TAMS17/TEN1 STATISTISK TEORI FK
TENTAMEN ONSDAG 10/1 2018 KL 08.00-13.00.

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 $X \sim \mathrm{U}(0,\theta)$, där $\theta > 0$ är en okänd parameter. En statistiker avser att genomföra en Bayesiansk analys, där θ betraktas som en stokastisk variabel Θ med apriorifördelningen $\Gamma(2,1)$. Det observerade värdet på X är x. Bestäm aposteriorifördelningen för θ, samt Bayesskattaren (den Bayesskattare som svarar mot den vanliga kvadratiska kostnadsfunktionen) av θ.

Uppgift 2

Låt $\{X_t; t = 1, 2, \ldots\}$ vara oberoende $\mathrm{Po}(\mu)$-fördelade stokastiska variabler, där $\mu > 0$ är okänd. För varje $n = 1, 2, \ldots$, låt $X = (X_1, \ldots, X_n)^T$. ML-skattaren av μ baserad på X är $\widehat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (detta behöver inte visas).

(a) Ange ML-skattaren $\widehat{\tau}(\mu)$ av $\tau(\mu) = \mu^{-1}$ baserad på X. Härledning krävs ej.

(b) Bestäm ett <u>deterministiskt tal</u> $c(\mu) > 0$ (OBS: deterministiskt = inte stokastiskt), som eventuellt beror av μ, så att

$$\frac{\sqrt{n}(\widehat{\tau}(\mu) - \tau(\mu))}{c(\mu)}$$

konvergerar i fördelning mot $Z \sim \mathrm{N}(0,1)$ då $n \to \infty$. *Ledning:* utnyttja centrala gränsvärdessatsen och deltametoden.

(c) Ange ett konfidensintervall för $\tau(\mu)$ baserat på X med asymptotisk konfidensgrad 0.95 då $n \to \infty$. Härledning krävs ej.

Uppgift 3

Låt $X = (X_1, \ldots, X_n)^T$, där $X_1, \ldots, X_n$ är oberoende likafördelade stokastiska variabler. X_1 har täthetsfunktion i familjen $\{f_{X_1}(x|\beta); \beta > 0\}$, där

$$f_{X_1}(x|\beta) = \begin{cases} \frac{1}{\Gamma(\beta)} x^{\beta-1} e^{-x}, & \text{för } x > 0; \\ 0, & \text{annars.} \end{cases}$$

(a) Bestäm en endimensionell tillräcklig statistika för β.

(b) Visa med hjälp av lämplig sats att det existerar ett likformigt starkaste test (UMP) av $H_0 : \beta \geq 1$ mot $H_1 : \beta < 1$, samt ange dess testfunktion.

Uppgift 4

Låt $X = (X_1, \ldots, X_n)^T$, där $X_1, \ldots, X_n$ är oberoende Be(θ)-fördelade stokastiska variabler, och parametern $\theta \in (0, 1)$ är okänd. X_1 har alltså sannolikhetsfunktionen

$$f_{X_1}(x|\theta) = \begin{cases} \theta^x (1-\theta)^{1-x}, & \text{för } x \in \{0, 1\}; \\ 0, & \text{annars.} \end{cases}$$

ML-skattaren av θ baserad på X är $\widehat{\theta} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ (detta behöver inte visas). Man vill testa hypotesen $H_0 : \theta = 0.5$ mot $H_1 : \theta \neq 0.5$ med ett likelihoodkvottest på den asymptotiska (d.v.s., då $n \to \infty$) signifikansnivån 0.05.

(a) Härled testfunktionen för detta test.

(b) Antag att $n = 75$, och att det observerade stickprovsmedelvärdet är $\overline{x} = \frac{1}{75} \sum_{i=1}^{75} x_i = 0.64$. Genomför testet och redovisa resultatet.

Uppgift 5

Låt X vara en stokastisk variabel med täthetsfunktion

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1}, & \text{för } 0 < x < 1; \\ 0, & \text{annars,} \end{cases}$$

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

(a) Visa att $Y = X^\theta$ är en pivotvariabel för θ. Vilken täthetsfunktion har Y?

(b) Härled ett ensidigt uppåt begränsat konfidensintervall för θ baserat på X, med (exakt) konfidensgrad $1 - \alpha$. Utnyttja pivotvariabeln Y.

