Wahrscheinlichkeit und Statistik Prüfungsnotizen

Fabian Bösiger

30.06.2020

Inhalt

Wahrscheinlichkeiten

Grundbegriffe

Ereignisraum Ω: Menge aller möglichen elementaren Ereignissen.

Beispiel: Bei einem Würfelwurf sind die Elementarereignisse $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Potenzmenge $\mathcal{P}(\Omega)$ oder 2^Ω: Menge aller Teilmengen von Ω.

Klasse aller beobachtbaren Ereignisse $\mathcal{F}$: $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ und $\mathcal{F}$ ist eine σ-Algebra. Bei diskreten, d.h. endlichen bzw. abzählbaren Wahrscheinlichkeitsräumen wird $\mathcal{F} = \mathcal{P}(\Omega)$ gewählt.

σ-Algebra: $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ ist eine σ-Algebra, wenn gilt:

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3. $(A_n)_{n\in\mathbb{N}}, A_n \in \mathcal{F} \implies \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$

Beispiel: Jemand wirft einen Würfel und teilt uns mit, ob die gewürfelte Zahl gerade oder ungerade ist.

Wir könnten den Grundraum $\Omega_1 = \{G, U\}$ wählen mit G für gerade und U für ungerade. In diesem Fall wäre $\mathcal{F} = \{\emptyset, \Omega_1, \{G\}, \{U\}\}$.

Jedoch könnten wir auch den Grundraum $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$ wählen. Dann wäre $\mathcal{F} = \{\emptyset, \Omega_2, \{2, 4, 6\}, \{1, 3, 5\}\} \neq \mathcal{P}(\Omega_2)$, da beispielsweise das prinzipielle Ereignis $\{1\}$ nie beobachtbar ist.

Wahrscheinlichkeitsmass $P : \mathcal{F} \to [0, 1]$: $P[A] \in \mathcal{F}] \in [0, 1]$ ist die Wahrscheinlichkeit, dass A eintritt. Dabei muss gelten:

1. $\forall A \in \mathcal{F} : P[A] \geq 0$
2. $P[\Omega] = 1$
3. $P[\bigcup\limits_{i=1}^{\infty} A_i] = \sum\limits_{i=1}^{\infty} P[A_i]$, sofern die $A_i \in \mathcal{F}$ paarweise disjunkt sind.

Folgende Rechenregeln lassen sich herleiten:

1. $P[A^c] = 1 - P[A]$
2. $P[\emptyset] = 0$
3. Für $A \subseteq B$ gilt $P[A] \leq P[B]$
4. Additionsregel: $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

Diskrete Wahrscheinlichkeitsräume

Bei diskreten, d.h. endlichen bzw. abzählbaren Wahrscheinlichkeitsräumen gilt $\mathcal{F} = \mathcal{P}(\Omega)$ und $P[A] = \sum\limits_{w_i \in A} P[\{w_i\}]$.

Laplace-Raum Ω ist endlich und ich alle Elementarereignisse $\Omega = \{w_1, \ldots, w_N\}$ sind gleich wahrscheinlich mit $p_1 = \cdots = p_N = \frac{1}{N}$.

Diskrete Gleichverteilung In einem Laplace-Raum gilt für beliebige $A \subseteq \Omega$: $P[A] = \frac{|A|}{|\Omega|}$.

Beispiel: Beim zweimaligen Münzwurf ist $\Omega = \{KK, KZ, ZK, ZZ\}$, also $|\Omega| = 4$ und damit $p_i = \frac{1}{4}$. Dann ist $P[\text{Mindestens einmal Kopf}] = P[\{KK, KZ, ZK\}] = \frac{3}{4}$.

Bedingte Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeit Wahrscheinlichkeit, dass B eintritt, unter der Bedingung, dass A eintritt:

$$\boxed{P[B \mid A] := \frac{P[B \cap A]}{P[A]} = \frac{P[A \mid B]P[B]}{P[A]}}$$

Satz von der totalen Wahrscheinlichkeit (Satz 1.1) Sei $A_1, \dots, A_n$ eine Zerlegung von Ω in paarweise disjunkte Ereignisse, dann gilt für beliebiege Ereignisse B:

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

Formel von Bayes (Satz 1.2) Ist $A_1, \dots, A_n$ eine Zerlegung von Ω mit $P[A_i] > 0$ und B ein Ereignis mit $P[B] > 0$, so gilt für jedes k:

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

Unabhängigkeit

Stochastische Unabhängigkeit Zwei Ereignisse A, B heissen stochastisch unabhängig, falls $P[A \cap B] = P[A]P[B]$.

