Uitwerkingen Tentamen Kres 4, 290307

1. (a) De pdf is te schrijven als

$$f_X(x;\theta) = \theta e^{-(\theta+1)\log x} I_{(1,\infty)}(x) = c(\theta)h(x)e^{q(\theta)t(x)},$$

met $c(\theta) = \theta$, $h(x) = I_{(1,\infty)}(x)$, $q(\theta) = -1-\theta$ en $t(x) = \log x$, dus het betreft hier een pdf uit de REC familie. De C&S statistic is $S := \sum_{i=1}^n t(X_i) = \sum_{i=1}^n \log X_i$. We bepalen eerst de verdeling van $Y := \log X$. $X = e^Y$, $J = e^y$, $f_Y(y;\theta) = e^y \theta e^{-(\theta+1)y} = \theta e^{-\theta y}$ voor $y > 0$. Omdat $Y \sim \text{EXP}(\theta^{-1})$ geldt nu $S = \sum_{i=1}^n Y_i \sim \text{GAM}(\theta^{-1}, n)$

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

In het bijzonder is dus $E[S^{-1}] = \frac{\theta}{n-1}$, en daarom is $\frac{n-1}{S}$ UMVUE voor θ.

(b) De CRLB voor zuivere schatters voor $\tau(\theta) = \theta$ is

$$\frac{(\tau'(\theta))^2}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X;\theta)\right]}$$

met $\tau'(\theta) = \theta' = 1$. We berekenen eerst de verwachting in de noemer

$$-E\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X;\theta)\right] = -E\left[\frac{\partial}{\partial\theta}\left(\frac{1}{\theta} - \log X\right)\right] = \frac{1}{\theta^2}.$$

We vinden dan voor de CRLB voor zuivere schatters de volgende uitdrukking:

$$\frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X;\theta)\right]} = \frac{\theta^2}{n}.$$

Vervolgens de variantie van de UMVUE berekenen: $\text{Var}\left(\frac{n-1}{S}\right) = (n-1)^2\text{Var}(S^{-1})$.

$$E[(S^{-1})^2] = E[S^{-2}] = \int_0^\infty \frac{\theta^n}{\Gamma(n)} s^{n-3} e^{-\theta s}\,\mathrm{d}s = \frac{\theta^2}{(n-1)(n-2)}.$$

$$\begin{aligned}
\text{Var}\left(\frac{n-1}{S}\right) &= (n-1)^2\left(\frac{\theta^2}{(n-1)(n-2)} - \left(\frac{\theta^2}{(n-2)^2}\right)^2\right) \\
&= \left(\frac{n-1}{n-2} - 1\right)\theta^2 = \frac{\theta^2}{n-2} > \frac{\theta^2}{n} = \text{CRLB},
\end{aligned}$$

dus de UMVUE voor θ bereikt de CRLB niet. Andere methode:

$$\left[\frac{\partial}{\partial\theta}\log f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta)\right] = \frac{n}{\theta} - \sum_{i=1}^{n}\log x_i = \frac{1}{n}\left(\frac{1}{\theta} - \bar{Y}\right).$$

Volgens Opm. 9.3.II kan een UMVUE de CRLB dus slechts bereiken voor $\tau(\theta) = \theta^{-1}$, of een lineaire combinatie daarvan. De UMVUE voor θ^{-1} is af te lezen uit bovenstaande uitdrukking, en is $\bar{Y} = \sum_{i=1}^{n}\log X_i$.

(c) We weten: $S \sim \mathrm{GAM}(\theta^{-1}, n)$, dus geldt $2\theta S \sim \chi^2(2n)$, en daarmee:

$$\begin{aligned} P\left[\chi^2_{\frac{1-\gamma}{2}}(2n) \le 2\theta S \le \chi^2_{\frac{1+\gamma}{2}}\right] &= \gamma \\ \Leftrightarrow P\left[\frac{\chi^2_{\frac{1-\gamma}{2}}(2n)}{2S} \le \theta \le \frac{\chi^2_{\frac{1+\gamma}{2}}}{2S}\right] &= \gamma. \end{aligned}$$

Het gevraagde BI is dus $\left(\frac{\chi^2_{\frac{1-\gamma}{2}}(2n)}{2S}, \frac{\chi^2_{\frac{1+\gamma}{2}}}{2S}\right)$

(d) $F_X(x;\theta) = \int_1^x \theta s^{-\theta-1}\,\mathrm{d}s = 1 - x^{-\theta}$ voor $x > 1$.

$$\begin{aligned} \pi(\theta) &= P\left[\log X_{1:n} \ge \frac{1}{n} \mid \theta\right] \\ &= P\left[X_{1:n} \ge e^{\frac{1}{n}} \mid \theta\right] \\ &= \left(1 - F_X(e^{\frac{1}{n}};\theta)\right)^n \\ &= \left(e^{\frac{1}{n}}\right)^{-\theta n} \\ &= e^{-\theta} \end{aligned}$$

Grootte: $\alpha = \max_{\theta\ge 3}\pi(\theta) = \pi(3) = e^{-3} \simeq 0.050$. Onderscheidingsvermogen (power) is $\pi(\theta) = e^{-\theta}$ voor $\theta < 3$. Merk op dat de power van deze toets dus niet toeneemt met n. Deze toets is dan ook niet meest onderscheidend.

