8 Asymptotik von Schätzfehlern

8.1 Problemstellung

Seien $X_1, X_2, \ldots, X_n \overset{uiv}{\sim} P_\vartheta$, mit $\vartheta \in \Theta \subset \mathbb{R}^k$.
Die Schätzfolge $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \ldots, X_n)$ sei konsistent, es gilt also

$$\hat{\vartheta}_n \overset{P_\vartheta}{\to} \vartheta \text{ für } n \to \infty \quad \forall \vartheta \in \Theta$$

Sei (a_n) eine reelle Folge mit $a_n > 0 \quad \forall n$ und $a_n \to \infty$ für $n \to \infty$.
Die Folge $(\hat{\vartheta}_n)_{n\geq 1}$ heißt a_n- **konsistent,** wenn

$$a_n(\hat{\vartheta}_n - \vartheta) = O_{P_\vartheta}(1) \quad \forall \vartheta \in \Theta.$$

Hierbei bedeutet $Y_n = O_P(1)$ für eine Folge (Y_n), dass für jedes $\varepsilon > 0$ eine kompakte Menge $K \subset \mathbb{R}^d$ existiert, so dass $P(Y_n \in K) \geq 1 - \varepsilon$ für alle $n \in \mathbb{N}$.[22]

Typischerweise liegt $\sqrt{n}-$Konsistenz vor, d.h. es gilt

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = O_{P_\vartheta}(1) \quad \forall \vartheta \in \Theta.$$

Zusätzlich kann man oftmals Aussagen über Konvergenz in Verteilung machen, insbesondere

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \overset{D_\vartheta}{\to} \mathcal{N}_k(0, \Sigma(\vartheta)), \quad \vartheta \in \Theta.$$

8.2 Multivariater Zentraler-Grenzwert-Satz (ZGWS)

Seien $Y_1, Y_2 \ldots \overset{uiv}{\sim} Y$ mit einer $\mathbb{R}^d-$wertigen Zufallsvariablen Y mit
$E\|Y\|^2 < \infty$. Mit $a := EY$ und $\Sigma := E(Y-a)(Y-a)^T$, gilt

$$\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n} Y_j - na\right) \overset{D}{\to} \mathcal{N}_d(0, \Sigma).$$

[22] vergleiche Stochastik II: Straffheit

8.3 δ-Methode

Seien $Z_1, Z_2, \ldots$ d-dimensionale Zufallsvariablen mit

$$\sqrt{n}\,(Z_n - a) \quad \xrightarrow{D} \quad \mathcal{N}_d(0, \Sigma),$$

mit $a := (a_1, \ldots, a_d) \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d \times d}$.

Sei $g := (g_1, \ldots, g_s)^T : \ \mathbb{R}^d \to \mathbb{R}^s$ differenzierbar und

$$\frac{dg}{da} := \left(\frac{\partial g_j}{\partial a_k} \right)_{\substack{1 \le j \le s, \\ 1 \le k \le d}}$$

dann gilt

$$\sqrt{n}\,(g(Z_n) - g(a)) \quad \xrightarrow{D} \quad \mathcal{N}_s \left(0, \frac{dg}{da}\,\Sigma \left(\frac{dg}{da} \right)^T \right).$$

Beweis:
Nach der Definition der Differenzierbarkeit gilt

$$\sqrt{n}\,(g(Z_n) - g(a)) \quad = \quad \underbrace{\frac{dg}{da}\sqrt{n}(Z_n - a)}_{=:U_n} + \underbrace{\|\sqrt{n}(Z_n - a)\| \cdot r(Z_n - a)}_{=:V_n},$$

mit $r(Z_n - a) \to 0$ für $Z_n \to a$.
Beachte, dass $\|\sqrt{n}(Z_n - a)\| \in O_p(1)$.

Aus $Z_n \xrightarrow{P} a$ folgt, dass $r(Z_n - a) \xrightarrow{P} 0$, und somit

$$V_n \quad \xrightarrow{P} \quad 0.$$

Aus der Voraussetzung folgt mit dem Abbildungssatz weiter

$$U_n \xrightarrow{D} \frac{dg}{da} \cdot T$$

mit $T \sim \mathcal{N}_d(0, \Sigma)$ und somit

$$U_n \quad \xrightarrow{D} \quad \mathcal{N}_s \left(0, \frac{dg}{da}\,\Sigma \left(\frac{dg}{da} \right)^T \right).$$

Die Behauptung folgt schließlich aus dem Lemma von Slutzky.

