4 Schätzmethoden

a) Maximum-Likelihood-Schätzer (ML-Schätzer)

ML-Methode (R. A. Fisher) setzt dominierte Verteilungsklasse
$\wp := \{P_\vartheta : \vartheta \in \Theta\}$ voraus. ($\Theta \subset \mathbb{R}^k$)
Im Folgenden: $(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$

4.1 Grundannahmen

1) $\exists\sigma$-endliches Maß μ auf $\mathcal{B}$ mit:

$$\forall N \in \mathcal{B} : \ \mu(N) = 0 \Rightarrow \ P_\vartheta(N) = 0 \ \forall\vartheta \in \Theta$$

d.h. P_ϑ stetig bzgl. μ $\forall\vartheta$.
($\Rightarrow$ P_ϑ besitzt Dichte bzgl. μ)

2) Im Folgenden stets

(i) $\mu = \lambda^n$ (Borel-Lebesgue-Maß)
($\rightarrow$ stetige Verteilung)

oder

(ii) μ =Zählmaß auf einer abzählbaren Menge $A \subset \mathbb{R}^n$.
($\rightarrow$ diskrete Verteilung)

Im Falle (i) bezeichne $f(x, \vartheta) = \frac{dP_\vartheta}{d\lambda^n}(x)$ die Lebesgue-Dichte[9] von X, also

$$P_\vartheta(X \in B) = \int_B f(x, \vartheta) d\lambda^n(x), \ B \in \mathcal{B}$$

Im Falle (ii) bezeichne $f(x, \vartheta) = \frac{dP_\vartheta}{d\mu}(x)$ die Zähldichte von X, also

$$f(x, \vartheta) = P_\vartheta(X = x), \ x \in A$$

$$P_\vartheta(X \in B) = \sum_{x \in B \cap A} f(x, \vartheta)$$

[9] Beachte Schreibweise aus Stochastik II!

4.2 Definition und Bemerkung

Für jedes $x \in \mathbb{R}^n$ heißt die Abbildung

$$L_x : \begin{cases} \Theta & \to & [0,\infty) \\ \vartheta & \mapsto & L_x(\vartheta) := f(x,\vartheta) \end{cases}$$

die Likelihood-Funktion zur Stichprobe x.

Jeder Wert $\hat{\vartheta}(x) \in \Theta$, der Lösung t von

$$L_x(t) = \sup_{\vartheta \in \Theta} L_x(\vartheta) \quad (*)$$

ist, heißt (ein) ML-Schätzwert für $\vartheta \in \Theta$

(i) Im Allgemeinen Existenz gesichert, falls Θ abgeschlossen ist.

(ii) Falls Θ nicht abgeschlossen, so häufig $\vartheta \mapsto f(x,\vartheta)$ auf $\bar{\Theta}$ fortsetzbar. Dann sieht man $\hat{\vartheta}(x)$ auch als Lösung an, wenn sup in $(*)$ im Punkt $\hat{\vartheta}(x) \in \bar{\Theta} \setminus \Theta$ angenommen wird.

Eine messbare Funktion $\hat{\vartheta} : (\mathbb{R}^n, \mathcal{B}^n) \to (\bar{\Theta}, \bar{\Theta} \cap \mathcal{B}^k)$ heißt **ML-Schätzer** für ϑ, wenn für jedes $x \in \mathfrak{X}$ gilt: $\hat{\vartheta}(x)$ ist Lösung von $(*)$ im obigen Sinn[10].

4.3 Bemerkungen

(i) Oft ist $L_x(\cdot) = f(x,\cdot)$ differenzierbar.
Dann liefert $\frac{\partial}{\partial\vartheta} f(x,\vartheta) \stackrel{!}{=} 0 \in \mathbb{R}^k$ die lokalen Maximalstellen von L_x im Inneren Θ^0 von Θ.

