Kapitel 14

Parameterschätzung

Modell
Es sei $\{P_\theta | \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$ eine Familie von Verteilungen auf χ (sog. Stichprobenraum), $x = (x_1, \ldots, x_n)$ sei eine Realisierung der Zufallsstichprobe $X = (X_1, \ldots, X_n)$ zu einer Verteilung $P \in \{P_\theta | \theta \in \Theta\}$, wobei der wahre Parameter θ unbekannt ist.

Problem
Schätze unbekanntes θ aus der konkreten Realisierung von X.

Definition 14.1
Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit Verteilung P.
Dann heißt der Zufallsvektor $(X_1, \ldots, X_n)$ ***Zufallsstichprobe zur Verteilung*** *P.*
Eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ nennt man ***Stichprobe****.*

Beispiel 14.1
Gegeben sei ein Würfel, den wir n-mal werfen dürfen. Aus der Beobachtung ist die Wahrscheinlichkeit zu schätzen eine 6 zu würfeln.
Modell: $\chi = \{0, 1\}$: $0 \mathrel{\hat=}$ keine 6 gewürftelt, $1 \mathrel{\hat=}$ 6 gewürfelt.
$P_\theta = B(1, \theta)$, $\Theta = [0, 1]$

Definition 14.2 *Eine messbare Abbildug $T : \chi^n \to \tilde{\Theta}$, $\tilde{\Theta} \supset \Theta$, heißt* ***Schätzer*** *für θ.*

Beispiel 14.2

a) Die Statistik $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ heißt **Stichprobenmittel**

b) Die Statistik $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ heißt **Stichprobenvarianz**

Beispiel 14.3 (Schätzung eines Fischbestandes)
Teich enthält unbekannte Zahl $N = \theta$ von Fischen
r Fische werden gefangen, (rot) markiert und wieder ausgesetzt.
In einem zweiten Zug werden m Fische gefangen x davon seien markiert.
Wie groß ist N?

Modell

Urne mit $N = \theta$ Kugeln, r rot, $N - r =: s$ schwarz
m Kugeln werden ohne Zurücklegen gezogen. Die Zufallsvariable X beschreibe die Anzalh der roten unter den gezogenen Kugeln.
Also
$\chi = \mathbb{N}_0$, $P_\theta \sim$ Hypergeom.(θ, r, m), $\Theta = \mathbb{N}$. Beachte: $n = 1$

14.1 Maximum-Likelihood-Methode

Idee:
Wir wählen für θ den Wert, unter dem die Wahrscheinlichkeit, dass die konkrete Stichprobe vorliegt maximiert wird.
Im folgenden sei P_θ diskret mit Zähldichte $p(x;\theta)$ oder stetig mit Dichte $f(x;\theta)$

Definition 14.3
Gegeben sei eine Stichprobe $x = (x_1, \ldots, x_n)$. Dann heißt die Funktion

$$L_x(\theta) := f(x_1;\theta) \cdot \cdots \cdot f(x_n;\theta) \quad \text{bzw.} \quad L_x(\theta) := \underbrace{p(x_1;\theta) \cdot \cdots \cdot p(x_n;\theta)}_{P_\theta(X_1,\ldots,X_n)=(x_1,\ldots,x_n)}$$

die ***Likelihood-Funktion*** *der Stichprobe.*
Eine Funktion $\hat{\theta}_{\mathrm{ML}}(x)$ heißt ***Maximum-Likelihood Schätzer (MLS)****, falls*

$$L_x(\hat{\theta}_{ML}(x)) = \sup_{\theta \in \Theta} L_x(\theta)$$

Bemerkung 14.1

a) Ist P_θ diskret, so gilt:

$$L_X(\theta) = p(x_1;\theta) \cdots p(x_n;\theta) = P_\theta(X_1 = x_1) \cdots P_\theta(X_n = x_n) \overset{X \text{ unabhängig}}{=} P_\theta(X = x)$$

b) Der MLS $\hat{\theta}_{\mathrm{ML}}(x)$ ist nicht immer eindeutig.

