Assignment 9

Adhvik Mani Sai Murarisetty - AI20BTECH11015

Download latex-tikz codes from

https://github.com/adhvik24/AI1103-PROBABILITY-AND-RANDOM-VARIABLES/blob/main/Assignment_9/AI1103_Assignment9.tex

1 GATE 2020 (ST), Q.23 (STATISTICS SECTION)

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n ($n \geq 2$) from an exponential distribution with the probability density function

$$f(x, \theta) = \begin{cases} e^{-(x-2\theta)}, & x > 2\theta \\ 0, & otherwise \end{cases} \tag{1.0.1}$$

where $\theta \in (0, \infty)$. If $X_{(1)} = min\{X_1, X_2, \cdots, X_n\}$ then the conditional expectation

$$E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|X_1 - X_2 = 2\right] = \_\_\_\_\_\_$$

2 SOLUTION

DEFINITIONS:

1) **Completeness:** The statistic T is said to be complete for the distribution of X if, for every measurable function g if

$$E(g(T)) = 0 \implies P(g(T) = 0) = 1 \; \forall\theta \tag{2.0.1}$$

2) **Sufficiency:** Let $f(x, \theta)$ be the joint pdf of the sample X. A statistic T is sufficient for θ iff there are functions h (does not depend on θ)and g (depends on θ) on the range of T such that

$$f(x, \theta) = g(T(x), \theta)\, h(x) \tag{2.0.2}$$

3) **Basu's Theorem:** If T(X) is complete and sufficient, and S(X) is ancillary, then S(X) and T(X) are independent for all θ.
$\implies$ complete sufficient statistic is independent of any ancillary statistic.

Given PDF of the distribution as,

$$f(x, \theta) = \begin{cases} e^{-(x-2\theta)}, & x > 2\theta \\ 0, & otherwise \end{cases} \tag{2.0.3}$$

Then CDF of the distribution given is,

$$F(x, \theta) = \int_{-\infty}^{x} f(x, \theta)\, dx \tag{2.0.4}$$

Using (2.0.3) in (2.0.4),

$$F(x, \theta) = \begin{cases} 0, & x < 2\theta \\ 1 - e^{-(x-2\theta)}, & x > 2\theta \end{cases} \tag{2.0.5}$$

As given $X_{(1)} = min\{X_1, X_2, \cdots, X_n\}$,
Let us find CDF of $X_{(1)}$,

$$\begin{aligned} F_{X_{(1)}}(x, \theta) &= \Pr(X_{(1)} \leq x) \\ &= \Pr(at\, least\, one\, of\, X_1, X_2, \cdots, X_n \leq x) \\ &= 1 - \Pr(X_{(1)} > x) \\ &= 1 - \Pr(X_1 > x, X_2 > x, \cdots, X_n > 1) \\ &= 1 - \Pr(X_1 > x) \cdots \Pr(X_n > x) \\ &= 1 - (1 - F(x, \theta))^n \end{aligned} \tag{2.0.6}$$

Using (2.0.5) in (2.0.6),

$$F_{X_{(1)}}(x, \theta) = \begin{cases} 0, & x < 2\theta \\ 1 - e^{-n(x-2\theta)}, & x > 2\theta \end{cases} \tag{2.0.7}$$

Using CDF of $X_{(1)}$ to find PDF of $X_{(1)}$,

$$f_{X_{(1)}}(x, \theta) = \frac{d}{dx} F_{X_{(1)}}(x, \theta) \tag{2.0.8}$$

Using (2.0.7) in (2.0.8), PDF of $X_{(1)}$ is

$$f_{X_{(1)}}(x, \theta) = \begin{cases} ne^{-n(x-2\theta)}, & x > 2\theta \\ 0, & otherwise \end{cases} \tag{2.0.9}$$

$X_{(1)}, \cdots, X_{(n)}$ are ordered statistics of $X_1, \cdots, X_n$.
Where $X_{(k)}$ is kth order statistic of $X_1, \cdots, X_n$.

$$\implies \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_{(i)} \tag{2.0.10}$$

Some results that we use in future:

1) Ordered statistics are complete and sufficient statistic of X.
Proof:

Let $E[g(X_{(1)})] = 0$,

$$\implies \int_{-\infty}^{\infty} g(x) f_{X_{(1)}}(x) dx = 0 \tag{2.0.11}$$

$$\int_{2\theta}^{\infty} g(x) n e^{-n(x-2\theta)} dx = 0 \tag{2.0.12}$$

$$\int_{2\theta}^{\infty} g(x) e^{-n(x-2\theta)} dx = 0 \tag{2.0.13}$$

differentiating w.r.t θ on both sides in (2.0.13),

$$\frac{d}{dx}\int_{2\theta}^{\infty} g(x) e^{-n(x-2\theta)} dx = 0$$

$$\frac{d}{dx}\left(\int_{2\theta}^{\infty} g(x) e^{-nx} dx\right) e^{2n\theta} = 0$$

$$2ne^{2n\theta}\int_{2\theta}^{\infty} g(x) e^{-nx} dx + e^{2n\theta}(2) g(2\theta) e^{-2n\theta} = 0$$