(e) We beschouwen twee parameterwaarden $\theta_2 > \theta_1 > 0$.

$$\frac{f(\boldsymbol{x};\theta_2)}{f(\boldsymbol{x};\theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^n e^{-(\theta_2-\theta_1)\sum_{i=1}^{n}\log x_i}.$$

Deze familie van pdf's bezit dus een MLR in $t(x_1,\ldots,x_n) = -\sum_{i=1}^{n}\log X_i = -S$. De UMP toets voor H_0: $\theta \le 1$ tegen H_a: $\theta > \theta_1$ bestaat eruit te verwerpen als S te klein is. Voor $\theta = 1$ geldt $2S \sim \chi^2(2n)$, dus verwerp als $2S \le \chi^2(2n)_\alpha$.

(a) De joint pdf is te schrijven als

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \left(\frac{5}{2}\right)^n \theta^{\frac{5}{2}n}\left(\prod x_i\right)^{-\frac{7}{2}} I_{[\theta,\infty)}(x_{1:n}) = g(x_{1:n};\theta)h(x_1,\ldots,x_n)$$

met $g(x_{1:n};\theta) = I_{[\theta,\infty)}(x_{1:n})$ en $h(x_1,\ldots,x_n) = \left(\frac{5}{2}\right)^n \theta^{\frac{5}{2}n} \left(\prod x_i\right)^{-\frac{7}{2}}$. Volgens het factorisatiecriterium is daarmee aangetoond dat $S := X_{1:n}$ sufficient is. Is S ook complete? We moeten daarvoor kijken of $E[u(S)] = 0$ $u(s) = 0$ impliceert voor (bijna) alle s. Eerst de verdeling van S bepalen met de CDF methode. $F_X(x) = \int_\theta^x f_X(s)\mathrm{d}s = 1 - \left(\frac{\theta}{x}\right)^{\frac{5}{2}}$. $F_S(s) = P[S \le s] = P[X_{1:n} \le s] = 1 - (1 - F_X(s))^n = 1 - \left(\frac{\theta}{s}\right)^{\frac{5}{2}n}$. Differentiëren geeft $f_S(s) = \frac{5}{2}n\theta^{\frac{5}{2}n}s^{-\frac{5}{2}n-1}I_{[\theta,\infty)}(s)$.

$$\begin{aligned} E[u(S)] &= \tfrac{5}{2}n\theta^{\frac{5}{2}n}\textstyle\int_\theta^\infty u(s)s^{-\frac{5}{2}n-1}\,\mathrm{d}s = 0 \\ \Leftrightarrow \textstyle\int_\theta^\infty u(s)s^{-\frac{5}{2}n-1}\,\mathrm{d}s = 0 & \\ \Rightarrow -u(\theta)\theta^{-\frac{5}{2}n-1} = 0 & \\ \Leftrightarrow u(\theta) = 0 & \end{aligned}$$

waar de op één na laatste implicatie volgt uit differentiëren naar θ, waarbij de regel van Leibniz gebruikt is. Omdat deze implicatie geldt voor alle $\theta > 0$ is hiermee aangetoond dat $E[u(S) = 0] \to u(s) = 0, \forall s > 0$, dus is $X_{1:n}$ tevens complete.

(b) Zoals berekend onder 2a. is de pdf van S gegeven door

$$f_S(s) = \frac{5}{2}n\theta^{\frac{5}{2}n}s^{-\frac{5}{2}n-1}I_{[\theta,\infty)}(s),$$

zodat

$$\begin{aligned} E[S^k] &= \frac{5}{2}n\theta^{\frac{5}{2}n}\int_\theta^\infty s^{-\frac{5}{2}n+k-1}\mathrm{d}s \\ E[S^k] &= \frac{\frac{5}{2}n\theta^{\frac{5}{2}n}}{-\frac{5}{2}n+k}\left[s^{-\frac{5}{2}n+k}\right]_{s=\theta}^\infty \\ &= \frac{5n}{5n-2k}\theta^k. \end{aligned}$$

Hieruit volgt (met $k = 1$) dat $E[S] = \frac{5n}{5n-2}\theta$, en de UMVUE voor θ is dus $\frac{5n-2}{5n}S$.