Beispiel 14.4 (vgl. Beispiel 14.3)
$n = 1$

$$\text{Likelihood-Funktion:} \quad L_x(\theta) = \frac{\binom{r}{x}\binom{\theta-r}{k-x}}{\binom{\theta}{k}}$$

Für welches θ ist $L_x(\theta)$ maximal?
Betrachte:

$$\frac{L_x(\theta)}{L_x(\theta-1)} = \frac{\binom{r}{x}\binom{\theta-r}{m-x}\binom{\theta-1}{m}}{\binom{\theta}{m}\binom{r}{x}\binom{\theta-1-r}{m-x}} = \frac{(\theta-r)(\theta-m)}{\theta(\theta-r-m+x)}$$

$$L_x(\theta) > L_x(\theta-1) \Leftrightarrow (\theta-r)(\theta-m) > \theta(\theta-r-m+x) \Leftrightarrow mr > \theta x \Leftrightarrow \theta < \frac{mr}{x}$$

Also ist $L_x(\theta)$ maximal für $\hat{\theta}(x) = \lfloor \frac{mr}{x} \rfloor$

$\hat{\theta}(x)$ ist eindeutig, falls $\frac{mr}{x} \notin \mathbb{N}$
Falls $\frac{mr}{x} \in \mathbb{N}$ sind $\hat{\theta}_1(x) = \frac{mr}{x}$ und $\hat{\theta}_2(x) = \frac{mr}{x} - 1$ MLS

Beispiel 14.5 (Schätzung einer Erfolgswahrscheinlichkeit)
Aus n Bernoulli-Experimenten liegen x Erfolge vor, gesucht ist die Erfolgswahrscheinlichkeit:
Modell: $\chi = \mathbb{N}$, $n = 1$, $P_\theta = B(m, \theta)$, $\Theta = (0, 1)$.

$$\text{Likelihood-Funktion:} \quad L_x(\theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x} \quad , \theta \in [0,1]$$

Statt $L_x(\theta)$ ist es oft einfacher, $\log(L_x(\theta))$ zu maximieren, die sogenannte **Log-Likelihoodfunktion**

$$\log(L_x(\theta)) = \log \binom{m}{x} + x \log \theta + (m-x) \log(1-\theta)$$

$$\frac{\partial}{\partial \theta} \log(L_x(\theta)) = \frac{x}{\theta} - \frac{(m-x)}{1-\theta} = 0 \quad \Leftrightarrow \quad \theta = \frac{x}{m}$$

θ ist tatsächlich Maximum-Stelle, dass heißt $\hat{\theta}_{\mathrm{ML}}(x) = \frac{x}{m}$

14.2 Momentenmethode

Idee:
Die ersten Momente von P_θ sollten mit den empirischen Momenten übereinstimmen. Aus diesen Gleichungssystemen bestimmen wir den Schätzer.
Es sei $X \sim P_\theta$. Dann ist das k-te Moment

$$\mu_k = \mu_k(\theta) = E_\theta X^k \quad , k = 1, 2, \ldots$$

Wir betrachten nun die **empirischen Momente** zur Stichprobe $x = (x_1, \ldots, x_n)$.

$$\overline{x_k} := \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

Es soll nun gelten: $\mu_k(\theta) = \overline{x_k}$ $k = 1, 2, \ldots, m$. Aufgelöst nach θ ergibt sich dann der Momentenschätzer $\hat{\theta}_{\mathrm{MM}}(x) \in \Theta$.

Beispiel 14.6
$P_\theta = N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$, $m = 2$

(1) $\mu_1 = \mu = E_\theta X = \overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$

(2) $\mu_2 = E_\theta X^2 = \mathrm{Var}_\theta(X) + (E_\theta X)^2 = \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$

Aus (1) folgt: $\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n x_i = \overline{x}$
Aus (2) folgt: $\hat{\sigma}^2(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\overline{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^2$

14.3 Wünschenswerte Eigenschaften von Punktschätzern

Im folgenden sei $X = (X_1, \dots, X_n)$ eine Zufallsstichprobe zur Verteilung P_θ und $T : \chi \to \tilde{\Theta}$ ein Schätzer von θ. Mit E_θ bezeichnen wir den Erwartungswert bezüglich P_θ.

