# BASU'S THEOREM

Adhvik Murarisetty

AI20BTECH11015

# DEFINITIONS

## Completeness of a Statistics

Consider a random variable $X \sim f(x, \theta)$ parametrized by $\theta$.
Say T is a statistic; that is, the composition of a measurable function with a random sample $X_1, \cdots, X_n$. The statistic T is said to be complete for the distribution of X if, for every measurable function g if

$$E[g(T)] = 0 \ \forall\theta \implies \Pr(g(T) = 0) = 1 \ \forall\theta \tag{1}$$

## Ancillary statistic

Let X is a R.V with PDF as f(x,$\theta$), then S(X) is an ancillary statistic if its distribution does not depends on $\theta$.

# DEFINITIONS Contd..

## Sufficient statistic

A statistic t = T(X) is sufficient for underlying parameter $\theta$ precisely if the conditional probability distribution of the data X, given the statistic t = T(X), does not depend on the parameter $\theta$.

$\implies f(X|t = T(X))$ does not depends on $\theta$

## Fisher–Neyman factorization theorem

Fisher's factorization theorem provides a convenient characterization of a sufficient statistic. If the probability density function is f(x), then T is sufficient for $\theta$ if and only if non negative functions g and h can be found such that

$$f(x) = h(x)\, g(T(X), \theta) \tag{2}$$

# THEOREM

**Basu's theorem**

Let$X \sim f\,(X|\theta)$ and T=T(X) be a complete and sufficient statistics (for $\theta$) where $T \sim h\,(t|\theta)$, and V(X) is ancillary, then V(X) and T(X) are independent for all $\theta$.

**Proof:** Let $V \sim k\,(V|\theta) = k\,(V)$ because V is ancillary.
Since T is complete, if $E[g\,(T)] = 0\ \forall\theta$, then $g\,(T) = 0\ \forall\theta$.
For a fixed V such that V=v, define a function,

$$\psi\,(T|V = v) = k\,(V = v) - k\,(V = v|\,T) \tag{3}$$

Then,

$$\begin{aligned} E[\psi\,(T|V = v)] &= E[k\,(V = v) - k\,(V = v|\,T)] \qquad (4) \\ &= E[k\,(V = v)] - E[k\,(V = v|\,T)] \qquad (5) \end{aligned}$$

# Proof Contd..

$$E[\psi(T|V=v)] = E[k(V=v)] - E[k(V=v|T)] \tag{6}$$

$$= k(V=v) - \int_{-\infty}^{\infty} k(V=v|T)\; h(t|\theta)\; dt \tag{7}$$

$$= k(V=v) - \int_{-\infty}^{\infty} \frac{k(V=v, T=t)}{h(t|\theta)}\; h(t|\theta)\; dt \tag{8}$$

$$= k(V=v) - \int_{-\infty}^{\infty} k(V=v, T=t)\, dt \tag{9}$$

$$= k(V=v) - k(V=v) = 0 \tag{10}$$

$\implies \psi(T|V=v) = 0 \implies k(V=v|T) = k(V=v)$.
$\therefore$ V(X) and T(X) are independent for all $\theta$.

# QUESTION

## 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

$$f(x, \theta) = \begin{cases} e^{-(x-2\theta)}, & x > 2\theta \\ 0, & otherwise \end{cases} \tag{11}$$

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

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

# SOLUTION

Given PDF of the distribution as,

$$f(x,\theta) = \begin{cases} e^{-(x-2\theta)}, & x > 2\theta \\ 0, & otherwise \end{cases} \tag{12}$$

Then CDF of the distribution given is,

$$F(x,\theta) = \int_{-\infty}^{x} f(x,\theta)\,dx \tag{13}$$

Using (12) in (13),

$$F(x,\theta) = \begin{cases} 0, & x < 2\theta \\ 1 - e^{-(x-2\theta)}, & x > 2\theta \end{cases} \tag{14}$$

# Solution Contd..

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

$$\begin{aligned} F_{X_{(1)}}(x,\theta) &= \Pr\left(X_{(1)} \le x\right) \\ &= \Pr\left(at\ least\ one\ of\ X_1, X_2, \cdots, X_n \le x\right) \\ &= 1 - \Pr\left(X_{(1)} > x\right) \\ &= 1 - \Pr\left(X_1 > x, X_2 > x, \cdots, X_n > x\right) \\ &= 1 - \Pr\left(X_1 > x\right) \cdots \Pr\left(X_n > x\right) \\ &= 1 - (1 - F(x,\theta))^n \end{aligned} \tag{15}$$

Using (14) in (15),

$$F_{X_{(1)}}(x,\theta) = \begin{cases} 0, & x < 2\theta \\ 1 - e^{-n(x-2\theta)}, & x > 2\theta \end{cases} \tag{16}$$

# Solution Contd..

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

$$f_{X_{(1)}}(x, \theta) = \frac{d}{dx} F_{X_{(1)}}(x, \theta) \tag{17}$$

Using (16) in (17), PDF of $X_{(1)}$ is

$$f_{X_{(1)}}(x, \theta) = \begin{cases} ne^{-n(x-2\theta)}, & x > 2\theta \\ 0, & otherwise \end{cases} \tag{18}$$