8.4 Asymptotik des Momentenschätzers (vgl. 4.8)

$X_1, \dots, X_n \overset{uiv}{\sim} X,\ X\ \mathbb{R}^1$ -wertig, $P^X \in \{P_\vartheta : \vartheta \in \Theta\},\ \Theta \subset \mathbb{R}^k$,
$\vartheta = (\vartheta_1, \dots, \vartheta_n)$
Sei $m_l := EX^l,\ \hat{m}_l = \frac{1}{n}\sum_{i=1}^n X_i^l\ (\hat{=} \bar{X}_n^l$ aus 4.8)
Voraussetzung:
$\vartheta = g(m_1, \dots, m_k)$ mit $g : \mathbb{R}^k \to \mathbb{R}^k$
Momentschätzer: $\hat{\vartheta} = g(\hat{\vartheta}_1, \dots, \hat{\vartheta}_k)$
Sei

$$Y_j := \begin{pmatrix} X_j \\ \vdots \\ X_j^k \end{pmatrix} Y := \begin{pmatrix} X \\ \vdots \\ X^k \end{pmatrix},\ a := EY = \begin{pmatrix} m_1 \\ \vdots \\ m_k \end{pmatrix}$$

$E\|Y\|^2 < \infty \Leftrightarrow EX^{2k} < \infty$

$$\begin{aligned} \Sigma := E[(Y-a)(Y-a)^T] &= (E[(X^i - m_i)(X^j - m_j)^T])_{i,j=(1,\dots,k)} \\ &= (EX^{i+j} - m_i m_j)_{i,j} \\ &= (m_{i+j} - m_i m_j)_{i,j} \end{aligned}$$

8.2 $\Rightarrow$

$$\begin{aligned} \frac{1}{\sqrt{n}}\left(\sum_{j=1}^n Y_j - na\right) &= \frac{1}{\sqrt{n}}\left(\begin{pmatrix} \sum_j X_j \\ \vdots \\ \sum_j X_j^k \end{pmatrix} - n \cdot \begin{pmatrix} m_1 \\ \vdots \\ m_k \end{pmatrix}\right) \\ &= \sqrt{n} \cdot \begin{pmatrix} \hat{m}_1 - m_1 \\ \vdots \\ \hat{m}_k - m_k \end{pmatrix} \qquad \overset{D_\vartheta}{\to} \mathcal{N}(0, \Sigma(\vartheta)) \end{aligned}$$

Aus 8.3 folgt: Falls $EX^{2k} < \infty$ und g differenzierbar, so gilt:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \overset{D_\vartheta}{\to} \mathcal{N}_k(0,\ \frac{dg}{da}\sum(\vartheta)(\frac{dg}{da})T)$$

Achtung: Σ hängt von $m_1, \dots, m_{2k}$ und somit von unbekanntem ϑ ab.

(Schreibweise „asymptotisch normalverteilt“:)
$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \approx \mathcal{N}_k(0, T) \Leftrightarrow \hat{\vartheta}_n \approx \mathcal{N}_k(\vartheta, \frac{T}{n}),\ \hat{\vartheta}_n \sim AN(\vartheta, \frac{\hat{T}}{n}),\ \hat{T} = T(\hat{\vartheta}_n)$

8.5 Asymptotik des ML-Schätzers

$X_1, \ldots, X_n \overset{uiv}{\sim} f_1(\xi,\ \vartheta)$ (Dichte bezüglich dominierendem Maß μ)
$\vartheta \in \Theta \subset \mathbb{R}^k,\ \Theta$ offen
Regularitätsvoraussetzungen: (a)-(e) aus 5.7- 5.9 seien erfüllt.
Zusätzlich gelte:
$\{\xi : f_1(\xi, \vartheta) > 0\}$ist unabhängig von ϑ!
$\forall i, j, l \in \{1, \ldots, k\}$ existiert $\frac{\partial^3 \log f_1(\xi,\vartheta)}{\partial\vartheta_i \partial\vartheta_j \partial\vartheta_l} = L_{ijl}(\xi, \vartheta)$
$\forall \vartheta \in \Theta\ \forall \delta > 0\ \forall i, j, l \in \{1, \ldots, k\}$ existiert eine Funktion $M_{i,j,l}(\xi) \geq 0$ mit

