TAMS17/TEN1 STATISTISK TEORI FK
TENTAMEN ONSDAG 24/8 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 $X_1, \ldots, X_n$ vara oberoende Ge($\frac{\theta}{1+\theta}$)-fördelade stokastiska variabler, med sannolikhetsfunktionen

$$f(x|\theta) = \begin{cases} \frac{1}{1+\theta}(\frac{\theta}{1+\theta})^x, & \text{för } x = 0, 1, 2, \ldots; \\ 0, & \text{annars,} \end{cases}$$

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

(a) Beräkna den nedre gräns för variansen av en väntevärdesriktig skattare av θ som ges av Cramér-Raos olikhet. OBS: regularitetsvillkoren för Cramér-Raos olikhet är uppfyllda, vilket <u>inte</u> behöver visas.

(b) ML-skattaren av θ är $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, vilket <u>inte</u> behöver visas. Visa att ML-skattaren har lägst varians bland alla väntevärdesriktiga skattare av θ.

Uppgift 2

Låt X vara en stokastisk variabel med täthetsfunktion

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

där $\theta > 0$ är en okänd parameter. Ange en pivotvariabel för θ, och utnyttja denna för att bestämma ett ensidigt uppåt begränsat konfidensintervall för θ av exakt konfidensgrad 95%.

Uppgift 3

Låt X vara en stokastisk variabel om vilken två hypoteser föreligger: H_0 : "$X \sim \text{Bin}(8, 0.25)$" respektive H_1 : "$X \sim \text{Bin}(8, 0.75)$". Bestäm det starkaste testet av H_0 mot H_1 på valfri signifikansnivå α, i intervallet $0.01 \leq \alpha \leq 0.05$. Ange den valda signifikansnivån, samt testets styrka mot H_1.

Uppgift 4

Låt $X_1, \ldots, X_n$ vara oberoende och likafördelade stokastiska variabler, vars täthetsfunktion tillhör läges–och skalfamiljen $\{f(x|\mu, \sigma); \mu \in \mathbb{R}, \sigma > 0\}$, där

$$f(x|\mu,\sigma) = \begin{cases} \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x-\mu)}, & \text{för } x > \mu; \\ 0, & \text{annars.} \end{cases}$$

(a) Bestäm en tvådimensionell tillräcklig statistika för (μ, σ).

(b) Beräkna ML-skattarna av μ och σ.

Uppgift 5

Låt $X_1, \ldots, X_n$ vara oberoende $\text{N}(0, \sigma)$-fördelade stokastiska variabler, vars varians $\sigma^2 > 0$ är okänd. ML-skattaren av σ^2 är $\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$, vilket <u>inte</u> behöver visas.

(a) Bestäm ett tal $b > 0$ (som eventuellt beror av σ^2) så att $\frac{\sqrt{n}}{b}(\widehat{\sigma}_n^2 - \sigma^2)$ konvergerar i fördelning mot $Z \sim \text{N}(0, 1)$ då $n \to \infty$. *Ledning:* $E(X_1^4) = 3\sigma^4$.

(b) Bestäm ett tal $c > 0$ (som eventuellt beror av σ^2) så att $\frac{\sqrt{n}}{c}(\exp(\widehat{\sigma}_n^2/2) - \exp(\sigma^2/2))$ konvergerar i fördelning mot $Z \sim \text{N}(0, 1)$ då $n \to \infty$.

(c) Bestäm ett konfidensintervall för $\exp(\sigma^2/2)$ med asymptotisk konfidensgrad 95% då $n \to \infty$.

Uppgift 6

Låt x vara det observerade värdet av en $\text{U}(0, \theta)$-fördelad stokastisk variabel, där $\theta > 0$ är en okänd parameter. En Bayesiansk analys genomförs, där θ betraktas som en stokastisk variabel Θ med apriorifördelningen $\Gamma(2, 1)$. Aposteriorifördelningens täthetsfunktion är då

$$f_{\Theta|X=x}(\theta) = \begin{cases} \exp(-(\theta - x)), & \text{för } \theta > x; \\ 0, & \text{annars,} \end{cases}$$

där $x > 0$. Detta behöver <u>inte</u> visas.

