Einführung in die Stochastik
13. Übung
Gruppenübung: 23./24.06.2008
Lösungsvorschlag

Gruppenübung

G 21 Wir betrachten ein statistisches Experiment, gegeben durch $(\Omega, \mathfrak{A}, P^\vartheta)_{\vartheta\in\Theta}$ mit P^ϑ und $X = (X_1, \ldots, X_n)$. Sei $x = (x_1, \ldots, x_n)$ eine Realisierung von X.
Analog zum Vorgehen in Abschnitt VIII.2. definieren wir die *Likelihood-Funktion* $L_x(\vartheta)$ gemäß

$$L_x(\vartheta) = P^\vartheta(\{X = x\}).$$

(i) Es gelte $X_1 \sim \mathbf{P}(\lambda)$. Geben Sie die Likelihood-Funktion an.

(ii) Wir betrachten das Schätzproblem $\gamma(\lambda) = \lambda$. Überlegen Sie, wie sich mit Hilfe der Likelihood-Funktion eine Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ für $\gamma(\lambda)$ bestimmen läßt und berechnen Sie diese.

(i) Aufgrund der Unabhängigkeit von $X_1, \ldots, X_n$ folgt

$$L_x(\lambda) = \prod_{i=1}^n P(\{X_i = x_i\}) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{\prod\limits_{i=1}^n x_i!}.$$

(ii) Wir verwenden als Schätzer für λ denjenigen Wert, für den die Wahrscheinlichkeit der beobachteten Realisierung maximal ist, d.h. gesucht ist $\lambda > 0$, so daß

$$L_x(\lambda) = P(\{X = (x_1, \ldots, {}_n)\}) = \prod_{i=1}^n P(\{X_i = x_i\})$$

maximal wird. Aufgrund der strengen Monotonie der ln*-Funktionen ist die Maximalstelle von $L_x(\lambda)$ gleich der Maximalstelle von $\ell_x(\lambda)$. Somit*

$$\frac{d}{d\lambda}\ell_x(\lambda) = -n + \sum_{i=1}^n x_i \stackrel{!}{=} 0$$

mit Lösung $\hat\lambda = \frac{1}{n}\sum_{i=1}^n x_i$.

Es bleibt zu überprüfen, daß $\hat\lambda$ Maximalstelle ist. Wegen

$$\frac{d^2}{d\lambda^2}\ell_x(\lambda)\big|_{\lambda=\hat\lambda} = -\sum_{i=1}^n \frac{x_i}{\lambda^2}\big|_{\lambda=\hat\lambda} < 0$$

folgt die Behauptung und es gilt

$$g_n(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i.$$

Hinweis: *Eine Schätzfunktion, welche auf diese Art und Weise bestimmt wird, heißt* Maximum-Likelihood-Schätzer.

G 22 Sei g_n eine Schätzfunktion für γ und $R^\vartheta(g_n)$ bzw. $B^\vartheta(g_n)$ der Quadratmittelfehler bzw. Bias. Zeigen Sie:

(i) $R^\vartheta(g_n) = Var^\vartheta(g_n) + B^\vartheta(g_n)^2$

Wir betrachten nun ein statistisches Modell mit n unabhängigen auf $[0, \vartheta]$ gleichverteilten Zufallsvariablen $X_1, \ldots, X_n$. Als Schätzfunktion für $\gamma(\vartheta) = \vartheta$ wählen wir

$$g_n(x) = \max_{i=1,\ldots,n} \{x_i\}.$$

(ii) Zeigen Sie, daß $g_n(x)$ nicht erwartungstreu für γ ist und definieren Sie ein $c_n > 0$, so daß $h_n := c_n \cdot g_n$ eine erwartungstreue Schätzfunktion ist.

(iii) Verwenden Sie das arithmetische Mittel, um eine erwartungstreue Schätzfunktion m_n für γ zu konstruieren.
Berechnen Sie für g_n, h_n und m_n den Quadratmittelfehler und vergleichen Sie diese! Nutzen Sie Aussage (i)!

(i)

$$\begin{aligned}
R^\vartheta(g_n) &= E^\vartheta(g_n(X) - \gamma(\vartheta))^2 \\
&= E^\vartheta((g_n(X))^2) - 2E^\vartheta(g_n(X) \cdot \gamma(\vartheta)) + E^\vartheta((\gamma(\vartheta))^2) \\
&= E^\vartheta((g_n(X))^2) - (E^\vartheta(g_n(X)))^2 + (E^\vartheta(g_n(X)))^2 \\
&\quad - 2\gamma(\vartheta)E^\vartheta(g_n(X)) + E^\vartheta((\gamma(\vartheta))^2) \\
&= Var(g_n(X)) + (E^\vartheta(g_n(X)) - \gamma(\vartheta))^2 = Var(g_n(X)) + \Big(B^\vartheta(g_n)\Big)^2
\end{aligned}$$

(ii)Es gilt

$$\begin{aligned}
F^\vartheta_{g_n}(z) &= P(\{\max\{X_1, \ldots, X_n\} \le z\}) \\
&= P(\{X_1 \le z, X_2 \le z, \ldots, X_n \le z\}) \\
&= \prod_{i=1}^n F_i^\vartheta(z) = \left(\frac{z}{\vartheta}\right)^n.
\end{aligned}$$

