MA 324 | Mid-semester Examination
Statistical Inference and Multivariate Analysis | 14:00–16:00 IST
IIT GUWAHATI | March 01, 2022

Model Answers of Mid-semester Examination

1. (2 points) Let $X_1,\, X_2 \overset{i.i.d.}{\sim} Poi(\lambda)$, where $\lambda > 0$ is unknown parameter. Is the family of distributions induced by the statistic $\boldsymbol{T} = (X_1,\, X_2)$ complete?

   Solution: Note that $E\left(X_1 - X_2\right) = 0$ for all $\lambda > 0$. Now,

   $$P\left(X_1 \neq X_2\right) \geq P\left(X_1 = 0,\, X_2 = 1\right) = \lambda e^{-2\lambda} > 0 \implies P\left(X_1 - X_2 = 0\right) = P\left(X_1 = X_2\right) < 1.$$

   Thus, the family of distributions induced by the statistics $(X_1,\, X_2)$ is not complete. [2 points]

2. (5 points) Let $X_1,\, X_2,\, \ldots,\, X_9$ be a random sample of size 9 form population having $U(\theta_1,\, \theta_2)$ distribution, where both θ_1 and θ_2 are unknown and $-\infty < \theta_1 < \theta_2 < \infty$. Derive the estimators of θ_1 and θ_2 using method of moment.

   Solution: Here $E(X_1) = \frac{\theta_1+\theta_2}{2}$ and $E\left(X_1^2\right) = \frac{\theta_1^2+\theta_1\theta_2+\theta_2^2}{3}$. [1 point]
   Let $M_1 = \frac{1}{9}\sum_{i=1}^{9} X_i$, $M_2 = \frac{1}{9}\sum_{i=1}^{9} X_i^2$ and $S^2 = \frac{1}{9}\sum_{i=1}^{9}\left(X_i - M_1\right)^2$. The method of moment estimators can be found by solving the following equations:

   $$\theta_1 + \theta_2 = 2m_1 \qquad \text{and} \qquad \theta_1^2 + \theta_1\theta_2 + \theta_2^2 = 3m_2. \; \text{[1 point]}$$

   The solutions for $(\theta_1,\, \theta_2)$ are $\left(m_1 - \sqrt{3}s,\, m_1 + \sqrt{3}s\right)$ and $\left(m_1 + \sqrt{3}s,\, m_1 - \sqrt{3}s\right)$, where S is the positive square root of S^2. As $\theta_1 < \theta_2$, the estimator for θ_1 and θ_2 are $\widehat{\theta_1} = M_1 - \sqrt{3}S$ and $\widehat{\theta_2} = M_1 + \sqrt{3}S$, respectively. [3 points]

3. Let $X_1,\, X_2,\, \ldots,\, X_n$ be a random sample of size $n\,(\geq 2)$ from a population having probability density function

   $$f(x,\, \theta) = \begin{cases} \frac{2}{\theta} x \exp\left[-\frac{x^2}{\theta}\right] & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

   where $\theta > 0$ is a unknown parameter. Consider the problem of estimation of $\tau(\theta) = \dfrac{1}{\sqrt{\theta}}$.

   (a) (5 points) Derive minimum variance unbiased estimator of $\tau(\theta)$.

   Solution: Note that $f(\cdot, \theta)$ belongs to a full rank exponential family. Thus, using property of exponential family of distributions, $T = \sum_{i=1}^{n} X_i^2$ is complete and sufficient statistic for θ. [2 points]
   Notice that X_i^2 follows exponential distribution with mean θ. Therefore,

   $$T \sim Gamma\left(n,\, \frac{1}{\theta}\right). \; \text{[1 point]}$$

   Thus, for $k > -n$,

   $$E\left(T^k\right) = \frac{1}{\theta^n \Gamma(n)} \int_0^\infty t^{k+n-1} e^{-\frac{t}{\theta}} dt = \frac{\Gamma(n+k)}{\Gamma(n)}\theta^k \implies E\left(\frac{\Gamma(n)}{\Gamma(n+k)} T^k\right) = \theta^k.$$

