Tutorial Letter 202/2/2016

Statistical Inference III

STA3702

Semester 2

Department of Statistics

SOLUTIONS TO ASSIGNMENT 02

BAR CODE

Question 1 **[20]**

(a)

$$\begin{aligned} f(x|\theta) &= g(x)\exp\{p(\theta)K(x) - q(\theta)\} \\ &= \exp\{p(\theta)K(x) + -q(\theta) + \ln(g(x))\} \\ &= \exp\left\{p(\theta)K(x) + q^*(\theta) + g^*(x)\right\} \end{aligned}$$

where $q^*(\theta) = -q(\theta)$ and $g^*(\theta) = \ln(g(x))$. **(4)**

(b) $\sum_{i=1}^n K(x_i)$ is the complete sufficient statistic for θ. **(4)**

(c) The likelihood function of θ is

$$L(\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n g(x_i)\exp\left\{p(\theta)\sum_{i=1}^n K(x_i) - nq(\theta)\right\}$$

and the log-likelihood function of θ is

$$l(\theta) = \sum_{i=1}^n \ln g(x_i) + p(\theta)\sum_{i=1}^n K(x_i) - nq(\theta).$$

The MLE of θ is $\hat{\theta}$ which solves the equation

$$0 = l'(\hat{\theta}) = p'(\hat{\theta})\sum_{i=1}^n K(X_i) - nq'(\hat{\theta}).$$

Clearly the solution will only depend on the sample through $\sum_{i=1}^n K(X_i)$. **(6)**

(d) $\frac{1}{n}\sum_{i=1}^n K(X_i)$ is the $MVUE$ of $q'(\theta)/p'(\theta)$ because $\frac{1}{n}\sum_{i=1}^n K(X_i)$ is a function of a complete sufficient statistic for θ and $E\left(\frac{1}{n}\sum_{i=1}^n K(X_i)\right) = \frac{1}{n}\sum_{i=1}^n E(K(X_i)) = q'(\theta)/p'(\theta)$ (unbiased).

(6)

Question 2 **[36]**

(a)

$$\begin{aligned} p(x|\theta) &= \theta(1-\theta)^{x-1} \\ &= \left(\frac{\theta}{1-\theta}\right)(1-\theta)^x \\ &= \exp\left(x\ln(1-\theta) + \ln\frac{\theta}{1-\theta}\right) \\ &= g(x)\exp\left(p(\theta)K(x) - q(\theta)\right) \end{aligned}$$

where $p(\theta) = \ln(1-\theta)$, $K(x) = x$, $q(\theta) = -\ln\frac{\theta}{1-\theta}$ and $g(x) = 1$. **(5)**

(b) $\sum_{i=1}^n K(X_i) = \sum_{i=1}^n X_i$ since $f(x|\theta)$ belongs to regular exponential family. **(4)**

(c) $E\left(\sum_{i=1}^{n} K(X_i)\right) = \sum_{i=1}^{n} E\left(K(X_i)\right) = n\dfrac{q'(\theta)}{p'(\theta)} = n\dfrac{-1/\theta - 1/(1-\theta)}{-1/(1-\theta)} = -n\dfrac{1/(\theta(1-\theta))}{-1/(1-\theta)} = \dfrac{n}{\theta}$. **(2)**

(d) $p(x|\theta)$ here is in the form of $f(x|\theta)$ in Question 1. It follows from Question 1 (c) that the MLE of θ here will depend on the sample through the complete sufficient statistic for θ. In particular, it can be shown that the MLE of θ is $\hat{\theta}$ which solves the equation

$$0 = l'(\hat{\theta}) = p'(\hat{\theta})\sum_{i=1}^{n} K(X_i) - nq'(\hat{\theta}) = -\frac{1}{1-\hat{\theta}}\sum_{i=1}^{n} X_i + \frac{n}{\hat{\theta}(1-\hat{\theta})}$$

and the solution is

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

(6)

(e) Yes. Because: (i) the MLE of $\dfrac{1}{\theta}$ is $\dfrac{1}{\hat{\theta}} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ by the invariance property of $MLE's$;

(ii) the MLE of $\dfrac{1}{\theta}$ is $\dfrac{1}{n}\sum_{i=1}^{n} X_i$ is a function of a complete sufficient statistic for θ; and

(iii) $E\left(\dfrac{1}{\hat{\theta}}\right) = \dfrac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \dfrac{1}{n}\dfrac{n}{\theta} = \dfrac{1}{\theta}$ (unbiased, see Question 1 (c)). **(7)**

(f) $l''(\theta) = -\frac{1}{(1-\theta)^2}\sum_{i=1}^{n} X_i - \frac{n}{\theta^2} + \frac{n}{(1-\theta)^2} = -\frac{n}{\hat{\theta}(1-\theta)^2} - \frac{n}{\theta^2} + \frac{n}{(1-\theta)^2} \Longrightarrow$

