16. Vorlesung: Parametrische Statistik

Nikolas Tapia

13. Juni 2024, Stochastik für Informatik(er)

Grundannahmen

- Große Menge von Beobachtungsdaten $(x_1, \dots, x_n)$ Datenvektor $\in \mathbb{R}^n$
- Modellierung als Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_n$, mit **unbekannter** Verteilung $\mathbb{P}_\theta$. Bsp: $f_\mu(x) = (2\pi\sigma^2)^{1/2} \exp\left(-(x-\mu)^2/2\sigma^2\right)$
- θ ist einen unbekannten Parameter, der die Verteilung $\mathbb{P}_\theta$ charakterisiert.
- **Ziel:** Schätzung von θ. Weitere Annahme: $\theta \in \mathbb{R}$, $\theta \in [0, 1]$ usw. $\theta \in \Theta$

Die Familie $(\mathbb{P}_\theta)_{\theta \in \Theta}$ ist ein statistische Modell.

Definition 16.1

Gegeben eine Schätzfunktion $\theta_n : \mathbb{R}^n \to \mathbb{R}$,

$$\bar{\theta}_n := \theta_n(x_1, \dots, x_n) \in \mathbb{R}$$

ist ein Schätzer für θ. $\theta_n(X_1, \dots, X_n)$ ZV.

Einige Schätzern

Zufallsvariablen

- $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^n x_i$, der empirische Mittelwert.
- $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{\mu}_n)^2$, die empirische Varianz.
- $\bar{m}_n$, der (empirische) Median.

$$\bar{\mu}_n(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^n X_i$$

$$\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{\mu}_n(X_1, \dots, X_n)\right)^2$$

Definition 16.2

Sei θ_n eine Schätzfunktion und $\bar{\theta}_n$ den entsprechenden Schätzer. Dann heißt $\bar{\theta}_n$

- **erwartungstreu**, falls $\mathbb{E}_\theta[\theta_n(X_1, \dots, X_n)] = \theta$,
- **konsistent**, falls $\bar{\theta}_n \to \theta$,
- **effizient**, falls $\mathbb{V}_\theta(\theta_n(X_1, \dots, X_n)) \to 0$.

Beispiel

Maximum-Likelihood-Schätzung（最大似然估计）

给定条件

设 $\Theta \subseteq \mathbb{R}$ 是所有理论上可能的参数的集合。

Definition 16.3

似然函数（Likelihood Function）是函数 $L : \mathbb{R}^n \times \Theta \to [0, \infty)$，定义如下：

- 对于离散分布，似然函数 $L((x_1, \dots, x_n); \theta) := \mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} p_\theta(x_i)$，
  其中 $p_\theta(x)$ 是给定参数 θ 下，离散随机变量 X_i 取值 x_i 的概率。
- 对于连续分布，似然函数 $L((x_1, \dots, x_n); \theta) = \prod_{i=1}^{n} f_\theta(x_i)$，
  其中 $f_\theta(x)$ 是给定参数 θ 下，连续随机变量 X_i 的概率密度函数。

Maximum-Likelihood-Schätzung

Sei $\Theta \subseteq \mathbb{R}$ die Menge aller theoretisch möglichen Parameter.

Definition 16.3

Die Likelihood-Funktion ist die Funktion $L\colon \mathbb{R}^n \times \Theta \to [0, \infty)$, gegeben durch

- $L((x_1, \ldots, x_n); \theta) := \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^n p_\theta(x_i)$ für diskrete Verteilungen, $\sim$ Verteilung(θ) $\quad p_\theta(x) = e^{-\theta} \frac{\theta^x}{x!}$
- $L((x_1, \ldots, x_n); \theta) = \prod_{i=1}^n f_\theta(x_i)$ für stetige Verteilungen.

Definition 16.4

Der Maximum-Likelihood-Schätzer für θ ist $\bar{\theta}_n^*$, welches die Likelihood-Funktion unter allen $\theta \in \Theta$ maximiert, d.h.

$$\bar{\theta}_n^* = \arg\max_{\theta \in \Theta} L((x_1, \ldots, x_n); \theta).$$

In anderen Worten,

$$L((x_1, \ldots, x_n); \bar{\theta}_n^*) \geq L((x_1, \ldots, x_n); \theta) \quad \forall \theta \in \Theta.$$

Log-Likelihood-Funktion

Definition 16.5

Die Log-Likelihood-Funktion ist die Funktion $\ell\colon \mathbb{R}^n \times \Theta \to \mathbb{R}$, gegeben durch

$$\ell((x_1, \dots, x_n); \theta) := \log L((x_1, \dots, x_n); \theta).$$

Aussage 16.1

Der Maximum-Likelihood-Schätzer $\bar{\theta}_n^*$ ist auch der Schätzer, der die Log-Likelihood-Funktion maximiert.