$$2n(0) + 2g(2\theta) = 0 \implies g(2\theta) = 0$$

$\implies$ $X_{(1)}$ is complete statistics.
Using (2.0.10) in (2.0.16)

$$f_X(x,\theta) = f(x_1,\theta) f(x_2,\theta)\cdots f(x_n,\theta) \tag{2.0.14}$$

$$= e^{-(x_1-2\theta)} e^{-(x_2-2\theta)} \cdots e^{-(x_n-2\theta)} \tag{2.0.15}$$

$$= e^{-\left(\sum_{i=1}^{n} x_i - 2n\theta\right)} = e^{-\left(\sum_{i=1}^{n} x_{(i)} - 2n\theta\right)} \tag{2.0.16}$$

$$= \underbrace{\prod_{j=1}^{n} e^{-(x_{(j)}-2\theta)}}_{\text{g}} \times \underbrace{(1)}_{\text{h}} \tag{2.0.17}$$

$\therefore$ Ordered statistics of X are sufficient statistics for θ.
$\therefore$ $X_{(1)}$ is complete and sufficient statistics of θ.

2) $X_1 - X_2$ is ancillary of θ.

Proof: Let U=$X_1 - X_2$ then,

$$F_U(x) = \Pr(X_1 - X_2 < x)$$
$$= \int_{-\infty}^{\infty} \Pr(X_1 < x+k) \Pr(X_2 > k)\, dk$$
$$= \int_{2\theta}^{\infty} \left(1 - e^{-(x+k-2\theta)}\right)\left(e^{-(k-2\theta)}\right) dk$$
$$= \int_{2\theta}^{\infty} e^{-(k-2\theta)} - e^{-(2k+x-2\theta)} dk$$
$$= \left[\frac{e^{-(k-2\theta)}}{-1} - \frac{e^{-(2k+x-2\theta)}}{-2}\right]_{2\theta}^{\infty}$$
$$= (0-0) - \left(-1 + \frac{e^{-x}}{2}\right)$$

$$F_U(x) = 1 - \frac{e^{-x}}{2} \tag{2.0.18}$$

$$\implies f_U(x) = \frac{d}{dx} F_U(x) \tag{2.0.19}$$

$$= \frac{e^{-x}}{2} \tag{2.0.20}$$

$\therefore$ $U = X_1 - X_2$ is an ancillary statistic of θ.
Let U be a random variable such that $U = X_1 - X_2$.

$$E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|X_1 - X_2 = 2\right]$$
$$= E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|U = 2\right] \tag{2.0.21}$$

As $X_1, X_2, \cdots, X_n$ are independent and from Basu's theorem $X_{(1)}$ and U are also independent.
As we know that if X and Y are independent then $E[X|Y] = E[X]$. Using this in (2.0.21)

$$E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|U = 2\right] = E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)\right] \tag{2.0.22}$$

$$= \frac{1}{\theta}\left(E[X_{(1)}] - \frac{1}{n}\right) \tag{2.0.23}$$

We have to find expectation of $X_{(1)}$,

$$E[X_{(1)}] = \int_{-\infty}^{\infty} x f_{X_{(1)}}(x,\theta) dx \tag{2.0.24}$$

Using (2.0.9) in (2.0.24).

$$E\left[X_{(1)}\right] = \int_{2\theta}^{\infty} nx\, e^{-(x-2\theta)n} dx$$
$$= e^{2n\theta} \int_{2\theta}^{\infty} nx\, e^{-nx} dx \tag{2.0.25}$$

Using integration by parts in (2.0.25),

$$E\left[X_{(1)}\right] = e^{2n\theta} \int_{2\theta}^{\infty} nx\, e^{-nx} dx$$
$$= e^{2n\theta} \left(\left[nx \frac{e^{-nx}}{-n} \right]_{2\theta}^{\infty} - \int_{2\theta}^{\infty} n \frac{e^{-nx}}{-n} dx \right)$$
$$= e^{2n\theta} \left(\left[nx \frac{e^{-nx}}{-n} \right]_{2\theta}^{\infty} + \left[\frac{e^{-nx}}{-n} \right]_{2\theta}^{\infty} \right)$$
$$= e^{2n\theta} \left(2\theta e^{-2n\theta} + \frac{e^{-2n\theta}}{n} \right)$$
$$E\left[X_{(1)}\right] = 2\theta + \frac{1}{n} \tag{2.0.26}$$

Use (2.0.26) in (2.0.23),

$$E\left[\frac{1}{\theta} \left(X_{(1)} - \frac{1}{n} \right) | U = 2 \right] = \frac{1}{\theta} \left(E\left[X_{(1)}\right] - \frac{1}{n} \right)$$
$$= \frac{1}{\theta} \left(2\theta + \frac{1}{n} - \frac{1}{n} \right)$$
$$E\left[\frac{1}{\theta} \left(X_{(1)} - \frac{1}{n} \right) | U = 2 \right] = 2 \tag{2.0.27}$$

Using (2.0.27) in (2.0.21),

$$\therefore E\left[\frac{1}{\theta} \left(X_{(1)} - \frac{1}{n} \right) | X_1 - X_2 = 2 \right] = 2$$