

# 14 Schätzmethoden

## Eigenschaften von Schätzungen $\hat{\theta}$

Sei  $\hat{\theta}_n$  eine Schätzung eines Parameters  $\theta$ , die auf  $n$  Beobachtungen beruht.

- $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$  “Konsistenz” (Minimalforderung)
- $E\hat{\theta}_n = \theta$  “Erwartungstreue”  
 $E\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$  “Asymptotische Erw.treue”

- $\text{var } \hat{\theta}_n$  möglichst klein: “gute”, “effiziente” Schätzung
- wenn  $\text{var } \hat{\theta}_n$  den kleinstmöglichen Wert annimmt für alle e-treuen Schätzungen:  
 $\hat{\theta}_n$ : “optimale Schätzung”

- $\text{MSE} = \text{var } \hat{\theta}_n + \text{bias}^2 \hat{\theta}_n$   
 $= \text{var } \hat{\theta}_n + (E\hat{\theta}_n - \theta)^2$   
→ minimal oder möglichst klein.
- Eigenschaften sollten “möglichst” auch bei (kleinen) Abweichungen von der (Normal-)Verteilungsannahme gelten  
→ robuste Schätzung.

## Schätzmethoden

- Momentenmethode

Man drückt den zu schätzenden Parameter durch die Momente, z.B.  $E(X)$ , aus.

Dann werden die Momente durch die entsprechenden *empirischen* Momente, z.B. durch  $\bar{X}$ , ersetzt.

- Maximum-Likelihood-Schätzung (ML-Schätzung)

Es wird der Schätzwert für den unbekannt Parameter ermittelt, der am meisten für diesen Parameter spricht (most likely).

- Kleinste-Quadrat-Schätzung (KQS)

Sei  $\theta$  der zu schätzende Parameter. Man geht aus von einem Modell, z.B.

$$Y_i = g(\theta, X_i) + \epsilon_i$$

Dann versucht man die Summe der Fehlerquadrate

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - g(\theta, X_i))^2.$$

zu minimieren (Kleinste Quadrate).

### **Bsp. 97 (Momentenschätzung bei Normalverteilung)**

Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma)$ .

$$\mu = \mathbf{E}X_i \quad \Longrightarrow \quad \hat{\mu} = \overline{X}$$

$$\sigma^2 = \mathbf{E}(X - \mathbf{E}X)^2 \quad \Longrightarrow \quad \hat{\sigma}^2 = \overline{(X_i - \overline{X})^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$$

### **Bsp. 98 (Momentenschätzung bei Exponentialverteilung)**

Seien  $X_1, \dots, X_n \sim \text{Exp}(\lambda)$ .

$$\lambda = \frac{1}{\mathbf{E}X_i} \quad \Longrightarrow \quad \hat{\lambda} = \frac{1}{\overline{X}}$$

### **Bsp. 99 (Momentenschätzung bei Binomialverteilung)**

*Seien  $X_1, \dots, X_n \sim Bi(1, p)$ .*

$$p = \mathbf{E}X_i \quad \Longrightarrow \quad \hat{p} = \overline{X}$$

*der relative Anteil der Realisierungen  $x_i = 1$ .*

### **Bsp. 100 (ML-Schätzung bei Binomialverteilung)**

*Beobachten  $n=1000$  Jugendliche. Stichprobe  $(X_1, \dots, X_n)$*

*$X_i = 1$  falls Übergewicht festgestellt*

*$X_i = 0$  sonst.*

*Die Wkt., daß die beobachtete Stichprobe auftritt, wenn der Parameter  $p$  vorliegt ist*

$$P(X_1 = x_1, \dots, X_n = x_n) = \binom{n}{k} p^k (1 - p)^{n-k},$$

*wobei  $k = \sum_{i=1}^n x_i$ .*

*Der ML-Schätzer ist der Wert, der diese Funktion,  $L_n(p)$ ,*

*Likelihood-Funktion genannt, bzgl.  $p$  maximiert. Maximieren*



*statt  $L_n(p)$ :  $\log L_n(p)$  (Arg.Max. ist dasselbe).*

$$\begin{aligned}\ln L_n(p) &= \ln\left(\binom{n}{k} p^k (1-p)^{n-k}\right) \\ &= \ln\left(\binom{n}{k}\right) + k \ln p + (n-k) \ln(1-p).\end{aligned}$$

