3.4 Asymptotische Evaluierung von Schätzer

3.4.1 Konsistenz

Bis jetzt haben wir Kriterien basierend auf endlichen Stichproben betrachtet. **Konsistenz** ist ein asymptotisches Kriterium ($n \to \infty$) und bezieht sich auf eine Folge von Schätzer $W_n = W_n(X_1, \ldots, X_n)$, z.B. auf die Folge empirischer Mittel $\overline{X}_1 = X_1$, $\overline{X}_2 = (X_1 + X_2)/2$, $\ldots$, $\overline{X}_n = \sum_i X_i/n, \ldots$.

Definition 3.4.1: Eine Folge von Schätzer $W_n = W(X_1, \ldots, X_n)$ ist eine **konsistente Folge von Schätzer** für den Parameter θ, falls für jedes $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt

$$\lim_{n\to\infty} P_\theta\left(|W_n - \theta| < \epsilon\right) = 1\,.$$

Falls $n \to \infty$ (und die Stichprobenvarianz damit besser wird), wird der Schätzer mit großer Wahrscheinlichkeit beliebig nahe dem Parameter sein. In anderen Worten ist die Wahrscheinlichkeit, dass ein konsistenter Schätzer den Parameter verfehlt, sehr gering.

Dies ist vergleichbar mit dem Begriff der **Konvergenz in Wahrscheinlichkeit**. Definition 3.4.1 sagt sogar, dass eine konsistente Folge von Schätzer in Wahrscheinlichkeit zum Parameter θ konvergiert, den sie schätzt.

Während wir für die Konvergenz in Wahrscheinlichkeit nur eine Folge von Zufallsvariablen aus einer Verteilung betrachtet hatten, behandelt Definition 3.4.1 eine gesamte Verteilungsfamilie indiziert durch θ.

Beispiel 3.4.1: Seien $X_1, X_2, \ldots$ iid Normal$(\theta, 1)$. Dann gilt für die Folge $\overline{X}_n$

$$\begin{aligned} P_\theta(|\overline{X}_n - \theta| < \epsilon) &= P_\theta(-\epsilon < \overline{X}_n - \theta < \epsilon) \\ &= P_\theta(-\epsilon\sqrt{n} < Z_n < \epsilon\sqrt{n}) \to 1 \qquad \text{für } n \to \infty\,, \end{aligned}$$

wobei

$$Z_n = \frac{\overline{X}_n - \theta}{1/\sqrt{n}} \sim \text{Normal}(0,1)\,.$$

Daher ist $\overline{X}_n$ konsistenter Schätzer für θ.

Markov Ugl:

$$P_\theta(|W_n - \theta| \geq \epsilon) \leq \frac{\mathsf{E}_\theta(W_n - \theta)^2}{\epsilon^2}\,.$$

Somit ist

$$\lim_{n\to\infty} \mathsf{E}_\theta(W_n - \theta)^2 = 0 \qquad \text{für alle } \theta \in \Theta$$

eine hinreichende Bedingung für die Konsistenz von W_n. Weiters ist

$$\mathsf{E}_\theta(W_n - \theta)^2 = \mathsf{var}_\theta(W_n) + \mathsf{bias}^2_\theta(W_n, \theta)$$

und wir erhalten:

Satz 3.4.1: Ist W_n eine Folge von Schätzer für θ, die für jedes $\theta \in \Theta$

(a) $\lim_{n\to\infty} \mathsf{var}_\theta(W_n) = 0$,

(b) $\lim_{n\to\infty} \mathsf{bias}_\theta(W_n, \theta) = 0$

erfüllt, so ist W_n eine konsistente Folge von Schätzer für θ.

Beispiel 3.4.1 (Fortsetzung): Seien $X_1, X_2, \ldots$ iid Normal$(\theta, 1)$. Da

$$\overline{X}_n \sim \text{Normal}(\theta, 1/n) \quad \Rightarrow \quad \text{bias}_\theta(\overline{X}_n, \theta) = 0\,, \ \lim_{n\to\infty} \text{var}_\theta(\overline{X}_n) = 0$$

die Voraussetzungen von Satz 3.4.1 erfüllt, ist $\overline{X}_n$ konsistenter Schätzer für θ.