Allgemeiner: Zwei Ereignisse A, B heissen stochastisch unabhängig, wenn für jede endliche Teilfamile $\{k_1, \dots, k_m\} \subseteq \{1, \dots, n\}$ gilt, dass $P\left[\bigcap_{i=1}^{m} A_{k_i}\right] = \prod_{i=1}^{m} P[A_{k_i}]$.

Ist $P[A] = 0$ oder $P[B] = 0$, so sind A und B immer unabhängig.

Für $P[A] \neq 0$ gilt: A, B unabhängig $\iff$ $P[B \mid A] = P[B]$.

Diskrete Zufallsvariablen und Verteilungen

Grundbegriffe

Diskrete Zufallsvariable Funktion $X : \Omega \to \mathbb{R}$, $W(X)$: Wertebereich von X.

Indikatorfunktion Für jede Teilmente $A \subset \Omega$ gilt: $I_A(w) := \begin{cases} 1 & w \in A \\ 0 & w \in A^c \end{cases}$

Verteilungsfunktion $F_x : \mathbb{R} \to [0, 1]$, $F_X(t) := P[X \leq t] = P[\{w \mid X(w) \leq t\}]$.

Gewichtsfunktion $p_x : W(X) \to [0, 1]$, $p_X(x_k) := P[X = x_k] = P[\{w \mid X(w) = x_k\}]$.

Es gilt $F_X(t) = P[X \leq t] = \sum_{x_k \leq t} p_X(x_k)$

Erwartungswerte

Erwartungswert $E[X] := \sum_{x_k \in W(X)} x_k p_X(x_k)$. Es gilt:

1. Monotonie: Ist $X \leq Y$ (d.h. $\forall w : X(w) \leq Y(w)$), so gilt auch $E[X] \leq E[Y]$
2. Linearität: $E[aX + b] = aE[X] + b$
3. Falls $X \geq 0$, so gilt $E[X] = \sum_{j=1}^{\infty} P[X \geq j]$

Varianz $\text{Var}[X] := E[(X - E[X])^2]$. Es gilt:

1. $\text{Var}[X] = E[X^2] - (E[X])^2$
2. $\text{Var}[aX + b] = a^2\text{Var}[X]$
3. Für unabhängige X, Y gilt: $\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y]$
4. Für abhängige X, Y gilt: $\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\text{Cov}[X, Y]$

Standardabweichung $\sigma(X) = \sqrt{\text{Var}[X]}$.

Gemeinsame Verteilungen und unabhängige Zufallsvariablen

Gemeinsame Verteilungsfunktion $F : \mathbb{R}^n \to [0,1]$, $F(x_1, \ldots, x_n) := P[X_1 \leq x_1, \ldots, X_n \leq x_n]$

Gemeinsame Gewichtsfunktion $p : \mathbb{R}^n \to [0,1]$, $p(x_1, \ldots, x_n) := P[X_1 = x_1, \ldots, X_n = x_n]$

$F(x_1, \ldots, x_n) = P[X_1 \leq x_1, \ldots, X_n \leq x_n] = \sum\limits_{y_1 \leq x_1, \ldots, y_n \leq x_n} P[X_1 = y_1, \ldots, X_n = y_n] = \sum\limits_{y_1 \leq x_1, \ldots, y_n \leq x_n} p(y_1, \ldots, y_n)$

Randverteilung Verteilungsfunktion der Randverteilung von X: $F_X(x) := P[X \leq x] = P[X \leq x, Y < \infty] = \lim\limits_{y \to \infty} F(x, y)$

Gewichtsfunktion der Randverteilung von X: $p_X(x) := P[X = x] = \sum\limits_{y_j \in W(Y)} P[X = x, Y = y_j] = \sum\limits_{y_j \in W(Y)} p(x, y_j)$

Unabhängigkeit $X_1, \ldots, X_n$ heissen unabhängig, falls $F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$ beziehungsweise $p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$

X und Y sind unabhängig, wenn gilt $E[XY] = E[X]E[Y]$.

Funktionen von mehreren Zufallsvariablen

Linearität (Satz 2.4) Seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen mit endlichen Erwartungswerten. Sei $Y = a + \sum\limits_{l=1}^{n} b_l X_l$ mit Konstanten $a, b_1, \ldots, b_n$. Dann gilt $E[Y] = a + \sum\limits_{l=1}^{n} b_l E[X_l]$.

