Tutorial letter 202/2/2017

Statistical Inference III

STA3702

Semester 2

Department of Statistics

SOLUTIONS TO ASSIGNMENT 02

Question 1 **[16]**

(a) The likelihood function of θ is

$$
\begin{aligned}
L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} f(x_i|\theta) = \left(\prod_{i=1}^{n} g(x_i)\right)\exp\left\{p(\theta)\sum_{i=1}^{n} K(x_i) - nq(\theta)\right\} \\
&= m_1(\mathbf{x}) \times m_2\left(\theta, \sum_{i=1}^{n} K(x_i)\right)
\end{aligned}
$$

where $m_1(\mathbf{x}) = \prod_{i=1}^{n} g(x_i)$ and $m_2\left(\theta, \sum_{i=1}^{n} K(x_i)\right) = \exp\left\{p(\theta)\sum_{i=1}^{n} K(x_i) - nq(\theta)\right\}$. This means $\sum_{i=1}^{n} K(X_i)$ is a sufficient statistic for θ by the factorisation theorem.

Now let $Y_1, Y_2, Y_3, ..., Y_n$ be another random sample from the same distribution. Then

$$
\frac{L(\theta|\mathbf{x})}{L(\theta|\mathbf{y})} = \left(\sum_{i=1}^{n} \frac{g(x_i)}{g(y_i)}\right)\exp\left\{p(\theta)\sum_{i=1}^{n}(K(x_i) - K(y_i))\right\}
$$

is independent of θ if $\sum_{i=1}^{n}(K(x_i) - K(y_i)) = 0$ equivalently if $\sum_{i=1}^{n} K(y_i) = \sum_{i=1}^{n} K(x_i)$. This means that $\sum_{i=1}^{n} K(X_i)$ is a minimal sufficient statistic for θ. **(6)**

(b) Yes $\sum_{i=1}^{n} K(X_i)$ is the complete sufficient statistic for θ - a property of the members of the exponential family of distributions. **(3)**

(c) From part (a) the log likelihood function of θ is

$$
l(\theta) = \ln L(\theta|\mathbf{x}) = \ln\left(\prod_{i=1}^{n} g(x_i)\right) + p(\theta)\sum_{i=1}^{n} K(x_i) - nq(\theta).
$$

Hence

$$
l'(\theta) = p'(\theta)\sum_{i=1}^{n} K(x_i) - nq'(\theta) \text{ and } l''(\theta) = p''(\theta)\sum_{i=1}^{n} K(x_i) - nq''(\theta).
$$

The Fisher information about θ is

$$
\begin{aligned}
I(\theta) &= -E\left[l''(\theta)\right] = -p''(\theta)\sum_{i=1}^{n} E\left[K(x_i)\right] + nq''(\theta) \\
&= nq''(\theta) - np''(\theta)\frac{q'(\theta)}{p'(\theta)} \\
&= \frac{n}{p'(\theta)}\left[q''(\theta)p'(\theta) - p''(\theta)q'(\theta)\right]
\end{aligned}
$$

The Cramer-Rao lower bound for an unbiased estimator of θ is

$$
\left[\frac{\partial\theta}{\partial\theta}\right]^2 / I(\theta) = \frac{p'(\theta)}{n\left[q''(\theta)p'(\theta) - p''(\theta)q'(\theta)\right]}.
$$

(7)

Question 2 **[30]**

(a) Yes because

$$\begin{aligned} p(x|\theta) &= \theta(1-\theta)^{x-1},\ x = 1, 2, 3, ...,\ 0 < \theta < 1, \\ &= \exp\{x\ln(1-\theta) - -\ln(\theta/(1-\theta))\} \\ &= g(x)\exp\{p(\theta)K(x) - q(\theta)\} \end{aligned}$$

where $p(\theta) = \ln(1-\theta)$, $K(x) = x$, $q(\theta) = -\ln(\theta/(1-\theta)) = \ln(1-\theta) - \ln(\theta)$ and $g(x) = 1$.

(5)

(b) Yes because $p(x|\theta)$ belongs to regular exponential family. **(4)**

(c) Use Question 1.

$$p'(\theta) = -\frac{1}{1-\theta},\ q'(\theta) = -\frac{1}{1-\theta} - \frac{1}{\theta} = -\frac{1}{\theta(1-\theta)} \Longrightarrow E[X] = -\frac{1}{\theta(1-\theta)} \times (-1)(1-\theta) = \frac{1}{\theta}.$$

Hence

$$E\left[\sum_{i=1}^{n} X_i\right] = \frac{n}{\theta}.$$

(4)

(d) Use Question 1 part(c).

$$p''(\theta) = -\frac{1}{(1-\theta)^2},\ q''(\theta) = -\frac{1}{(1-\theta)^2} + \frac{1}{\theta^2} = \frac{1-2\theta}{\theta^2(1-\theta)^2} \Longrightarrow$$

