KAPITEL 6

Güteeigenschaften von Schätzern

Wir erinnern an die Definition des parametrischen Modells. Sei $\{h_\theta : \theta \in \Theta\}$, wobei $\Theta \subset \mathbb{R}^m$, eine Familie von Dichten oder Zähldichten. Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte oder Zähldichte h_θ. Dabei ist θ der zu schätzende Parameter. Um die Notation zu vereinfachen, werden wir in diesem Kapitel nicht zwischen den Zufallsvariablen $(X_1, \ldots, X_n)$ und deren Realisierung $(x_1, \ldots, x_n)$ unterscheiden. Ein Schätzer ist eine beliebige (Borel–messbare) Funktion

$$\hat{\theta} : \mathbb{R}^n \to \Theta, \quad (X_1, \ldots, X_n) \mapsto \hat{\theta}(X_1, \ldots, X_n).$$

Die Aufgabe eines Schätzers ist es, den richtigen Wert von θ möglichst gut zu erraten. Im Folgenden definieren wir einige Eigenschaften von Schätzern, die uns erlauben, "gute" Schätzer von "schlechten" Schätzern zu unterscheiden.

6.1. Erwartungstreue, Konsistenz, asymptotische Normalverteiltheit

In diesem Kapitel sei der Parameterraum Ω eine Teilmenge von $\mathbb{R}^m$.

Definition 6.1.1. Ein Schätzer $\hat{\theta}$ heißt *erwartungstreu* (oder *unverzerrt*), falls

$$\mathbb{E}_\theta[\hat{\theta}(X_1, \ldots, X_n)] = \theta \text{ für alle } \theta \in \Theta.$$

Bemerkung 6.1.2. Damit diese Definition Sinn macht, muss man voraussetzen, dass die Zufallsvariable (bzw. Zufallsvektor) $\hat{\theta}(X_1, \ldots, X_n)$ integrierbar ist.

Definition 6.1.3. Der *Bias* (die *Verzerrung*) eines Schätzers $\hat{\theta}$ ist

$$\operatorname{Bias}_\theta(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}(X_1, \ldots, X_n)] - \theta.$$

Wir betrachten $\operatorname{Bias}_\theta(\hat{\theta})$ als eine Funktion von $\theta \in \Theta$.

Bemerkung 6.1.4. Ein Schätzer $\hat{\theta}$ ist genau dann erwartungstreu, wenn $\operatorname{Bias}_\theta(\hat{\theta}) = 0$ für alle $\theta \in \Theta$.

Beispiel 6.1.5. In diesem Beispiel werden wir verschiedene Schätzer für den Endpunkt der Gleichverteilung konstruieren. Es seien $X_1, \ldots, X_n$ unabhängige und auf dem Intervall $[0, \theta]$ gleichverteilte Zufallsvariablen, wobei $\theta > 0$ der zu schätzende Parameter sei. Es seien $X_{(1)} < \ldots < X_{(n)}$ die Ordnungsstatistiken von $X_1, \ldots, X_n$. Folgende Schätzer für θ erscheinen natürlich.

1. Der Maximum-Likelihood-Schätzer

$$\hat{\theta}_1(X_1, \ldots, X_n) = X_{(n)} = \max\{X_1, \ldots, X_n\}.$$

Es ist offensichtlich, dass $\hat{\theta}_1 < \theta$. Somit wird θ von diesem Schätzer immer unterschätzt.

2. Wir versuchen nun den Schätzer $\hat{\theta}_1$ zu verbessern, indem wir ihn vergrößern. Wir würden ihn gerne um $\theta - X_{(n)}$ vergößern, allerdings ist θ unbekannt. Deshalb machen wir den folgenden Ansatz. Wir gehen davon aus, dass die beiden Intervalle $(0, X_{(1)})$ und $(X_{(n)}, \theta)$ ungefähr gleich lang sind, d.h.

$$X_{(1)} \stackrel{!}{=} \theta - X_{(n)}.$$

Lösen wir diese Gleichung bzgl. θ, so erhalten wir den Schätzer

$$\hat{\theta}_2(X_1, \ldots, X_n) = X_{(n)} + X_{(1)}.$$

3. Es gibt aber auch einen anderen natürlichen Ansatz. Wir können davon ausgehen, dass die Intervalle

$$(0, X_{(1)}), (X_{(1)}, X_{(2)}), \ldots, (X_{(n)}, \theta)$$

ungefähr gleich lang sind. Dann kann man die Länge des letzten Intervalls durch das arithmetische Mittel der Längen aller vorherigen Intervalle schätzen, was zu folgender Gleichung führt:

$$\theta - X_{(n)} \stackrel{!}{=} \frac{1}{n}(X_{(1)} + (X_{(2)} - X_{(1)}) + (X_{(3)} - X_{(2)}) + \ldots + (X_{(n)} - X_{(n-1)})).$$

Da auf der rechten Seite eine Teleskop-Summe steht, erhalten wir die Gleichung

$$\theta - X_{(n)} \stackrel{!}{=} \frac{1}{n} X_{(n)}.$$

Auf diese Weise ergibt sich der Schätzer

$$\hat{\theta}_3(X_1, \ldots, X_n) = \frac{n+1}{n} X_{(n)}.$$

4. Wir können auch den Momentenschätzer betrachten. Setzen wir den Erwartungswert von X_i dem empirischen Mittelwert gleich, so erhalten wir

$$\mathbb{E}_\theta[X_i] = \frac{\theta}{2} \stackrel{!}{=} \bar{X}_n.$$

Dies führt zum Schätzer

$$\hat{\theta}_4(X_1, \ldots, X_n) = 2\bar{X}_n.$$

Aufgabe 6.1.6. Zeigen Sie, dass $\hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ erwartungstreu sind, $\hat{\theta}_1$ jedoch nicht.

Man sieht an diesem Beispiel, dass es für ein parametrisches Problem mehrere natürliche erwartungstreue Schätzer geben kann. Die Frage ist nun, welcher Schätzer der beste ist.

Definition 6.1.7. Sei $\Theta = (a, b) \subset \mathbb{R}$ ein Intervall. Der *mittlere quadratische Fehler* (*mean square error*, MSE) eines Schätzers $\hat{\theta} : \mathbb{R}^n \to \mathbb{R}$ ist definiert durch

$$\mathrm{MSE}_\theta(\hat{\theta}) = \mathbb{E}_\theta[(\hat{\theta}(X_1, \ldots, X_n) - \theta)^2].$$

Wir fassen $\mathrm{MSE}_\theta(\hat{\theta})$ als eine Funktion von $\theta \in (a, b)$ auf. Damit die obige Definition Sinn hat, muss man voraussetzen, dass $\hat{\theta}(X_1, \ldots, X_n)$ eine quadratisch integrierbare Zufallsvariable ist.