Kovarianz $\operatorname{Cov}(X_1, X_2) := E[X_1 X_2] - E[X_1]E[X_2]$

$\operatorname{Cov}(X, X) = Var[X]$

Unkorreliertheit X_1 und X_2 sind unkorreliert, wenn gilt $\operatorname{Cov}(X_1, X_2) = 0$.

Unabhängigkeit impliziert Unkorreliertheit, die andere Richtung folgt jedoch nicht.

Produkte von Zufallsvariablen (Satz 2.5) Seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen mit endlichen Erwartungswerten. Falls $X_1, \ldots, X_n$ unabhängig sind, so gilt: $E\left[\prod\limits_{i=1}^{n} X_i\right] = \prod\limits_{i=1}^{n} E[X_i]$. Ausserdem sind dann $X_1, \ldots, X_n$ unkorreliert und es gilt: $\operatorname{Var}\left[\prod\limits_{i=1}^{n} X_i\right] = \prod\limits_{i=1}^{n} \operatorname{Var}[X_i]$.

Bedingte Verteilungen

Bedingte Gewichtsfunktion Seien X and Y diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion $p(x, y)$. Die bedingte Gewichtsfunktion von X, gegeben dass $Y = y$, is te definiert durch $p_{X|Y}(x \mid y) := P[X = x \mid Y = y] = \frac{P[X=x, Y=y]}{P[Y=y]} = \frac{p(x,y)}{p_Y(y)}$ für $p_Y(y) > 0$ und 0 sonst.

Wichtige diskrete Verteilungen

Verteilung	Gewichtsfunktion $p_X(k) = P[X = k]$	Verteilungsfunktion $F_X(t) = P(X \leq t)$	Erwartungswert $E[X]$	Varianz $\operatorname{Var}[X]$
Diskrete Gleichverteilung	$\frac{1}{n}$	$\frac{\lfloor t \rfloor}{n}$	$\frac{a+b}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Bernoulli-Verteilung $X \sim \operatorname{Be}(p)$	$p^k(1-p)^{1-k}$	$1-p$	p	$p(1-p)$
Binomialverteilung $X \sim \operatorname{Bin}(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$		np	$np(1-p)$

Verteilung	Gewichtsfunktion $p_X(k) = P[X = k]$	Verteilungsfunktion $F_X(t) = P(X \leq t)$	Erwartungswert $E[X]$	Varianz $\mathrm{Var}[X]$
Geometrische Verteilung $X \sim \mathrm{Geom}(p)$	$p(1-p)^{k-1}$	$1-(1-p)^t$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negativbinomiale Verteilung $X \sim \mathrm{NB}(r,p)$	$\binom{k-1}{r-1}p^r(1-p)^{k-r}$		$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometrische Verteilung	$\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$		$m\frac{r}{n}$	$m\frac{r}{n}(1-\frac{r}{n})\frac{n-m}{n-1}$
Poisson-Verteilung $X \sim \mathcal{P}(\lambda)$	$e^{-\lambda}\frac{\lambda^k}{k!}$		λ	λ