Definition 14.4

*a) Der Schätzer T heißt **erwartungstreu** (unbiased), falls*

$$E_\theta T(X_1, \dots, X_n) = \theta \qquad \forall\, \theta \in \Theta$$

*b) $b_T(\theta) := E_\theta T(X_1, \dots, X_n) - \theta$ heißt **Verzerrung** (Bias) des Schätzers T. Ein erwartungstreuer Schätzer ist unverzerrt.*

Beispiel 14.7 (vgl. Bsp.14.6)

- $T(x) = \overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ ist ein erwartungstreuer Schätzer für $\theta = E_\theta X_i$, denn $E_\theta(T(X)) = E_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E_\theta X_i = \theta$
- Ein erwartungstreuer Schätzer für $\theta = \mathrm{Var}_\theta(X_i)$ ist $[S^2(x) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$

Definition 14.5
Sei T ein Schätzer für θ Dann heißt

$$\mathbf{MSE(T)} := E_\theta[(T(X_1, \dots, X_n) - \theta)^2]$$

*(mittlerer) **quadratischer Fehler** ("mean-squared-error")*

Beispiel 14.8
Sei $P_\theta = U(0, \theta)$, $\Theta = \mathbb{R}$, $\chi = \mathbb{R}_+$ und $X = (X_1, \dots, X_n)$ eine Zufallsstichprobe zur Verteilung $U(0, \theta)$
Momentenmethode: $\overline{x} = E_\theta X_i = \frac{\theta}{2} \Rightarrow \hat{\theta}_{\mathrm{MM}} = 2 \cdot \overline{x}$
Maximum-Likelihood-Methode:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{sonst} \end{cases}$$

$$\Rightarrow L_x(\theta) = L_{(x_1,\dots,x_n)}(\theta) = f(x_1; \theta) \cdot \dots \cdot f(x_n; \theta) = \begin{cases} \frac{1}{\theta^n}, & 0 \le x_1, \dots, x_n \le \theta \\ 0, & \text{sonst} \end{cases}$$

Maximiere $L_x(\theta)$ in θ: $\hat{\theta}_{\mathrm{ML}}(x) = \max(x_1, \dots, x_n)$
Welcher Schätzer ist besser?

$$E_\theta[\hat{\theta}_{\mathrm{MM}}(X)] = 2E_\theta\overline{X} = \theta \quad \text{, also ist } \hat{\theta}_{\mathrm{MM}} \text{ erwartungstreu}$$

Verteilungsfunktion von $\hat{\theta}_{\mathrm{ML}}(X)$ ist

$$F_\theta(x) = P_\theta(\max(X_1, \dots, X_n) \le x) = P_\theta(X_1 \le x, \dots, X_n \le x)$$

$$= P_\theta(X_1 \leq x) \cdot \dots \cdot P_\theta(X_n \leq x) = \left(\frac{x}{\theta}\right)^n \qquad ,\text{falls} \quad 0 \leq x \leq \theta$$

Also Dichte von $\hat{\theta}_{\text{ML}}(X)$:

$$f_{\hat{\theta}_{\text{ML}}}(x) = \begin{cases} \frac{n}{\theta}(\frac{x}{\theta})^{n-1} & , \text{ falls } \ 0 \leq x \leq \theta \\ 0 & , \text{ sonst} \end{cases}$$

und

$$E_\theta[\hat{\theta}_{\text{ML}}(X)] = \int_0^\theta x f_{\hat{\theta}_{\text{ML}}}(x)\,dx = \int_0^\theta n\left(\frac{x}{\theta}\right)^n dx = \frac{n}{n+1}\theta$$

also <u>nicht</u> erwartungstreu.

