Tentamen Kres 4, 270308, 09:00–12:00

Antwoorden motiveren; kritieke gebieden en betrouwbaarheidsintervallen zo uitvoerig mogelijk beschrijven. Ieder onderdeel is 10 punten waard. U dient uit te gaan van een steekproef van omvang n uit de betreffende pdf. Bij het tentamen mag alleen gebruik worden gemaakt van een (zelf mee te nemen) onbeschreven kopie van Appendix B van Bain en Engelhardt!

1.

$$f_X(x;\theta) = \frac{1}{\sqrt{2\pi\theta x}} e^{-\frac{x}{2\theta}} I_{[0,\infty)}(x), \qquad \theta > 0.$$

(a) Geef een statistische grootheid die 'complete & sufficient' is voor θ. Geef tevens de UMVUE van θ.

(b) Bepaal de CRLB voor zuivere schatters van θ.

(c) Geef de uniform meest onderscheidende (UMP) toets, met onbetrouwbaarheidsdrempel (grootte) α, voor H_0: $\theta \leq 2$ tegen H_a: $\theta > 2$.

(d) Toon aan dat $\frac{\sum_{i=1}^n X_i}{\theta}$ een spilfunctie is en geef een 90% betrouwbaarheidsondergrens voor θ gebaseerd op deze grootheid.

(e) Bepaal de UMVUE van θ^{-1}.

2.

$$f_X(x;\theta) = 2e^{2(\theta - x)} I_{[\theta,\infty)}(x), \qquad \theta > 1.$$

(a) Toon aan dat $S = X_{1:n}$ 'complete & sufficient' is voor θ.

(b) Bepaal de UMVUE van θ.

(c) Geef een "gelijke staarten" $100\gamma\%$ betrouwbaarheidsinterval voor θ, gebaseerd op S.

(d) Een onderzoeker verwerpt H_0: $\theta \leq 1$ tegen H_a: $\theta > 1$ als $S \geq 1 + \frac{1}{n}$. Bereken *en schets* de powerfunctie $\pi(\theta)$. Hoe gedraagt $\pi(\theta)$ zich als $\theta \to 0$? Geef een uitdrukking voor de size α van de toets (numerieke waarde niet noodzakelijk).

3. Aan 25 aselect gekozen jongens en 25 aselect gekozen meisjes werd gevraagd of zij mobiele telefoons van merk A of B beter vinden. De waargenomen aantallen voorkeuren zijn

merk	A	B
jongens	14	11
meisjes	6	19

Toets, met een onbetrouwbaarheidsdrempel $\alpha = 0.05$, de nulhypothese dat de voorkeur onder deze twee merken niet van het geslacht afhangt.

Succes!!!

Uitwerkingen

1. (a) Het is een dichtheid uit de REC. De pdf is te schrijven als

$$f_X(x;\theta) = g(\theta)s(x)e^{h(\theta)t(x)}$$

met $g(\theta) = \frac{1}{\sqrt{2\pi\theta}}$, $s(x) = 1/\sqrt{x}$, $h(\theta) = -\frac{1}{2\theta}$ en $t(x) = x$. Uit deze vorm valt af te lezen dat $S \equiv \sum_{i=1}^{n} t(X_i) = \sum_{i=1}^{n} X_i$ C&S voor θ is. Omdat $X \sim \text{GAM}(2\theta, \frac{1}{2})$ is $E(X_i) = 2\theta/2 = \theta$, dus is $\frac{1}{n}\sum X_i$ UMVUE voor θ.

(b) We hebben

$$\frac{\partial}{\partial\theta}\log f(X;\theta) = -\frac{1}{2\theta} + \frac{X}{2\theta^2}$$

en

$$-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) = -\frac{1}{2\theta^2} + \frac{X}{\theta^3}.$$

Nu kunnen we op twee manieren de verwachting in de noemer van de CRLB bepalen:

$$E\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2 = \text{Var}\frac{X}{2\theta^2} = \frac{1}{4\theta^4}\text{Var}(X) = \frac{1}{2\theta^2}$$

waarbij gebruikt is dat $\text{Var}X = (2\theta)^2\frac{1}{2}$, of:

$$-E\left(\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right) = -\frac{1}{2\theta^2} + \frac{E(X)}{\theta^3} = \frac{1}{2\theta^2}.$$

Dit geeft

$$\text{CRLB} = \frac{\tau'(\theta)}{n\frac{1}{2\theta^2}} = \frac{2\theta^2}{n},$$

voor $\tau(\theta) = 1$.

