Tutorial letter 103/2/2018

Statistical Inference III

STA3702

Semester 2

Department of Statistics

TRIAL EXAMINATION PAPER

1. Trial Examination Papar
2. Trial Examination Paper Solutions

1 Trial Examination Paper

INSTRUCTIONS

1. Answer all questions.
2. Show intermediate steps.

Abbreviations	
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
pdf	probability density function
cdf	cumulative distribution function
$MVUE$	Minimum Variance Unbiased Estimator
MME	Method of Moments Estimator
$CRLB$	Cramer Rao Lower Bound
pmf	probability mass function
mgf	moment generating function
re	relative efficiency

Question 1 **[33]**

A distribution belongs to the regular 1-parameter exponential family if among other regularity conditions its pdf or pmf has the form:

$$f(y|\theta) = a(\theta)g(y)\exp\{b(\theta)r(y)\},\ y \in \chi \subset (-\infty, \infty) \text{ and } \theta \in \Theta \subset (-\infty, \infty)$$

where $g(y) > 0$, $r(y)$ is a function of y which does not depend on θ, and $a(\theta) > 0$ and $b(\theta) \neq 0$ are real valued functions of θ. Furthermore, for this distribution (assuming the first and second derivatives of $a(\theta)$ and $b(\theta)$ exist)

$$E[r(Y)] = -\frac{a'(\theta)}{a(\theta)b'(\theta)}.$$

Suppose that $Y_1, Y_2, ..., Y_n$ is a random sample from this distribution.

(a) Show that $f(y|\theta)$ can also be expressed as:

$$f(y|\theta) = g(y)\exp\{b(\theta)r(y) - q(\theta)\},\ y \in \chi \subset (-\infty, \infty) \text{ and } \theta \in \Theta \subset (-\infty, \infty).$$

(2)

(b) Prove that $\sum_{i=1}^{n} r(Y_i)$ is a minimal sufficient statistic for θ. **(5)**

(c) Justify that $\sum_{i=1}^{n} r(Y_i)$ is also a complete sufficient statistic for θ. **(1)**

(d) Write down a function of the complete sufficient statistic for θ that is a minimum variance unbiased estimator ($MVUE$) of $E[r(Y)]$. **Justify your answer.** **(3)**

(e) Suppose that $X_1, X_2, ..., X_n$ is random sample from a distribution with pdf :

$$f(x|\theta) = \begin{cases} \dfrac{\theta}{(1+x)^{1+\theta}} & \text{if } x > 0 \text{ and } \theta > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $f(x|\theta)$ belongs to the regular 1-parameter exponential family by showing that the pdf can be expressed as:

$$f(x|\theta) = a(\theta)g(x)\exp\{b(\theta)r(x)\},\ x > 0 \text{ and } \theta > 0.$$

(5)

(ii) What is the mean of the complete sufficient statistic for θ? **Justify your answer.** **(6)**

(iv) What is the maximum likelihood estimator (MLE) of θ? **Justify your answer.** **(6)**

(v) What is the method of moments estimator (MME) of θ? **Justify your answer.** **(2)**

(vi) What is the minimum variance unbiased estimator ($MVUE$) of $\frac{1}{\theta}$? **Justify your answer.** **(3)**

Question 2 **[30]**

Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with pdf :

$$f(x|\theta) = \begin{cases} \dfrac{1}{\theta} & \text{if } 0 < x \leq \theta, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Let $Y_n = \max\{X_1, X_2, ..., X_n\}$. The probability density function of Y_n is

$$g(y|\theta) = \begin{cases} ny^{n-1}\theta^{-n} & \text{if } 0 < y \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that the ten independent observations,

$$23,\ 23,\ 24,\ 23,\ 23,\ 23,\ 24,\ 22,\ 24,\ 25,$$

are from a population with probability density function $f(x|\theta)$.

