TAMS17/TEN1 STATISTISK TEORI FK
TENTAMEN FREDAG 1/4 2016 KL 08.00-12.00.

Examinator och jourhavande lärare: Torkel Erhardsson, tel. 28 14 78.

Tillåtna hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen (utan egna anteckningar). Ett A4-ark med egna anteckningar på båda sidor. Räknedosa med tömda minnen.

Tentamen består av 6 uppgifter värda 3 poäng vardera. Betygsgränser: 8 poäng för 3, 11.5 poäng för 4, 15 poäng för 5. Resultatet meddelas via e-post.

Uppgift 1

Låt $X = (X_1, \ldots, X_n)$, där $X_1, \ldots, X_n$ är oberoende likafördelade stokastiska variabler, med täthetsfunktionen

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta^2}, & \text{för } 0 < x < \theta; \\ 0, & \text{annars,} \end{cases}$$

där $\theta > 0$ är en okänd parameter.

(a) Är $\{f(x|\theta); \theta > 0\}$ en exponentialfamilj? Svaret måste motiveras.

(b) Är $\{f(x|\theta); \theta > 0\}$ en lägesfamilj, en skalfamilj, eller ingetdera? Svaret måste motiveras.

(c) Beräkna en (endimensionell) tillräcklig statistika för θ.

Uppgift 2

Låt $X = (X_1, \ldots, X_n)$, där $X_1, \ldots, X_n$ är oberoende likafördelade stokastiska variabler, med täthetsfunktionen

$$f(x|\lambda) = \sqrt{\frac{\lambda}{\pi x}}\, e^{-\lambda x} \qquad \forall x > 0,$$

där parametern $\lambda > 0$ är okänd. Beräkna ML-skattaren av λ baserad på $X_1, \ldots, X_n$, samt ett konfidensintervall för λ med ML-skattaren som mittpunkt och med asymptotisk konfidensgrad 95% (då $n \to \infty$). OBS: de villkor som garanterar att ML-skattaren är konsistent och asymptotiskt normalfördelad är uppfyllda, vilket inte behöver visas.

Uppgift 3

Låt $X \sim \mathrm{Po}(\mu)$, där parametern $\mu > 0$ är okänd. Visa att familjen $\{\mathrm{Po}(\mu); \mu > 0\}$ har monoton likelihoodkvot (MLR), samt utnyttja detta för att bestämma det likformigt starkaste testet (UMP) av $H_0 : \mu \leq 3$ mot $H_1 : \mu > 3$ baserat på X, på en valfri signifikansnivå i intervallet $[0.01, 0.05]$. Ange den valda signifikansnivån, samt testets styrka mot alternativet $\mu = 6$.

Uppgift 4

Låt X vara en stokastisk variabel med täthetsfunktion

$$f(x|\theta) = \begin{cases} \theta^2 x e^{-\theta x}, & \text{om } x > 0; \\ 0, & \text{annars,} \end{cases}$$

där parametern $\theta > 1$ är okänd. Låt θ ha en apriorifördelning med täthetsfunktion $f(\theta) = \frac{1}{\theta^2}$ för $\theta > 1$ (och $f(\theta) = 0$ annars).

(a) Bestäm aposteriorifördelningen (= dess täthetsfunktion) för θ.

(b) Beräkna det kortaste “Bayesianska konfidensintervallet” (credible set) för θ av grad $1 - \alpha$.

Uppgift 5

Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(0, \sigma)$-fördelade stokastiska variabler, där standardavvikelsen σ är okänd. ML-skattaren av σ är $\widehat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} X_i^2}$, vilket inte behöver visas. Härled likelihoodkvottestet av $H_0 : \sigma = 1$ mot $H_1 : \sigma \neq 1$, och visa att testet består i att H_0 avvisas om $\widehat{\sigma} < a$ eller $\widehat{\sigma} > b$, där $0 < a < b$ är konstanter som bestämmer testets signifikansnivå.
Ledning: funktionen $H(z) = z^n \exp(-\frac{n}{2}(z^2 - 1))$ har derivatan $H'(z) = nz^{n-1}(1 - z^2) \exp(-\frac{n}{2}(z^2 - 1))$ för $z \geq 0$.

Uppgift 6

Låt $X = (X_1, \ldots, X_n)$, där $X_1, \ldots, X_n$ är oberoende $\mathrm{Be}(\theta)$-fördelade stokastiska variabler, och där $\theta \in (0, 1)$ är en okänd parameter. (X_1 antar alltså värdet 1 med sannolikhet θ, och värdet 0 med sannolikhet $1 - \theta$.) Låt $Y = \sum_{i=1}^{n} X_i$. Då är $Y \sim \mathrm{Bin}(n, \theta)$, vilket inte behöver visas.