Lemma 6.1.8. *Es gilt folgender Zusammenhang zwischen dem mittleren quadratischen Fehler und dem Bias:*

$$\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{Var}_\theta \hat{\theta} + (\mathrm{Bias}_\theta(\hat{\theta}))^2.$$

Beweis. Um die Notation zu vereinfachen, schreiben wir in diesem Beweis $\hat{\theta}$ für $\hat{\theta}(X_1, \ldots, X_n)$. Wir benutzen die Definition des mittleren quadratischen Fehlers, erweitern mit $\mathbb{E}_\theta[\hat{\theta}]$ und quadrieren:

$$\begin{aligned}
\mathrm{MSE}_\theta(\hat{\theta}) &= \mathbb{E}_\theta[(\hat{\theta} - \theta)^2] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}] + \mathbb{E}_\theta[\hat{\theta}] - \theta)^2] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])^2] + 2\mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}]) \cdot (\mathbb{E}_\theta[\hat{\theta}] - \theta)] + \mathbb{E}_\theta[(\mathbb{E}_\theta[\hat{\theta}] - \theta)^2] \\
&= \mathrm{Var}_\theta(\hat{\theta}) + 2(\mathbb{E}_\theta[\hat{\theta}] - \theta) \cdot \mathbb{E}_\theta[\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}]] + (\mathrm{Bias}_\theta(\hat{\theta}))^2.
\end{aligned}$$

Dabei haben wir benutzt, dass $\mathbb{E}_\theta[\hat{\theta}] - \theta$ nicht zufällig ist. Der mittlere Term auf der rechten Seite verschwindet, denn $\mathbb{E}_\theta[\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}]] = \mathbb{E}_\theta[\hat{\theta}] - \mathbb{E}_\theta[\hat{\theta}] = 0$. Daraus ergibt sich die gewünschte Identität. $\square$

Bemerkung 6.1.9. Ist $\hat{\theta}$ erwartungstreu, so gilt $\mathrm{Bias}_\theta(\hat{\theta}) = 0$ für alle $\theta \in \Theta$ und somit vereinfacht sich Lemma 6.1.8 zu

$$\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{Var}_\theta(\hat{\theta}).$$

Definition 6.1.10. Seien $\hat{\theta}_1$ und $\hat{\theta}_2$ zwei Schätzer. Wir sagen, dass $\hat{\theta}_1$ *besser* als $\hat{\theta}_2$ ist, falls

$$\mathrm{MSE}_\theta(\hat{\theta}_1) < \mathrm{MSE}_\theta(\hat{\theta}_2) \text{ für alle } \theta \in \Theta.$$

Bemerkung 6.1.11. Falls $\hat{\theta}_1$ und $\hat{\theta}_2$ erwartungstreu sind, dann ist $\hat{\theta}_1$ besser als $\hat{\theta}_2$, wenn

$$\mathrm{Var}_\theta(\hat{\theta}_1) < \mathrm{Var}_\theta(\hat{\theta}_2) \text{ für alle } \theta \in \Theta.$$

Bemerkung 6.1.12. In Beispiel 6.1.5 ist $\hat{\theta}_3 = \frac{n+1}{n} X_{(n)}$ der beste Schätzer unter allen erwartungstreuen Schätzern. Der Beweis hierfür folgt später.

In der Statistik ist der Stichprobenumfang n typischerweise groß. Wir schauen uns deshalb die asymptotischen Güteeigenschaften von Schätzern an. Wir betrachten eine Folge von Schätzern

$$\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), \ldots, \hat{\theta}_n(X_1, \ldots, X_n), \ldots.$$

Sei im Folgenden Θ eine Teilmenge von $\mathbb{R}^m$.

Definition 6.1.13. Eine Folge von Schätzern $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ heißt *asymptotisch erwartungstreu*, falls

$$\lim_{n\to\infty} \mathbb{E}_\theta \hat{\theta}_n(X_1, \ldots, X_n) = \theta \text{ für alle } \theta \in \Theta.$$

Beispiel 6.1.14. In Beispiel 6.1.5 ist $X_{(n)}$ eine asymptotisch erwartungstreue (aber nicht erwartungstreue) Folge von Schätzern, denn (Übungsaufgabe)

$$\lim_{n\to\infty} \mathbb{E}_\theta X_{(n)} = \lim_{n\to\infty} \frac{n}{n+1}\theta = \theta.$$

Definition 6.1.15. Eine Folge von Schätzern $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ heißt *schwach konsistent*, falls

$$\hat{\theta}_n(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{P} \theta \text{ unter } \mathbb{P}_\theta \text{ für alle } \theta \in \Theta.$$

Mit anderen Worten, für jedes $\varepsilon > 0$ und jedes $\theta \in \Theta$ soll gelten:

$$\lim_{n\to\infty} \mathbb{P}_\theta[|\hat{\theta}_n(X_1,\dots,X_n) - \theta| > \varepsilon] = 0.$$

Definition 6.1.16. Eine Folge von Schätzern $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ heißt *stark konsistent*, falls

$$\hat{\theta}_n(X_1,\dots,X_n) \xrightarrow[n\to\infty]{f.s.} \theta \text{ unter } \mathbb{P}_\theta \text{ für alle } \theta \in \Theta.$$

Mit anderen Worten, es soll für alle $\theta \in \Theta$ gelten:

$$\mathbb{P}_\theta\left[\lim_{n\to\infty} \hat{\theta}_n(X_1,\dots,X_n) = \theta\right] = 1.$$

Bemerkung 6.1.17. Eine fast sicher konvergente Folge von Zufallsvariablen konvergiert auch in Wahrscheinlichkeit. Aus der starken Konsistenz folgt somit die schwache Konsistenz.

Definition 6.1.18. Eine Folge von Schätzern $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ heißt *L^2-konsistent*, falls

$$\hat{\theta}_n(X_1,\dots,X_n) \overset{L^2}{\to} \theta \text{ für alle } \theta \in \Theta.$$

Mit anderen Worten, es soll für alle $\theta \in \Theta$ gelten:

$$\lim_{n\to\infty} \mathbb{E}_\theta|\hat{\theta}_n(X_1,\dots,X_n) - \theta|^2 = 0.$$

Bemerkung 6.1.19. Aus der L^2-Konsistenz folgt die schwache Konsistenz.

Beispiel 6.1.20. Bei vielen Familien von Verteilungen, z.B. $\mathrm{Bern}(\theta)$, $\mathrm{Poi}(\theta)$ oder $\mathrm{N}(\theta,\sigma^2)$ stimmt der Parameter θ mit dem Erwartungswert der entsprechenden Verteilung überein. In diesem Fall ist die Folge von Schätzern $\hat{\theta}_n = \bar{X}_n$ stark konsistent, denn für jedes θ gilt

$$\bar{X}_n \xrightarrow[n\to\infty]{f.s.} \mathbb{E}_\theta X_1 = \theta \text{ unter } \mathbb{P}_\theta$$

nach dem starken Gesetz der großen Zahlen.