Therefore, using Lehmann-Scheffee theorem and taking $k = -\frac{1}{2}$, we have

$$\widehat{\tau} = \frac{\Gamma(n)}{\Gamma(n-\frac{1}{2})}\left(\sum_{i=1}^{n} X_i^2\right)^{-\frac{1}{2}}$$

is the UMVUE of $\tau(\theta) = \theta^{-\frac{1}{2}}$. [2 points]

(b) (3 points) Show that the estimator that you obtained in (a) is consistent. You may use Stirling's approximation for $\Gamma(n)$: $\Gamma(n) \sim \sqrt{2\pi}\,(n-1)^{n-\frac{1}{2}}\,e^{-n+1}$.

Solution: Here $E\left(X_1^2\right) = \theta$. Therefore, using WLLN,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \longrightarrow \theta$$

in probability. [1 point]
Now,

$$\begin{aligned}
\widehat{\tau}_n &= \frac{\Gamma(n)}{\Gamma(n-\frac{1}{2})}\left(\sum_{i=1}^{n} X_i^2\right)^{-\frac{1}{2}} \\
&= \frac{\Gamma(n)}{\sqrt{n}\Gamma(n-\frac{1}{2})}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-\frac{1}{2}} \\
&\sim e^{-\frac{1}{2}}\left(1-\frac{1}{n}\right)^{\frac{1}{2}} \frac{1}{\left(1-\frac{1}{2(n-1)}\right)^{n-1}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-\frac{1}{2}} \longrightarrow \tau(\theta)
\end{aligned}$$

in probability. Therefore $\widehat{\theta}$ is a consistent estimator of $\tau(\theta)$. [2 points]

(c) (3 points) Compute Cramer-Rao lower bound of an unbiased estimator of $\tau(\theta)$.

Solution: Here $\tau'(\theta) = -\frac{1}{2}\theta^{-\frac{3}{2}}$. The Fisher information present in X_1 is

$$\begin{aligned}
\mathcal{I}_{X_1}(\theta) &= -E\left[\frac{d^2}{d\theta^2}\ln f(X,\,\theta)\right] \\
&= -E\left[\frac{d^2}{d\theta^2}\ln\left\{\frac{2X_1}{\theta}e^{-\frac{X_1^2}{\theta}}\right\}\right] \\
&= -E\left[\frac{d^2}{d\theta^2}\left\{\ln 2 - \ln\theta + \ln X_1 - \frac{X_1^2}{\theta}\right\}\right] \\
&= -E\left[\frac{1}{\theta^2} - \frac{2X_1^2}{\theta^3}\right] \\
&= \frac{1}{\theta^2}.
\end{aligned}$$

[1 point]

Therefore, CRLB is

$$\frac{\left(\tau'(\theta)\right)^2}{n\mathcal{I}_{X_1}(\theta)} = \frac{1}{4n\theta}.$$

[2 points]

4. (5 points) Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with success probability $p = \frac{1}{1+e^\theta}$, where $\theta \in \mathbb{R}$. Find the maximum likelihood estimator of θ. [Hint: Investigate the existence and non-existence of maximum likelihood estimator.]

Solution: The likelihood function of θ is

$$L(\theta) = \left(\frac{1}{1+e^\theta}\right)^m \left(1 - \frac{1}{1+e^\theta}\right)^{n-m},$$

where m is the realized value of $\sum_{i=1}^n X_i$. [1 point]

Now, consider the following cases.