(a) Visa att Y är en tillräcklig och fullständig statistika för θ.

(b) Beräkna den betingade sannolikheten $P(X_1 X_2 = 1 | Y = y)$. *Ledning:* utnyttja oberoendet mellan $X_1, \ldots, X_n$, samt att $\{X_1 X_2 = 1\} = \{X_1 = 1\} \cap \{X_2 = 1\}$.

(c) Bestäm den bästa väntevärdesriktiga skattningen (UMVUE) av θ^2 (OBS!) baserad på $X_1, \ldots, X_n$. *Ledning:* vad är $E(X_1 X_2)$?

KORTFATTADE LÖSNINGAR TILL TENTAMEN I TAMS17 STATISTISK TEORI FK, FREDAG 1/4 2016 KL 08.00-12.00.

1. (a) Nej, ty området $\{x \in \mathbb{R}; f(x|\theta) > 0\}$ beror av θ, vilket det inte kan göra om $\{f(x|\theta); \theta > 0\}$ är en exponentialfamilj.

   (b) Den är en skalfamilj, eftersom

$$f(x|\theta) = \frac{1}{\theta} f(\frac{x}{\theta}) \qquad \forall x > 0, \theta > 0,$$

   där $f(x) = 2x$ för $0 < x < 1$, och $f(x) = 0$ annars.

   (c) $L(\theta|x) = \Pi_{i=1}^n f(x_i|\theta) = \Pi_{i=1}^n \frac{2x_i}{\theta^2} I\{0 < x_i < \theta\} = \frac{2^n \prod_{i=1}^n x_i}{\theta^{2n}} I\{0 < \min(x_1, \ldots, x_n)\} I\{\max(x_1, \ldots, x_n) < \theta\}$. Enligt Fisher-Neymans faktoriseringssats är $T(X) = \max(X_1, \ldots, X_n)$ tillräcklig för θ.

2. X har likelihoodfunktionen $L(\lambda|x) = C\lambda^{n/2} \exp(-\lambda \sum_{i=1}^n x_i)$, så $l(\lambda) = \ln L(\lambda|x) = \ln C + \frac{n}{2} \ln \lambda - \lambda \sum_{i=1}^n x_i$. Ekvationen $l'(\lambda) = \frac{n}{2\lambda} - \sum_{i=1}^n x_i = 0$ har lösningen $\lambda = \frac{1}{2\bar{x}}$, vilket är ett maximum, ty $l''(\lambda) = -\frac{n}{2\lambda^2} < 0$. ML-skattaren av λ är därför $\widehat{\lambda} = \frac{1}{2\bar{X}}$. Vidare gäller enligt uppgiften att

$$\frac{\sqrt{n}(\widehat{\lambda} - \lambda)}{\mathcal{I}_{X_1}(\lambda)^{-1/2}} \xrightarrow{d} Z \sim \mathrm{N}(0,1)$$

   då $n \to \infty$, där $\mathcal{I}_{X_1}(\lambda) = -E(\frac{d^2}{d\lambda^2} \ln f(X_1|\lambda)) = \frac{1}{2\lambda^2}$. Ett konfidensintervall för λ av Wald-typ, med ML-skattaren som mittpunkt och med asymptotisk konfidensgrad 95%, är då: $C(x) = \widehat{\lambda} \pm 1.96\sqrt{2}\widehat{\lambda}\frac{1}{\sqrt{n}} = \widehat{\lambda} \pm \widehat{\lambda}\frac{2.77}{\sqrt{n}}$.

3. Låt $0 < \mu_1 < \mu_2$. Då är

$$\frac{f(x|\mu_2)}{f(x|\mu_1)} = \frac{\mu_2^x e^{-\mu_2}}{\mu_1^x e^{-\mu_1}} = (\frac{\mu_2}{\mu_1})^x e^{-(\mu_2 - \mu_1)},$$

   vilket är en växande funktion i x, så familjen har MLR. Enligt Karlin-Rubins sats är därför det likformigt starkaste testet $\phi(x) = I\{x > k\}$, där k väljs så att $P_{\mu=3}(X > k) \in [0.01, 0.05]$. Enligt tabellen i F.S. kan vi t.ex. välja $k = 6$, vilket ger signifikansnivån $\alpha = 0.0335$, och styrkan $\beta(6) = 0.394$.