Uppgift 6

Låt $X = (X_1, \ldots, X_n)^T$, där $X_1, \ldots, X_n$ är oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler, med väntevärde $\mu \in \mathbb{R}$ och standardavvikelse $\sigma > 0$ (båda okända). X har alltså täthetsfunktion i familjen $\{f(x|\mu, \sigma); \mu \in \mathbb{R}, \sigma > 0\}$, där

$$f(x|\mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\bigl(-n(\overline{x} - \mu)^2 - \frac{(n-1)s^2}{2\sigma^2}\bigr) \qquad \forall x \in \mathbb{R}^n,$$

och där $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ och $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$ (detta behöver inte visas).

(a) Bestäm en tvådimensionell tillräcklig och fullständig statistika för (μ, σ). Svaret måste motiveras.

(b) Visa att

$$\widehat{\sigma}_n := \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \sqrt{\frac{n-1}{2}} S,$$

där $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ och $\Gamma(\cdot)$ är gammafunktionen, är en väntevärdesriktig skattare av σ. Är $\widehat{\sigma}_n$ också den bästa väntevärdesriktiga skattaren (UMVUE) av σ baserad på X? Svaret måste motiveras. *Ledning:* Utnyttja formelsamlingen, och det faktum (som inte behöver visas!) att

$$\widehat{\sigma}_n = \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \sqrt{\frac{T}{2}} \sigma,$$

där

$$T = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1).$$

KORTFATTADE LÖSNINGAR TILL TENTAMEN I TAMS17 STATISTISK TEORI FK, ONSDAG 10/1 2018 KL 08.00-13.00.

1. $f_{X,\Theta}(x,\theta) = \frac{1}{\theta}\theta\exp(-\theta) = \exp(-\theta)$ för $0 < x < \theta$, vilket ger:

$$f_X(x) = \int_x^\infty \frac{1}{\theta}\theta\exp(-\theta)d\theta = \exp(-x) \quad \forall x > 0.$$

Alltså är aposteriorifördelningens täthetsfunktion:

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

Bayesskattaren av θ är aposteriorifördelningens väntevärde:

$$E(\Theta|X=x) = \int_x^\infty \theta\exp(-(\theta-x))d\theta = \int_0^\infty (t+x)\exp(-t)dt = \underline{\underline{x+1}}.$$

2. (a) ML-skattarens invariansegenskap ger att $\underline{\underline{\widehat{\tau}(\mu) = \overline{X}^{-1}}}$.

(b) Centrala gränsvärdessatsen ger att

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sqrt{\mu}} \xrightarrow{d} Z \sim \mathrm{N}(0,1)$$

då $n \to \infty$. Enligt deltametoden fås den önskade konvergensen om $c(\mu) = |\tau'(\mu)|\sqrt{\mu}$, där $\tau'(\mu) = -\mu^{-2}$. Vi väljer således: $c(\mu) = \underline{\underline{\mu^{-3/2}}}$.

(c) Ett konfidensintervall av Wald-typ för $\tau(\mu)$ med asymptotisk konfidensgrad 0.95 är:

$$C(X_1,\ldots,X_n) = \underline{\underline{\overline{X}^{-1} \pm 1.96\frac{\overline{X}^{-3/2}}{\sqrt{n}}}}.$$

3. (a) X har täthetsfunktionen

$$f(x|\beta) = \prod_{i=1}^n f_{X_1}(x_i|\beta) = \frac{1}{\Gamma(\beta)^n}(\prod_{i=1}^n x_i)^{\beta-1}e^{-\sum_{i=1}^n x_i}I\{x \in (0,\infty)^n\},$$

så enligt faktoriseringssatsen är $T(X) = \underline{\underline{\prod_{i=1}^n X_i}}$ en tillräcklig statistika för β.

(b) Låt $0 < \beta_1 < \beta_2$. Då gäller:

$$\frac{f(x|\beta_2)}{f(x|\beta_1)} = \frac{\Gamma(\beta_1)^n}{\Gamma(\beta_2)^n}\Big(\prod_{i=1}^n x_i\Big)^{\beta_2-\beta_1},$$

vilket är en strängt växande funktion i $T(x)$, så familjen har monoton likelihoodkvot (MLR). Det likformigt starkaste testet av $H_0 : \beta \geq 1$ mot $H_1 : \beta < 1$ ges därför av: $\underline{\underline{\phi(X) = I\{\prod_{i=1}^n X_i < k\}}}$, där k väljs så att testet får rätt signifikansnivå.

