Konstruktion von Schätzern – Maximum-Likelihood-Schätzer

Grundannahme: Verteilung $(P_\vartheta)_{\vartheta\in\Theta}$ besitze Dichte oder Wahrscheinlichkeitsfunktion $p_\vartheta(x)$

Erinnerung:

bei stetiger Verteilung: Dichte $\rightsquigarrow$ $P_\vartheta(X \le x) = \int_{-\infty}^{x} p_\vartheta(t)\,dt$

bei diskreter Verteilung: W.-Funktion $\rightsquigarrow$ $P_\vartheta(X \le x) = \sum_{\substack{y\in\mathcal{X}\\ y\le x}} p_\vartheta(y)$

insb. $P_\vartheta(X = x) = p_\vartheta(x)$

Definition (Likelihood-Funktion & Maximum-Likelihood-Schätzer)

Es sei $x \in \mathcal{X}$ eine Beobachtung.

Dann nennt man $L: \Theta \to [0,1]$, $\vartheta \mapsto L(\vartheta) = p_\vartheta(x)$

Dichte/Wahrscheinlichkeitsfunktion (klein p)

die Likelihood-Funktion zur Beobachtung x.

Der Maximum-Likelihood-Schätzer $\hat{\vartheta}_{ML}$ für ϑ ist der Wert von ϑ für den L ihr Maximum annimmt

Beispiel: $X \sim B(n,\vartheta)$

$$\Rightarrow P(X=k) = \begin{cases} \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}, & k\in\{0,\dots,n\} \\ 0, & k\notin\{0,\dots,n\} \end{cases}$$

in W-Rechnung: n, ϑ bekannt

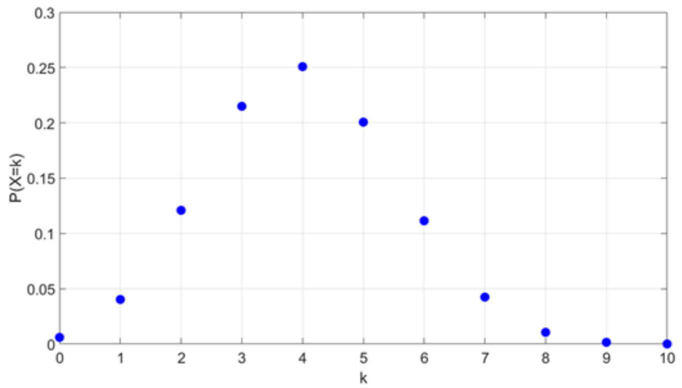
z.B. $n = 10$, $\vartheta = 0.4$: $\cdot\ P(X = 3) = \binom{10}{3}\cdot 0{,}4^3\ 0{,}6^7 \approx 0{,}2150$

bekannt

$\cdot\ P(X = 8) = \binom{10}{8}\cdot 0{,}4^8\cdot 0{,}6^2 \approx 0{,}0106$

Wahrscheinlichkeiten

variierbar



- in der Statistik

bekannt / fixiert: n, x (Wert der Zufallsvariable)

unbekannt: $\vartheta \in [0,1]$

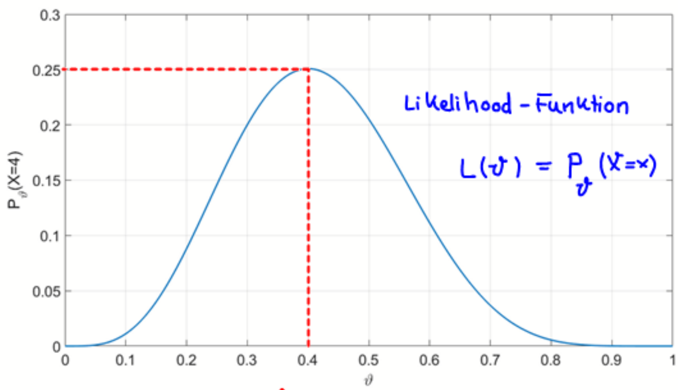
$$\boxed{P_{\vartheta}(X = x)} = \binom{n}{x} \cdot \vartheta^{x} \, (1-\vartheta)^{n-x} \quad \text{wenn } x \in \{0,1,..,n\} = \mathcal{X}$$

Wahrscheinlichkeit abhängig von ϑ — variierbar — variierbar

z.B. $n = 10$, $x = 4$

- $\vartheta = 0$: $P_0(X=4) = \binom{10}{4} \cdot 0^4 \cdot 1^6 = 0$
- $\vartheta = 0{,}1$: $P_{0,1}(X=4) = \binom{10}{4} \cdot 0{,}1^4 \cdot 0{,}9^6 \approx 0{,}0112$
- $\vartheta = 0{,}3$: $P_{0,3}(X=4) = \binom{10}{4} \cdot 0{,}3^4 \cdot 0{,}7^6 \approx 0{,}2001$