Definition 6.1.21. Eine Folge von Schätzern $\hat{\theta}_n : \mathbb{R}^n \to \Theta \subset \mathbb{R}$ heißt *asymptotisch normalverteilt*, wenn es zwei Folgen $a_n(\theta) \in \mathbb{R}$ und $b_n(\theta) > 0$ gibt, sodass für alle $\theta \in \Theta$

$$\frac{\hat{\theta}(X_1,\dots,X_n) - a_n(\theta)}{b_n(\theta)} \xrightarrow[n\to\infty]{d} \mathrm{N}(0,1) \text{ unter } \mathbb{P}_\theta.$$

Normalerweise wählt man $a_n(\theta) = \mathbb{E}_\theta\hat{\theta}(X_1,\dots,X_n)$ und $b_n^2(\theta) = \operatorname{Var}_\theta \hat{\theta}(X_1,\dots,X_n)$, sodass die Bedingung folgendermaßen lautet: Für alle $\theta \in \Theta$

$$\frac{\hat{\theta}(X_1,\dots,X_n) - \mathbb{E}_\theta\hat{\theta}(X_1,\dots,X_n)}{\sqrt{\operatorname{Var}_\theta \hat{\theta}(X_1,\dots,X_n)}} \xrightarrow[n\to\infty]{d} \mathrm{N}(0,1) \text{ unter } \mathbb{P}_\theta.$$

Beispiel 6.1.22. Es seien $X_1,\dots,X_n$ unabhängig und $\mathrm{Bern}(\theta)$-verteilt, mit $\theta \in (0,1)$. Dann ist der Schätzer $\hat{\theta} = \bar{X}_n$ asymptotisch normalverteilt, denn nach dem Satz von de Moivre–Laplace gilt

$$\frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta(a-\theta)}{n}}} = \frac{X_1 + \ldots + X_n - n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow[n\to\infty]{d} \mathrm{N}(0,1) \text{ unter } \mathbb{P}_\theta.$$

6.2. Güteeigenschaften des ML-Schätzers

In diesem Abschnitt sei $\Theta = (a, b)$ ein Intervall. Sei $\{h_\theta : \theta \in \Theta\}$ eine Familie von Dichten oder Zähldichten und sei $\theta_0 \in \Theta$ fest. Seien $X, X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte h_{θ_0}, wobei θ_0 als der "wahre Wert des Parameters" aufgefasst wird. Uns ist der wahre Wert allerdings unbekannt und wir schätzen ihn mit dem Maximum-Likelihood-Schätzer $\hat{\theta}_{ML} = \hat{\theta}_{ML}(X_1, \ldots, X_n)$. In diesem Abschnitt wollen wir die Güteeigenschaften des Maximum-Likelihood-Schätzers untersuchen. Um die Güteeigenschaften von $\hat{\theta}_{ML}$ zu beweisen, muss man gewisse Regularitätsbedingungen an die Familie $\{h_\theta : \theta \in \Theta\}$ stellen. Leider sind diese Bedingungen nicht besonders schön. Deshalb werden wir nur die Ideen der jeweiligen Beweise zeigen. Wir werden hier nur eine der vielen Regularitätsbedingungen formulieren: Alle Dichten (oder Zähldichten) h_θ sollen den gleichen Träger haben, d.h. die Menge

$$J := \{x \in \mathbb{R} : h_\theta(x) \neq 0\}$$

soll nicht von θ abhängen.

Konsistenz des ML-Schätzers. Zuerst fragen wir, ob der Maximum-Likelihood-Schätzer stark konsistent ist, d.h. ob

$$\hat{\theta}_{ML}(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{f.s.} \theta_0.$$

Wir werden zeigen, dass das stimmt. Hierfür betrachten wir die durch n geteilte log-Likelihood-Funktion

$$L_n(X_1, \ldots, X_n; \theta) := \frac{1}{n} \log L(X_1, \ldots, X_n; \theta) = \frac{1}{n} \sum_{i=1}^{n} \log h_\theta(X_i).$$

Nach dem Gesetz der großen Zahlen gilt

$$L_n(X_1, \ldots, X_n; \theta) \xrightarrow[n\to\infty]{f.s.} \mathbb{E}_{\theta_0} \log h_\theta(X) = L_\infty(\theta) = \int_J \log h_\theta(t) h_{\theta_0}(t) dt.$$

Lemma 6.2.1. *Für alle $\theta \in \Theta$ gilt:*

$$L_\infty(\theta) \leq L_\infty(\theta_0).$$

Beweis. Mit der Definition von $L_\infty(\theta)$ ergibt sich

$$L_\infty(\theta) - L_\infty(\theta_0) = \mathbb{E}_{\theta_0}[\log h_\theta(X) - \log h_{\theta_0}(X)] = \mathbb{E}_{\theta_0}\left[\log \frac{h_\theta(X)}{h_{\theta_0}(X)}\right].$$

Nun wenden wir auf die rechte Seite die Ungleichung $\log t \leq t - 1$ (wobei $t > 0$) an:

$$\begin{aligned}
L_\infty(\theta) - L_\infty(\theta_0) &\leq \mathbb{E}_{\theta_0}\left[\frac{h_\theta(X)}{h_{\theta_0}(X)} - 1\right] \\
&= \int_J \left(\frac{h_\theta(t)}{h_{\theta_0}(t)} - 1\right) h_{\theta_0}(t) dt \\
&= \int_J h_\theta(t) dt - \int_J h_{\theta_0}(t) dt \\
&= 0,
\end{aligned}$$

denn $\int_J h_\theta(t)dt = \int_J h_{\theta_0}(t)dt = 1$. $\square$

SATZ 6.2.2. *Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte oder Zähldichte h_{θ_0}. Unter Regularitätsbedingungen an die Familie $\{h_\theta : \theta \in \Theta\}$ gilt, dass*

$$\hat{\theta}_{ML}(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{f.s.} \theta_0.$$

BEWEISIDEE. Per Definition des Maximum-Likelihood-Schätzers ist

$$\hat{\theta}_{ML}(X_1, \ldots, X_n) = \text{argmax } L_n(X_1, \ldots, X_n; \theta).$$

Indem wir zum Grenzwert für $n \to \infty$ übergehen, erhalten wir

$$\begin{aligned}\lim_{n\to\infty} \hat{\theta}_{ML}(X_1, \ldots, X_n) &= \lim_{n\to\infty} \text{argmax } L_n(X_1, \ldots, X_n; \theta)\\ &= \text{argmax } \lim_{n\to\infty} L_n(X_1, \ldots, X_n; \theta)\\ &= \text{argmax } L_\infty(X_1, \ldots, X_n; \theta)\\ &= \theta_0,\end{aligned}$$

wobei der letzte Schritt aus Lemma 6.2.1 folgt. $\square$

Der obige Beweis ist nicht streng. Insbesondere bedarf der Schritt $\lim_{n\to\infty}$ argmax $=$ argmax $\lim_{n\to\infty}$ einer Begründung.

Asymptotische Normalverteiltheit des ML-Schätzers. Wir werden zeigen, dass unter gewissen Regularitätsbedingungen der Maximum-Likelihood-Schätzer asymptotisch normalverteilt ist:

$$\sqrt{n}(\hat{\theta}_{\text{ML}}(X_1, \ldots, X_n) - \theta_0) \xrightarrow[n\to\infty]{d} \text{N}(0, \sigma^2_{\text{ML}}) \text{ unter } \mathbb{P}_{\theta_0},$$

wobei die Varianz σ^2_{ML} später identifiziert werden soll. Wir bezeichnen mit

$$l_\theta(x) = \log h_\theta(x)$$

die log-Likelihood einer einzelnen Beobachtung x. Die Ableitung nach θ wird mit

$$D = \frac{d}{d\theta}$$

bezeichnet. Insbesondere schreiben wir

$$Dl_\theta(x) = \frac{d}{d\theta} l_\theta(x), \quad D^2 l_\theta(x) = \frac{d^2}{d\theta^2} l_\theta(x).$$

DEFINITION 6.2.3. Sei $\{h_\theta : \theta \in \Theta\}$, wobei $\Theta = (a, b)$ ein Intervall ist, eine Familie von Dichten oder Zähldichten. Die *Fisher–Information* ist eine Funktion $I : \Theta \to \mathbb{R}$ mit

$$I(\theta) = \mathbb{E}_\theta (Dl_\theta(X))^2.$$

LEMMA 6.2.4. *Unter Regularitätsbedingungen an die Familie $\{h_\theta : \theta \in \Theta\}$ gilt für jedes $\theta \in (a, b)$, dass*

(1) $\mathbb{E}_\theta Dl_\theta(X) = 0$.
(2) $\mathbb{E}_\theta D^2 l_\theta(X) = -I(\theta)$.