Allgemeine Zufallsvariablen

Übersicht

	Diskrete Zufallsvariablen	Allgemeine Zufallsvariablen
Verteilungsfunktion	$F_X : \mathbb{R} \to [0,1]$ $F_X(t) := P[X \leq t] := P[\{w \mid X(w) \leq t\}] = \sum\limits_{x_k \leq t} p_X(x_k)$	$F_X : \mathbb{R} \to [0,1]$ $F_X(t) := P[X \leq t] := P[\{w \mid X(w) \leq t\}] = \int\limits_{-\infty}^{t} f_X(s)ds$
Gemeinsame Verteilungsfunktion	$F : \mathbb{R}^n \to [0,1]$ $F(x_1,\ldots,x_n) = P[X_1 \leq x_1,\ldots,X_n \leq x_n] = \sum\limits_{y_1 \leq x_1,\ldots,y_n \leq x_n} p(y_1,\ldots,y_n)$	$F : \mathbb{R}^n \to [0,1]$ $F(x_1,\ldots,x_n) = P[X_1 \leq x_1,\ldots,X_n \leq x_n] = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} f(t_1,\ldots,t_n)dt_n \cdots dt_1$
Monoton wachsend *Rechtsstetig*	$\forall s \leq t : F_X(s) \leq F_X(t)$ $\forall u > t, u \to t : F_X(u) \to F_X(t)$ $\lim\limits_{t\to\infty} F_X(t) = 1,\ \lim\limits_{t\to-\infty} F_X(t) = 0$	Analog Analog Analog
Verteilung	$\mu_X(B) := P[X \in B] = \sum\limits_{x_k \in B} p_X(x_k)$	$\mu_X(B) := P[X \in B] = \int\limits_B f_X(s)ds$
Randverteilung	$F_X(x) := P[X \leq x] = P[X \leq x, Y \leq \infty] = \lim\limits_{y\to\infty} F(x,y)$	Analog
Gewichtsfunktion, Dichtefunktion	$p_X(x_k) := P[X = x_k] = P[\{w \mid X(w) = x_k\}]$	$f_X(t) = \frac{d}{dt}F_X(t)$ $f_X \geq 0,\ f_X = 0$ ausserhalb von $W(X)$ $\int\limits_{-\infty}^{\infty} f_X(s)ds = 1$
Gemeinsame Gewichtsfunktion, Dichtefunktion	$p_X(x_1,\ldots,x_n) := P[X_1 = x_1,\ldots,X_n = x_n]$	$f(t_1,\ldots,t_n)$ $f(x_1,\ldots,x_n) \geq 0$, $f(x_1,\ldots,x_n) = 0$ ausserhalb von $W(X_1,\ldots,X_n)$ $\int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(x_1,\ldots,x_n)dx_n \cdots dx_1 = 1$
Bedingte Gewichtsfunktion	$p_{X\|Y}(x \mid y) := P[X = x \mid Y = y] = \frac{P[X=x,Y=y]}{P[Y=y]} = \frac{p(x,y)}{p_Y(y)}$	$f_{X\|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$
Randdichte		$f_X(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy$, $f_Y(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx$
Unabhängigkeit	$p(x_1,\ldots,x_n) = p_{X_1}(x_1)\cdots p_{X_n}(x_n)$	$f(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$

	Diskrete Zufallsvariablen	Allgemeine Zufallsvariablen
Erwartungswert	$E[X] = \sum_{-\infty}^{\infty} xp(x)$, für nichtnegative ganzzahlige Zufallsvariablen: $E[X] = \sum_{i=1}^{\infty} P[X \geq i]$	$E[X] = \int_{-\infty}^{\infty} xf(x)dx$
Bedingter Erwartungswert		$E[X \mid Y = y] = \int_{-\infty}^{\infty} xf_{X\mid Y=y}(y)dx = \int_{-\infty}^{\infty} x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx$, weiter vereinfachen falls X, Y unabhängig, siehe Unabhängigkeit
Varianz	$\text{Var}(X) = E[X^2] - E[X]^2$	Analog

Wichtige stetige Verteilungen

Verteilung	Dichtefunktion $f_X(t)$	Verteilungsfunktion $F_X(t)$	Erwartungswert $E[X]$	Varianz $\text{Var}[X]$
Gleichverteilung $X \sim \mathcal{U}(a, b)$	$\begin{cases} \frac{1}{b-a} & a \leq t \leq b \\ 0 & \text{sonst} \end{cases}$	$\begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ 1 & t > b \end{cases}$	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$
Exponentialverteilung $X \sim \text{Exp}(\lambda)$	$\begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$	$\begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normalverteilung $X \sim \mathcal{N}(\lambda, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(t-\mu)^2}{2\sigma^2})$		μ	σ^2

Standard-Normalverteilung Für die Standard-Normalverteilung $\mathcal{N}(0, 1)$ gilt $F_X(t) = \Phi(t)$. Es gilt $\Phi(-t) = 1 - \Phi(t)$.

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

Abhängige Zufallsvariablen (Satz 4.1) Sei X eine Zufallsvariable und $Y = g(X)$ eine weitere Zufallsvarable. Ist X stetig mit Dichte $f_X(x)$, so ist $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$.