(a) What **are** the method of moments **estimator** and **estimate** of θ? **(4)**

(b) What **are** the maximum likelihood **estimator** and **estimate** of θ? **(8)**

(c) Find the mean and the variance, in terms of θ and n, of the maximum likelihood **estimator** of θ. Is the estimator consistent? **(10)**

(d) What is an estimate of the standard error of the maximum likelihood estimate of θ? **(3)**

(e) Find an approximate 95% confidence interval for θ. **(5)**

Question 3 [18]

Let $X_1, X_2, ..., X_n$ be a random sample from the distribution with pdf :

$$f(x|\theta_1,\theta_2) = \begin{cases} \dfrac{\theta_1}{\theta_2^{\theta_1}} x^{\theta_1 - 1} & \text{if } 0 < x \leq \theta_2 \text{ and } \theta_1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the likelihood function and the log-likelihood function of θ_1 and θ_2. **(6)**

(b) Show that the respective maximum likelihood estimators ($MLE's$) of θ_1 and θ_2 are

$$\hat{\theta}_1 = \frac{n}{n \ln X_{(n)} - \sum_{i=1}^{n} \ln X_i} \text{ and } \hat{\theta}_2 = X_{(n)} = \max\{X_1, X_2, ..., X_n\},$$

respectively. **Hint: First find the MLE of θ_2 and then that of θ_1.** **(12)**

Question 4 [19]

A random variable X has probability density function

$$f(x|\theta) = \frac{\exp\{x-\theta\}}{(1+\exp\{x-\theta\})^2}, \quad -\infty < x < \infty \text{ and } -\infty < \theta < \infty,$$

and **cumulative** distribution function

$$F(x|\theta) = 1 - \frac{1}{1+\exp\{x-\theta\}}, \quad -\infty < x < \infty \text{ and } -\infty < \theta < \infty.$$

It is desired to test the null hypothesis $H_0 : \ \theta = 0$ against the alternative $H_1 : \ \theta = 1$ using one random observation X.

(a) Calculate the level of significance and the power of the test which rejects H_0 if $X > 1$. **(12)**

(b) Consider the test which rejects H_0 if $X_1 \geq c$. Find c for which the level of significance of the test is 0.1.

(7)

TOTAL [100]

2 Trial Examination Solutions

Question 1 **[Total marks=30]**

(a)

$$\begin{aligned} f(y|\theta) &= a(\theta)g(y)\exp\{b(\theta)R(y)\} \\ &= g(y)\exp\{b(\theta)r(y) - -\ln[a(\theta)]\} \\ &= g(y)\exp\{b(\theta)r(y) - q(\theta)\} \end{aligned}$$

where $q(\theta) = -\ln[a(\theta)]$. **(2)**

(b) The likelihood function of θ is

$$\begin{aligned} L(\theta|y) &= \prod_{i=1}^{n} f(y_i|\theta) = \left(\prod_{i=1}^{n} g(y_i)\right)\exp\left\{b(\theta)\sum_{i=1}^{n} r(y_i) - nq(\theta)\right\} \\ &= m_1(y) \times m_2\left(\theta, \sum_{i=1}^{n} r(y_i)\right) \end{aligned}$$

where $m_1(\mathbf{y}) = \prod_{i=1}^{n} g(y_i)$ and $m_2\left(\theta, \sum_{i=1}^{n} r(y_i)\right) = \exp\left\{b(\theta)\sum_{i=1}^{n} r(y_i) - nq(\theta)\right\}$. This means $\sum_{i=1}^{n} r(Y_i)$ is sufficient statistic for θ by the factorisation theorem.

Now let $X_1, X_2, ..., X_n$ be another random sample from the same distribution. Then

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

is independent of θ if $\sum_{i=1}^{n}(r(y_i) - r(x_i)) = 0$ equivalently if $\sum_{i=1}^{n} r(y_i) = \sum_{i=1}^{n} r(x_i)$. This means that $\sum_{i=1}^{n} r(Y_i)$ is a minimal sufficient statistic for θ. **(5)**

(c) It is because $f(y|\theta)$ belongs to the 1–parameter exponential family. **(1)**

(d) $\frac{1}{n}\sum_{i=1}^{n} r(Y_i)$ because it is a function of the complete sufficient statistic for θ and $E\left[\frac{1}{n}\sum_{i=1}^{n} r(Y_i)\right] = E[r(Y)]$. **(3)**