Beweisidee. Für den Beweis formen wir zuerst $Dl_\theta(x)$ und $D^2 l_\theta(x)$ wie folgt um:

$$Dl_\theta(x) = D \log h_\theta(x) = \frac{Dh_\theta(x)}{h_\theta(x)},$$
$$D^2 l_\theta(x) = \frac{(D^2 h_\theta(x))h_\theta(x) - (Dh_\theta(x))^2}{h_\theta^2(x)} = \frac{D^2 h_\theta(x)}{h_\theta(x)} - (Dl_\theta(x))^2.$$

Außerdem gilt für alle θ, dass $\int_J h_\theta(t)dt = 1$, denn h_θ ist eine Dichte. Wir können nun diese Identität nach θ ableiten:

$$D \int_J h_\theta(t)dt = 0 \text{ und somit } \int_J Dh_\theta(t)dt = 0,$$
$$D^2 \int_J h_\theta(t)dt = 0 \text{ und somit } \int_J D^2 h_\theta(t)dt = 0.$$

Dabei haben wir die Ableitung und das Integral vertauscht, was unter gewissen Regularitätsbedingungen möglich ist. Mit diesen Resultaten erhalten wir, dass

$$\mathbb{E}_\theta Dl_\theta(X) = \int_J \frac{Dh_\theta(X)}{h_\theta(X)} h_\theta(t)dt = \int_J Dh_\theta(X)dt = 0.$$

Somit ist die erste Behauptung des Lemmas bewiesen. Die zweite Behauptung des Lemmas kann man wie folgt zeigen:

$$\begin{aligned}
\mathbb{E}_\theta D^2 l_\theta(X) &= \int_J (D^2 l_\theta(t))h_\theta(t)dt \\
&= \int_J \left(\frac{D^2 h_\theta(t)}{h_\theta(t)} - (Dl_\theta(t))^2 \right) h_\theta(t)dt \\
&= \int_J D^2 h_\theta(t)dt - \mathbb{E}_\theta (Dl_\theta(X))^2 \\
&= -\mathbb{E}_\theta (Dl_\theta(X))^2 \\
&= -I(\theta),
\end{aligned}$$

wobei der letzte Schritt aus der Definition der Fisher–Information folgt. □

Indem wir die Notation $L_\infty(\theta) = \mathbb{E}_{\theta_0} \log h_\theta(X)$ verwenden, können wir das obige Lemma wie folgt formulieren.

Lemma 6.2.5. *Unter Regularitätsbedingungen an die Familie $\{h_\theta : \theta \in \Theta\}$ gilt, dass*

(1) $\mathbb{E}_{\theta_0} Dl_{\theta_0}(X) = DL_\infty(\theta_0) = 0.$
(2) $\mathbb{E}_{\theta_0} D^2 l_{\theta_0}(X) = D^2 L_\infty(\theta_0) = -I(\theta_0).$

Beweis. Unter Regularitätsbedingungen kann man den Erwartungswert $\mathbb{E}$ und die Ableitung D vertauschen. Somit gilt $\mathbb{E}_\theta Dl_\theta(X) = D\mathbb{E}_\theta l_\theta(X) = DL_\infty(X)$ und $\mathbb{E}_\theta D^2 l_\theta(X) = D^2 \mathbb{E}_\theta l_\theta(X) = D^2 L_\infty(X)$. □

Satz 6.2.6. *Sei $\{h_\theta : \theta \in \Theta\}$ mit $\Theta = (a, b)$ eine Familie von Dichten oder Zähldichten. Sei $\theta_0 \in (a, b)$ und seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen mit Dichte oder Zähldichte h_{θ_0}.*

Unter Regularitätsbedingungen gilt für den Maximum-Likelihood-Schätzer $\hat{\theta}_{ML}(X_1, \dots, X_n)$, dass

$$\sqrt{n}(\hat{\theta}_{ML}(X_1, \dots, X_n) - \theta_0) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{I(\theta_0)}\right) \text{ unter } \mathbb{P}_{\theta_0}.$$

Beweisidee. Schritt 1. Für den Maximum-Likelihood-Schätzer gilt

$$\hat{\theta}_{ML} = \operatorname{argmax}_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log h_\theta(X_i) = \operatorname{argmax}_{\theta\in\Theta} L_n(\theta).$$

Somit gilt

$$DL_n(\hat{\theta}_{ML}) = 0.$$

Der Mittelwertsatz aus der Analysis besagt, dass wenn f eine differenzierbare Funktion auf einem Intervall $[x, y]$ ist, dann lässt sich ein c in diesem Intervall finden mit

$$f(y) = f(x) + f'(c)(y - x).$$

Wir wenden nun diesen Satz auf die Funktion $f(\theta) = DL_n(\theta)$ und auf das Intervall mit den Endpunkten θ_0 und $\hat{\theta}_{ML}$ an. Es lässt sich also ein ξ_n in diesem Intervall finden mit

$$0 = DL_n(\hat{\theta}_{ML}) = DL_n(\theta_0) + D^2L_n(\xi_n)(\hat{\theta}_{ML} - \theta_0).$$

Daraus ergibt sich, dass

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) = -\frac{\sqrt{n}DL_n(\theta_0)}{DL_n(\xi_n)}.$$

Schritt 2. Anwendung von Lemma 6.2.5 führt zu $\mathbb{E}_{\theta_0} Dl_{\theta_0}(X) = 0$. Somit gilt

$$\sqrt{n}DL_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Dl_{\theta_0}(X_i) - 0) = \frac{\sum_{i=1}^{n} Dl_{\theta_0}(X_i) - n\mathbb{E}_{\theta_0} Dl_{\theta_0}(X_i)}{\sqrt{n}}.$$

Indem wir nun den zentralen Grenzwertsatz anwenden, erhalten wir, dass

$$\sqrt{n}DL_n(\theta_0) \xrightarrow[n\to\infty]{d} N \sim \mathrm{N}(0, \operatorname{Var}_{\theta_0} Dl_{\theta_0}(X)) = \mathrm{N}(0, I(\theta_0)),$$

denn $\operatorname{Var}_{\theta_0} Dl_{\theta_0}(X) = I(\theta_0)$ nach Lemma 6.2.5.