The Cramer-Rao lower bound for an unbiased estimator of θ is

$$\frac{\theta^2(1-\theta)}{n}.$$

Alternatively from first principles. The likelihood function is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta) = \left[\theta/(1-\theta)\right]^n (1-\theta)^{\sum_{i=1}^{n} x_i}$$

and the log likelihood function is

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

Now,

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

The Fisher information about θ in the sample is

$$I(\theta) = -E\left[l''(\theta)\right] = \frac{n}{\theta^2} - \frac{n}{(1-\theta)^2} + \frac{1}{(1-\theta)^2}\sum_{i=1}^{n} E\left[X_i\right]$$
$$E[X_i] = \frac{n}{\theta^2} - \frac{n}{(1-\theta)^2} + \frac{n}{\theta(1-\theta)^2} = \frac{n}{\theta^2(1-\theta)}.$$

Hence the required Cramer-Rao lower bound is

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

(7)

(e) From part (d),

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

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

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

Clearly the solution depends on the complete sufficient statistic $\sum_{i=1}^{n} X_i$. **(4)**

(f) From parts (b) and (c), the $MVUE$ of $\frac{1}{\theta}$ is $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ because it is a function of a complete sufficient statistic for θ which is also an unbiased estimator of $\frac{1}{\theta}$. **(6)**

Question 3 **[14]**

(a) The level of significance of the test is

$$\alpha = P\left(X = 2|\theta = \frac{1}{2}\right) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

The probability of the Type II Error is

$$\beta = P\left(X \neq 2|\theta = \frac{3}{4}\right) = 1 - P\left(X = 2|\theta = \frac{3}{4}\right) = 1 - \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{2-1} = \frac{13}{16}.$$

Hence, the power of the test is

$$Power = 1 - \beta = 1 - \frac{13}{16} = \frac{3}{16}.$$

(5)

(b) The level of significance of the test is

$$\alpha = P\left(X = 1 \textbf{ or } 2|\theta = \frac{1}{2}\right) = P\left(X = 1|\theta = \frac{1}{2}\right) + P\left(X = 2|\theta = \frac{1}{2}\right) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

The probability of the Type II Error is

$$\begin{aligned}\beta &= P\left(X \neq 1 \textbf{ or } 2|\theta = \frac{3}{4}\right) = 1 - P\left(X = 1|\theta = \frac{3}{4}\right) - P\left(X = 2|\theta = \frac{3}{4}\right) \\ &= 1 - \frac{3}{4} - \frac{3}{16} = \frac{1}{16}.\end{aligned}$$

Hence, the power of the test is

$$Power = 1 - \beta = 1 - \frac{1}{16} = \frac{15}{16}.$$

(6)

(c) The test is powerful but the probability of committing the Type I Error is too large. **(3)**

Question 4 **[15]**

(a) The level of significance of the test is

$$\begin{aligned}0.1 = P\left(X_1 \geq c|\theta = 2\right) &= \int_c^1 (\ln 2)2^x dx \\ &= \left[2^x\right]_c^1 = 2 - 2^c.\end{aligned}$$

This means $2^c = 1.9 \Longrightarrow c = \ln 1.9/\ln 2 = 0.923.$ **(5)**

(b) The likelihood function of θ is

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

$$\begin{aligned}r(2,3|\mathbf{x}) &= \frac{L(2|\mathbf{x})}{L(3|\mathbf{x})} \\ &= (\ln 2)^n 2^{\sum_{i=1}^n x_i} \frac{2^n}{(\ln 3)^n 3^{\sum_{i=1}^n x_i}} = \left(\frac{2\ln 2}{\ln 3}\right)^n \left(\frac{2}{3}\right)^{\sum_{i=1}^n x_i}\end{aligned}$$

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

$$\begin{aligned}r(2,3|\mathbf{x}) &= \left(\frac{2\ln 2}{\ln 3}\right)^n \left(\frac{2}{3}\right)^{\sum_{i=1}^n x_i} \leq k_{0.05} \in (0,1] \\ &\equiv \text{ if } \left(\frac{2}{3}\right)^{\sum_{i=1}^n x_i} \leq \left(\frac{2\ln 2}{\ln 3}\right)^{-n} k_{0.05} = k^*_{0.05} \\ &\equiv \text{ if } \sum_{i=1}^n x_i \geq \frac{\ln k^*_{0.05}}{\ln(2/3)} = c_{0.05} \text{ (}\textbf{since } \ln(2/3) \textbf{ is negative}\text{)}\end{aligned}$$

where $c_{0.05}$ solves the probability equation

$$0.05 = P\left(\sum_{i=1}^{n} X_i \geq c_{0.05} | \theta = 2\right).$$

(10)

TOTAL: [75]