(e) (i)

$$\begin{aligned} f(x|\theta) &= \frac{\theta}{(1+x)^{1+\theta}} \\ &= \theta(1+x)^{-1}(1+x)^{-\theta} \\ &= \theta(1+x)^{-1}\exp\{-\theta\ln(1+x)\} \\ &= a(\theta)\,g(x)\exp\{b(\theta)\,r(x)\} \end{aligned}$$

where $a(\theta)=\theta;\ g(x)=(1+x)^{-1};\ b(\theta)=\theta;$ and $r(x)=-\ln(1+x)$. **(5)**

(ii) The complete sufficient statistic is $\sum_{i=1}^{n} r(X_i) = -\sum_{i=1}^{n}\ln(1+X_i)$ since $f(x|\theta)$ belongs to the 1–parameter exponential family. $E[r(X)] = -\frac{a'(\theta)}{a(\theta)\,b'(\theta)} = \frac{1}{\theta\times 1} = \frac{1}{\theta} \Rightarrow E\left[-\sum_{i=1}^{n}\ln(1+X_i)\right] = -\frac{n}{\theta}$. **(6)**

(iii) The likelihood function is

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

and the log-likelihood function is

$$l(\theta) = n\ln\theta + \ln\left[\prod_{i=1}^{n}(1+x_i)^{-1}\right] - \theta\sum_{i=1}^{n}\ln(1+x_i).$$

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

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

The solution is $\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\ln(1+X_i)}$. **(6)**

(iv) $E[\ln(1+X)] = \frac{1}{\theta} \Longrightarrow \theta = \frac{1}{E[\ln(1+X)]} \Longrightarrow$ the MME of θ is $\widetilde{\theta} = \frac{n}{\sum_{i=1}^{n}\ln(1+X_i)}$. **(2)**

(v) The minimum variance unbiased estimator $(MVUE)$ of $\frac{1}{\theta}$ is $\frac{1}{n}\sum_{i=1}^{n}\ln(1+X_i)$ because the estimator is a function of a complete sufficient statistic for θ and $E\left[\frac{1}{n}\sum_{i=1}^{n}\ln(1+X_i)\right] = \frac{1}{\theta}$. **(3)**

Question 2 **[Total marks=30]**

(a) $E(X) = \int_0^\theta x/\theta dx \left[\frac{x^2}{2\theta}\right]_0^\theta = \frac{\theta}{2} \Rightarrow \theta = 2E[X]$. This means the method of moments estimate (MME) of θ is $\widetilde{\theta} = 2\bar{x} = 2 \times 23.2 = 46.4$. **(4)**

(b) The likelihood function of θ

$$L(\theta|\mathbf{x}) = \begin{cases} \theta^{-n} & \text{if } \theta \geq x_{(n)} = \max\{x_1, x_2, ..., x_n\}. \\ 0 & \text{otherwise,} \end{cases}$$

decreases as θ increase. Thus the function attains its maximum at the smallest value of θ which is $x_{(n)}$. This means the maximum likelihood estimate of θ is $x_{(10)} = 25$. **(8)**

(c) $E[Y_n] = \int_0^\theta yny^{n-1}\theta^{-n}dy = \left[\frac{n}{n+1}y^{n+1}\theta^{-n}\right]_0^\theta = \frac{n}{n+1}\theta$. **(2)**

$E\left[Y_n^2\right] = \int_0^\theta y^2ny^{n-1}\theta^{-n}dy = \left[\frac{n}{n+2}y^{n+2}\theta^{-n}\right]_0^\theta = \frac{n}{n+2}\theta^2 \Rightarrow$

$Var(Y_n) = E\left[Y_n^2\right] - (E[Y_n])^2 = \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$. **(4)**

That the bias of the estimate is $\theta - E[Y_n] = \frac{1}{n+1}\theta \to 0$ as $n \to \infty$ and $\lim_{n\to\infty} Var(Y_n) = 0$ means Y_n is a consistent estimator of θ. **(4)**

(d) The estimate is $\sqrt{Var(Y_n)|_{n=10,\ \theta=x_{(10)}=25}} = \sqrt{\frac{10}{12 \times 11^2} \times 25^2\widetilde{\theta}} = 2.0747$. **(3)**