Ungleichungen und Grenzwertsätze

Ungleichungen

Markov-Ungleichung Sei X eine Zufallsvariable und $g : W(X)- > [0, \infty)$ eine wachsende Funktion. Dann gilt für jedes $c \in \mathbb{R}$ mit $g(c) > 0$:

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

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

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

Oder äquivalänt dazu:

$$P[|Y - E[Y]| < b] \geq 1 - \frac{Var[Y]}{b^2}$$

Das Gesetz der grossen Zahlen

Schwaches Gesetz der grossen Zahlen Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit dem gleichen Erwartungswert $E[X_i] = \mu$ und der gleichen Varianz $\text{Var}[X_i] = \sigma^2$. Sei $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Dann konvergiert $\overline{X}_n$ für $n \to \infty$ stochastisch gegen $\mu = E[X_i]$, d.h.:

$$\boxed{\forall \epsilon > 0 : P[|\bar{X}_n - \mu| > \varepsilon] \xrightarrow{n\to\infty} 0}$$

Beweis mit Chebyshev-Ungleichung.

Starkes Gesetz der grossen Zahlen Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit dem gleichen **endlichen** Erwartungswert $E[X_i] = \mu$ und der gleichen Verteilung. Sei $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Dann konvergiert $\overline{X}_n$ für $n \to \infty$ fastsicher gegen $\mu = E[X_i]$, d.h.:

$$\boxed{P[\{\omega \in \Omega \mid \bar{X}_n(\omega) \xrightarrow{n\to\infty} \mu\}] = 1}$$

Der Zentrale Grenzwertsatz

Zentraler Genzwertsatz (Satz 5.5) Sei $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen mit $E[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$. Für die Summe $S_n = \sum_{i=1}^{n} X_i$ gilt dann:

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

S_n hat den Erwartungswert $E[S_n] = n\mu$ und Varianz $\text{Var}[S_n] = n\sigma^2$. Also ist $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}}$ die Standartisierung von S_n mit $E[S_n^*] = 0$ und $\text{Var}[S_n] = 1$. Deshalb gilt für grosse n:

$$\boxed{P[S_n^* \le x] \approx \Phi(x) \quad S_n^* \sim \mathcal{N}(0,1)}$$

Kontinuitätsstruktur Für $S_n \sim \text{Bin}(n,p)$ ist $\mathcal{N}(np, np(1-p))$ und:

$$P[a < S_n \le b] = P\left[\frac{a - np}{\sqrt{np(1-p)}} < S_n^* \le \frac{b - np}{\sqrt{np(1-p)}}\right] \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Grosse Abweichungen und Chernoff-Schranken

Momenterzeugende Funktion Die momenterzeugende Funktion einer Zufallsvariable X ist $M_X(t) := E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$ für $t \in \mathbb{R}$.

(Satz 5.6) Seien $X_1, \ldots, X_n$ Zufallsvariablen, für weliche die momenterzeugende Funktion $M_X(t)$ für alle $t \in \mathbb{R}$ endlich ist. Für jedes $b \in \mathbb{R}$ gilt dann:

$$P[S_n \ge b] \le \exp(\inf_{t \in \mathbb{R}} (n \log M_X(t) - tb))$$

Chernoff-Schranke (Satz 5.7) Seien $X_1, \ldots, X_n$ unabhängig mit $_i \sim \text{Be}(p_i)$ und $S_n = \sum_{i=1}^{n} X_i$. Sei ferner $\mu_n = E[S_n] = \sum_{i=1}^{n} p_i$ und $\delta > 0$. Dann gilt:

$$\boxed{P[S_n \ge (1+\delta)\mu_n] \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mu_n}}$$

Schätzer

Grundbegriffe

Stichprobe Gesamtheit der Beobachungen $x_1, \dots, x_n$ oder Zufallsvariablen $X_1, \dots, X_n$. Die Anzahl n heisst Stichprobenanzahl.

Stochastischer Mechanismus P_ϑ ist ein konkreter stochastischer Mechanismus, der besagt, wie sich $X_1, \dots, X_n$ verhalten. Dabei wird der Parameter ϑ zu bestimmen versucht.

Schätzer Die Schätzer $T_1, \dots, T_m$ schätzen die Parameter $\vartheta_1, \dots, \vartheta_m$. Sie sind Zufallsvariablen mit der Form $T_j = t_j(X_1, \dots, X_n)$.

Schätzwert $T_j(\omega) = t_j(X_1(\omega), \dots, X_n(\omega))$ eines konkreten Experiments ω.