Satz 3.4.2: Sei W_n eine konsistente Folge von Schätzer für θ und $a_1, a_2, \ldots$ und $b_1, b_2, \ldots$ Folgen von Konstanten, die

$\lim_{n\to\infty} a_n = 1$,

$\lim_{n\to\infty} b_n = 0$

genügen. Dann ist $U_n = a_n W_n + b_n$ eine konsistente Folge von Schätzer für θ.

Wir wollen nun ein allgemeines Resultat betreffs der **Konsistenz von MLE's** skizzieren. Dieses Resultat zeigt, dass MLE's konsistente Schätzer ihrer Parameter sind. Somit garantiert diese Methode zum Finden von Schätzer eine derartige Optimalitätseigenschaft.

Satz 3.4.3: (Konsistenz von MLE's) Seien $X_1, X_2, \ldots$ iid mit Dichte $f(x|\theta)$, und sei $L(\theta|\mathbf{x}) = \prod_i f(x_i|\theta)$ die Likelihood Funktion. Bezeichne $\hat{\theta}$ den MLE für θ. Sei $\tau(\theta)$ eine stetige Funktion in θ. Unter gewissen Regularitätsbedingungen bezüglich $f(x|\theta)$, und daher auch bezüglich $L(\theta|\mathbf{x})$, gilt für jedes $\epsilon > 0$ und für jedes $\theta \in \Theta$

$$\lim_{n\to\infty} P_\theta\left(|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon\right) = 0\,.$$

Dies bedeutet, dass $\tau(\hat{\theta})$ ein konsistenter Schätzer für $\tau(\theta)$ ist.

Regularitätsbedingungen:

(A1) Die Elemente der Stichprobe X_i seien iid aus $f(x|\theta)$

(A2) Der Parameter sei identifizierbar, d.h. für $\theta \neq \theta'$ gelte $f(x|\theta) \neq f(x|\theta')$

(A3) Die Dichten $f(x|\theta)$ haben einen gemeinsamen Träger und seien diff.bar in θ

(A4) Der Parameterraum Θ enthalte eine offene Menge in deren Innerem der wahre Parameterwert θ_0 liege.

3.4.2 Effizienz

Konsistenz beschreibt die asymptotische Genauigkeit eines Schätzers (konvergiert der Schätzer zum Parameter den er schätzt?). Effizienz ist damit verwandt und bezieht sich auf die asymptotische Varianz eines Schätzers.

Sei $T_n = T(X_1, \ldots, X_n)$ ein Schätzer. Um dessen asymptotische Varianz zu bekommen, betrachte dessen endliche $\mathsf{var}_\theta(T_n)$ und berechnen $\lim_{n\to\infty} k_n \mathsf{var}_\theta(T_n)$, wobei k_n eine normalisierende Konstante ist. Beachte, dass in vielen Fällen $\mathsf{var}(T_n) \to 0$ für $n \to \infty$; deshalb benötigen wir den Faktor k_n um für die Varianz einen existierenden Grenzwert zu bekommen.

Definition 3.4.2: Falls für einen Schätzer T_n gilt, dass

$$\lim_{n\to\infty} k_n \mathsf{var}_\theta(T_n) = \tau^2 < \infty\,,$$

mit einer Folge von positiven Normalisierungskonstanten $\{k_n\}$, dann nennt man τ^2 die **Grenzvarianz** (Grenzwert der Varianzen).

Beispiel 3.4.2: Seien $X_1, X_2, \ldots$ iid Normal(μ, σ^2). Falls wir $T_n = \overline{X}_n$ betrachten, so ist $\lim_{n\to\infty} n\mathsf{var}(\overline{X}_n) = \sigma^2$ die Grenzvarianz von T_n.

Aber wie sieht es aus mit $T_n = 1/\overline{X}_n$ (als Schätzer für $1/\mu$)? Hierfür ist $\mathsf{var}(T_n) = \infty$ und die Grenzvarianz existiert nicht.

Verwende **approximative Momente:** Sei Y Zufallsvariable mit $\mathsf{E}(Y) = \mu \neq 0$. Wir wollen $g(\mu)$ schätzen. Dafür liefert eine lineare Approximation

$$g(Y) = g(\mu) + g'(\mu)(Y - \mu)\,.$$

Falls wir $g(Y)$ als Schätzer für $g(\mu)$ verwenden, so gilt dafür approximativ

$$\mathsf{E}(g(Y)) \approx g(\mu)\,, \qquad \mathsf{var}(g(Y)) \approx \left[g'(\mu)\right]^2 \mathsf{var}(Y)\,.$$

In unserem Fall schätzen wir $1/\mu$ durch $1/\overline{X}_n$ und erhalten

$$\mathsf{E}(1/\overline{X}_n) \approx 1/\mu\,, \qquad \mathsf{var}(1/\overline{X}_n) \approx \left(-1/\mu^2\right)^2 \mathsf{var}(\overline{X}_n) = \sigma^2/n\mu^4 < \infty\,.$$

Dieses Beispiel zeigt die Probleme auf, wenn man den Grenzwert der Varianzen als Maß bei großen Stichproben verwendet.