4. (a) $f_{X,\Theta}(x,\theta) = f_{X|\Theta=\theta}(x) f_\Theta(\theta) = \theta^2 x e^{-\theta x} \frac{1}{\theta^2} = x e^{-\theta x}$ för $x > 0$, $\theta > 1$. Detta ger att

$$f_X(x) = \int_1^\infty x e^{-\theta x} d\theta = \left[-e^{-\theta x}\right]_{\theta=1}^{\theta=\infty} = e^{-x} \qquad \forall x > 0.$$

Alltså är aposteriorifördelningens täthetsfunktion

$$f_{\Theta|X=x}(\theta) = \frac{f_{X,\Theta}(x,\theta)}{f_X(x)} = \underline{\underline{x e^{-x(\theta-1)} \quad \forall \theta > 1}}.$$

(b) Eftersom aposteriorifördelningens täthetsfunktion är strängt avtagande för $\theta > 1$, så är det kortaste "Bayesianska konfidensintervallet" $\underline{\underline{C(x) = [1, u(x)]}}$ där $u(x)$ är lösningen till ekvationen

$$\int_1^{u(x)} x e^{-x(\theta-1)} d\theta = \left[-e^{-x(\theta-1)}\right]_{\theta=1}^{\theta=u(x)} = 1 - e^{-x(u(x)-1)} = 1 - \alpha,$$

d.v.s., $\underline{\underline{u(x) = 1 - \frac{1}{x}\ln\alpha}}$.

5. Likelihoodkvotteststatistikan är:

$$\lambda(x) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x_i^2)}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp(-\frac{1}{2\widehat{\sigma}^2}x_i^2)} = \widehat{\sigma}^n \exp(-\frac{1}{2}(1 - \frac{1}{\widehat{\sigma}^2})\sum_{i=1}^n x_i^2)$$

$$= \widehat{\sigma}^n \exp(-\frac{1}{2}(1 - \frac{1}{\widehat{\sigma}^2}) n\widehat{\sigma}^2) = \widehat{\sigma}^n \exp(-\frac{n}{2}(\widehat{\sigma}^2 - 1)) = H(\widehat{\sigma}).$$

Testet har testfunktionen $\phi(x) = I\{\lambda(x) < k\} = I\{H(\widehat{\sigma}) < k\}$, där $k \in (0,1)$ är en konstant. Ett teckenstudium av derivatan ger att funktionen H är strängt växande för $0 \le z \le 1$ och strängt avtagande för $z \ge 1$, så $H(\widehat{\sigma}) < k$ om och endast om $\underline{\underline{\widehat{\sigma} < a \text{ eller } \widehat{\sigma} > b}}$, där $0 < a < b$ är lösningarna till $H(a) = H(b) = k$.

6. (a) Eftersom $f_X(x|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^y(1-\theta)^{n-y}$, så är Y tillräcklig enligt faktoriseringssatsen. Vidare är

$$\theta^y(1-\theta)^{n-y} = (1-\theta)^n \exp(y\ln\theta - y\ln(1-\theta)) = (1-\theta)^n \exp(y \ln\frac{\theta}{1-\theta}),$$

så $\{f_X(x|\theta); \theta \in (0,1)\}$ är en exponentialfamilj sådan att $\{\ln\frac{\theta}{1-\theta}; \theta \in (0,1)\} = \mathbb{R}$. Alltså Y är fullständig.

(b)

$$P(X_1X_2 = 1|Y = y) = \frac{P(\{X_1X_2 = 1\} \cap \{\sum_{i=3}^n X_i = y - 2\})}{P(\sum_{i=1}^n X_i = y)}$$

$$= \frac{P(X_1X_2 = 1)P(\sum_{i=3}^n X_i = y - 2)}{P(\sum_{i=1}^n X_i = y)} = \frac{\theta^2\binom{n-2}{y-2}\theta^{y-2}(1-\theta)^{n-y}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} = \underline{\underline{\frac{y(y-1)}{n(n-1)}}}.$$

(c) X_1X_2 är en väntevärdesriktig skattare av θ^2, ty $E(X_1X_2) = P(X_1X_2 = 1) = \theta^2$. UMVUE ges därför av:

$$E(X_1X_2|Y) = P(X_1X_2 = 1|Y = y)|_{y=Y} = \underline{\underline{\frac{Y(Y-1)}{n(n-1)}}}.$$