(c) We tonen aan dat de familie van pdf's een MLR eigenschap heeft. Daartoe kijken we voor $\theta_2 > \theta_1 > 0$ naar de ratio

$$\frac{f(x_1,\ldots x_n;\theta_2)}{f(x_1,\ldots x_n;\theta_1)} = \left(\frac{\theta_1}{\theta_2}\right)^{\frac{n}{2}} e^{\frac{1}{2}(\frac{1}{\theta_1}-\frac{1}{\theta_2})\sum_{i=1}^{n} x_i}.$$

Omdat $\frac{1}{\theta_1} - \frac{1}{\theta_2} > 0$ heeft de familie van pdf's de MLR eigenschap in $S = \sum_{i=1}^{n} X_i$. De betreffende hypothese dient verworpen te worden voor te grote waarden van S. Omdat $S \sim \text{GAM}(2\theta, n/2)$ geldt dat $2\frac{S}{2\theta} \sim \chi^2(n)$. De toets eruit H_0 te verwerpen als

$$2\frac{S}{2\theta_0} = \frac{S}{\theta_0} \geq \chi^2_{1-\alpha}(n),$$

$\theta_0 = 2$, dus verwerpen als $S \geq 2\chi^2_{1-\alpha}(n)$.

(d) Laat $Y = X/\theta$, dan is $S \equiv \sum_{i=1}^n X_i$ te schrijven als $S = \theta \sum_{i=1}^n Y_i$. De pdf van Y is te vinden via de transformatieregels van pdf's. De inverse relatie tussen de uitkomsten voor X en Y is $x = \theta y$, dus is de Jacobiaan $|J| = \theta$, en

$$f_Y(y) = \theta \frac{1}{\sqrt{2\pi\theta^2 y}} e^{-y/2} = \frac{1}{\sqrt{2\pi y}} e^{-y/2},$$

voor $y > 0$. We vinden $Y \sim \chi^2(1)$, en dus $\sum_{i=1}^n Y_i \sim \chi^2(n)$. Andere manier: De algemene regel zegt dat we van een $V \sim \mathrm{GAM}(a, b)$ naar een χ^2-verdeling komen via $2V/a \sim \chi^2(2b)$. In dit geval geeft dat $2\frac{S}{2\theta} \sim \chi^2(2\frac{n}{2})$, dus is $S/\theta \sim \chi^2(2\frac{n}{2})$. We hebben dus

$$\begin{aligned} P\left(\tfrac{S}{\theta} < \chi^2_{0.90}(n)\right) &= 90\% \\ \Leftrightarrow P\left(\tfrac{S}{\chi^2_{0.90}(n)} < \theta\right) &= 90\% \end{aligned}$$

dus is de gevraagde betrouwbaarheidsondergrens voor θ gegeven door $\frac{S}{\chi^2_{0.90}(n)}$.

(e) We bepalen eerst de verwachting van S^{-1}. Dat kan direct met de $\mathrm{GAM}(2\theta, \frac{n}{2})$-verdeling van S, of met behulp van $S = \theta W$, met $W = \sum_{i=1}^n Y_i \sim \chi^2(n)$ (zie onderdeel 1c). De directe methode geeft:

$$\begin{aligned} E[S^{-1}] &= \textstyle\int_0^\infty \frac{1}{(2\theta)^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-2} e^{-\frac{x}{2\theta}}\, \mathrm{d}x \\ &= \tfrac{1}{2\theta}\tfrac{\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n}{2})} \textstyle\int_0^\infty \frac{1}{(2\theta)^{\frac{n-2}{2}}\Gamma(\frac{n-2}{2})} x^{\frac{n-2}{2}-1} e^{-\frac{x}{2\theta}}\, \mathrm{d}x \\ &= \tfrac{1}{2\theta}\tfrac{\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n}{2})} = \tfrac{1}{2\theta(\frac{n}{2}-1)} = \tfrac{1}{\theta(n-2)}. \end{aligned}$$

Hieruit volgt dat $\frac{n-2}{S}$ UMVUE voor θ^{-1} is.