(ii) Oft: $X = (X_1, \ldots, X_n)$, $X_1, \ldots, X_n \stackrel{uiv}{\sim} f_1(\xi,\vartheta)$ [Dichte von X_1.]
Dann:

$$f(x,\vartheta) = \prod_{j=1}^{n} f_1(x_j,\vartheta), \; x = (x_1, \ldots, x_n)$$

<u>Log-Likelihood-Funktion</u>

$$\log L_x(\vartheta) = \sum_{j=1}^{n} \log f_1(x_j,\vartheta)$$

$\frac{\partial}{\partial\vartheta} \log L_x(\vartheta) \stackrel{!}{=} 0 \;\to$ Maximalstellen von L_x in Θ^0

[10] siehe Punkt (ii)

4.4 **Satz** (Invarianzprinzip für ML-Schätzer)

Sei $g: \mathbb{R}^k \to \mathbb{R}^l$ messbar und

$$M_x(\gamma) := \sup_{\vartheta:\ g(\vartheta)=\gamma} L_x(\vartheta)$$

(sogenannte von g induzierte Likelihood-Funktion)

Ist $\hat{\vartheta}$ ML-Schätzer für $\vartheta \in \Theta$, so ist $\hat{\gamma} := g(\hat{\vartheta})$ der ML-Schätzer für $\gamma = g(\vartheta) \in \Gamma := g(\Theta)$, es gilt also $M(\hat{\gamma}) \geq M(\gamma)\ \forall \gamma \in \Gamma$.
(Plug-In-Methode)

Beweis:[11]
Aus

$$M_x(g(\hat{\vartheta})) = \sup_{\vartheta:\ g(\vartheta)=g(\hat{\vartheta})} L_x(\vartheta) \geq L_x(\hat{\vartheta})$$

und

$$M_x(g(\hat{\vartheta})) \leq \sup_{\gamma\in\Gamma} M_x(\gamma) = L_x(\hat{\vartheta})$$

folgt

$$M_x(g(\hat{\vartheta})) = L_x(\hat{\vartheta}) \geq M_x(\gamma) \quad \forall \gamma \in \Gamma$$

4.5 Beispiel

$X_1, \ldots, X_n \overset{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$

$$\hat{\vartheta}(x) = (\bar{X}_n, \hat{\sigma}_n^2)$$

$\Rightarrow$ $\bar{X}_n$ ist ML-Schätzer für μ
$\hat{\sigma}_n^2$ ist ML-Schätzer für σ^2
$\hat{\sigma}_n = +\sqrt{\hat{\sigma}_n^2}$ ist ML-Schätzer für σ

b) Minimum-Quadrat-Schätzer (MQ-Schätzer)

4.6 Situation

Seien $X_1, \ldots, X_n$ stochastisch unabhängig.

Annahme:
$EX_j = \mu_j(\vartheta)$, wobei $\vartheta \in \mathbb{R}^p$ **unbekannt**, $\mu_j : \mathbb{R}^p \to \mathbb{R}$ $(j = 1, \ldots, n)$ **bekannte Regressionsfunktionen**.

[11] *In der 1. Zeile gilt eigentlich bereits Gleichheit.*

Für $\varepsilon_j := X_j - EX_j$ gilt dann:
$\varepsilon_1, \ldots, \varepsilon_n$ unabhängig, $E(\varepsilon_j) = 0 \ \forall j$, $X_j = \mu_j(\vartheta) + \varepsilon_j$ $(j = 1, \ldots, n)$ bzw.

$$X = \mu(\vartheta) + \varepsilon$$

wobei

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \ \mu(\vartheta) = \begin{pmatrix} \mu_1(\vartheta) \\ \vdots \\ \mu_n(\vartheta) \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Schätzmethode von ϑ durch Methode der **kleinsten Quadrate**, d.h. durch Minimierung der Fehlerquadratsumme

$$Q(\vartheta) := \sum_{j=1}^{n} (X_j - \mu_j(\vartheta))^2 = \|X - \mu(\vartheta)\|^2$$