Demzufolge lautet die Dichte

$$f^\vartheta_{g_n}(z) = \begin{cases} n\frac{z^{n-1}}{\vartheta^n} & z \in [0, \vartheta] \\ 0 & \text{sonst.} \end{cases}$$

Somit

$$E^\vartheta(g_n(X)) = \int_0^\vartheta z \cdot n\frac{z^{n-1}}{\vartheta^n} dz = \frac{n}{n+1}\vartheta \ne \vartheta.$$

Betrachten wir demnach die Funktion $h_n(x) = \frac{n+1}{n} g_n(x)$, so ist h_n erwartungstreu.

(iii) Wir wissen, daß für $X \sim \mathbf{U}[0, \vartheta]$ gilt: $E(X) = \frac{\vartheta}{2}$. Außerdem ist $\frac{1}{n}\sum_{i=1}^n x_i$ eine erwartungstreue Schätzfunktion für $E(X) = \frac{\vartheta}{2}$. Also ist doch

$$m_n(x) = \frac{2}{n}\sum_{i=1}^n x_i$$

eine erwartungstreue Schätzfunktion für $\gamma(\vartheta) = \vartheta$.

Für die Berechnung des Quadratmittelfehlers nutzen wir die Formel aus (i). Dazu

$$E^\vartheta((g_n(X))^2) = \int_0^\vartheta z^2 \cdot n \frac{z^{n-1}}{\vartheta^n} dz = \frac{n}{n+2}\vartheta^2$$

$$Var^\vartheta(g_n(X)) = E^\vartheta((g_n(X))^2) - (E^\vartheta(g_n(X)))^2 = \frac{n}{(n+1)^2(n+2)}\vartheta^2$$

$$\begin{aligned} R^\vartheta(g_n) &= Var^\vartheta(g_n(X)) + B^\vartheta(g_n;\gamma) \\ &= \frac{n}{(n+1)^2(n+2)}\vartheta^2 + (E^\vartheta(g_n(X)) - \gamma(\vartheta))^2 \\ &= \frac{n}{(n+1)^2(n+2)}\vartheta^2 + \Big(\frac{n}{n+1}\vartheta - \vartheta\Big)^2 \\ &= \frac{2}{(n+1)(n+2)}\vartheta^2 \end{aligned}$$

Da $h_n(x)$ ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ ist, folgt mit Bemerkung VIII.2.14

$$\begin{aligned} R^\vartheta(h_n) &= Var^\vartheta(h_n(X)) = Var^\vartheta(\frac{n+1}{n}g_n(X)) \\ &= (\frac{n+1}{n})^2 Var^\vartheta(g_n(X)) = (\frac{n+1}{n})^2 \frac{n}{(n+1)^2(n+2)\vartheta^2} \\ &= \frac{1}{n(n+2)}\vartheta^2 \end{aligned}$$

$$R^\vartheta(m_n) = Var^\vartheta(m_n(X)) = \frac{4}{n^2}Var^\vartheta(\sum_{i=1}^n X_i)$$

$$= \frac{1}{3n}\vartheta^2$$

Für $n > 1$ gilt $R^\vartheta(h_n) < R^\vartheta(g_n)$ und $R^\vartheta(h_n) < R^\vartheta(m_n)$. Für $n > 3$ gilt $R^\vartheta(g_n) < R^\vartheta(m_n)$.

G 23 Vor einer Theaterkasse warten in einer Schlange 40 Personen, deren Bedienung im Mittel jeweils 50 Sekunden dauert. Es wird angenommen, daß sich die Bedienungszeiten durch unabhängige, identisch exponentialverteilte Zufallsvariablen beschreiben lassen. Berechnen Sie näherungsweise die Wahrscheinlichkeit dafür, daß alle 40 Personen innerhalb von 35 Minuten bedient werden.

Die Zufallsvariable X_i beschreibe die Bedienungsdauer der i-ten Person ($i = 1, \ldots, 40$). Dann kann $(X_i)_{i\in\mathbb{N}}$ als unabhängige, identisch verteilte Folge angenommen werden. Aus $X_1 \sim$ **Exp**(λ) *und $E(X_1) = 5/6$ folgt $\lambda = 6/5$ (s. Beispiel VI.6). Die Summe $S := X_1 + \ldots + X_{40}$ ist nach dem Zentralen Grenzwertsatz näherungsweise normalverteilt, und zwar mit den Parametern (verwende Beispiel VI.6)*

$$\mu = E(S) = 40E(X_1) = 33\,\frac{1}{3},$$

$$\sigma^2 = Var(S) = 40 \cdot Var(X_1) = 27\,\frac{7}{9}.$$

Damit folgt

$$\begin{aligned} P(\{S \le 35\}) &\approx P\left(\left\{S^* \le \frac{35 - 33\,\frac{1}{3}}{\sqrt{27\,\frac{7}{9}}}\right\}\right) \\ &\approx \Phi(0.32) = 0.6255. \end{aligned}$$