Case I: $m = 0$. In this case, the likelihood function of θ is

$$L(\theta) = \left(1 - \frac{1}{1+e^\theta}\right)^n,$$

which is an increasing function in $\theta \in \mathbb{R}$. Therefore, in this case the MLE of θ does not exist. [1 point]

Case II: $m = n$. In this case the likelihood function of θ is

$$L(\theta) = \left(\frac{1}{1+e^\theta}\right)^n,$$

which is a decreasing function in $\theta \in \mathbb{R}$. Thus, the MLE of θ does not exist in this case also. [1 point]

Case III: $m = 1, 2, \ldots, n-1$. In this case, the log-likelihood function is

$$l(\theta) = -n \ln\left(1 + e^\theta\right) + (n-m)\theta.$$

Taking first derivative with respect to θ and equate it to zero, we obtain

$$-\frac{ne^\theta}{1+e^\theta} + n - m = 0 \implies \theta = \ln\left(\frac{n}{m} - 1\right).$$

Moreover,

$$\frac{d^2}{d\theta^2} l(\theta) = -\frac{ne^\theta}{\left(1+e^\theta\right)^2} < 0$$

for all $\theta \in \mathbb{R}$. Therefore, $l(\theta)$ attains it's maximum at $\theta = \ln\left(\frac{n}{m} - 1\right)$. Thus, MLE of θ exists in this case and the MLE is $\widehat{\theta} = \ln\left(\frac{n}{\sum_{i=1}^n X_i} - 1\right)$. [2 points]

5. (7 points) Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n\,(\geq 2)$ from a population with probability density function

$$f(x;\, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is unknown. With preassigned $\alpha \in (0, 1)$, derive a level α likelihood ratio test for $H_0 : \theta = \theta_0 (> 0)$ against $H_1 : \theta \neq \theta_0$.

Solution: Here $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = (0, \infty) \setminus \Theta_0$. The likelihood function is

$$L(\theta) = \frac{1}{\theta^n} \exp\left[-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right]. \quad \text{[1 point]}$$

Therefore,

$$\sup_{\theta \in \Theta_0} L(\theta) = L(\theta_0) = \frac{1}{\theta_0^n} \exp\left[-\frac{1}{\theta_0}\sum_{i=1}^{n} x_i\right].$$

To find $\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)$, we need to find MLE of $\theta > 0$. Now, standard calculation shows that MLE of θ is $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Therefore,

$$\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta) = \sup_{\theta \in (0,\infty)} L(\theta) = L\left(\overline{x}\right) = \frac{1}{\overline{x}^n} \exp\left[-\frac{1}{\overline{x}}\sum_{i=1}^{n} x_i\right] = \frac{1}{\overline{x}^n} e^{-n}. \quad \text{[1 point]}$$

The likelihood ratio test statistics is

$$\Lambda = \left(\frac{\overline{x}}{\theta_0}\right)^n \exp\left[-n\left(\frac{\overline{x}}{\theta_0} - 1\right)\right]. \quad \text{[1 point]}$$

A LRT rejects null hypothesis if

$$\begin{aligned} & \Lambda < k \quad \text{[1 point]} \\ \iff & \left(\frac{\overline{x}}{\theta_0}\right)^n \exp\left[-\frac{n\overline{x}}{\theta_0}\right] < k. \end{aligned}$$

Here k is used as an generic constant. Now consider the function

$$f(y) = y^n e^{-ny} \quad \text{for } y > 0.$$

It is easy to see that f has unique maximum at $y = 1$, f is strictly increasing for $0 < y < 1$ and strictly decreasing for $y > 1$. Moreover, $f(0) = 0$ and $\lim_{y\to\infty} f(y) = 0$. Therefore,

$$\Lambda < k \iff \left(\frac{\overline{x}}{\theta_0}\right)^n \exp\left[-\frac{n\overline{x}}{\theta_0}\right] < k \iff \frac{n\overline{x}}{\theta_0} < k_1 \text{ or } \frac{n\overline{x}}{\theta_0} > k_2, \quad \text{[2 points]}$$

for $k_1 < k_2$. Now, under null hypothesis,

$$\frac{n\overline{X}}{\theta_0} \sim Gamma(n, 1).$$

Thus, the test function of level α LRT is

$$\psi(x) = \begin{cases} 1 & \text{if } \frac{n\overline{x}}{\theta_0} < G_{1-\frac{\alpha}{2}} \text{ or } \frac{n\overline{x}}{\theta_0} > G_{\frac{\alpha}{2}} \\ 0 & \text{otherwise,} \end{cases}$$

where G_α is upper α-point of a $Gamma(n, 1)$ distribution. [1 point]