$$|L_{i,j,l}(\xi, \eta)| \leq M_{i,j,l}(\xi),\ \|\eta - \vartheta\| \leq \delta$$

und $E_\vartheta M_{i,j,l}(X_1) < \infty$

Sei

$$\mathcal{U}_n(\vartheta) := \sum_{j=1}^{n} \frac{d}{d\vartheta} \log f_1(X_j, \vartheta),\ E_\vartheta \mathcal{U}_n(\vartheta) = 0$$

$$I_n(\vartheta) = E[\mathcal{U}_n(\vartheta)\mathcal{U}_n(\vartheta)^T] = nI_1(\vartheta)$$

$$W_n(\vartheta) = \frac{d}{d\vartheta^T}\mathcal{U}_n(\vartheta),\ E_\vartheta[W_n(\vartheta)] = -I_n(\vartheta)$$

8.5.1 Satz

Es gelte $\mathcal{U}_n(\hat{\vartheta}_n) = 0$ (d.h. $\hat{\vartheta}_n$ ist Lösung der Likelihood-Gleichung $\mathcal{U}_n(\vartheta) = 0$) und $\hat{\vartheta}_n \overset{P_\vartheta}{\to} \vartheta,\ \vartheta \in \Theta$ (d.h. $(\hat{\vartheta}_n)_n$ ist konsistent). Dann folgt:

$$\boxed{\sqrt{n}(\hat{\vartheta}_n - \vartheta) \overset{D_\vartheta}{\to} \mathcal{N}_k(0, I_1(\vartheta)^{-1})}$$

$\Leftrightarrow \hat{\vartheta}_n \sim AN(\vartheta, \frac{I_n^{-1}}{n})$

Beweisskizze:

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\mathcal{U}_n(\hat{\vartheta}_n) \\ &\overset{\text{Taylor}}{=} \frac{1}{\sqrt{n}}\mathcal{U}_n(\vartheta) + \frac{1}{\sqrt{n}}W_n(\vartheta)(\hat{\vartheta}_n - \vartheta) + \underbrace{\frac{1}{\sqrt{n}}R_n(\vartheta, \hat{\vartheta}_n - \vartheta)}_{z.z.:\ = o_{P_\vartheta}(1)} \end{aligned}$$

$$\Rightarrow \underbrace{\frac{1}{n}W_n(\vartheta)}_{\overset{P_\vartheta - f.s}{\to} -I_1(\vartheta)(SGGZ)} \quad \sqrt{n}(\hat{\vartheta}_n - \vartheta) = \underbrace{\underbrace{-\frac{1}{\sqrt{n}}\mathcal{U}_n(\vartheta)}_{\overset{D_\vartheta}{\to}\mathcal{N}_k(0,I_1(\vartheta))\ \text{(ZGWS 8.2)}} \quad + o_{P_\vartheta}(1)}_{\overset{D_\vartheta}{\to}\mathcal{N}_k(0,I_1(\vartheta))} \ (*)$$

$$\Rightarrow \sqrt{n}(\hat{\vartheta}_n - \vartheta) \overset{D_\vartheta}{\to} \mathcal{N}_k(0, -I_1(\vartheta)^{-1} I_1(\vartheta)(-I_1(\vartheta)^{-1}))$$

(z.B. Knight, 249 oder Lehmann/Casella, 443-468)[23]