(i) $\hat{\theta} = \frac{200}{400} = 0.5$. Hence

$$I(\mathbf{x}) = -l''(\hat{\theta}) = \frac{n}{\hat{\theta}(1-\hat{\theta})^2} + \frac{n}{\hat{\theta}^2} - \frac{n}{(1-\hat{\theta})^2} = 1600;$$

and **(6)**

(ii) $se(\hat{\theta}) \approx \sqrt{1/I(\mathbf{x})} = \sqrt{1/1600} = \frac{1}{40} = 0.025$. Thus, the approximate 95% confidence interval for θ is

$$\hat{\theta} \pm 1.96 \times se(\hat{\theta}) = [0.451; 0.549].$$

(6)

Question 3 **[23]**

(a)

$$\begin{aligned} f(x|\theta) &= \theta(1+x)^{-(1+\theta)} \\ &= (1+x)^{-1}\theta(1+x)^{-\theta} \\ &= (1+x)^{-1}\exp\left[-\theta\ln(1+x) + \ln\theta\right] \\ &= g(x)\exp\left[p(\theta)K(x) - q(\theta)\right] \end{aligned}$$

where $p(\theta) = \theta$, $K(x) = -\ln(1+x)$, $q(\theta) = -\ln(\theta)$ and $g(x) = (1+x)^{-1}$. **(4)**

(b) $\sum_{i=1}^{n} K(X_i) = -\sum_{i=1}^{n}\ln(1+X_i)$ since $f(x|\theta)$ belongs to regular exponential family. **(3)**

(c) Note that $-E\left[\sum_{i=1}^n \ln(1+X_i)\right] = -n\frac{q'(\theta)}{p'(\theta)} = -\frac{n}{\theta}$.
The likelihood function of θ is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta) = \theta^n \prod_{i=1}^n (1+x_i)^{-(1+\theta)}$$

and the log likelihood function of θ is

$$l(\theta) = n\ln(\theta) - (1+\theta)\sum_{i=1}^n \ln(1+x_i).$$

Hence

$$l'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n \ln(1+x_i) \text{ and } l''(\theta) = -\frac{n}{\theta^2}.$$

The Fisher information for θ is

$$I_n(\theta) = -E(l''(\theta)) = \frac{n}{\theta^2}.$$

Hence the $CRLB$ for an unbiased estimator of $\frac{1}{\theta}$ is

$$CRLB = \left[\frac{\partial 1/\theta}{\partial\theta}\right]^2 / I_n(\theta) = \frac{(\theta^{-2})^2}{n/\theta^2} = \frac{1}{n\theta^2}.$$

(8)

(d) $\frac{1}{n}\sum_{i=1}^n \ln(1+X_i)$ is a function of a complete sufficient statistic. Furthermore

$$\frac{1}{n}\sum_{i=1}^n E[\ln(1+X_i)] = \frac{n}{n}E[\ln(1+X)] = -\frac{q'(\theta)}{p'(\theta)} = \frac{1}{\theta}$$

(unbiased). Hence, $\frac{1}{n}\sum_{i=1}^n \ln(1+X_i)$ is the $MVUE$ of $\frac{1}{\theta}$. **(8)**

Question 4 **[21]**

(a) (i) The level of significance of the test is

$$\begin{aligned}\alpha &= P\left(X_1 \leq \frac{1}{2}|\theta = 2\right) = P\left(0 < X_1 < \frac{1}{2}|\theta = 2\right) \\ &= \int_0^{1/2} 2x\,dx \\ &= \left[x^2\right]_0^{1/2} = 1/4.\end{aligned}$$

The probability of the type III error is

$$\begin{aligned}\beta &= P\left(X_1 > \frac{1}{2}|\theta = 1\right) = 1 - P\left(0 < X_1 < \frac{1}{2}|\theta = 1\right) \\ &= 1 - \int_0^{1/2} dx \\ &= 1 - [x]_0^{1/2} = 1/2.\end{aligned}$$

Hence, the power of the test is

$$Power = 1 - \beta = 1 - 1/2 = 1/2.$$

(8)

(ii) The level of significance of the test is

$$\begin{aligned} 0.1 &= P(X_1 \leq c|\theta = 2) = P(0 < X_1 < c|\theta = 2) \\ &= \int_0^c 2x dx \\ &= \left[x^2\right]_0^c = c^2. \end{aligned}$$

This means $c = \sqrt{0.1} = 0.3162$ **(5)**

(b) The likelihood function of θ is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \theta x_i^{(\theta-1)} = \theta^n \prod_{i=1}^{n} x_i^{(\theta-1)}.$$

Hence

$$\begin{aligned} r(1, 2|\mathbf{x}) &= \frac{L(2|\mathbf{x})}{L(1|\mathbf{x})} \\ &= 2^n \prod_{i=1}^{n} x_i. \end{aligned}$$

According to the Neyman-Pearson theorem, the best test of the hypotheses rejects H_0 if

$$\begin{aligned} r(1, 2|\mathbf{x}) &= 2^n \prod_{i=1}^{n} x_i \leq k_\alpha \in (0, 1] \\ &\equiv \text{ if } \prod_{i=1}^{n} x_i \leq 2^{-n} k_\alpha = c \end{aligned}$$

where c solves the probability equation

$$\alpha = P\left(\prod_{i=1}^{n} x_i \leq c|\theta = 2\right).$$

(8)

TOTAL: [100]