Sind $\mu_1, \ldots, \mu_n$ stetig differenzierbar, so gilt mit

$$M(\vartheta) := \begin{pmatrix} \frac{\partial}{\partial\vartheta_1}\mu_1(\vartheta) & \cdots & \frac{\partial}{\partial\vartheta_p}\mu_1(\vartheta) \\ \vdots & & \vdots \\ \frac{\partial}{\partial\vartheta_1}\mu_n(\vartheta) & \cdots & \frac{\partial}{\partial\vartheta_p}\mu_n(\vartheta) \end{pmatrix}$$

$$\frac{\partial}{\partial\vartheta} Q(\vartheta) = -2 \cdot M^T(\vartheta) \cdot (X - \mu(\vartheta))$$

Beweis:
Sei allgemein $f, g: \ \mathbb{R}^p \to \mathbb{R}^q$.
Jacobi-Matrix

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_p} \\ \vdots & & \vdots \\ \frac{\partial f_q}{\partial x_1} & \cdots & \frac{\partial f_q}{\partial x_p} \end{pmatrix} \in \mathbb{R}^{q \times p}$$

$$\Rightarrow \underbrace{J^T_{f^T \cdot g}}_{\in \mathbb{R}^p} = \underbrace{\underbrace{J^T_f}_{\in \mathbb{R}^{p \times q}} \cdot g}_{\in \mathbb{R}^p} + \underbrace{J^T_g \cdot f}_{\in \mathbb{R}^p} \qquad (*)$$

[Beachte: f,g vektorwertig!]

Hier speziell: $f(\vartheta) = g(\vartheta) = X - \mu(\vartheta)$

$$\Rightarrow Q(\vartheta) = f^T(\vartheta) \cdot f(\vartheta)$$

$$\Rightarrow J_f = -M(\vartheta) = J_g$$

$$\begin{aligned}\overset{(*)}{\Rightarrow}\ \frac{\partial}{\partial\vartheta}[f^T(\vartheta)\cdot f(\vartheta)] &= -M^T(\vartheta)g - M^T(\vartheta)f \\ &= -2M^T(\vartheta)(X-\mu(\vartheta))\end{aligned}$$

Die Lösungen $\hat{\vartheta}$ von

$$Q(\hat{\vartheta}) = \min_{\vartheta\in\mathbb{R}^p} Q(\vartheta)$$

(sogenannte MQ-Schätzer) befinden sich also unter den Lösungen ϑ der sogenannten **Normalengleichung**

$$M^T(\vartheta)\cdot\mu(\vartheta) = M^T(\vartheta)\cdot X$$

4.7 Beispiel (Einfach lineare Regression)

$\vartheta = (\vartheta_0, \vartheta_1) \in \mathbb{R}$

$$\mu_i(\vartheta) = \vartheta_0 + \vartheta_1 t_i \ (i = 1, \dots, n)$$

t_i bekannt, nicht alle gleich.

$$Q(\vartheta) = \sum_{i=1}^n (X_i - \vartheta_0 - \vartheta_1 t_i)^2 = \min_{\vartheta_0,\vartheta_1}!$$

$$M^T(\vartheta) = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix}$$

Normalengleichung:

$$\begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix}\begin{pmatrix} \vartheta_0+\vartheta_1 t_1 \\ \vdots \\ \vartheta_0+\vartheta_1 t_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\Leftrightarrow n\vartheta_0 + \vartheta_1\sum_{i=1}^n t_i = \sum_{i=1}^n x_i$$

$$\vartheta_0\sum_{i=1}^n t_i + \vartheta_1\sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i x_i$$

Mit $\bar{t} = \frac{1}{n}\sum_{i=1}^n t_i,\ \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ folgt

$$\boxed{\begin{aligned}\hat{\vartheta}_0 &= \bar{x} - \hat{\vartheta}_1\bar{t} \\ \hat{\vartheta}_1 &= \frac{\sum_{i=1}^n t_i x_i - n\bar{t}\bar{x}}{\sum_{i=1}^n (t_i-\bar{t})^2}\end{aligned}}$$