Aber:

$$\text{MSE}(\hat{\theta}_{\text{MM}}(X)) = E_\theta\left([2\frac{1}{n}\sum_{i=1}^n X_i - \theta]^2\right) = \frac{\theta^2}{3n}$$

$$\text{MSE}(\hat{\theta}_{\text{ML}}(X)) = E_\theta([\max(X_1, \dots, X_n) - \theta]^2) = \frac{2\theta^2}{(n+2)(n+1)}$$

$$\frac{\text{MSE}(\hat{\theta}_{\text{MM}}(X))}{\text{MSE}(\hat{\theta}_{\text{ML}}(X))} = \frac{2}{3}\frac{n}{(n+2)(n+1)} \qquad \textbf{relative Effizienz}$$

Bei großem n ist $\text{MSE}(\hat{\theta}_{\text{ML}}(X))$ kleiner als $\text{MSE}(\hat{\theta}_{\text{MM}}(X))$.

Bemerkung 14.2
Falls T erwartungstreu ist, gilt $MSE(T) = \text{Var}_\theta(T)$

Für $\text{Var}_\theta(T)$ kann man die folgende untere Abschätzung angeben.

Satz 14.1 (Ungleichung von Rao-Cramér)
Sei $X = (X_1, \dots, X_n)$ eine Zufallsstichprobe zur Verteilung P_θ. T sei ein Schätzer für θ. Dann gilt:

$$\text{Var}_\theta(T(X)) \geq \frac{(1 + \frac{\partial}{\partial\theta}b_T(\theta))^2}{E_\theta[(\frac{\partial}{\partial\theta}\log L_X(\theta))^2]}$$

Bemerkung 14.3

(i) $I(\theta) := E_\theta[(\frac{\partial}{\partial\theta}\log L_X(\theta))^2]$ heißt **Fisher-Information**

(ii) Ist T erwartungstreu, so ist $b_T(\theta) = 0$ und $\text{Var}_\theta(T(X)) \geq \frac{1}{I(\theta)}$

Beweis
Wir nehmen an, dass $L_x(\theta) > 0 \;\forall x \in \chi^n \;\forall\theta \in \Theta$, Θ sei ein offenes Intervall in $\mathbb{R}$ und P_θ sei diskret. Es gilt:

$$\frac{\partial}{\partial\theta}\log L_x(\theta) = \frac{L'_x(\theta)}{L_x(\theta)}$$

Weiter gilt:

(1) $\sum_{x \in \chi^n} L_x(\theta) = \sum_{x \in \chi^n} P_\theta(X = x) = 1$

(2) $\theta + b_T(\theta) = E_\theta T(X) = \sum_{x \in \chi^n} T(x) \cdot L_x(\theta)$

Wir differenzieren (1) und (2) nach θ, und nehmen an, dass wir $\frac{\partial}{\partial \theta}$ und $\sum$ vertauschen könne.

(1')

$$0 = \sum_{x \in \chi^n} L'_x(\theta) = \sum_{x \in \chi^n} \frac{\partial}{\partial \theta} \log(L_x(\theta)) \cdot L_x(\theta) = E_\theta \left[\frac{\partial}{\partial \theta} \log L_X(\theta) \right]$$

(2')

$$1 + b'_T(\theta) = \sum_{x \in \chi^n} T(x) L'_x(\theta)$$

$$\sum_{x \in \chi^n} T(x) \frac{\partial}{\partial \theta} \log(L_x(\theta)) \cdot L_x(\theta) = E_\theta \left[T(X) \cdot \frac{\partial}{\partial \theta} \log L_X(\theta) \right]$$

(2') - (1')$E_\theta T(X)$

$$1 + b'_T(\theta) = E_\theta[(T(X) - E_\theta T(X)) \frac{\partial}{\partial \theta} \log L_X(\theta)]$$

Mit der Ungleichung von Cauchy-Schwarz folgt:

$$\left(1 + b'_T(\theta)\right)^2 = \left(E_\theta[(T(X) - E_\theta T(X)) \frac{\partial}{\partial \theta} \log L_X(\theta)] \right)^2$$

$$\leq E_\theta[(T(X) - E_\theta T(X))^2] \cdot E_\theta \left[\frac{\partial}{\partial \theta} \log L_X(\theta)^2 \right] = \mathrm{Var}_\theta T \cdot I(\theta)$$