# Solution Contd..

Some results that we use in future:
$X_{(1)}$ is complete and sufficient statistic of X.
**Proof:**
Let $E[g(X_{(1)})] = 0$,

$$\implies \int_{-\infty}^{\infty} g(x) f_{X_{(1)}}(x) dx = 0 \tag{19}$$

$$\int_{2\theta}^{\infty} g(x) n e^{-n(x-2\theta)} dx = 0 \tag{20}$$

$$\int_{2\theta}^{\infty} g(x) e^{-n(x-2\theta)} dx = 0 \tag{21}$$

## Solution Contd..

differentiating w.r.t $\theta$ on both sides in (21),

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

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

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

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

$\implies$ $X_{(1)}$ is complete statistics.

# Solution Contd..

Firstly We define a function,

$$I(a,b) = \begin{cases} 1 & b > a \\ 0 & otherwise \end{cases} \tag{22}$$

Using (22) in (25)

$$f_X(x,\theta) = f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta) \tag{23}$$

$$= e^{-(x_1-2\theta)}e^{-(x_2-2\theta)}\cdots e^{-(x_n-2\theta)} \tag{24}$$

$$= e^{-\left(\sum_{i=1}^{n} x_i - 2n\theta\right)} I(2\theta, X_{(1)}) \tag{25}$$

$$= \underbrace{I\left(2\theta, X_{(1)}\right)}_{\text{g}} \times \underbrace{e^{-\left(\sum_{i=1}^{n} x_i - 2n\theta\right)}}_{\text{h}} \tag{26}$$

$\therefore$ Ordered statistics of X are sufficient statistics for $\theta$.

$\therefore$ $X_{(1)}$ is complete and sufficient statistics of $\theta$.

# Solution Contd..

$X_1 - X_2$ is ancillary of $\theta$.

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

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

## Solution Contd..

$$
\begin{aligned}
F_U(x) &= (0-0) - \left(-1 + \frac{e^{-x}}{2}\right) \\
F_U(x) &= 1 - \frac{e^{-x}}{2} \qquad (27) \\
\implies f_U(x) &= \frac{d}{dx} F_U(x) \qquad (28) \\
&= \frac{e^{-x}}{2} \qquad (29)
\end{aligned}
$$

$\therefore\ U = X_1 - X_2$ is an ancillary statistic of $\theta$.

# Solution Contd..

Let U be a random variable such that $U = X_1 - X_2$.

$$E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|X_1 - X_2 = 2\right] = E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|U = 2\right] \tag{30}$$

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 (30)

$$\begin{aligned} E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)|U = 2\right] &= E\left[\frac{1}{\theta}\left(X_{(1)} - \frac{1}{n}\right)\right] \\ &= \frac{1}{\theta}\left(E\left[X_{(1)}\right] - \frac{1}{n}\right) \end{aligned} \tag{31}$$

# Solution Contd..

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

$$E\left[X_{(1)}\right] = \int_{-\infty}^{\infty} x f_{X_{(1)}}(x,\theta)dx \tag{32}$$

Using (18) in (32).

$$\begin{aligned} E\left[X_{(1)}\right] &= \int_{2\theta}^{\infty} nx\, e^{-(x-2\theta)n} dx \\ &= e^{2n\theta} \int_{2\theta}^{\infty} nx\, e^{-nx} dx \end{aligned} \tag{33}$$

# Solution Contd..

Using integration by parts in (33),

$$\begin{aligned} E\left[X_{(1)}\right] &= e^{2n\theta}\int_{2\theta}^{\infty} nx\, e^{-nx} dx \\ &= e^{2n\theta}\left(\left[nx\frac{e^{-nx}}{-n}\right]_{2\theta}^{\infty} - \int_{2\theta}^{\infty} n\frac{e^{-nx}}{-n}dx\right) \\ &= e^{2n\theta}\left(\left[nx\frac{e^{-nx}}{-n}\right]_{2\theta}^{\infty} + \left[\frac{e^{-nx}}{-n}\right]_{2\theta}^{\infty}\right) \\ &= e^{2n\theta}\left(2\theta e^{-2n\theta} + \frac{e^{-2n\theta}}{n}\right) \\ E\left[X_{(1)}\right] &= 2\theta + \frac{1}{n} \end{aligned} \tag{34}$$

## Solution Contd..

Use (34) in (31),

$$E\left[\frac{1}{\theta}\left(X_{(1)}-\frac{1}{n}\right)|U=2\right]=\frac{1}{\theta}\left(E\left[X_{(1)}\right]-\frac{1}{n}\right)$$
$$=\frac{1}{\theta}\left(2\theta+\frac{1}{n}-\frac{1}{n}\right)$$
$$E\left[\frac{1}{\theta}\left(X_{(1)}-\frac{1}{n}\right)|U=2\right]=2 \tag{35}$$

Using (35) in (30),

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

# THANK YOU