2. (a) De joint pdf is

$$f(x_1, \ldots, x_n; \theta) = 2^n e^{2n\theta} e^{-2\sum_{i=1}^n x_i} I_{[\theta,\infty)}(x_{1:n}).$$

Uit de factorisatiestelling volgt dan dat $S = X_{1:n}$ sufficient is voor θ. We dienen de 'completeness' nog aan te tonen. Daarvoor hebben we de pdf van $X_{1:n}$ nodig. De CDF van X is $F_X(x) = 1 - e^{2(\theta - x)}$ waarmee die van S (KReS 3) wordt $F_S(s) = 1 - e^{2n(\theta - s)}$, met als afgeleide $f_S(s) = 2ne^{2n(\theta - s)}$, voor $s \geq \eta$. De verwachting van een functie van S gelijkstellen aan nul geeft

$$\begin{aligned} E[u(S)] = 2n \textstyle\int_\theta^\infty u(s) e^{2n(\theta - s)}\, \mathrm{d}s &= 0 \\ \Leftrightarrow \textstyle\int_\theta^\infty u(s) e^{-2ns}\, \mathrm{d}s &= 0 \\ \Rightarrow -u(\theta) e^{-2n\theta} &= 0 \quad \forall \theta > 0. \end{aligned}$$

De laatste stap volgt uit differentiëren van de integraal naar η (Leibniz). Omdat de e-macht positief is volgt dus $u(\theta) = 0$ voor $\theta > 0$, waarmee completeness is aangetoond.

(b) $F_X(x) = 1 - e^{2(\theta - x)}$, $F_S(s) = 1 - e^{2n(\theta - s)}$, dus $f_s(s) = 2ne^{2n(\theta - s)}$.

$$\begin{aligned} E[S] &= 2ne^{2n\theta} \int_\theta^\infty se^{-2ns}\, \mathrm{d}s \\ &= e^{2n\theta} \int_\theta^\infty s\left(-e^{-2ns}\right)'\, \mathrm{d}s \\ &= \left[-e^{2n\theta} se^{-2ns}\right]_{s=\theta}^\infty + e^{-2n\theta} \int_\theta^\infty e^{-2ns}\, \mathrm{d}s \\ &= \theta - \tfrac{1}{2n}\left[e^{-2n\theta} e^{-2ns}\right]_{s=\theta}^\infty \\ &= \theta + \tfrac{1}{2n}. \end{aligned}$$

Dus $S - \frac{1}{2n}$ is zuiver voor θ en een functie van de C&S, dus is het een UMVUE voor θ.

(c) Los op, naar θ

$$F_S(s) = 1 - e^{2n(\theta - s)} = \frac{1 \pm \gamma}{2}.$$

De oplossingen zijn

$$\theta_{1,2} = s + \frac{1}{2n} \ln \frac{1 \pm \gamma}{2},$$

dus wordt het gevraagde $100\gamma\%$ BI voor θ gegeven door

$$\left(s + \frac{1}{2n} \ln \frac{1 - \gamma}{2}, s + \frac{1}{2n} \ln \frac{1 + \gamma}{2}\right).$$

(d) De powerfunctie is

$$\begin{aligned} \pi(\theta) &= P(S > 1 + \tfrac{1}{n} | \theta) \\ &= 1 - P(S \leq 1 + \tfrac{1}{n} | \theta) \\ &= \exp\left(2n(\theta - 1 - \tfrac{1}{n})\right) \end{aligned}$$

voor $\theta \leq 1 + \frac{1}{n}$, voor $\theta > 1 + \frac{1}{n}$ is $\pi(\theta) = 1$. De powerfunctie is niet-dalend in θ, dus is de size α gegeven door $\max_{0<\theta\leq 1} \pi(\theta) = \pi(1) = e^{-2} \approx 0.1353$ (numeriek antwoord niet noodzakelijk). De powerfunctie gaat voor $\theta \to 0$ niet naar 0 maar naar de waarde e^{-2n-2}.

3. De steekproeffractie van de mensen die voorkeur geven aan A is 40%. We verwachten onder de nulhypothese dus in beide groepen $0.4 \times 25 = 10$ mensen met voorkeur voor A, en 15 voor B. De χ^2 toetsgrooetheid is

$$2\frac{4^2}{10} + 2\frac{4^2}{15} = 5.33$$

Omdat de steekproefgrootten van de jongens resp. meisjes vastliggen is er sprake van $(c-1)(r-1) = 1$ vrijheidsgraad. De kritieke waarde bij $\alpha = 0.05$ en twee vrijheidsgraden is 3.841, dus H_0 wordt verworpen.