Beispiel

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(x-\mu)^2}{2\theta}\right)$$

bekannter μ!

σ^2

$$L((x_1,\dots,x_n);\theta) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$= \frac{1}{\left(\sqrt{2\pi\theta}\right)^n} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\theta}\right)$$

$$\ell((x_1,\dots,x_n);\theta) = -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2$$

Beispiel

$$\ell\left((x_1,\dots,x_n);\theta\right) = -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2$$

Optimierung:

$$\frac{\partial}{\partial\theta}\ell\left((x_1,\dots,x_n);\theta\right) = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial\theta}\log(2\pi\theta) = \frac{\partial}{\partial\theta}\left(\log 2\pi + \log\theta\right) = 0 + \frac{1}{\theta}$$

$$\frac{\partial}{\partial\theta}\ell = 0 \Rightarrow \frac{n}{2\theta} = \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow \bar{\theta}_n^* = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \neq \bar{\sigma}_n^2$$

$$\mathbb{E}\left[\bar{\theta}_n^*\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

$$\ell((x_1,\dots,x_n);\theta) = -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial^2}{\partial\theta^2}\ell((x_1,\dots,x_n);\theta) = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow \left.\frac{\partial^2 \ell}{\partial\theta^2}\right|_{\theta=\bar\theta_n^*} = \frac{n}{2\left(\frac{1}{n}\sum(x_i-\mu)^2\right)^2} - \frac{1}{\left(\frac{1}{n}\sum(x_i-\mu)^2\right)^3}\times\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= \frac{n^3}{2\left(\sum(x_i-\mu)^2\right)^2} - \frac{n^3}{\left(\sum(x_i-\mu)^2\right)^2} < 0$$

Beispiel

Binomialverteilung (n, θ) (bekannter n)

Beispiel

$$P_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad \theta \in [0,1] = \Theta, \quad x \in \{0, \dots, n\}.$$

$$L((x_1, \dots, x_N); \theta) = \prod_{i=1}^N P_\theta(x_i)$$

$$= \prod_{i=1}^N \binom{n}{x_i} \theta^{x_i} (1-\theta)^{n-x_i}$$

$$= \prod_{i=1}^N \binom{n}{x_i} \times \theta^{\sum_{i=1}^N x_i} (1-\theta)^{nN - \sum_{i=1}^N x_i}$$

$$\ell((x_1, \dots, x_N); \theta) = \log \theta \left(\sum_{i=1}^N x_i\right) + \log(1-\theta)\left(nN - \sum_{i=1}^N x_i\right) + \log(\dots)$$

Beispiel

$$\frac{\partial}{\partial\theta}\ell\big((x_1,\dots,x_N);\theta\big) = \frac{1}{\theta}\Big(\sum_{i=1}^{N} x_i\Big) - \frac{1}{1-\theta}\Big(nN - \sum_{i=1}^{N} x_i\Big)$$

$$S_N := \sum_{i=1}^{N} x_i$$

$$\Rightarrow \frac{\partial}{\partial\theta}\ell = 0 \Rightarrow \frac{1}{\theta} S_N = \frac{1}{1-\theta}\big(nN - S_N\big)$$

$$\Rightarrow (1-\theta) S_N = \theta\big(nN - S_N\big)$$

$$\Rightarrow S_N = \theta\big(nN - \cancel{S_N} + \cancel{S_N}\big)$$

$$\Rightarrow \theta^*_N = \frac{1}{nN} S_N = \frac{1}{n}\bar{\mu}_N$$

$$\bar{\mu}_N = \frac{1}{N} S_N$$

$$X \sim \mathrm{Binom}(n,p) \Rightarrow \mathbb{E}[X] = np \Rightarrow p = \frac{1}{n}\mathbb{E}[X] \approx \frac{1}{n}\bar{\mu}_N$$

$$\frac{\partial^2}{\partial \theta^2} \ell = -\frac{1}{\theta^2} S_N + \frac{1}{(1-\theta)^2} (Nn - S_N)$$

$$\Rightarrow \frac{\partial^2}{\partial \theta^2} \ell \Big|_{\theta = \theta_N^*} < 0 .$$

Beispiel

Bei der 6LP-Variante sind die Inhalte bis hierher Klausurrelevant.