Wegen[12]

$$\sum_i a_i b_i - n\bar{a}\bar{b} = \sum_i (a_i - \bar{a})(b_i - \bar{b}) = \sum_i (a_i - \bar{a}) b_i$$

folgt

$$\hat{\vartheta}_1 = \hat{\vartheta}_1(X) = \frac{\sum_{i=1}^n (t_i - \bar{t}) x_i}{\sum_{i=1}^n (t_i - \bar{t})^2}$$

und somit

$$E(\vartheta_1) = \frac{1}{\sum_i (t_i - \bar{t})^2} \sum_i (t_i - \bar{t})(\vartheta_0 + \vartheta_1 t_i) = \vartheta_1$$

Falls $\mathrm{Var}(X_i) = \sigma^2 \ \forall i$, so gilt:

$$\mathrm{Var}(\hat{\vartheta}_1) = \frac{1}{(\sum_i (t_i - \bar{t})^2)^2} \sum_i (t_i - \bar{t})^2 \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^n (t_i - \bar{t})^2}$$

[$\mathrm{Var}(\hat{\vartheta}_1) = \mathrm{MQA}$, da erwartungstreu; t_i so wählen, dass $\mathrm{Var}(\hat{\vartheta}_1)$ klein wird, also möglichst weit auseinander.]
Weiter gilt

$$E(\hat{\vartheta}_0) = E\bar{X} - \bar{t}E(\hat{\vartheta}_1) = \frac{1}{n}\sum_{i=1}^n (\vartheta_0 + \vartheta_1 t_i) - \bar{t}\vartheta_1 = \vartheta_0 + \vartheta_1 \bar{t} - \vartheta_1 \bar{t} = \vartheta_0$$

Bemerkungen:

(i) Falls $\mathrm{Var}(X_i) = \sigma^2 \ \forall i$ ($\mathrm{Cov}(X_i, X_j) = 0 \ \forall i \neq j$ wegen Unabhängigkeit[13]), so gilt mit $\bar{t^2} = \frac{1}{n}\sum_{i=1}^n t_i^2$

$$\mathrm{Var}(\hat{\vartheta}_0) = \frac{\sigma^2 \bar{t^2}}{n(\bar{t^2} - (\bar{t})^2)}$$

(ii) Falls $X_i \sim \mathcal{N}(\mu, \sigma^2) \ \forall i$, so ist der MQ-Schätzer auch ML-Schätzer für ϑ

[12] $\sum (a_i - \bar{a})\bar{b} = \bar{b} \sum (a_i - \bar{a}) = 0$
[13] Voraussetzung!

c) Momentenmethode

4.8 Definition

Es seien $X_1, \dots, X_n \overset{uiv}{\sim} X$, X reellwertig, $P^X \in \{P_\vartheta : \vartheta \in \Theta\},\ \Theta \subset \mathbb{R}^k$

$$\vartheta = (\vartheta_1, \dots, \vartheta_k)$$

Annahme

(i) $E\left|X^k\right| < \infty$

(ii) Es gibt Funktionen $g_1, \dots, g_k : \mathbb{R}^k \to \mathbb{R}$ mit

$$\vartheta_1 = g_1(EX, \dots, EX^k)$$

$$\vdots$$

$$\vartheta_k = g_k(EX, \dots, EX^k)$$

Sei $\bar{X_n^l} = \frac{1}{n} \sum_{j=1}^n X_j^l \ (l = 1, \dots, k)$.
Dann ist der Momentenschätzer für ϑ

$$\hat{\vartheta} := \begin{pmatrix} g_1(\bar{X_n^1}, \dots, \bar{X_n^k}) \\ \vdots \\ g_k(\bar{X_n^1}, \dots, \bar{X_n^k}) \end{pmatrix}$$

Beispiel:
$X_i \overset{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\mu = EX$, $\sigma^2 = EX^2 - (EX)^2$
$\Rightarrow \hat{\mu}_n = \bar{X_n^1},\ \hat{\sigma}_n^2 = \bar{X_n^2} - (\bar{X_n^1})^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

Probleme:
$g_1, \dots, g_k$ nicht explizit gegeben.
Ausreißeranfälligkeit.
Momentenschäter sind nicht „robust“.