Die exakte (endliche) Stichprobenvarianz von $1/\overline{X}$ ist ∞. Jedoch hat für $\mu \neq 0$ der Bereich in dem $1/\overline{X}$ groß wird Wahrscheinlichkeit die gegen Null geht.

Daher ist die Approximation aus Beispiel 3.4.2 realistischer (natürlich auch hilfreicher). Wir adaptieren diesen Ansatz zur Berechnung von (großen Stichproben) Varianzen. Vergleiche auch mit der **Delta-Methode** aus Abschnitt 1.3.

Definition 3.4.3: Angenommen für einen Schätzer T_n gilt

$$k_n\Big(T_n - \tau(\theta)\Big) \stackrel{D}{\to} N(0, \sigma^2)\,,$$

dann bezeichnet man den Parameter σ^2 als **asymptotische Varianz** (Varianz der Grenzverteilung) von T_n.

Im Sinne der Cramér-Rao Schranke gibt es eine optimale asymptotische Varianz.

Definition 3.4.4: Eine Folge von Schätzer W_n ist **asymptotisch effizient** für einen Parameter $\tau(\theta)$, falls

$$\sqrt{n}\Big(W_n - \tau(\theta)\Big) \stackrel{D}{\to} N(0, v(\theta))\,,$$

und

$$v(\theta) = \frac{(\tau'(\theta))^2}{\mathsf{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right)^2}\,,$$

also die asymptotische Varianz von W_n die Cramér-Rao Schranke erreicht.

Satz 3.4.3 sagte aus, dass unter recht allgemeinen Bedingungen die MLE's konsistente Schätzer sind. Unter etwas stärkeren Annahmen gilt dies auch bzgl. der asymptotischen Effizienz. Daher können wir MLE's als **konsistent und asymptotisch effizient** betrachten.

Satz 3.4.4: (Asymptotische Effizienz des MLE) Seien $X_1, X_2, \ldots$ iid aus $f(x|\theta)$, und $\hat{\theta}$ der MLE für θ, sowie $\tau(\theta)$ stetig in θ. Unter gewissen Regularitätsbedingungen bezüglich $f(x|\theta)$, und daher auch bezüglich $L(\theta|\mathbf{x})$, gilt

$$\sqrt{n}\left(\tau(\hat{\theta}) - \tau(\theta)\right) \xrightarrow{D} N(0, v(\theta))\,,$$

wobei $v(\theta)$ die Cramér-Rao Schranke bezeichnet. Dies bedeutet, dass $\tau(\hat{\theta})$ ein konsistenter und asymptotisch effizienter Schätzer für $\tau(\theta)$ ist.

Zusätzliche Regularitätsbedingungen:

(A5) Für jedes $x \in \mathcal{X}$ sei $f(x|\theta)$ dreimal stetig differenzierbar in θ und die dritte Ableitung sei stetig in θ

(A6) Für jedes $\theta_0 \in \Theta$ existiere eine positive Zahl c und eine Funktion $M(x)$ mit $\mathsf{E}_{\theta_0}(M(X)) < \infty$ (beide dürfen von θ_0 abhängen), so dass

$$\left|\frac{\partial^3}{\partial\theta^3} \log f(x|\theta)\right| \leq M(x)\,, \quad \text{für alle } x \in \mathcal{X}, \quad \theta_0 - c < \theta < \theta_0 + c\,.$$

Definition 3.4.5: Falls für zwei Folgen von Schätzer gilt, dass

$$\begin{aligned}\sqrt{n}\,(W_n - \tau(\theta)) &\overset{D}{\to} N(0, \sigma_W^2)\,,\\ \sqrt{n}\,(V_n - \tau(\theta)) &\overset{D}{\to} N(0, \sigma_V^2)\,,\end{aligned}$$

dann ist die **asymptotische relative Effizienz** (ARE) von V_n bezüglich W_n gleich

$$\mathsf{ARE}(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}\,.$$