Schritt 3. Da sich ξ_n zwischen $\hat{\theta}_{ML}$ und θ_0 befindet und $\lim_{n\to\infty} \hat{\theta}_{ML} = \theta_0$ wegen Satz 6.2.2 (Konsistenz) gilt, erhalten wir, dass $\lim_{n\to\infty} \xi_n = \theta_0$. Nach dem Gesetz der großen Zahlen gilt somit

$$D^2L_n(\xi_n) = \frac{1}{n}\sum_{i=1}^{n} D^2 l_{\xi_n}(X_i) \underset{n\to\infty}{\longrightarrow} \mathbb{E}_{\theta_0} D^2 l_{\theta_0}(X) = -I(\theta_0),$$

wobei wir im letzten Schritt Lemma 6.2.5 benutzt haben.

Schritt 4. Kombiniert man nun diese Eigenschaften, so führt dies zu

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) = -\frac{\sqrt{n}L_n(\theta_0)}{D^2L_n(\xi_n)} \xrightarrow[n\to\infty]{d} \frac{N}{I(\theta_0)} \sim \mathrm{N}\left(0, \frac{1}{I(\theta_0)}\right),$$

wobei $N \sim \mathrm{N}(0, I(\theta_0))$. □

Beispiel 6.2.7. In diesem Beispiel betrachten wir die Familie der Bernoulli-Verteilungen mit Parameter $\theta \in (0,1)$. Die Zähldichte ist gegeben durch

$$h_\theta(1) = \theta, \quad h_\theta(0) = 1 - \theta.$$

Eine andere Schreibweise dafür ist diese:

$$h_\theta(x) = \theta^x (1-\theta)^{1-x}, \quad x \in \{0,1\}.$$

Die log-Likelihood einer einzelnen Beobachtung $x \in \{0,1\}$ ist

$$l_\theta(x) = \log h_\theta(x) = x \log \theta + (1-x) \log(1-\theta).$$

Ableiten nach θ führt zu

$$Dl_\theta(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta}, \quad D^2 l_\theta(x) = -\left(\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}\right).$$

Sei X eine mit Parameter θ Bernoulli-verteilte Zufallsvariable. Somit ist die Fisher-Information gegeben durch

$$I(\theta) = -\mathbb{E}_\theta D^2 l_\theta(X) = \mathbb{E}_\theta \left[\frac{X}{\theta^2} + \frac{1-X}{(1-\theta)^2}\right] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)}.$$

Seien nun $X_1, X_2, \ldots, X_n$ unabhängige, mit Parameter θ Bernoulli-verteilte Zufallsvariablen. Dann ist der Maximum-Likelihood-Schätzer für den Parameter θ gegeben durch

$$\hat\theta_{ML}(X_1, \ldots, X_n) = \frac{X_1 + \ldots + X_n}{n} = \bar X_n.$$

Mit Satz 6.2.6 erhalten wir die asymptotische Normalverteiltheit von $\hat\theta_{ML} = \bar X_n$:

$$\sqrt{n}(\bar X_n - \theta) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathrm{N}(0, \theta(1-\theta)) \text{ unter } \mathbb{P}_\theta.$$

Diese Aussage können wir auch aus dem Zentralen Grenzwertsatz herleiten, denn

$$\sqrt{n}(\bar X_n - \theta) = \frac{X_1 + \ldots + X_n - n\theta}{\sqrt{n}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathrm{N}(0, \theta(1-\theta)) \text{ unter } \mathbb{P}_\theta,$$

da $\mathbb{E} X_i = \theta$ und $\operatorname{Var} X_i = \theta(1-\theta)$.

Beispiel 6.2.8. Nun betrachten wir ein Beipsiel, in dem der Maximum-Likelihood-Schätzer *nicht* asymptotisch normalverteilt ist. Der Grund hierfür ist, dass eine Regularitätsbedingung verletzt ist. Im folgenden Beispiel sind nämlich die Träger der Verteilungen, die zu verschiedenen Werten des Parameters gehören, nicht gleich.
Wir betrachten die Familie der Gleichverteilungen auf den Intervallen der Form $[0,\theta]$ mit $\theta > 0$. Die Dichte ist gegeben durch

$$h_\theta(x) = \frac{1}{\theta} \mathbb{1}_{x \in [0,\theta]}.$$

Seien $X_1, X_2, \ldots$ unabhängige und auf dem Intervall $[0,\theta]$ gleichverteilte Zufallsvariablen. Der Maximum-Likelihood-Schätzer für θ ist gegeben durch

$$\hat\theta_{ML}(X_1, \ldots, X_n) = \max\{X_1, \ldots, X_n\} =: M_n.$$

Wir zeigen nun, dass dieser Schätzer nicht asymptotisch normalverteilt, sondern asymptotisch exponentialverteilt ist.

Satz 6.2.9. *Es seien $X_1, X_2, \ldots$ unabhängige und auf dem Intervall $[0,\theta]$ gleichverteilte Zufallsvariablen. Dann gilt für $M_n = \max\{X_1, \ldots, X_n\}$, dass*

$$n\left(1 - \frac{M_n}{\theta}\right) \xrightarrow[n\to\infty]{d} Exp(1).$$

Beweis. Sei $x \geq 0$. Es gilt

$$\mathbb{P}\left[n\left(1-\frac{M_n}{\theta}\right) > x\right] = \mathbb{P}\left[\frac{M_n}{\theta} < 1 - \frac{x}{n}\right] = \mathbb{P}\left[X_1 < \theta\left(1-\frac{x}{n}\right), \ldots, X_n < \theta\left(1-\frac{x}{n}\right)\right].$$

Für genügend großes n ist $0 \leq \theta(1-\frac{x}{n}) \leq \theta$ und somit

$$\mathbb{P}\left[X_i < \theta\left(1-\frac{x}{n}\right)\right] = 1 - \frac{x}{n},$$

denn $X_i \sim \mathrm{U}[0,\theta]$. Wegen der Unabhängigkeit von $X_1, \ldots, X_n$ erhalten wir, dass

$$\lim_{n\to\infty} \mathbb{P}\left[n\left(1-\frac{M_n}{\theta}\right) > x\right] = \lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n = e^{-x}.$$

Somit erhalten wir, dass

$$\lim_{n\to\infty} \mathbb{P}\left[n\left(1-\frac{M_n}{\theta}\right) \leq x\right] = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Für $x < 0$ ist der Grenzwert gleich 0, denn das Ereignis $M_n > \theta$ ist unmöglich. Daraus ergibt sich die zu beweisende Aussage. □

Beispiel 6.2.10. In diesem Beispiel betrachten wir die Familie der Exponentialverteilungen. Die Dichte der Exponentialverteilung mit Parameter $\theta > 0$ ist gegeben durch

$$h_\theta(x) = \theta \exp(-\theta x), \quad x > 0.$$

Die log-Likelihood einer einzelnen Beobachtung $x > 0$ ist

$$l_\theta(x) = \log h_\theta(x) = \log\theta - \theta x.$$

Zweimaliges Ableiten nach θ führt zu

$$D^2 l_\theta(x) = -\frac{1}{\theta^2}.$$

Das Ergebnis ist übrigens unabhängig von x. Sei $X \sim \mathrm{Exp}(\theta)$. Die Fisher–Information ist gegeben durch

$$I(\theta) = -\mathbb{E}_\theta D^2 l_\theta(X) = \frac{1}{\theta^2}, \quad \theta > 0.$$