*Ableiten nach  $p$  und Nullsetzen liefert:*

$$\frac{k}{p} - \frac{n-k}{1-p} = 0$$

*Die einzige Lösung ist:*

$$\hat{p} = \frac{k}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

*Für ein relatives Extremum in  $(0,1)$  kommt nur dieser Wert in Betracht.*

*Müssen aber noch die Likelihood-Funktion an den Rändern betrachten:*

*Für  $p = 0$  und  $p = 1$  wird  $\ln L(p) = -\infty$ . Also:*

$$\hat{p}_{ML} = \frac{k}{n}.$$

**Bsp. 101 (ML-Schätzung bei Normalverteilung)**

*Likelihood:  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ , die gemeinsame Dichtefunktion der  $X_i$ .*

*Seien  $X_1, \dots, X_n$  unabhängig,  $X_i \sim \mathcal{N}(\mu, 1)$ .*

*Likelihood:*

$$\begin{aligned} L_n(\mu) &= \prod_{i=1}^n f_{X_i}(x_i) \quad (\text{Unabhängigkeit}) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu)^2 / 2} \end{aligned}$$

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

$$\frac{\partial L_n(\mu)}{\partial \mu} = 2 \sum_{i=1}^n (x_i - \mu)$$

*Nullsetzen liefert:*

$$\hat{\mu} = \bar{X}.$$

### Bsp. 102 (ML-Schätzung bei Gleichverteilung auf $(0, \theta)$ )

*Likelihood:*  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ ,

*die gemeinsame Dichtefunktion der  $X_i$ .*

*Seien  $X_1, \dots, X_n$  unabhängig,  $X_i \sim \mathbb{R}(0, \theta)$ , d.h.*

$$f_{X_i}(x_i) = \begin{cases} \frac{1}{\theta} & \text{falls } 0 \leq x_i \leq \theta \\ 0 & \text{sonst} \end{cases}$$

*Likelihood:*

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n f_{X_i}(x_i) && \text{(Unabhängigkeit)} \\ &= \begin{cases} \frac{1}{\theta^n} & \text{falls } 0 \leq x_i \leq \theta \quad \forall x_i \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

*Maximal, wenn  $\theta \geq x_1, \dots, x_n$ , und wenn  $\theta$  möglichst klein, also*

$$\hat{\theta} = \max(x_1, \dots, x_n).$$

**Bsp. 103 ( KQS des Lageparameters) Modell:**

$$Y_i = \mu + \epsilon_i$$

*Die Summe der Fehlerquadrate*

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \mu)^2.$$

*minimieren: Differenzieren und Nullsetzen liefert:*

$$\hat{\mu}_{KQS} = \bar{Y}.$$

## Bsp. 104 (KQS im einfachen linearen Regressionsmodell)

$$Y_i = \theta_1 + \theta_2 X_i + \epsilon_i$$

$$f(X, \theta_1, \theta_2) = \theta_1 X + \theta_2$$

$$\frac{\partial f}{\partial \theta_1} = X \qquad \frac{\partial f}{\partial \theta_2} = 1$$

⇒

$$\frac{1}{n} \sum_{i=1}^n (Y_i - (\theta_1 X_i + \theta_2)) \cdot X_i = 0$$

$$\frac{1}{n} \sum_{i=1}^n (Y_i - (\theta_1 X_i + \theta_2)) \cdot 1 = 0$$



⇒

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \theta_2 \sum_i X_i = 0$$
$$\sum_i Y_i - \theta_1 \sum_i X_i - \theta_2 \cdot n = 0$$