Erwartungstreuheit Ein Schätzer T heisst erwartungstreu für $\vartheta \in \Theta$, falls gilt $E_\vartheta[T] = \vartheta$

Bias $E_\vartheta[T] - \vartheta$

Mittlerer quardratischer Schätzfehler $\mathrm{MSE}_\vartheta[T] := E_\vartheta[(T-\vartheta)^2] = Var_\vartheta[T] + (E_\vartheta[T] - \vartheta)^2$

Konsistenz Eine Folge von Schätzern $T^{(n)}$ heisst konsistent für ϑ, falls für jedes $\vartheta \in \Theta$ und jedes $\varepsilon > 0$ gilt: $\lim\limits_{n\to\infty} P_\vartheta[|T^{(n)} - \vartheta| > \varepsilon] = 0$. *Beweisen mit Chebyshev-Ungleichung.*

Die Maximum-Likelihood-Methode

Likelihood-Funktion $L(x_1, \dots, x_n; \vartheta) := \begin{cases} p(x_1, \dots, x_n; \vartheta) = \prod\limits_{i=1}^n p_X(x_i; \vartheta) & \text{Im diskreten Fall} \\ f(x_1, \dots, x_n; \vartheta) = \prod\limits_{i=1}^n f_X(x_i; \vartheta) & \text{Im stetigen Fall} \end{cases}$

log-**Likelihood-Funktion** $l(x_1, \dots, x_n; \vartheta) := \log L(x_1, \dots, x_n; \vartheta)$

ML-Schätzer Der ML-Schätzer T für ϑ wird definiert als Maximierung von $L(X_1, \dots, X_n; \vartheta)$ als Funktion von ϑ.

Um den ML-Schätzer zu finden:

1. Bilde log-Likelihood-Funktion, da sie meistens einfacher ist zu maximieren
2. Bilde Ableitung $\frac{\partial}{\partial\vartheta} \log L(x_1, \dots, x_n; \vartheta)$
3. ML-Schätzer ϑ kann durch das Nullsetzen von $\frac{\partial}{\partial\vartheta} \log L(x_1, \dots, x_n; \vartheta)$ gefunden werden

Momentenschätzer Der ML-Schätzer für $\vartheta = (\mu, \sigma^2) = (E_\vartheta[X], \mathrm{Var}_\vartheta[X])$, genannt Momentenschätzer, ist:

$$\boxed{T = (T_1, T_2) \quad T_1 = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n \quad T_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \sum_{i=1}^n X_i^2 - \overline{X}_n^2}$$

Dieser Schätzer ist nicht erwartungstreu.

Um einen Erwartungstreuen Schätzer für $(E_\vartheta[X], \mathrm{Var}_\vartheta[X])$ zu haben, verwendet man meistens:

$$\boxed{T = (T_1', T_2') \quad T_1' = T_1 = \overline{X}_n \quad T_2' = \frac{1}{n-1}T_2 = \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{1}{n-1}\overline{X}_n^2}$$

$T_1 = T_1'$ wird *Stichprobenmittel*, $S^2 := T_2' = \frac{1}{n-1}\sum\limits_{i=1}^n (X_i - \overline{X}_n)^2$ wird *Stichprobenvarianz* genannt.

Um den Momentenschätzer zu finden:

1. Berechne T

2. Setze $T = (E[X], \mathrm{Var}[X])$, wobei $E[X] = \int\limits_{-\infty}^{\infty} x f(x) dx$
3. Löse auf nach der zu schätzenden Variable ϑ

Verteilungsaussagen

Mehrere Normalverteilte Variablen (Satz 7.1) Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

1. $\overline{X}_n$ ist normalverteilt gemäss $\sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$, und $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
2. $\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum\limits_{i=1}^{n} (X_i - \overline{X}_n)^2$ ist $\mathcal{X}^2$-verteilt mit $n-1$ Freiheitsgraden.
3. $\overline{X}_n$ und S^2 sind unabhänging.
4. Der Quotient $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ ist t-verteilt mit $n-1$ Freiheitsgraden.

Tests

Grundbegriffe

Niveau Je kleiner das Niveau α, desto kleiner ist die Wahrscheinlichkeit, dass die Nullhypothese H_0 abgelehnt wird, obwohl sie richtig ist.

p-Wert Auch Signifikanzwert genannt, entspricht dem kleinsten Signifikanzniveau α, bei dem die Nullhypothese H_0 gerade noch verworfen werden kann.

Nullhypothese $H_0 : \vartheta \in \Theta_0$

Alternativhypothese $H_A : \vartheta \in \Theta_A$, wobei $\Theta_A = \Theta_0^c$, falls keine explizite Alternative spezifizert wurde.