Avsikten är nu att testa $H_0 : \theta \leq 5$ mot $H_1 : \theta > 5$. Detta är ett beslutsproblem med beslutsmängden $\mathcal{A} = \{0, 1\}$, där 1="avvisa H_0" och 0="avvisa ej H_0". Som kostnadsfunktion (loss function) används

$$L(a,\theta) = \begin{cases} 2, & \text{om } \theta \leq 5 \text{ och } a = 1; \\ 1, & \text{om } \theta > 5 \text{ och } a = 0; \\ 0, & \text{annars.} \end{cases}$$

(a) Ange Bayesregeln (d.v.s., Bayestestet) svarande mot ovanstående kostnadsfunktion och apriori- och aposteriorifördelning. Uttrycket för testfunktionen ska förenklas så långt som möjligt.

(b) Beräkna den "klassiska" (d.v.s., den vanliga) signifikansnivån (size) för Bayestestet.

KORTFATTADE LÖSNINGAR TILL TENTAMEN I TAMS17 STATISTISK TEORI FK, ONSDAG 24/8 2016 KL 08.00-12.00.

1. (a) Den nedre gränsen är $\frac{1}{n\mathcal{I}_{X_1}(\theta)}$, där $\mathcal{I}_{X_1}(\theta) = -E_\theta(\frac{d^2}{d\theta^2}\ln f(X_1|\theta))$ för $\theta > 0$. Eftersom $\ln f(x|\theta) = -(1+x)\ln(1+\theta) + x\ln\theta$, fås att $\frac{d^2}{d\theta^2}\ln f(x|\theta) = \frac{1+x}{(1+\theta)^2} - \frac{x}{\theta^2}$. Enligt formelsamlingen är $E_\theta(X_1) = \frac{1-\frac{1}{1+\theta}}{\frac{1}{1+\theta}} = \theta$, så $\mathcal{I}_{X_1}(\theta) = -\frac{1}{1+\theta} + \frac{1}{\theta} = \frac{1}{\theta(1+\theta)}$. Variansen är alltså nedåt begränsad av $\underline{\underline{\frac{\theta(1+\theta)}{n}}}$.

   (b) Enligt formelsamlingen är $E_\theta(\overline{X}) = E_\theta(X_1) = \theta$, så ML-skattaren är väntevärdesriktig. Vidare är $V(X_1) = \frac{1-\frac{1}{1+\theta}}{\frac{1}{(1+\theta)^2}} = \theta(1+\theta)$, så $V_\theta(\overline{X}) = \frac{\theta(1+\theta)}{n}$, vilket är lika med Cramér-Rao-gränsen. ML-skattaren har alltså lägst varians bland alla väntevärdesriktiga skattare.

2. Eftersom $\{f(x|\theta); \theta > 0\}$ är en skalfamilj, så är $\underline{\underline{Y = \frac{X}{\theta}}}$ en pivotvariabel för θ, med täthetsfunktion $f(x|1)$. Vi väljer nu $c > 0$ så att

$$0.95 = P_\theta(c \leq \frac{X}{\theta}) = \int_c^\infty x e^{-x^2/2} dx = [-e^{-x^2/2}]_{x=c}^{x=\infty} = e^{-c^2/2},$$

   d.v.s., $\underline{\underline{c = \sqrt{-2\ln 0.95} \approx 0.32}}$. Eftersom $P_\theta(c \leq \frac{X}{\theta}) = P_\theta(\theta \leq \frac{X}{c}) = 0.95$, så är $\underline{\underline{C(X) = [0, \frac{X}{c})}}$ ett konfidensintervall av konfidensgrad 0.95.

3. Enligt Neyman-Pearsons lemma är det starkaste testet $\phi(x) = I\{f_1(x) > kf_0(x)\}$, där $k \geq 0$, och där $f_0(x)$ och $f_1(x)$ är sannolikhetsfunktionerna för X under H_0 respektive under H_1, som man hittar i binomialfördelningstabellen i formelsamlingen. De nio möjliga värdena på kvoten $\frac{f_1(x)}{f_0(x)}$ (då $x = 0, 1, 2, ..., 8$) är:

   0.00 0.00 0.01 0.11 1.00 9.00 81.0 729 6561

   Om vi väljer $\underline{\underline{k = 5}}$, så får testet signifikansnivån $\underline{\underline{\alpha \approx 0.027}}$ och styrkan $\underline{\underline{\beta \approx 0.89}}$.

4. (a) $L(\mu, \sigma|x) = \Pi_{i=1}^n \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x_i-\mu)} I\{x_i \geq \mu\} = \frac{1}{\sigma^n} e^{-\frac{n}{\sigma}(\overline{x}-\mu)} I\{x_{(1)} \geq \mu\}$. En tvådimensionell tillräcklig statistika är därför $\underline{\underline{(\overline{X}, X_{(1)})}}$.