Beachte
Momentenschätzer sind konsistent, falls $g_1, \dots, g_k$ stetig sind an der Stelle $(EX, \dots, EX^k)$.

4.9 Beispiel (Gamma-Verteilung)

$X_1, \ldots, X_n \overset{uiv}{\sim} \Gamma(\alpha, \beta)$, Dichte

$$f(x, \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \ (x > 0)$$

$\vartheta = (\alpha, \beta) \in \Theta = \mathbb{R}_{>0} \times \mathbb{R}_{>0}$
$X \sim f(x, \alpha, \beta) \Rightarrow EX = \frac{\alpha}{\beta},\ EX^2 = \frac{\alpha(\alpha+1)}{\beta^2} \ \Rightarrow$

$$\alpha = \frac{(EX)^2}{EX^2 - (EX)^2} =: g_1(EX, EX^2)$$

$$\beta = \frac{EX}{EX^2 - (EX)^2} =: g_2(EX, EX^2)$$

$\Rightarrow$ Momentenschätzer

$$\hat{\alpha} = \frac{(\bar{X_n^1})^2}{\bar{X_n^2} - (\bar{X_n^1})^2} = \frac{\bar{X_n}^2}{\hat{\sigma_n^2}}$$

$$\hat{\beta} = \frac{\bar{X_n}}{\hat{\sigma_n^2}}$$

d) <u>Ein nichtparametrisches Schätzprinzip</u>
Seien $X_1, \ldots, X_n \overset{uiv}{\sim} F,\ F(t) = P(X \leq t),\ t \in \mathbb{R}$, F unbekannt

4.10 Definition

Die durch

$$\hat{F_n}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i \leq t\},\ t \in \mathbb{R}$$

definierte Funktion heißt **empirische Verteilungsfunktion** (EVF) von $X_1, \ldots, X_n$.
Die Realisierungen von $\hat{F_n}$ sind Treppenfunktionen.

$$\hat{F_n}(t_0) \overset{f.s.}{\to} E[\mathbf{1}\{X_1 \leq t_0\}] = P(X_1 \leq t_0) = F(t_0)$$

4.11 Satz von Glivenko-Cantelli

Sei $\hat{F_n^\omega}(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i(\omega) \leq t\},\ \omega \in \Omega$.
Falls $X_1, \ldots, X_n \overset{uiv}{\sim} F$ auf Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, so gilt

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} \underbrace{\sup_{t\in\mathbb{R}} \left|\hat{F_n^\omega}(t) - F(t)\right|}_{=:\left\|\hat{F_n^\omega} - F\right\|_\infty} = 0\}) = 1$$

Kurz: $\left\|\hat{F_n^\omega} - F\right\|_\infty \to 0 \quad \mathbb{P} - f.s.$
(Stochastik II, Henze)