$P_{\vartheta}(X=4)$ bestimmbar für jedes $\vartheta \in [0,1]$



$\hat{\vartheta}_{ML} = 0{,}4$

▶ 3 Bemerkungen

i) kompakt: $L(\hat{\vartheta}_{ML}) = \max\limits_{\vartheta \in \Theta} L(\vartheta)$

ii) oft alternativ verwendet: Loglikelihoodfunktion $\ell(\vartheta) = \ln L(\vartheta) = \ln P_{\vartheta}(x)$

iii) Bedeutung: $\hat{\vartheta}_{ML}$ ist der Parameterwert für den die gemachte Beobachtung am wahrscheinlichsten ist.

Beispiel

i) Binomialverteilung $X \sim B(n, \vartheta)$ ▶ 4

$$L(\vartheta) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} \qquad (x \in \{0, \dots, n\})$$

$$\ell(\vartheta) = \ln(L(\vartheta)) = \ln\left(\binom{n}{x}\right) + x \cdot \ln(\vartheta) + (n-x) \ln(1-\vartheta)$$

man kann zeigen: $\hat{\vartheta}_{ML} = \frac{x}{n}$

ii) Normalverteilung $X = (X_1, \dots, X_n)$ mit $X_i \sim N(\mu, \sigma^2)$ ▶ 5

Dichte: $$p_{\mu,\sigma^2}(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Likelihood-Funktion: $L(\mu, \sigma^2) = p_{\mu,\sigma^2}(x_1, \dots, x_n)$

Loglikelihood-Funktion

$$\ell(\mu, \sigma^2) = \ln(p_{\mu,\sigma^2}(x_1, \dots, x_n)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

▶ 6 gesucht: Stelle $(\hat{\mu}, \widehat{\sigma^2})$ an der die (Log-) Likelihood-Funktion ein Maximum annimmt.

Vorgehen: Extremalstellensuche

▶ 7 1.) notwendige Bedingung:

mit Substitution: $s = \sigma^2$ ergibt sich

$$\ell(\mu, s) = -\frac{n}{2} \ln(2\pi \cdot s) - \frac{1}{2s} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$$

gesucht $(\hat{\mu}, \hat{s})$ mit $\nabla \ell(\hat{\mu}, \hat{s}) = \vec{0} \in \mathbb{R}^2$

$$\nabla \ell(\hat{\mu}, \hat{s}) = \left(\frac{\partial \ell}{\partial \mu}(\hat{\mu}, \hat{s}), \frac{\partial \ell}{\partial s}(\hat{\mu}, \hat{s}) \right)^T \stackrel{!}{=} (0, 0)^T$$

▶ 8

- $$\frac{\partial \ell}{\partial \mu}(\hat{\mu}, \hat{s}) = -\frac{1}{2\hat{s}} \cdot \sum_{i=1}^{n} 2 \cdot (x_i - \hat{\mu}) \cdot (-1) = \boxed{\frac{1}{\hat{s}} \cdot \sum_{i=1}^{n} (x_i - \hat{\mu}) \stackrel{!}{=} 0}$$

▶ 9

- $$\frac{\partial \ell}{\partial s}(\hat{\mu}, \hat{s}) = -\frac{n}{2} \cdot \frac{2\pi}{2\pi \cdot \hat{s}} - (-1) \cdot \frac{1}{2\hat{s}^2} \cdot \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

$$= \boxed{-\frac{n}{2 \cdot \hat{s}} + \frac{1}{2\hat{s}^2} \cdot \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \stackrel{!}{=} 0}$$

Bestimmungsgleichungen: ▶ 10

① $$\frac{1}{\hat{s}} \cdot \sum_{i=1}^{n} (x_i - \hat{\mu}) = 0$$

② $$-\frac{n}{2 \cdot \hat{s}} + \frac{1}{2\hat{s}^2} \cdot \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = 0$$

aus ①: $$0 = \sum_{i=1}^{n} (x_i - \hat{\mu}) = \left(\sum_{i=1}^{n} x_i\right) - \left(\sum_{i=1}^{n} \hat{\mu}\right) = \left(\sum_{i=1}^{n} x_i\right) - n \cdot \hat{\mu}$$

$$\Leftrightarrow \hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

aus ② $\cdot (2\hat{s}^2)$: $$-n \cdot \hat{s} + \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = 0$$

$$\Leftrightarrow \hat{s} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

▶ 11 Der Maximum-Likelihood-Schätzer für Erwartungswert & Varianz ist:

$$\boxed{\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad \& \qquad \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2}$$

▶ 12 .... nur im Video: Anmerkung zu anderem, gebräuchlichem Schätzer für die Varianz der Normalverteilung