Bemerkung: (asymptotische Linearisierbarkeit des Schätzfehlers)

$$\begin{aligned}(*) \Rightarrow \sqrt{n}(\hat{\vartheta}_n - \vartheta) &= I_1(\vartheta)^{-1} \frac{1}{\sqrt{n}} \mathcal{U}_n(\vartheta) + o_{P_\vartheta}(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \underbrace{I_1(\vartheta)^{-1} \frac{d}{d\vartheta} \log f_1(X_j, \vartheta)}_{=:\widetilde{l}(X_j,\vartheta)} + o_{P_\vartheta}(1)\end{aligned}$$

mit $E_\vartheta \widetilde{l}(X_1, \vartheta) = 0$

8.5.2 Satz

Unter den obigen Voraussetzungen existiert eine Folge $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \ldots, X_n)$ mit:
Ist ϑ_0 der wahre Parameter, so gilt:

$$\lim P_{\vartheta_0}(\mathcal{U}_n(\hat{\vartheta}_n) = 0, \left|\hat{\vartheta}_n - \vartheta_0\right| \leq \varepsilon) = 1 \ \forall \varepsilon > 0$$

Korollar
Besitzt die Likelihood-Gleichung $\mathcal{U}_n(\vartheta) = 0$ für jedes n eine eindeutige Lösung $\hat{\vartheta}_n$, so gilt:

$$\hat{\vartheta}_n \overset{P_\vartheta}{\to} \vartheta, \ \vartheta \in \Theta$$

Anmerkung:

(1) $\{P_\vartheta : \vartheta \in \Theta\}$ mit Dichte $f(x, \vartheta)$ bzgl. dem Maß μ. Dann $\forall \vartheta \neq \vartheta_0$:

$$\begin{aligned} E_{\vartheta_0}\left[\log \frac{f(X,\vartheta)}{f(X,\vartheta_0)}\right] &\overset{\text{Jensensche Ungl.}}{<} \log \underbrace{E_{\vartheta_0}\left[\frac{f(X,\vartheta)}{f(X,\vartheta_0)}\right]}_{\int f(x,\vartheta)dx=1} \\ &= 0 \end{aligned}$$

$$\Leftrightarrow E_{\vartheta_0}[\log f(X, \vartheta_0)] > E_{\vartheta_0}[\log f(X, \vartheta)] \quad \forall \vartheta \neq \vartheta_0$$

d.h. $\boxed{\vartheta_0 \text{ maximiert } E_{\vartheta_0}[\log f(X,\vartheta)] \text{ bezüglich } \vartheta!}$

[23] $o_{P_\vartheta}(1)$ bedeutet stochastische Konvergenz gegen 0

(2) Funktional in $(*)$ ist nicht auswertbar, da ϑ_0 unbekannt!
Aber:

$$\frac{1}{n} \underbrace{\sum_{i=1}^{n} \log f(X_i, \vartheta)}_{l(X,\vartheta),\ \text{„Log-Likelihood Funktion“}} \overset{P_\vartheta - f.s.}{\to} E_{\vartheta_0}[\log f(X,\vartheta)] \forall \vartheta \in \Theta$$

Maximierung von $l(X,\vartheta)$ als „Ersatz“ für $(*)$.

(3) f,g μ-Dichten:

$$E_f\left[\log \frac{f(X)}{g(X)}\right] \geq 0$$

$$\text{“=“} \Leftrightarrow f = g$$

„Entropie“ zwischen f und g, Kullbach-Leibler-Information von g bezüglich f, Kullbach-Leibler-Abstand zwischen f und g

(4) Was tun, falls Lösung von $\mathcal{U}_n(\vartheta) = 0$ nicht eindeutig?

(i) Oft ist die Folge von globalen Maxima konsistent.
(Theorie von Wald 1949, Le Cam 1953)

(ii) Sei (δ_n) konsistent. Wähle Folge ϑ_n^*, die am nächsten zu δ_n liegt $\Rightarrow (\vartheta_n^*)$ konsistent, 8.5.1 anwendbar.

(iii) 1-Schritt-MLE verwenden:
$(\vartheta_n^{(0)})$ sei $\sqrt{n}$-konsistent. Mache einen Newton-Schritt zur Lösung von $\mathcal{U}_n(\vartheta) = 0$:

$$\vartheta_n^{(1)} = \vartheta_n^{(0)} - \frac{\mathcal{U}_n'(\vartheta_n^{(0)})}{\mathcal{U}_n''(\vartheta_n^{(0)})}$$

Dann hat $(\vartheta_n^{(1)})_{n\geq 1}$ dasselbe asymptotische Verhalten wie in 8.5.1.