Nachhol-Klausur zur Vorlesung Schätzen und Testen I

26. April 2011

Christian Heumann, Volker Schmid, Julia Kärcher

Aufgabe	1	2	3	4	5	$\sum$	Note
Punkte							

Hinweise:

- Überprüfen Sie zunächst, ob Ihr Klausurexemplar vollständig ist. Die Klausur sollte aus drei Blättern (ohne dieses Deckblatt) mit fünf Aufgaben bestehen.
- Verwenden Sie für Ihre Lösungen ausschließlich die Ihnen zur Verfügung gestellten Papierbögen. Beginnen Sie für jede Aufgabe eine neue Seite.
- Kennzeichnen Sie jedes abgegebene Blatt mit Ihrem Namen und Ihrer Matrikelnummer.
- Als Hilfsmittel sind zugelassen:
  - drei handschriftlich beschriebene DIN-A4-Blätter,
  - die in der Übung verteilte Übersicht konjugierter Verteilungen,
  - die in der Übung verteilten Verteilungsübersichten aus dem Buch von Gelman, Carlin, Stern und Rubin (2004),
  - ein nichtprogrammierbarer Taschenrechner,
  - ein Wörterbuch.
- Maximal können 100 Punkte erreicht werden. Die Klausur ist mit 45 Punkten bestanden.
- Die Bearbeitungszeit beträgt 120 Minuten.
- Bei Unterschleif gilt die Klausur als nicht bestanden und es erfolgt eine Meldung an das Prüfungsamt.

Bitte ausfüllen und unterschreiben!!!

Name, Vorname: ______________________________

Matrikelnummer: ____________________ *Studiengang:* ____________________

Abschluss: ____________________ *PO-Version (Jahr):* ____________________

Ich bestätige, dass ich obige Hinweise zur Kenntnis genommen habe und sie befolgen werde.

Unterschrift: ______________________________

Viel Erfolg!

Aufgabe 1	20 Punkte

Seien $X_1, \dots, X_n \overset{i.i.d.}{\sim} U[0, \theta]$ mit unbekanntem Parameter $\theta > 0$, d.h. die X_i haben die Dichte

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

für $x \in \mathbb{R}$ ($\mathbb{1}$ bezeichnet die Indikatorfunktion). Im Folgenden soll θ geschätzt werden.

(a) Liegt hier eine Fisher-reguläre Verteilungsfamilie vor? Begründen Sie Ihre Antwort.

(b) Zeigen Sie: $T = \max\{X_1, \dots, X_n\}$ ist suffizient für θ.

(c) Zeigen Sie: $V = 2\bar{X} = \frac{2}{n}\sum_{i=1}^n X_i$ ist erwartungstreu für θ.

(d) Berechnen Sie aus V einen verbesserten erwartungstreuen Schätzer V^* für θ in dem Sinne, dass $\mathrm{Var}_\theta(V^*) \leq \mathrm{Var}_\theta(V)$. Begründen Sie Ihr Vorgehen.

Aufgabe 2	21 Punkte

Seien $X_1, \dots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ mit unbekanntem Parameter $\mu \in \mathbb{R}$ und bekanntem $\sigma^2 \in \mathbb{R}_+$. Betrachten Sie für $\mu_0 > \mu_1$ die Hypothesen

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu = \mu_1.$$

(a) Zeigen Sie, dass der entsprechende Likelihood-Quotient die Form

$$\Lambda(x) = \Lambda(x_1, \dots, x_n) = \exp\left(-\frac{1}{2\sigma^2}\Big(n(\mu_1^2 - \mu_0^2) - 2(\mu_1 - \mu_0)z\Big)\right)$$

für $x_1, \dots, x_n \in \mathbb{R}$ und $z = \sum_{i=1}^n x_i$ hat.

(b) Begründen Sie, weshalb ein Test zum exakten Niveau α, der auf der Alternative die größte Güte hat, äquivalent ist zu

$$\phi(z) \quad = \quad \begin{cases} 1 & \text{, falls } z < c_\alpha \\ 0 & \text{, falls } z \geq c_\alpha. \end{cases}$$

Wieso ist keine Randomisierung notwendig?

(c) Bestimmen Sie den kritischen Wert c_α.

(d) Sind die Hypothesen H_0 und H_1 in obigem Test als gleichwertig anzusehen? Begründen Sie Ihre Antwort.

Aufgabe 3	31 Punkte

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen aus der verallgemeinerten Exponentialverteilung mit Dichte

$$g(x) = \begin{cases} \frac{1}{b} \exp\left(-\frac{x-a}{b}\right) & \text{für } x \geq a \\ 0 & \text{sonst} \end{cases}$$

für Parameter $a \in \mathbb{R}$ und $b > 0$. Es gilt $\mathrm{E}(X_i) = a + b$ und $\mathrm{Var}(X_i) = b^2$ für alle i.