*Die zweite Gleichung nach  $\theta_2$  auflösen:*

$$\theta_2 = \frac{1}{n} \sum_i Y_i - \theta_1 \frac{1}{n} \sum_i X_i$$

*und in die erste einsetzen:*

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \frac{1}{n} \sum_i Y_i \sum_i X_i + \theta_1 \frac{1}{n} \sum_i X_i \sum_i X_i = 0$$

$$\sum_i X_i Y_i - \frac{1}{n} \sum_i Y_i \sum_i X_i - \theta_1 \left( \sum_i X_i^2 - \frac{1}{n} \sum_i X_i \sum_i X_i \right) = 0$$

$\Rightarrow$

$$\hat{\theta}_1 = \frac{\sum_i X_i Y_i - \frac{1}{n} \sum_i X_i \sum_i Y_i}{\sum_i X_i^2 - \frac{1}{n} (\sum_i X_i)^2} = \frac{S_{XY}}{S_X^2}$$

$$\hat{\theta}_2 = \frac{1}{n} \left( \sum_i Y_i - \hat{\theta}_1 \sum_i X_i \right)$$

Sei  $\theta$  ein zu schätzender Parameter einer Population mit Dichte  $f$ .

Sei  $\hat{\theta} = \theta_n$  eine erwartungstreue Schätzung von  $\theta$ .

Dann gilt die Cramer-Rao-Ungleichung:

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)},$$

wobei

$$\begin{aligned} I(f, \theta) &= \mathbf{E} \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 \\ &= \int \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx \end{aligned}$$

die sogenannte Fisher-Information ist.

**Beispiele a)  $f$  normal,**

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\ln f(x, \mu) = -\ln(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x, \mu)}{\partial \mu} = \frac{x-\mu}{\sigma} \cdot \frac{1}{\sigma}$$

$$I(f, \mu) = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 \cdot f(x, \mu) dx = \frac{1}{\sigma^2}.$$

Also:

$$\text{var} \bar{X} \geq \frac{\sigma^2}{n}.$$

Vergleichen Sie mit:

$$\text{var} \bar{X} = \frac{1}{n^2} \sum_{i=1}^n \text{var} X_i = \frac{\sigma^2}{n}.$$

b)  $f$  exponential

$$f(x, \lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Es gilt:

$$I(f, \lambda) = \frac{1}{\lambda^2} \quad (\text{ÜA, 2 P.})$$

Die Cramer-Rao-Schranke ist also:

$$\frac{1}{nI(\lambda)} = \frac{\lambda^2}{n}.$$

Andererseits:

$$\text{var} \bar{X} = \frac{\lambda^2}{n}.$$

c)  $F$  Doppelsexponential (=Laplace)

$$f(x, \lambda) = \frac{1}{2} \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} & \text{falls } x \geq 0 \\ \frac{1}{\lambda} e^{\frac{1}{\lambda}x} & \text{falls } x < 0 \end{cases}$$

$$\ln f((x, \mu)) = -\ln 2 - \ln \lambda + \frac{x}{\lambda} \begin{cases} -1 & \text{falls } x \geq 0 \\ 1 & \text{falls } x < 0 \end{cases}$$

$$\frac{\partial \ln f(x, \lambda)}{\partial \lambda} = -\frac{1}{\lambda} + \frac{x}{\lambda^2} \begin{cases} 1 & \text{falls } x \geq 0 \\ -1 & \text{falls } x < 0 \end{cases}$$

$$\begin{aligned} I(f, \lambda) &= \frac{1}{2\lambda} \left( \int_0^{\infty} \left(-\frac{1}{\lambda} + \frac{x}{\lambda^2}\right)^2 \cdot e^{-\frac{x}{\lambda}} dx + \right. \\ &\quad \left. \int_{-\infty}^0 \left(-\frac{1}{\lambda} - \frac{x}{\lambda^2}\right)^2 \cdot e^{-\frac{x}{\lambda}} dx \right) \\ &= \frac{1}{2\lambda^2} + \frac{1}{2\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$