Seien $X_1, X_2, \ldots$ unabhängige, mit Parameter θ exponentialverteilte Zufallsvariablen. Der Maximum-Likelihood-Schätzer (und auch der Momentenschätzer) ist in diesem Beispiel

$$\hat\theta_{ML} = \frac{1}{\bar X_n}$$

Mit Satz 6.2.6 erhalten wir die asymptotische Normalverteiltheit von $\hat\theta_{ML}$:

$$\sqrt{n}\left(\frac{1}{\bar X_n} - \theta\right) \xrightarrow[n\to\infty]{d} \mathrm{N}(0,\theta^2) \text{ unter } \mathbb{P}_\theta. \tag{6.2.1}$$

Auf der anderen Seite, ergibt sich aus dem zentralen Grenzwertsatz, dass

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\theta}\right) \xrightarrow[n\to\infty]{d} \mathrm{N}\left(0, \frac{1}{\theta^2}\right) \text{ unter } \mathbb{P}_\theta, \tag{6.2.2}$$

denn $\mathbb{E}X_i = \frac{1}{\theta}$ und $\operatorname{Var} X_i = \frac{1}{\theta^2}$.
Sind nun (6.2.1) und (6.2.2) äquivalent? Etwas allgemeiner kann man auch fragen: Wenn ein Schätzer asymptotisch normalverteilt ist, muss dann auch eine Funktion von diesem Schätzer asymptotisch normalverteilt sein? Wir werden nun zeigen, dass unter gewissen Voraussetzungen an die Funktion die Antwort positiv ist.

Lemma 6.2.11. *Seien $Z_1, Z_2, \dots$ Zufallsvariablen und $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ Zahlen mit*

$$\sqrt{n}(Z_n - \mu) \xrightarrow[n\to\infty]{d} N(0, \sigma^2).$$

Außerdem sei φ eine differenzierbare Funktion mit $\varphi'(\mu) \neq 0$. Dann gilt:

$$\sqrt{n}(\varphi(Z_n) - \varphi(\mu)) \xrightarrow[n\to\infty]{d} N(0, (\varphi'(\mu)\sigma)^2).$$

Beweisidee. Durch die Taylorentwicklung von φ um den Punkt μ gilt

$$\varphi(Z_n) = \varphi(\mu) + \varphi'(\mu)(Z_n - \mu) + \text{Rest}.$$

Multipliziert man nun beide Seiten mit $\sqrt{n}$, so führt dies zu

$$\sqrt{n}(\varphi(Z_n) - \varphi(\mu)) = \varphi'(\mu)\sqrt{n}(Z_n - \mu) + \text{Rest}.$$

Nach Voraussetzung gilt für den ersten Term auf der rechten Seite, dass

$$\varphi'(\mu)\sqrt{n}(Z_n - \mu) \xrightarrow[n\to\infty]{d} \mathrm{N}(0, (\varphi'(\mu)\sigma)^2).$$

Der Restterm hat eine kleinere Ordnung als dieser Term, geht also gegen 0. Daraus folgt die Behauptung. □

Beispiel 6.2.12. Als Spezialfall von Lemma 6.2.11 mit $\varphi(x) = \frac{1}{x}$ und $\varphi'(x) = -\frac{1}{x^2}$ ergibt sich die folgende Implikation:

$$\sqrt{n}(Z_n - \mu) \xrightarrow[n\to\infty]{d} \mathrm{N}(0, \sigma^2) \Longrightarrow \sqrt{n}\left(\frac{1}{Z_n} - \frac{1}{\mu}\right) \xrightarrow[n\to\infty]{d} \mathrm{N}\left(0, \frac{\sigma^2}{\mu^4}\right).$$

Daraus ergibt sich die Äquivalenz von (6.2.1) und (6.2.2).

6.3. Cramér–Rao–Ungleichung

Sei $\{h_\theta(x) : \theta \in \Theta\}$, wobei $\Theta = (a, b)$, eine Familie von Dichten oder Zähldichten. Wir haben bereits gesehen, dass es mehrere erwartungstreue Schätzer für den Parameter θ geben kann. Unter diesen Schätzern versucht man einen Schätzer mit einer möglichst kleinen Varianz zu finden. Kann man vielleicht sogar für jedes vorgegebene $\varepsilon > 0$ einen erwartungstreuen Schätzer konstruieren, dessen Varianz kleiner als ε ist? Der nächste Satz zeigt, dass die Antwort negativ ist. Er gibt eine untere Schranke an die Varianz eines erwartungstreuen Schätzers.

Satz 6.3.1 (Cramér–Rao). *Sei $\{h_\theta(x) : \theta \in \Theta\}$, wobei $\Theta = (a, b)$, eine Familie von Dichten oder Zähldichten. Seien weiterhin $X, X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte $h_\theta(x)$. Sei $\hat{\theta}(X_1, \ldots, X_n)$ ein erwartungstreuer Schätzer für θ. Unter Regularitätsbedingungen gilt die folgende Ungleichung:*

$$\operatorname{Var}_\theta \hat{\theta}(X_1, \ldots, X_n) \geq \frac{1}{nI(\theta)}.$$

Beweisidee. Da $\hat{\theta}$ ein erwartungstreuer Schätzer ist, gilt für alle $\theta \in (a, b)$, dass

$$\theta = \mathbb{E}_\theta \hat{\theta}(X_1, \ldots, X_n) = \int_{\mathbb{R}^n} \hat{\theta}(x_1, \ldots, x_n) h_\theta(x_1) \ldots h_\theta(x_n) dx_1 \ldots dx_n.$$

Nun leiten wir nach θ ab:

$$\begin{aligned} 1 &= D \int_{\mathbb{R}^n} \hat{\theta}(x_1, \ldots, x_n) h_\theta(x_1) \ldots h_\theta(x_n) dx_1 \ldots dx_n \\ &= \int_{\mathbb{R}^n} \hat{\theta}(x_1, \ldots, x_n) D[h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n)] dx_1 \ldots dx_n. \end{aligned}$$

Indem wir nun die Formel $D \log f(\theta) = \frac{Df(\theta)}{f(\theta)}$ mit $f(\theta) = h_\theta(x_1) \cdot \ldots \cdot h_\theta(x_n)$ benutzen, erhalten wir, dass

$$\begin{aligned} 1 &= \int_{\mathbb{R}^n} \hat{\theta}(x_1, \ldots, x_n) h_\theta(x_1) \ldots h_\theta(x_n) \left(\sum_{i=1}^n D \log h_\theta(x_i) \right) dx_1 \ldots dx_n \\ &= \mathbb{E}_\theta[\hat{\theta}(X_1, \ldots, X_n) U_\theta(X_1, \ldots, X_n)], \end{aligned}$$

wobei

$$U_\theta(x_1, \ldots, x_n) := \sum_{i=1}^n D \log h_\theta(x_i).$$

Es sei bemerkt, dass U_θ die Ableitung der log-Likelihood-Funktion ist. Für den Erwartungswert von U_θ gilt nach Lemma 6.2.4, dass

$$\mathbb{E}_\theta U_\theta(X_1, \ldots, X_n) = \sum_{i=1}^n \mathbb{E}_\theta D \log h_\theta(X_i) = 0.$$

Für die Varianz von U_θ erhalten wir wegen der Unabhängigkeit von $X_1, \ldots, X_n$, dass

$$\mathbb{E}_\theta[U_\theta^2(X_1, \ldots, X_n)] = \operatorname{Var}_\theta U_\theta(X_1, \ldots, X_n) = \sum_{i=1}^n \operatorname{Var}_\theta[D \log h_\theta(X_i)] = nI(\theta),$$

denn

$$\operatorname{Var}_\theta[D \log h_\theta(X_i)] = \mathbb{E}_\theta(D \log h_\theta(X_i))^2 = I(\theta),$$

da $\mathbb{E}_\theta D \log h_\theta(X_i) = 0$ nach Lemma 6.2.4.