Teststatistik $T = t(X_1, \dots, X_n)$, wobei t eine Abbildung $t : \mathbb{R}^n \to \mathbb{R}$ ist.

Kritischer Bereich Der kritische Bereich oder Verwerfungsbereich hat die Form $K \subseteq \mathbb{R}$. Die Hypothese wird genau dann verworfen, wenn der realisierte Wert $t(x_1, \dots, x_n) = T(\omega)$ in K liegt.

Vertrauensintervall

Fehler erster Art $\vartheta \in \Theta_0$ und $T \in K$.

Fehler zweiter Art $\vartheta \in \Theta_A$ und $T \notin K$.

Vorgehen

1. *Minimierung des Fehlers erster Art.* Wähle Signifikanzniveau $\alpha \in (0,1)$ so dass $\sup\limits_{\vartheta \in \Theta_0} P_\vartheta[T \in K] \leq \alpha$.
2. *Minimierung des Fehlers zweiter Art.* Maximiere die Macht des Tests $\beta(\vartheta) := P_\vartheta[T \in K]$.

Konstruktion von Tests

Likelihood-Quotient $R(x_1, \dots, x_n; \vartheta_0, \vartheta_A) := \frac{L(x_1, \dots, x_n; \vartheta_0)}{L(x_1, \dots, x_n; \vartheta_A)}$

Ist der Likelihood-Quotient klein, sprechen die Daten gegen die Nullhypothese und für die Alternativhypothese.

Neyman-Pearson-Lemma (Satz 9.1) Sei $\Theta_0 = \{\vartheta_0\}$ und $\Theta_A = \{\vartheta_A\}$. Sei $T = R(X_1, \dots X_n, \vartheta_0, \vartheta_A)$ und $K = [0, c)$ sowie $\alpha^* := P_{\vartheta_0}[T \in K] = P_{\vartheta_0}[T < c]$. Der Likelihood-Quotienten-Test mit T und K ist dann optimal im Sinn dass jeder andere Test mit Signifikanzniveau $\alpha \leq \alpha^*$ eine kleinere Macht hat.

Beispiel zur Konstruktion von Tests:

Ist eine Münze gezinkt? Vermutung, dass für einen Wurf X_i gilt, dass $p > 0$ ist. Test mit $n = 10$, $\alpha = 2.5\%$. Die Resultate der Würfe sind $6 \times X_i = 1$ und somit $4 \times X_i = 0$.

1. Modell: $X_i \sim \mathrm{Ber}(p)$
2. Nullhypothese: $H_0 : p_0 = p = 0.5$

3. Alternativhypothese: $H_A : p > p_0$
4. Teststatistik: $T = \sum_{i=1}^n X_i$, denn der Likelihood-Quotiont kleiner als 1 und somit wird $R(x_1, \dots, x_10; p_0, p_A)$ klein genau dann wenn $T = \sum_{i=1}^n x_i$ gross ist.
5. Verteilung der Teststatistik unter der Nullhypothese: $p_0 = p = 0.5 \implies T \sim \text{Bin}(10, 0.5)$
6. Verwerfungsbereich: Da gilt, dass die Summe T gross wird bei einem kleinen Likelihood-Quotient, hat der Verwerfungsbereich die Form $K = (k, \infty)$. Zudem muss $P_{p_0}[T \leq k] \geq (1 - \alpha)$ gelten, woraus folgt, dass $k = 8$. Somit ist $K = [8, \infty)$.
7. Beobachteter Wert der Teststatistik: $T(\omega) = 6$
8. Testentscheid: Da $T(\omega) = 6 \notin K = [8, \infty)$, wird die Nullhypothese nicht verworfen.

z-Test Seien $X_1, \dots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$ Normalverteilt mit bekannter Varianz σ^2 und unbekanntem Erwartungsert ϑ. Wir wollen die Hypothese $H_0 : \vartheta = \vartheta_0$ testen.

Die Teststatistik ist $T := \frac{\overline{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \sim N(0, 1)$.

