Dan Mark Orapa

$2020-05453$

Math 150.2 AY 2022-2023 2nd Semester

March 22, 2023

W4 Assessment 2

1. Let X_1, X_2, ..., X_n denote a random sample from the distribution with common pdf

$$f(x; \theta) = e^{-(x-\theta)} \mathbb{1}_{(\theta, +\infty)}(x), \quad \theta \in \mathbb{R}$$

   Let $Y_1 = \min \{X_1, X_2, ..., X_n\}$. Is Y_1 a consistent estimator of θ? Justify your answer.

   Proof. Consider the following

   $$F_X(x) = \int_\theta^x f(t; \theta) dt$$
   $$= \int_\theta^x e^{-(t-\theta)} dt$$
   $$= 1 - e^{-(x-\theta)}$$

   Hence, we have the resulting CDF of f. We can now define the CDF of Y_1 which can be given by

   $$F_{Y_1}(x) = 1 - (1 - F_X(x))^n = 1 - e^{-n(x-\theta)}$$

   To prove the consistency of Y_1, we need to show that $\mathbb{P}[|Y_1 - \theta| \geq \varepsilon]$ converges to 0 as $n \longrightarrow \infty$. Let $\varepsilon > 0$. We have

   $$\mathbb{P}[|Y_1 - \theta| \geq \varepsilon] = 1 - \mathbb{P}[|Y_1 - \theta| < \varepsilon]$$
   $$= 1 - \mathbb{P}[\theta - \varepsilon < Y_1 < \theta + \varepsilon]$$
   $$= 1 - (F_{Y_1}(\theta + \varepsilon) - F_{Y_1}(\theta - \varepsilon))$$

   since f supports only the lower bound θ, then $F(\theta - \varepsilon) = 0$. It follows that

   $$\mathbb{P}[|Y_1 - \theta| \geq \epsilon] = 1 - F_{Y_1}(\theta + \varepsilon)$$
   $$= 1 - \left(1 - e^{-n(\varepsilon)}\right)$$
   $$= 1 - \left(1 - \frac{1}{e^{n\varepsilon}}\right)$$

   and since $\varepsilon > 0$, we have $e^{n\varepsilon} \longrightarrow \infty$ as $n \longrightarrow \infty$. Therefore,

   $$\mathbb{P}[|Y_1 - \theta| \geq \epsilon] = 1 - 1 = 0$$

   We have shown that indeed $\mathbb{P}[|Y_1 - \theta|] \geq \varepsilon] \longrightarrow 0$ as $n \longrightarrow \infty$. Therefore, Y_1 is a consistent estimator. $\qquad \square$

2. In a College-wide orientation, students were asked which province do they come from. Three interviewers have asked 18, 46, and 23 people before finding someone not from Metro Manila. What is the maximum likelihood estimate for the percentage of students from Manila?

Proof. The setup for the problem requires the geometric distribution with parameter p. We let the parameter p be the probability that a student is not from Metro Manila. That is,

$$P_X(x) = (1-p)^x p$$

We solve for the likelihood $\mathcal{L}(p)$ by

$$\begin{aligned}
\mathcal{L}(p) &= \mathbb{P}[(X_1, X_2, X_3) = (x_1, x_2, x_3)] \\
&= \mathbb{P}(X_1 = 18) \cdot \mathbb{P}(X_2 = 46) \cdot \mathbb{P}(X_3 = 23) \\
&= p^3(1-p)^{87}
\end{aligned}$$

We solve for $\ell'(p)$.

$$\ell(p) = 3\ln(p) + 87\ln(1-p)$$
$$\ell'(p) = \frac{3}{p} - \frac{87}{1-p}$$

We then equate the resulting derivative to zero. Moreover, $p \neq 0$ and so is $1-p$. We have

$$3 - 3p = 87p \longrightarrow p = \frac{1}{30}.$$

This resulting parameter is then denoted as $\hat{p}$ which defines the likelihood estimate for the percentage of students not in Metro Manila. Therefore, the maximum likelihood estimate for the percentage of students from Metro Manila is $1 - \hat{p} = \boxed{\dfrac{29}{30}}$. $\qquad\square$

3. Let X_1, X_2, ..., X_n be a random sample from a gamma distribution with parameters α and β. Suppose β is known and α is not.

(a) Use the factorization theorem to show that $T = \displaystyle\sum_{i=1}^{n} \ln(X_i)$ is a sufficient statistic for α

Proof. It is known that X is gamma distributed with parameters α and β. We have the pdf

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Hence,

$$\begin{aligned}
\prod_{i=1}^{n} f(x_i; \alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x_1^{\alpha-1} e^{-\beta x_1} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} x_2^{\alpha-1} e^{-\beta x_2} \cdot \ldots \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} x_n^{\alpha-1} e^{-\beta x_n} \\
&= \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \prod_{i=1}^{n} x_i^{\alpha-1} e^{-\sum \beta x_i}
\end{aligned}$$

Consider the following

$$\begin{aligned}
\prod_{i=1}^{n} f(x_i; \alpha, \beta) &= \exp\left[\ln\left(\prod_{i=1}^{n} f(x_i; \alpha, \beta) \right) \right] \\
&= \exp\left[n\ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right) + (\alpha - 1)\sum_{i=1}^{n} \ln(x_i) - \beta \sum_{i=1}^{n} x_i \right] \\
&= \exp\left[n\ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right) + (\alpha - 1)\sum_{i=1}^{n} \ln(x_i) \right] \cdot \frac{1}{\exp\left(\beta \sum_{i=1}^{n} x_i \right)}
\end{aligned}$$

We let the functions g and h be

$$g(\cdot; \alpha, \beta) := \exp\left[n \ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) + (\alpha - 1)\sum_{i=1}^{n} \ln(x_i)\right] \quad \text{and} \quad h(\cdot; \beta) := \frac{1}{\exp\left(\beta \sum_{i=1}^{n} x_i\right)}$$

where both g and h are nonnegative functions. Moreover, g is a function of T. By the factorization theorem, we conclude that T is indeed a sufficient statistic of α. $\quad\square$

(b) Is $\displaystyle\prod_{i=1}^{n} X_i$ also sufficient? Justify your answer.

Proof. Consider the same distribution as above and we have the following

$$\prod_{i=1}^{n} f(x_i; \alpha, \beta) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \prod_{i=1}^{n} x_i^{\alpha-1} e^{-\sum \beta x_i}$$

$$= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\sum \beta x_i}$$

We let the functions g and h by

$$g(\cdot; \alpha, \beta) := \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \quad \text{and} \quad h(\cdot; \beta) := e^{-\sum \beta x_i}$$

Both of which are nonnegative functions. We can see that g is indeed a function of $T = \displaystyle\prod_{i=1}^{n} X_i$ and h is independent of the parameter α. Hence, by the factorization theorem, $T = \displaystyle\prod_{i=1}^{n} X_i$ is also sufficient. $\quad\square$