Nun erweitern wir mit dem Erwartungswert und wenden die Cauchy–Schwarz–Ungleichung an:

$$\begin{aligned}1 &= \mathbb{E}_\theta[(\hat\theta(X_1,\dots,X_n)-\theta)\cdot U_\theta(X_1,\dots,X_n)]\\ &\le \sqrt{\operatorname{Var}_\theta[\hat\theta(X_1,\dots,X_n)]\cdot \operatorname{Var}_\theta[U_\theta(X_1,\dots,X_n)]}\\ &= \sqrt{\operatorname{Var}_\theta[\hat\theta(X_1,\dots,X_n)]\cdot nI(\theta)}.\end{aligned}$$

Umgestellt führt dies zu $\operatorname{Var}_\theta \hat\theta(X_1,\dots,X_n) \ge \frac{1}{nI(\theta)}$. □

Definition 6.3.2. Ein erwartungstreuer Schätzer $\hat\theta : \mathbb{R}^n \to \Theta$ heißt *Cramér–Rao–effizient*, falls für jedes $\theta \in \Theta$

$$\operatorname{Var}_\theta \hat\theta(X_1,\dots,X_n) = \frac{1}{nI(\theta)}.$$

Beispiel 6.3.3. Es seien $X_1,\dots,X_n$ unabhängig und Bernoulli-verteilt mit Parameter θ. Der Maximum-Likelihood-Schätzer und gleichzeitig der Momentenschätzer für θ ist der empirische Mittelwert:

$$\hat\theta(X_1,\dots,X_n) = \bar X_n = \frac{X_1+\ldots+X_n}{n}.$$

Die Varianz von $\hat\theta$ lässt sich wie folgt berechnen

$$\operatorname{Var}_\theta \hat\theta(X_1,\dots,X_n) = \operatorname{Var}_\theta\left[\frac{X_1+\ldots+X_n}{n}\right] = \frac{n}{n^2}\operatorname{Var}_\theta X_1 = \frac{\theta(1-\theta)}{n} = \frac{1}{nI(\theta)},$$

denn wir haben in Beispiel 6.2.7 gezeigt, dass $I(\theta) = \frac{1}{\theta(1-\theta)}$. Somit ist der Schätzer $\bar X_n$ Cramér–Rao–effizient. Es ist also unmöglich, einen erwartungstreuen Schätzer mit einer kleineren Varianz als die von $\bar X_n$ zu konstruieren. Somit ist $\bar X_n$ der beste erwartungstreue Schätzer für den Parameter der Bernoulli–Verteilung.

Aufgabe 6.3.4. Zeigen Sie, dass der Schätzer $\hat\theta = \bar X_n$ für die folgenden Familien von Verteilungen Cramér–Rao–effizient ist:

(1) $\{\mathrm{Poi}(\theta) : \theta > 0\}$.
(2) $\{\mathrm{N}(\theta,\sigma^2) : \theta \in \mathbb{R}\}$, wobei σ^2 bekannt ist.

Somit ist $\bar X_n$ in beiden Fällen der beste erwartungstreue Schätzer.

Bemerkung 6.3.5. Der Maximum–Likelihood–Schätzer muss nicht immer Cramér–Rao–effizient (und sogar nicht einmal erwartungstreu) sein. Wir wollen allerdings zeigen, dass bei einem großen Stichprobenumfang n der Maximum-Likelihood-Schätzer $\hat\theta_{ML}$ die Cramér–Rao–Schranke asymptotisch erreicht. Nach Satz 6.2.6 gilt unter $\mathbb{P}_\theta$, dass

$$\sqrt{n}(\hat\theta_{ML}-\theta) \xrightarrow[n\to\infty]{d} \mathrm{N}\left(0, \frac{1}{I(\theta)}\right).$$

Die Verteilung von $\hat\theta_{ML}$ ist also approximativ $\mathrm{N}(\theta, \frac{1}{nI(\theta)})$ und die Varianz ist approximativ $\frac{1}{nI(\theta)}$, bei großem n. Somit nähert sich der Maximum-Likelihood-Schätzer der Cramér–Rao–Schranke asymptotisch an.

6.4. Asymptotische Normalverteiltheit der empirischen Quantile

Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte $h(x)$ und Verteilungsfunktion $F(x)$. Sei $\alpha \in (0,1)$. Das theoretische α-Quantil Q_α der Verteilungsfunktion F ist definiert als die Lösung der Gleichung $F(Q_\alpha) = \alpha$. Wir nehmen an, dass es eine eindeutige Lösung gibt. Das empirische α-Quantil der Stichprobe $X_1, \ldots, X_n$ ist $X_{([\alpha n])}$, wobei $X_{(1)} < \ldots < X_{(n)}$ die Ordnungsstatistiken von $X_1, \ldots, X_n$ seien. Wir haben hier die Definition des empirischen Quantils etwas vereinfacht, alle nachfolgenden Ergebnisse stimmen aber auch für die alte Definition. Das empirische Quantil $X_{[\alpha n]}$ ist ein Schätzer für das theoretische Quantil Q_α. Wir zeigen nun, dass dieser Schätzer asymptotisch normalverteilt ist.

Satz 6.4.1. *Sei $\alpha \in (0,1)$ und seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte h, wobei h stetig in einer Umgebung von Q_α sei und $h(Q_\alpha) > 0$ gelte. Dann gilt*

$$\sqrt{n}(X_{([\alpha n])} - Q_\alpha) \xrightarrow[n\to\infty]{d} N\left(0, \frac{\alpha(1-\alpha)}{h^2(Q_\alpha)}\right).$$

Beweisidee. Sei $t \in \mathbb{R}$. Wir betrachten die Verteilungsfunktion

$$F_n(t) := \mathbb{P}[\sqrt{n}(X_{([\alpha n])} - Q_\alpha) \leq t] = \mathbb{P}\left[X_{([\alpha n])} \leq Q_\alpha + \frac{t}{\sqrt{n}}\right] = \mathbb{P}[K_n \geq \alpha n],$$

wobei die Zufallsvariable K_n wie folgt definiert wird:

$$K_n = \sum_{i=1}^n \mathbb{1}_{X_i \leq Q_\alpha + \frac{t}{\sqrt{n}}} = \#\left\{i \in \{1, \ldots, n\} : X_i \leq Q_\alpha + \frac{t}{\sqrt{n}}\right\}.$$