4.12 $\hat{F}_n$ als nichtparametrischer ML-Schätzer

Sei $\mathfrak{F}$ die Menge aller Verteilungsfunktionen auf $\mathbb{R}$, $X_1, \ldots, X_n \overset{uiv}{\sim} F \in \mathfrak{F}$.
Sei P_F das zu F gehörende Wahrscheinlichkeitsmaß auf $\mathcal{B}^1$, also

$$P_F([a,b]) = F(b) - F(a),\ a < b$$

$$P_F(\{x\}) = F(x) - F(x-0),\ x \in \mathbb{R}$$

Sei $(x_1, \ldots, x_n)$ Realisierung von $(X_1, \ldots, X_n)$. Die durch

$$L_x : \begin{array}{ccc} \mathfrak{F} & \to & [0,\infty) \\ G & \mapsto & L_x(G) := \prod_{i=1}^n P_G(\{x_i\}) \end{array}$$

definierte Funktion heißt nichtparametrische Likelihood-Funktion zu $x = (x_1, \ldots, x_n)$.
Beachte: $L_x(G) = 0$, falls $P_G(\{x_i\}) = 0$ für ein i.[14]

Behauptung:
$L_x(\cdot)$ wird maximal für $G(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{x_i \leq t\}$.

Beweis:
Seien $z_1, \ldots, z_k$ die unterschiedlichen Werte unter $x_1, \ldots, x_n$, $n_1, \ldots, n_k$ die entsprechenden Vielfachheiten.

$$L_x(G) = \prod_{i=1}^n P_G(\{x_i\}) = \prod_{j=1}^k \underbrace{P_G(\{z_j\})}_{=:p_j}{}^{n_j} = \prod_{j=1}^k p_j^{n_j}$$

Setze $\hat{p}_j := \frac{n_j}{n}, j = 1, \ldots, k$, Verteilungsfunktion ist $\hat{F}_n$.

[14] z.B. für G stetig

F sei beliebige Verteilungsfunktion mit $p_j := F(z_j) - F(z_j - 0) > 0, j = 1, \dots, k$ mit $p_j \neq \hat{p}_j$ für mindestens ein j.
Es gilt für $x > 0$:

$$\log x \leq x - 1 \quad (*)$$

$\log x = x - 1$ nur für $x = 1$.

$$\begin{aligned}
\log\left(\frac{L_x(F)}{L_x(\hat{F}_n)}\right) &= \sum_{j=1}^{k} n_j \cdot \log(\frac{p_j}{\hat{p}_j}) \\
&= n \sum_{j=1}^{k} \hat{p}_j \cdot \log(\frac{p_j}{\hat{p}_j}) \\
&\overset{(*)}{<} n \sum_{j=1}^{k} \hat{p}_j (\frac{p_j}{\hat{p}_j} - 1) \\
&= n(\underbrace{\sum_{j=1}^{k} p_j}_{\leq 1} - \underbrace{\sum_{j=1}^{k} \hat{p}_j}_{=1}) \\
&\leq 0
\end{aligned}$$

$\Rightarrow L_x(F) < L_x(\hat{F}_n)$ ■

4.13 Nichtparametrisches Schätzprinzip

Seien $X_1, \dots, X_n \overset{uiv}{\sim} F$, $F \in \mathfrak{F}$, $\mathfrak{F}$ Menge von Verteilungsfunktionen (Verteilungsannahme).
Sei $\gamma : \ \mathfrak{F} \to \mathbb{R}$ Funktional.

Interessierender Parameter sei $\gamma(F)$.
„Rezept“: $\boxed{\text{Schätze } \gamma(F) \text{ durch } \gamma(\hat{F}_n)}$

4.14 Beispiele

a) $\mathfrak{F} := \{F : \underbrace{\int |x| F(dx)}_{=E|X_1|} < \infty\}$

$$\gamma(F) := \int x F(dx) (= EX_1)$$

$$\gamma(\hat{F}_n) = \int x \hat{F}_n(dx) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n$$

b) $\mathfrak{F} := \{F : \int x^2 F(dx) < \infty\}$

$$\gamma(F) := \int (x - \int y dF(y))^2 dF(x) = \mathrm{Var}(X_1)$$

$$\gamma(\hat{F}_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

c) $\mathfrak{F} := \{F : F \text{ hat Lebesgue-Dichte } f\}$

$$\gamma(F) = F'(t_0) = f(t_0)$$

$$\gamma(\hat{F}_n) = ?$$