4. (a) Likelihoodkvotteststatistikan är:

$$-2\ln\lambda(x) = -2\ln\Big(\frac{f_X(x|0.5)}{f_X(x|\overline{x})}\Big) = -2\ln\Big(\frac{0.5^n}{\overline{x}^{n\overline{x}}(1-\overline{x})^{n(1-\overline{x})}}\Big)$$
$$= -2n(\ln 0.5 - \overline{x}\ln\overline{x} - (1-\overline{x})\ln(1-\overline{x})).$$

Testet $\phi(X) = \underline{\underline{I\{-2\ln\lambda(X) > \chi^2_{0.05}(1)\}}}$ (där $\chi^2_{0.05}(1) \approx 3.84$) har asymptotisk signifikansnivå 0.05 då $n \to \infty$.

(b) Med data givna i uppgiften fås:

$$-2\ln\lambda(x) = -150\big(\ln 0.5 - 0.64\ln 0.64 - (1-0.64)\ln(1-0.64)\big) \approx 5.96.$$

Slutsatsen av testet blir att $\underline{\underline{H_0 \text{ avvisas}}}$.

5. (a) Y är en funktion av X och θ, och vi ser direkt att $P(0 < Y < 1) = 1$. För $0 < y < 1$ fås:

$$F_Y(y) = P(Y \leq y) = P(X^\theta \leq y) = P(X \leq y^{1/\theta})$$
$$= \int_0^{y^{1/\theta}} f(x|\theta)dx = \big[x^\theta\big]_{x=0}^{x=y^{1/\theta}} = y.$$

Alltså är $\underline{\underline{Y \sim \mathrm{U}(0,1)}}$, vilket inte beror av θ.

(b) Välj $a > 0$ så att $P(Y \geq a) = 1 - \alpha$. Detta åstadkoms då $a = \alpha$, eftersom $Y \sim \mathrm{U}(0,1)$. En omskrivning ger sedan:

$$1 - \alpha = P(Y \geq \alpha) = P(X^\theta \geq \alpha) = P(\theta\ln X \geq \ln\alpha) = P(\theta \leq \frac{\ln\alpha}{\ln X}),$$

så det sökta konfidensintervallet är:

$$C(X) = \underline{\underline{\Big[0, \frac{\ln\alpha}{\ln X}\Big]}}.$$

6. (a) Utvecklas kvadraten $(\overline{x}-\mu)^2$ i det givna uttrycket, så ser man att $\{f(x|\mu,\sigma); \mu \in \mathbb{R}, \sigma > 0\}$ är en exponentialfamilj, med naturlig tillräcklig statistika $(\overline{X}, S^2)$ och viktsfunktioner $w_1(\mu,\sigma) = 2n\mu$ och $w_2(\mu,\sigma) = -\frac{(n-1)}{2\sigma^2}$. Eftersom $\{(w_1(\mu,\sigma), w_2(\mu,\sigma); \mu \in \mathbb{R}, \sigma > 0\} = \{(u,v); u \in \mathbb{R}, v < 0\}$ är en öppen delmängd av $\mathbb{R}^2$, så är $(\overline{X}, S^2)$ även $\underline{\underline{\text{fullständig}}}$.

(b) Det gäller att $E(\widehat{\sigma}_n) = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})}\sigma E(\sqrt{T})$, där

$$E(\sqrt{T}) = \int_0^\infty t^{1/2} f_T(t)dt = \int_0^\infty \frac{t^{n/2-1}e^{-t/2}}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})}dt$$

$$= \frac{2^{n/2}\Gamma(\frac{n}{2})}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})}\int_0^\infty \frac{t^{n/2-1}e^{-t/2}}{2^{n/2}\Gamma(\frac{n}{2})}dt = \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$

så $E(\widehat{\sigma}_n) = \underline{\underline{\sigma}}$. Eftersom $\widehat{\sigma}_n$ är en funktion av den fullständiga och tillräckliga statistikan $(\overline{X}, S^2)$, så är den $\underline{\underline{\text{UMVUE}}}$ av σ.