Somit ist $K_n \sim \mathrm{Bin}(n, F(Q_\alpha + \frac{t}{\sqrt{n}}))$. Für den Erwartungswert und die Varianz von K_n gilt somit

$$\mathbb{E}K_n = nF\left(Q_\alpha + \frac{t}{\sqrt{n}}\right), \quad \operatorname{Var} K_n = nF\left(Q_\alpha + \frac{t}{\sqrt{n}}\right)\left(1 - F\left(Q_\alpha + \frac{t}{\sqrt{n}}\right)\right).$$

Indem wir nun die Taylor–Entwicklung der Funktion F benutzen, erhalten wir, dass

$$\begin{aligned} \mathbb{E}K_n &= n\left(F(Q_\alpha) + F'(Q_\alpha)\frac{t}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right) = \alpha n + \sqrt{n}h(Q_\alpha)t + o(\sqrt{n}), \\ \operatorname{Var} K_n &= n\alpha(1-\alpha) + o(n). \end{aligned}$$

Nun kann die Verteilungsfunktion $F_n(t)$ wie folgt berechnet werden

$$F_n(t) = \mathbb{P}[K_n \geq \alpha n] = \mathbb{P}\left[\frac{K_n - \mathbb{E}[K_n]}{\sqrt{\operatorname{Var}(K_n)}} \geq \frac{\alpha n - \mathbb{E}[K_n]}{\sqrt{\operatorname{Var}(K_n)}}\right].$$

Benutzen wir nun die Entwicklungen von $\mathbb{E}K_n$ und $\operatorname{Var} K_n$, so erhalten wir

$$F_n(t) = \mathbb{P}\left[\frac{K_n - \mathbb{E}[K_n]}{\sqrt{\operatorname{Var}(K_n)}} \geq \frac{\alpha n - \alpha n - \sqrt{n}h(Q_\alpha)t - o(\sqrt{n})}{\sqrt{n\alpha(1-\alpha) + o(n)}}\right].$$

Und nun benutzen wir den zentralen Grenzwertsatz für K_n. Mit N standardnormalverteilt, erhalten wir, dass

$$\lim_{n\to\infty} F_n(t) = \mathbb{P}\left[N \geq -\frac{h(Q_\alpha)}{\sqrt{\alpha(1-\alpha)}}t\right] = \mathbb{P}\left[\frac{N\sqrt{\alpha(1-\alpha)}}{h(Q_\alpha)} \leq t\right],$$

wobei wir im letzten Schritt die Symmetrie der Standardnormalverteilung, also die Formel $\mathbb{P}[N \geq -x] = \mathbb{P}[N \leq x]$ benutzt haben.

Die Behauptung des Satzes folgt nun aus der Tatsache, dass $\frac{N\sqrt{\alpha(1-\alpha)}}{h(Q_\alpha)} \sim \mathrm{N}\left(0, \frac{\alpha(1-\alpha)}{h^2(Q_\alpha)}\right)$. □

Wir werden nun den obigen Satz benutzen, um den Median und den Mittelwert als Schätzer für den Lageparameter einer Dichte auf Effizienz hin zu untersuchen und zu vergleichen. Sei $h(x)$ eine Dichte. Wir machen folgende Annahmen:

(1) h ist symmetrisch, d.h. $h(x) = h(-x)$.
(2) h ist stetig in einer Umgebung von 0.
(3) $h(0) > 0$.

Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte

$$h_\theta(x) = h(x-\theta), \quad \theta \in \mathbb{R}.$$

Die Aufgabe besteht nun darin, θ zu schätzen. Dabei sei die Dichte h bekannt. Da sowohl der theoretische Median $Q_{1/2}$ als auch der Erwartungswert der Dichte h_θ wegen der Symmetrie gleich θ sind, können wir folgende natürliche Schätzer für θ betrachten:

(1) den empirischen Median $X_{([n/2])}$.
(2) den empirischen Mittelwert $\bar{X}_n$.

Welcher Schätzer ist nun besser und um wieviel? Um diese Frage zu beantworten, benutzen wir den Begriff der relativen Effizienz.

Definition 6.4.2. Seien $\hat{\theta}_n^{(1)}$ und $\hat{\theta}_n^{(2)}$ zwei Folgen von Schätzern für einen Parameter θ, wobei n den Stichprobenumfang bezeichnet. Beide Schätzer seien asymptotisch normalverteilt mit

$$\sqrt{n}(\hat{\theta}_n^{(1)} - \theta) \xrightarrow[n\to\infty]{d} \mathrm{N}(0, \sigma_1^2(\theta)) \text{ und } \sqrt{n}(\hat{\theta}_n^{(2)} - \theta) \xrightarrow[n\to\infty]{d} \mathrm{N}(0, \sigma_2^2(\theta)) \text{ unter } \mathbb{P}_\theta.$$

Die *relative Effizienz* der beiden Schätzer ist definiert durch

$$e_{\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}}(\theta) = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}.$$

Bemerkung 6.4.3. Ist z.B. $e_{\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}} > 1$ so heißt es, dass $\hat{\theta}_n^{(1)}$ besser als $\hat{\theta}_n^{(2)}$ ist.

Kommen wir nun wieder zurück zu unserer Frage, welcher Schätzer, $\hat{\theta}_n^{(1)} = X_{([n/2])}$ oder $\hat{\theta}_n^{(2)} = \bar{X}_n$, besser ist. Nach Satz 6.4.1 und nach dem zentralen Grenzwertsatz, sind beide Schätzer asymptotisch normalverteilt mit

$$\sqrt{n}(X_{([n/2])} - \theta) \xrightarrow[n\to\infty]{d} \mathrm{N}\left(0, \frac{1}{4h^2(0)}\right) \text{ und } \sqrt{n}(\bar{X}_n - \theta) \xrightarrow[n\to\infty]{d} \mathrm{N}(0, \mathrm{Var}_\theta X_1) \text{ unter } \mathbb{P}_\theta.$$

Die asymptotischen Varianzen sind also unabhängig von θ und gegeben durch

$$\sigma_1^2 = \frac{1}{4h^2(0)}, \quad \sigma_2^2 = \operatorname{Var}_\theta X_1 = \int_{\mathbb{R}} x^2 h^2(x) dx.$$

Nun betrachten wir zwei Beispiele.

Beispiel 6.4.4. Die Gauß–Dichte $h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Es gilt $h(0) = 0$, $\operatorname{Var}_\theta X_1 = 1$ und somit

$$\sigma_1^2 = \frac{\pi}{2}, \quad \sigma_2^2 = 1, \quad e_{Med,MW} = \frac{2}{\pi} \approx 0.6366 < 1.$$

Somit ist der empirische Mittelwert besser als der empirische Median. Das Ergebnis kann man so interpretieren: Der Median erreicht bei einer Stichprobe vom Umfang 100 in etwa die gleiche Präzision, wie der Mittelwert bei einer Stichprobe vom Umfamg 64.

Beispiel 6.4.5. Die Laplace–Dichte $h(x) = \frac{1}{2} e^{-|x|}$. Es gilt $h(0) = \frac{1}{2}$, $\operatorname{Var}_\theta X_1 = 2$ und somit

$$\sigma_1^2 = 1, \quad \sigma_2^2 = 2, \quad e_{Med,MW} = 2 > 1.$$

In diesem Beispiel ist also der Median besser: Bei einer Stichprobe vom Umfang 100 erreicht der Median in etwa die gleiche Präzision, wie der Mittelwert bei einer Stichprobe vom Umfang 200.