(b) Vi fixerar först $\mu \leq x_{(1)}$ och maximerar $L(\mu, \sigma | x)$ m.a.p. σ. Låt $l(\sigma) = \ln L(\mu, \sigma | x) = -n \ln \sigma - \frac{n}{\sigma}(\overline{x} - \mu)$. Villkoret $l'(\sigma) = -\frac{n}{\sigma} + \frac{n}{\sigma^2}(\overline{x} - \mu) = 0$ medför att $\sigma = \widehat{\sigma} = \overline{x} - \mu$ (detta är ett maximum, ty $l''(\widehat{\sigma}) = -\frac{n}{\widehat{\sigma}^2} < 0$). Vi maximerar sedan $L(\mu, \widehat{\sigma} | x) = \frac{1}{\widehat{\sigma}^n} e^{-n} = \frac{1}{(\overline{x} - \mu)^n} e^{-n}$ m.a.p. $\mu \leq x_{(1)}$. Denna funktion är strängt växande i μ, så maximum uppnås för $\underline{\underline{\widehat{\mu} = x_{(1)}}}$ och därmed för $\underline{\underline{\widehat{\sigma} = \overline{x} - x_{(1)}}}$.

5. (a) Vi vet att $E(X_1^2) = \sigma^2$, och enligt ledningen gäller att $V(X_1^2) = E(X_1^4) - E(X_1^2)^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4$. Centrala gränsvärdessatsen ger att

$$\frac{\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2)}{\sqrt{2}\sigma^2} \xrightarrow{d} Z \sim \mathrm{N}(0, 1)$$

då $n \to \infty$, så $b = \underline{\underline{\sqrt{2}\sigma^2}}$.

(b) Låt $\tau(\sigma^2) = e^{\sigma^2/2}$. Enligt deltametoden erhålls den önskade konvergensen om vi väljer $c = |\tau'(\sigma^2)| \sqrt{2}\sigma^2$, där $\tau'(\sigma^2) = \frac{1}{2} e^{\sigma^2/2}$. Alltså fås: $c = \underline{\underline{\frac{1}{\sqrt{2}} \sigma^2 e^{\sigma^2/2}}}$.

(c) Ett konfidensintervall av Wald-typ för $e^{\sigma^2/2}$ med asymptotisk konfidensgrad 95% är $\underline{\underline{C(x_1, \ldots, x_n) = e^{\widehat{\sigma}_n^2/2} \pm 1.96 \frac{\widehat{\sigma}_n^2 e^{\widehat{\sigma}_n^2/2}}{\sqrt{2n}}}}$.

6. (a) Bayesregeln är att avvisa H_0 om och endast om $2P(\Theta \leq 5 | X = x) \leq P(\Theta > 5 | X = x) = 1 - P(\Theta \leq 5 | X = x)$. Testfunktionen är: $\phi(x) = I\{P(\Theta \leq 5 | X = x) \leq \frac{1}{1+2} = \frac{1}{3}\}$. Enligt uppgiften gäller att $P(\Theta \leq 5 | X = x) = 0$ om $x > 5$, och

$$P(\Theta \leq 5 | X = x) = \int_x^5 \exp(-(\theta - x)) d\theta$$

$$= [-\exp(-(\theta - x))]_{\theta=x}^{\theta=5} = 1 - e^{-(5-x)} \quad \forall x \leq 5.$$

Alltså gäller: $\phi(x) = I\{e^{-(5-x)} \geq \frac{2}{3}\} = \underline{\underline{I\{x \geq 5 + \log \frac{2}{3}\}}}$.

(b) Det gäller att $P_\theta(X \geq 5 + \log \frac{2}{3}) = 0$ om $\theta < 5 + \log \frac{2}{3}$, och att

$$P_\theta(X \geq 5 + \log \frac{2}{3}) = \frac{\theta - (5 + \log \frac{2}{3})}{\theta} \qquad \forall \theta \geq 5 + \log \frac{2}{3}.$$

Bayestestets "klassiska" signifikansnivå är:

$$\alpha = \max_{0 \leq \theta \leq 5} P_\theta(X \geq 5 + \log \frac{2}{3}) = \frac{-\log \frac{2}{3}}{5} \approx \underline{\underline{0.081}}.$$