14 Schätzmethoden

Eigenschaften von Schätzungen $\hat{\theta}$

Sei $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ eine Schätzung eines Parameters θ, die auf n Beobachtungen beruht.

- $\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$ "Konsistenz" (Minimalforderung)

- $E\hat{\theta}_n = \theta$ "Erwartungstreue"

  $E\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$ "Asymptotische Erw.treue"

- var $\hat{\theta}_n$ möglichst klein: “gute”, “effiziente” Schätzung

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

- MSE = var $\hat{\theta}_n$ + bias^2 $\hat{\theta}_n$
  = var $\hat{\theta}_n + (E\hat{\theta}_n - \theta)^2$
  $\longrightarrow$ minimal oder möglichst klein.

- Eigenschaften sollten “möglichst” auch bei (kleinen)
  Abweichungen von der
  (Normal-)Verteilungsannahme gelten
  $\longrightarrow$ robuste Schätzung.

Schätzmethoden

- Momentenmethode

  Man drückt den zu schätzenden Parameter durch die Momente, z.B. $\mathbf{E}(X)$, aus.
  Dann werden die Momente durch die entsprechenden *empirischen* Momente, z.B. durch $\overline{X}$, ersetzt.

- Maximum-Likelihood-Schätzung (ML-Schätzung)

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

- Kleinste-Quadrat-Schätzung (KQS)

  Sei θ der zu schätzende Parameter. Man geht aus von einem Modell, z.B.

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

  Dannn 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 $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^)$.

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

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

Bsp. 98 (Momentenschätzung bei Exponentialverteilung)
Seien $X_1, \ldots, X_n \sim Exp(\lambda)$.

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

Bsp. 99 (Momentenschätzung bei Binomialverteilung)
Seien $X_1, \ldots, X_n \sim Bi(1,p)$.

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

der relative Anteil der Realisierungen $x_i = 1$.

Bsp. 100 (ML-Schätzung bei Binomialverteilung)

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

$X_i = 1$ *falls Übergewicht festgestellt*

$X_i = 0$ *sonst.*

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

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

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).*

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

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

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

Likelihood:

$$\begin{aligned}
L_n(\mu) &= \prod_{i=1}^{n} f_{X_i}(x_i) \qquad \textit{(Unabhängigkeit)} \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(x_i-\mu)^2/2} \\
\ln L_n(\mu) &= -n\ln(\sqrt{2\pi}) + \sum_{i=1}^{n}(-\frac{(x_i-\mu)^2}{2}) \\
\frac{\partial L_n(\mu)}{\partial \mu} &= \sum_{i=1}^{n}(x_i-\mu)
\end{aligned}$$

Nullsetzen liefert:

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

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

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

die gemeinsame Dichtefunktion der X_i.

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

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

Likelihood:

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

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

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

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:

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

Bsp. 104 (KQS im einfachen linearen Regressionsmodell)

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

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

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

$\Rightarrow$

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

$\Rightarrow$

$$\begin{aligned}\sum_i X_iY_i - \theta_1 \sum_i X_i^2 - \theta_2 \sum_i X_i &= 0\\ \sum_i Y_i - \theta_1 \sum_i X_i - \theta_2 \cdot n &= 0\end{aligned}$$

Die zweite Gleichung nach θ_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:

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

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

$\Rightarrow$

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

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

Sei θ ein zu schätzender Parameter einer Population mit Dichte f.

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

Dann gilt die Cramer-Rao-Ungleichung:

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

wobei

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

die sogenannte Fisher-Information ist.

Beispiele a) f normal,

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

$$\begin{aligned}
\ln f(x,\mu) &= -\ln(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2} \\
\frac{\partial \ln f(x,\mu)}{\partial \mu} &= \frac{x-\mu}{\sigma} \cdot \frac{1}{\sigma} \\
I(f,\mu) &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} (\frac{x-\mu}{\sigma})^2 \cdot f(x,\mu)\, dx = \frac{1}{\sigma^2}.
\end{aligned}$$

Also:

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

Vergleichen Sie mit:

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

b) f exponential

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

Es gilt:

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

Die Cramer-Rao-Schranke ist also:

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

Andererseits:

$$var\overline{X} = \frac{1}{n\lambda^2} = \frac{1}{nI(\lambda^{-1})}.$$

c) F Doppelexponential (=Laplace)

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

$$
\begin{aligned}
\ln f(x,\mu) &= -\ln 2 + \ln\lambda + \lambda x \begin{cases} -1 & \text{falls} \quad x \geq 0 \\ 1 & \text{falls} \quad x < 0 \end{cases} \\
\frac{\partial \ln f(x,\lambda)}{\partial\lambda} &= \frac{1}{\lambda} - x \begin{cases} 1 & \text{falls} \quad x \geq 0 \\ -1 & \text{falls} \quad x < 0 \end{cases} \\
I(f,\lambda) &= \frac{1}{2}\Bigg(\int_0^\infty \big(\frac{1}{\lambda} - x\big)^2 \cdot \lambda e^{-\lambda x}\,dx + \\
&\qquad \int_{-\infty}^0 \big(\frac{1}{\lambda} + x\big)^2 \cdot \lambda e^{\lambda x}\,dx\Bigg) \\
&= \int_0^\infty \big(\frac{1}{\lambda^2} - \frac{2x}{\lambda} + x^2\big) \cdot \lambda e^{-\lambda x}\,dx \\
&= \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} = \frac{1}{\lambda^2}.
\end{aligned}
$$

Cramer-Rao-Schranke

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

Vergleichen Sie mit (ÜA)

$$var\overline{X} = \frac{1}{n^2}\sum_{i=1}^{n} varX_i = \frac{2}{\lambda^2 n}.$$

Bem. 23 *Finden Sie eine bessere Schätzung für den Parameter λ!*

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

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

wobei

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

falls der Erwartungswert existiert. *Beweis:* Sei $\mathbf{x} = (x_1, \ldots, x_n)$ 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,

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

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

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

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