Der kritische Bereich ist von der folgenden Form mit Niveau α:

- $H_A : \vartheta > \vartheta_0$ mit $K_> = (c_>, \infty)$ und $c_> = \Phi^{-1}(1 - \alpha) := z_{1-\alpha}$ wegen der Bedingung $\alpha = P_{\vartheta_0}[T \in K_>] = P_{\vartheta_0}[T > c_>] = 1 - P_{\vartheta_0}[T \leq c_>] = 1 - \Phi(c_>)$.
- $H_A : \vartheta < \vartheta_0$ mit $K_< = (-\infty, c_<)$ und $c_< = -\Phi^{-1}(1 - \alpha) := -z_{1-\alpha}$ wegen der Bedingung $\alpha = P_{\vartheta_0}[T \in K_<] = P_{\vartheta_0}[T < c_<] = \Phi(c_<)$.
- $H_A : \vartheta \neq \vartheta_0$ mit $K_{\neq} = (-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$ und $c_{\neq} = \Phi^{-1}(1 - \frac{\alpha}{2}) := z_{1-\frac{\alpha}{2}}$ wegen der Bedingung $\alpha = P_{\vartheta_0}[T \in K_{\neq}] = P_{\vartheta_0}[T < -c_{\neq}] + P_{\vartheta_0}[T > c_{\neq}] = \Phi(-c_{\neq}) + 1 - \Phi(c_{\neq}) = 2(1 - \Phi(c_{\neq}))$.

Die Nullhypothese wird verworfen, wenn für die Realisierung der Teststatistik gilt $t \in H_a$.

t-Test Seien $X_1, \dots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$ Normalverteilt mit unbekannter Varianz und unbekanntem Erwartungswert $\vartheta = (\mu, \sigma^2)$. Wir wollen die Hypothese $H_0 : \mu = \mu_0$ testen.

Die Teststatistik ist $T := \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$.

Wir ersetzen die unbekannte Varianz durch den Schätzer $S^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \overline{X_n})^2$.

Der kritische Bereich ist von der folgenden Form mit Niveau α:

- $H_A : \mu > \mu_0$ mit $K_> = (c_>, \infty)$ und $c_> = t_{n-1,1-\alpha}$.
- $H_A : \mu < \mu_0$ mit $K_< = (-\infty, c_<)$ und $c_< = t_{n-1,\alpha}$.
- $H_A : \mu \neq \mu_0$ mit $K_{\neq} = (-\infty, c_{\neq}) \cup (c_{\neq}, \infty)$ und $c_{\neq} = t_{n-1,1-\frac{\alpha}{2}}$.

Die Nullhypothese wird verworfen, wenn für die Realisierung der Teststatistik gilt $t \in H_a$.

Konfidenzbereiche

Konfidenzintervall

$C(X_1, \dots, X_n)$ heisst der Konfidenzbereich zum Niveau $1 - \alpha$, falls gilt $P_\vartheta[\vartheta \in C(X_1, \dots, X_n)] \geq 1 - \alpha$.

Beispiel für eine Verteilung mit bekannter Standardabweichung und $\overline{X} \sim \mathcal{N}(\mu, \sigma^2)$, es gilt:

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

Durch Umformen erhalten wir:

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

Somit ist das Vertrauensintervall gegeben durch:

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

Konfidenzintervall des Erwartungswerts

$$\boxed{C(X_1, \ldots, X_n) = [\overline{X_n} - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X_n} + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]}$$

Konfidenzintervall der Varianz

$$\boxed{C(X_1, \ldots, X_n) = \left[\frac{(n-1)S^2}{\mathcal{X}^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\mathcal{X}^2_{n-1,\frac{\alpha}{2}}}\right]}$$

Kombinatorik kurz und knapp

Name	Fragestellung	Antwort
Anzahl Permutationen ohne Wiederholung von n Elementen.	Auf wieviele Arten kann man n Objekte nebeneinander anordnen?	$n! = n \times (n-1) \times \cdots \times 2 \times 1$
Anzahl Kombinationen ohne Wiederholung.	Auf wie viele Arten kann man k aus den n Objekten auswählen ohne Zurücklegen?	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$
Anzahl der Variationen mit Wiederholung.	Wie viele Sequenzen der Länge m kann man mit den n Symbolen bilden?	n^m

Anhang

Integration

Integration durch Substitution $x = \varphi(u)$

$\int\limits_{\varphi(a)}^{\varphi(b)} f(x)dx = \int\limits_a^b f(\varphi(u))\varphi'(u)du$

Partielle Integration $\int\limits_a^b f'(x)g(x)dx = [f(x)g(x)]_a^b - \int\limits_a^b f(x)g'(x)dx$

Integration von Indikatorfunktionen $\int\limits_{-\infty}^{\infty} 1_{(a,b)}(h)dh = \int\limits_a^b dh$

$\int\limits_0^1 1_{(0,p)}(h)dp \int\limits_h^1 dp$