Cramer-Rao-Schranke

$$\frac{\lambda^2}{n}.$$

Vergleichen Sie mit (ÜA)

$$\text{var} \bar{X} = \frac{1}{n^2} \sum_{i=1}^n \text{var} X_i = \frac{2 \cdot \lambda^2}{n}.$$

**Bem. 23** Für den Median  $x_{0.5}$  gilt:

$$\text{var}(x_{0.5}) \sim \frac{1 \cdot \lambda^2}{n}.$$

*(zusätzliche freiwillige ÜA, 10 P.)*

**Satz:** Sei  $f$  Dichte der Population, und  $\hat{\theta}$  eine erwartungstreue Schätzung des Parameters  $\theta$ . Dann gilt:

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)},$$

wobei

$$I(f, \theta) = \mathbf{E} \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2$$

falls der Erwartungswert existiert. *Beweis:* Sei  $\mathbf{x} = (x_1, \dots, x_n)$  eine unabhängige Stichprobe und

$$L(\mathbf{x}, \theta) := \prod_{i=1}^n f(x_i, \theta)$$

die sogenannte Likelihood der Stichprobe.

Offenbar gilt

$$\int_{\mathbb{R}^n} L(\mathbf{x}, \theta) d\mathbf{x} = 1.$$

und demzufolge (wir setzen voraus, Differentiation und Integration dürfen vertauscht werden.)

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(\mathbf{x}, \theta) d\mathbf{x} &= 0 \\ \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} &= 0 \end{aligned}$$

Weiter gilt, da  $\hat{\theta}$  erwartungstreu,

$$\begin{aligned}\mathbf{E}\hat{\theta} &= \theta \\ \int_{\mathbb{R}^n} \hat{\theta} L(\mathbf{x}, \theta) d\mathbf{x} &= \theta \\ \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \hat{\theta} L(\mathbf{x}, \theta) d\mathbf{x} &= 1 \\ \int_{\mathbb{R}^n} \hat{\theta} \underbrace{\frac{\partial L(\mathbf{x}, \theta)}{\partial \theta}} d\mathbf{x} &= 1 \\ \int_{\mathbb{R}^n} \hat{\theta} \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} &= 1 \\ \mathbf{E}\left(\hat{\theta} \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta}\right) &= 1\end{aligned}$$

Auf den linken Term in der vorletzten Gleichung wenden wir die Cauchy-Schwarzsche Ungleichung an,

$$\begin{aligned}
1 &= \int_{\mathbb{R}^n} \hat{\theta} \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} - \underbrace{\theta \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x}}_{=0} \\
&= \int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} \\
&\leq \int_{\mathbb{R}^n} (\hat{\theta} - \theta)^2 L(\mathbf{x}, \theta) d\mathbf{x} \cdot \int_{\mathbb{R}^n} \left( \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} \right)^2 L(\mathbf{x}, \theta) d\mathbf{x} \\
&= \text{var}(\hat{\theta}) \cdot \int_{\mathbb{R}^n} \left( \frac{\partial \sum_{i=1}^n \ln f(x_i, \theta)}{\partial \theta} \right)^2 L(\mathbf{x}, \theta) d\mathbf{x} \\
&= \text{var}(\hat{\theta}) \cdot \sum_{i=1}^n \int_{\mathbb{R}^n} \left( \frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right)^2 L(\mathbf{x}, \theta) d\mathbf{x} \\
&= \text{var}\hat{\theta} \cdot n \cdot I(f).
\end{aligned}$$

= 0

Die zu den gemischten Summanden gehörenden Integrale sind alle Null, ( $i \neq j$ ):

$$\int_{\mathbb{R}^2} \left( \frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right) \left( \frac{\partial \ln f(x_j, \theta)}{\partial \theta} \right) f(x_i, \theta) f(x_j, \theta) dx_i dx_j$$
$$= \int_{\mathbb{R}^2} \frac{\partial f(x_i, \theta)}{\partial \theta} \frac{\partial f(x_j, \theta)}{\partial \theta} dx_i dx_j = 0.$$