Assignment 6 - Q54, June 2018

Tarandeep Singh

Github Link

https://github.com/Tarandeep97/AI5030

1 Problem

(Q54, June 2018) Let $X_1, X_2, X_3..X_n$ be i.i.d. uniform (θ_1, θ_2) random variables, where $\theta_1 < \theta_2$ are unknown parameters. Which of the following is an ancillary statistic?

1) $\frac{X_{(k)}}{X_{(n)}}$ for any $k < n$
2) $\frac{X_{(n)}-X_{(k)}}{X_{(n)}}$ for any $k < n$
3) $\frac{X_{(k)}}{X_{(n)}-X_{(k)}}$ for any $k < n$
4) $\frac{X_{(k)}-X_{(1)}}{X_{(n)}-X_{(k)}}$ for any $k < n$ where $1 < k < n$

2 Solution

Definitions

Statistic: A statistic is any function of the observations in a sample.

Ancillary statistic: A statistic S(X) is an ancillary statistic if its distribution does not depend on θ.

Order Statistics: If $X_1, X_2, X_3..X_n$ are observations of a random sample of size n from a continuous distribution, we arrange the random variables as: $X_1' < X_2' < X_3' ... < X_n'$

where X_k' is the kth order statistic i.e. kth smallest among all observations.

Clearly, $X_i \sim U(\theta_1, \theta_2)$ is not an ancillary statistic as it depends on θ_1 and θ_2 i.e. its pdf and cdf is given by,

$$f_{X_i}(x) = \frac{1}{\theta_2 - \theta_1} \tag{2.0.1}$$

$$F_{X_i}(x) = P(X_i < x) = \frac{x - \theta_1}{\theta_2 - \theta_1} \tag{2.0.2}$$

for $\theta_1 \leq x_i \leq \theta_2$

These Random variables$X_i \sim U(\theta_1, \theta_2)$ can be converted to Standard Uniform Random Variables as follows,

As we know,

$$\theta_1 \leq x_i \leq \theta_2 \tag{2.0.3}$$

$$0 \leq x_i - \theta_1 \leq \theta_2 - \theta_1 \tag{2.0.4}$$

$$0 \leq \frac{x_i - \theta_1}{\theta_2 - \theta_1} \leq 1 \tag{2.0.5}$$

Considering,

$$y_i = \frac{x_i - \theta_1}{\theta_2 - \theta_1} \tag{2.0.6}$$

Hence, given distribution of $X_i \sim U(\theta_1, \theta_2)$ can be converted to $Y_i \sim U(0, 1)$ which is independent of given parameters.

This also implies that each of these Y_i and their combination will be an ancillary statistic as its pdf and cdf are as follows,

$$f_{Y_i}(x) = 1 \tag{2.0.7}$$

$$F_{Y_i}(x) = P(Y \leq x) \tag{2.0.8}$$

$$F_{Y_i}(x) = P\left(\frac{x_i - \theta_1}{\theta_2 - \theta_1} \leq x\right) \tag{2.0.9}$$

$$F_{Y_i}(x) = P\left[x_i \leq x(\theta_2 - \theta_1) + \theta_1\right] \tag{2.0.10}$$

Using 2.0.2,

$$F_{Y_i}(x) = x \tag{2.0.11}$$

Option 1 can be rewritten as,

$$\frac{X_{(k)}}{X_{(n)}} = \frac{X_{(k)} - \theta_1 + \theta_1}{X_{(n)} - \theta_1 + \theta_1} \tag{2.0.12}$$

$$= \frac{\frac{X_{(k)}-\theta_1}{\theta_2-\theta_1} + \frac{\theta_1}{\theta_2-\theta_1}}{\frac{X_{(n)}-\theta_1}{\theta_2-\theta_1} + \frac{\theta_1}{\theta_2-\theta_1}} \tag{2.0.13}$$

$$= \frac{Y_k + \frac{\theta_1}{\theta_2-\theta_1}}{Y_n + \frac{\theta_1}{\theta_2-\theta_1}} \tag{2.0.14}$$

Option 2 can be rewritten as,

$$\frac{X_{(n)} - X_{(k)}}{X_{(n)}} = \frac{X_{(n)} - \theta_1 - (X_{(k)} - \theta_1)}{X_{(n)} - \theta_1 + \theta_1} \tag{2.0.15}$$

$$= \frac{\frac{X_{(n)}-\theta_1}{\theta_2-\theta_1} + \frac{X_{(k)}-\theta_1}{\theta_2-\theta_1}}{\frac{X_{(n)}-\theta_1}{\theta_2-\theta_1} + \frac{\theta_1}{\theta_2-\theta_1}} \tag{2.0.16}$$

$$= \frac{Y_n + Y_k}{Y_n + \frac{\theta_1}{\theta_2-\theta_1}} \tag{2.0.17}$$

Option 3 can be rewritten as,

$$\frac{X_{(k)}}{X_{(n)} - X_{(k)}} = \frac{X_{(k)} - \theta_1 + \theta_1}{X_{(n)} - \theta_1 - (X_{(k)} - \theta_1)} \tag{2.0.18}$$

$$= \frac{\frac{X_{(k)}-\theta_1}{\theta_2-\theta_1} + \frac{\theta_1}{\theta_2-\theta_1}}{\frac{X_{(n)}-\theta_1}{\theta_2-\theta_1} + \frac{X_{(k)}-\theta_1}{\theta_2-\theta_1}} \tag{2.0.19}$$

$$= \frac{Y_k + \frac{\theta_1}{\theta_2-\theta_1}}{Y_n + Y_k} \tag{2.0.20}$$

Option 4 can be rewritten as,

$$\frac{X_{(k)} - X_{(1)}}{X_{(n)} - X_{(k)}} = \frac{X_{(k)} - \theta_1 - (X_{(1)} - \theta_1)}{X_{(n)} - \theta_1 - (X_{(k)} - \theta_1)} \tag{2.0.21}$$

$$= \frac{\frac{X_{(k)}-\theta_1}{\theta_2-\theta_1} + \frac{X_{(1)}-\theta_1}{\theta_2-\theta_1}}{\frac{X_{(n)}-\theta_1}{\theta_2-\theta_1} + \frac{X_{(k)}-\theta_1}{\theta_2-\theta_1}} \tag{2.0.22}$$

$$= \frac{Y_k + Y_1}{Y_n + Y_k} \tag{2.0.23}$$

Since, only option 4 does not depend on θ_1 and θ_2, it an ancillary statistic.