelten Nullstellen (z.B. $(x+1)^2$) folgende Nenner wählen: $x+1$ und $(x+1)^2$.

Hat der Nenner Grad $n > 1$, für den Zähler ein Polynom mit Grad $n-1$ wählen: $(x^2+1) \implies \frac{ax+b}{x^2+1}$

Wenn der Zähler grösseren Grad hat als der Nenner: Polynomdivision (Zähler durch Nenner):

$$\frac{x^3+2x^2-4x+2}{(x-1)(x+3)} = \mathbf{x} + \frac{\mathbf{(2\text{-}x)}}{(x-1)(x+3)}$$
$$\text{weil } x^3+2x^2-4x+2 = \mathbf{x} \cdot (x-1)(x+3) + \mathbf{(2\text{-}x)}$$

Komplexe Nullstelle: $x^2+px+q \implies \frac{Ax+B}{x^2+px+q}$

Beispiel:

$$\frac{1}{x^2-1} = \frac{1}{(x-1)(x+1)} = \frac{a}{x-1} + \frac{b}{x+1}$$
$$= \frac{a(x+1)+b(x-1)}{(x-1)(x+1)}$$
$$\implies a(x+1) + b(x-1) = 1$$
$$\implies a = \frac{1}{2}, b = -\frac{1}{2} \text{ (Nullstellen einsetzen)}$$

Beispiel:

$$\frac{9x^3-3x+1}{x^3-x^2} = 9 + \frac{9x^2-3x+1}{x^2(x-1)} \text{ (Polynomdivision)}$$
$$\implies \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x-1} = \frac{9x^2-3x+1}{x^2(x-1)} \text{ (Nullstellen)}$$
$$\implies A(x)(x-1) + B(x-1) + C(x^2) = 9x^2 - 3x + 1$$
$$\implies B = -1, C = 7 \implies A = 2$$

11.3 Differentials and Derivatives

f'(x)	**f**(x)	**F**(x)
$\cosh(x)$	$\sinh(x)$	$\cosh(x)$ + C
$\cos(x)$	$\sin(x)$	$-\cos(x)$
$2\sin(x)\cos(x)$	$\sin^2(x)$	$\frac{1}{2}\left(x - \frac{1}{2}\sin(2x)\right)$
$-\sin(x)$	$\cos(x)$	$\sin(x)$
$-2\sin(x)\cos(x)$	$\cos^2(x)$	$\frac{1}{2}\left(x + \frac{1}{2}\sin(2x)\right)$
$\frac{1}{\cos^2(x)} = 1 + \tan^2(x)$	$\tan(x)$	$-\ln\lvert\cos(x)\rvert$
$x(\ln\lvert x\rvert - 1)$	$\ln\lvert x\rvert$	$\frac{1}{x}$

12.5 Diskrete Verteilungen

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

12.6 Stetige Verteilungen

Verteilung	Parameter	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	$f_X(t)$	$F_X(t)$
Gleichverteilung	$[a,b]$: Intervall	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$	$\frac{1}{b-a}$	$\frac{t-a}{b-a}$
Exponentialverteilung	$\lambda : \frac{1}{\mathbb{E}[X]}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\begin{cases}\lambda e^{-\lambda t} & t \ge 0\\ 0 & t < 0\end{cases}$	$\begin{cases}1-e^{-\lambda t} & t > 0\\ 0 & t \le 0\end{cases}$
Normalverteilung	σ^2: Varianz $\mu : \mathbb{E}[X]$	μ	σ^2	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad -\infty < t < \infty$	$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}\mathrm{d}y$
χ^2-Verteilung	n: Freiheitsgrad	n	$2n$	$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}t^{\frac{n}{2}-1}e^{-\frac{t}{2}} \quad t > 0$	$\mathrm{Gamma}\left(\frac{n}{2}, \frac{t}{2}\right)$
t-Verteilung	n: Freiheitsgrad	$\begin{cases}0 & n > 1\\ \text{undef.} & \text{sonst}\end{cases}$	$\begin{cases}\frac{n}{n-2} & n > 2\\ \infty & 1 < n \le 2\\ \text{undef.} & \text{sonst}\end{cases}$	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$	I'd rather not