(e) The 95% confidence interval is $25 \pm 1.960 \times 2.0747 = [20.9334;\ 29.0664]$. **(5)**

Question 3 **[Total marks=18]**

(a) The likelihood function of θ_1 and θ_2 is

$$L(\theta_1, \theta_2|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta_1, \theta_2) = \begin{cases} \left(\prod_{i=1}^n x_i\right)^{-1} \frac{\theta_1^n}{\theta_2^{n\theta_1}} \left(\prod_{i=1}^n x_i\right)^{\theta_1} & \text{if } \theta_1 > 0,\ \theta_2 \geq \text{ all the } x_i's, \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently,

$$L(\theta_1, \theta_2|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta_1, \theta_2) = \begin{cases} \left(\prod_{i=1}^n x_i\right)^{-1} \frac{\theta_1^n}{\theta_2^{n\theta_1}} \left(\prod_{i=1}^n x_i\right)^{\theta_1} & \text{if } \theta_1 > 0,\ \theta_2 \geq x_{(n)}, \\ 0 & \text{otherwise.} \end{cases}$$

where $x_{(n)} = \max\{x_1, x_2, x_3, ..., x_n\}$. **(3)**

The log likelihood function is

$$\begin{aligned} l(\theta_1,\theta_2) &= \ln L(\theta_1,\theta_2) \\ &= \begin{cases} \ln\left(\prod_{i=1}^{n} x_i\right)^{-1} + n\ln\theta_1 - n\theta_1\ln\theta_2 + \theta_1\ln\left(\prod_{i=1}^{n} x_i\right) & \text{if } \theta_1 > 0,\ \theta_2 \ge x_{(n)}, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

(3)

(b) For any $\theta_1 > 0$, $L(\theta_1,\theta_2|\mathbf{x})$ decreases as θ_2 increases. Thus the function attains its maximum at the smallest value of θ_2 which is $x_{(n)}$. This means the maximum likelihood estimator of θ_2 is $X_{(n)}$. **(6)**

Consider the log likelihood

$$l\left(\theta_1, x_{(n)}\right) = \ln\left(\prod_{i=1}^{n}\right)^{-1} + n\ln\theta_1 - n\theta_1\ln x_{(n)} + \theta_1\ln\left(\prod_{i=1}^{n} x_i\right)$$

whose derivative with respect to θ_1 is

$$l'\left(\theta_1 x_{(n)}\right) = \frac{n}{\theta_1} - n\ln x_{(n)} + \ln\left(\prod_{i=1}^{n} x_i\right) = \frac{n}{\theta_1} - n\ln x_{(n)} + \sum_{i=1}^{n}\ln x_i.$$

The MLE of θ_1 is $\hat{\theta}_1$ which solves the equation

$$0 = l'\left(\hat{\theta}_1, X_{(n)}\right) = \frac{n}{\hat{\theta}_1} - n\ln X_{(n)} + \ln\left(\prod_{i=1}^{n} X_i\right) = \frac{n}{\hat{\theta}_1} - n\ln X_{(n)} + \sum_{i=1}^{n}\ln X_i.$$

The solution is

$$\hat{\theta}_1 = \frac{n}{n\ln X_{(n)} - \sum_{i=1}^{n}\ln X_i} = \frac{n}{n\ln X_{(n)} - \ln\left(\prod_{i=1}^{n} X_i\right)}.$$

(6)

Question 4 **[Total marks=19]**

(a) $\alpha = P(X > 1|\theta = 0) = 1 - P(X < 1|\theta = 0) = 1 - F(1|0) = \dfrac{1}{1+\exp(1)} = 0.2689.$ **(5)**

$\beta = P(X < 1|\theta = 1) = F(1|1) = 1 - \dfrac{1}{1+\exp(0)} = \dfrac{1}{2}.$ **(5)**

Hence the power $= 1 - \beta = \dfrac{1}{2}.$ **(2)**

(b) $0.1 = P(X \ge c|\theta = 0) = 1 - F(c|0) = \dfrac{1}{1+\exp(c)} \Rightarrow$

$10 = 1 + \exp(c) \Rightarrow 9 = \exp(c) \Rightarrow c = \ln 9 = 2.1972.$ **(7)**

TOTAL [100]