(a) Beschreiben Sie die verallgemeinerte Exponentialverteilung als Lokations- und Skalenfamilie.

(b) Für Beobachtungen $x_1, \ldots, x_n \geq a$: Berechnen Sie Likelihood- und Log-Likelihoodfunktion von a und b. Zeigen Sie, dass die Scorefunktion von a und b wie folgt aussieht:

$$s(a, b) = \begin{pmatrix} \frac{n}{b} \\ -\frac{n}{b} + \frac{n(\bar{x} - a)}{b^2} \end{pmatrix},$$

wobei $\bar{x} = (\sum_{i=1}^n x_i)/n$.

(c) Sei a bekannt. Wie lautet der Maximum-Likelihood-Schätzer für b? (Ein Nachweis durch zweite Ableitungen wird nicht verlangt!)

(d) Sei b bekannt. Wie lautet der Maximum-Likelihood-Schätzer für a?

(e) Berechnen Sie die erwartete Fisher-Information von b.

(f) Geben Sie eine asymptotische Verteilung für die Folge von Maximum Likelihood-Schätzer $\hat{b}_n$ für b an. Verwenden Sie eine geeignete Normierung und begründen Sie Ihr Vorgehen.

Im Folgenden sei $a = 0$, und der Parameter b soll geschätzt werden. Bei der Schätzung von b werde das Modell nun allerdings fehlspezifiziert: Für $X_1, \ldots, X_n$ wird die Dichtefunktion f_{σ^2} der trunkierten Normalverteilung angenommen, d.h.

$$f_{\sigma^2}(x) = \begin{cases} \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

für $\sigma^2 > 0$.

(g) Zeigen Sie, dass gilt:

$$\mathrm{E}_g\big(\log f_{\sigma^2}(X)\big) = \frac{1}{2} \log\left(\frac{2}{\pi}\right) - \frac{1}{2} \log \sigma^2 - \frac{b^2}{\sigma^2}.$$

(h) Berechnen Sie dasjenige $\sigma^2 > 0$, das die Kullback-Leibler-Distanz $D(g, f_{\sigma^2})$ minimiert. (Ein Nachweis durch zweite Ableitungen wird nicht verlangt!)

Aufgabe 4	16 Punkte

Betrachten Sie das bayesianische lineare Modell

$$\boldsymbol{y} \sim \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\ , \quad \boldsymbol{\varepsilon} \sim \mathrm{MVN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$

mit $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma^2 > 0$. A priori seien

$$p(\boldsymbol{\beta}|\boldsymbol{X}) \propto 1 \qquad \text{und} \qquad \sigma^2|\boldsymbol{X} \sim \mathrm{IG}(a, b)\ ,\ a, b \in \mathbb{R}_+.$$

(a) Berechnen Sie die gemeinsame Posteriori $f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{X})$ und erläutern Sie Ihr Vorgehen.

(b) Berechnen Sie die vollständig bedingten Dichten von $\boldsymbol{\beta}$ und σ^2 und stellen Sie sie (soweit möglich) als Dichten von bekannten Verteilungen dar.
Hinweis: Die vollständige bedingte Dichte von $\boldsymbol{\beta}$ ist proportional zum Kern einer multivariaten Normalverteilung, die vollständige bedingte Dichte von σ^2 ist proportional zum Kern einer inversen Gamma-Verteilung.

(c) Wie kann man hier aus der Posteriori-Randdichte $f(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X})$ simulieren? Erläutern Sie.

Aufgabe 5	12 Punkte

Die folgende Tabelle enthält sechs verschiedene Aussagen. Geben Sie für jede dieser Aussagen an, ob sie wahr oder falsch ist, indem Sie das entsprechende Feld ankreuzen. Sie erhalten für jedes richtig angekreuzte Feld zwei Punkte, **für jedes falsch angekreuzte zwei Minuspunkte**. Sie können in dieser Aufgabe maximal 12 und minimal 0 Punkte erreichen.

Aussage	**Wahr**	**falsch**
Es ist immer möglich den MSE-optimalen Schätzer für einen Parameter zu bestimmen.	☐	☐
Ein zulässiger Schätzer ist ein erwartungstreuer Schätzer, dessen MSE von keinem anderen Schätzer strikt dominiert wird.	☐	☐
Unter Fisher-Regularität ist eine Statistik genau dann suffizient, wenn die Fisher-Information der Statistik gleich der Fisher-Information der Stichprobe ist.	☐	☐
Ein erwartungstreuer Schätzer ist genau dann UMVU-Schätzer, wenn die Varianz des Schätzers die Rao-Cramer-Schranke annimmt.	☐	☐
Mit der Bonferroni-Prozedur wird in einem multiplen Testproblem die FWER (family-wise error rate) stark kontrolliert.	☐	☐
Jeffrey's Priori heißt nicht-informativ, da sie größtmögliche Varianz hat.	☐	☐