(c) MME:

$$\begin{aligned} E[X] &= \frac{5}{2}\theta^{\frac{5}{2}}\int_\theta^\infty x^{-\frac{5}{2}}\,\mathrm{d}x \\ &= \frac{5}{2}\theta^{\frac{5}{2}}\left[-\frac{2}{3}x^{-\frac{3}{2}}\right]_{x=\theta}^\infty \\ &= \frac{5}{3}\theta. \end{aligned}$$

De MME is de oplossing van de vergelijking $\frac{5}{3}\theta = \bar{X}$ naar θ, te weten $\tilde{\theta} = \frac{3}{5}\bar{X}$. De MME is zuiver omdat $E[\tilde{\theta}] = \frac{3}{5}E[\bar{X}] = \frac{3}{5}E[X] = \theta$. De variantie is $\mathrm{Var}(\tilde{\theta}) = \frac{9}{25}\mathrm{Var}(\bar{X}) = \frac{9}{25n}\mathrm{Var}(X)$. $\mathrm{Var}(X)$ bepalen: $E[X^2] = \frac{5}{2}\theta^{\frac{5}{2}}\int_\theta^\infty x^{-\frac{3}{2}}\,\mathrm{d}x = \ldots = 5\theta^2$. $\mathrm{Var}(X) = 5\theta^2 - \frac{25}{9}\theta^2 = \frac{20}{9}\theta^2$, dus $\mathrm{MSE}(\tilde{\theta}) = \mathrm{Var}(\theta) = \frac{9}{25n} \times \frac{20}{9}\theta^2 = \frac{4}{5n}\theta^2 \underset{n\to\infty}{\longrightarrow} 0$ waarmee is aangetoond dat $\tilde{\theta}$ MSE consistent is (en dus ook simple consistent).

MLE: De likelihoodfunctie is $L(\theta) = f(x_1, \ldots, x_n; \theta) = \left(\frac{5}{2}\right)^n \theta^{\frac{5}{2}n} \left(\prod x_i\right)^{-\frac{7}{2}} I_{[\theta,\infty)}(x_{1:n})$. Deze is positief en dalend in θ zolang $x_{1:n} \geq \theta$, en anders nul. De functie heeft dus een randmaximum voor $\widehat{\theta} = x_{1:n}$.

Er geldt $E[S] = \frac{5n}{5n-2}\theta$. De MSE is als volgt op te splitsen: $\mathrm{MSE}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}) + (b(\widehat{\theta}))^2$ waarbij $b(\widehat{\theta}) = E[\widehat{\theta}] - \theta$. In dit geval vinden we $b(\widehat{\theta}) = \frac{-2}{5n-2}\theta \underset{n\to\infty}{\longrightarrow} 0$ (de MLE is dus niet zuiver, maar wel asymptotisch zuiver) en $\mathrm{Var}(\widehat{\theta}) = \left(\frac{5n}{5n-4} - \left(\frac{5n}{5n-2}\right)^2\right) \underset{n\to\infty}{\longrightarrow} 0$ omdat beide termen tussen de haken naar 1 convergeren. De MSE van $\widehat{\theta}$ gaat dus naar nul als n toeneemt, waarmee is aangetoond dat $\widehat{\theta}$ MSE consistent is.

(d) De CDF van S is $F_s(s) = P[S \leq s] = 1 - \left(\frac{\theta}{s}\right)^{\frac{5n}{2}}$. De oplossing van $F_S(S \leq s) = \gamma$ is $s = \theta(1-\gamma)^{-\frac{2}{5n}}$, zodat

$$P\left[S < \theta(1-\gamma)^{-\frac{2}{5n}}\right] = \gamma,$$

oftewel

$$P\left[S(1-\gamma)^{\frac{2}{5n}} < \theta\right] = \gamma,$$

waaruit valt af te lezen dat $S(1-\gamma)^{\frac{2}{5n}}$ een $100\gamma\%$ betrouwbaarheidsondergrens is voor θ.

(e) Definieer $Y_i = X_i/\theta$. Bepaal de pdf van Y: $x = \theta y$, $J = \theta$, $f_Y(y) = \theta\frac{5}{2}\theta^{\frac{5}{2}}\theta y^{-\frac{7}{2}} I_{[\theta,\infty)}(\theta y) = \frac{5}{2}y^{-\frac{7}{2}} I_{[1,\infty)}(y)$.De verdeling van Y hangt dus niet af van welke onbekende parameter dan ook. Omdat de verdeling van $\frac{X_1+X_2}{X_3}$ uitgedrukt kan worden in termen van Y_i, via

$$\frac{X_1+X_2}{X_3} = \frac{Y_1+Y_2}{Y_3},$$

geldt dat ook de verdeling van $\frac{X_1+X_2}{X_3}$ niet van onbekende parameters afhangt. De stelling van Basu garandeert nu dat deze grootheid en de C&S statistic $X_{1:n}$ stochastisch onafhankelijk zijn.