Ans The formal definition of sufficiency says that :-

$P(X = x \mid T(x) = t)$ does not depend on θ.

$$P(X=x \mid T(x)=t) = \frac{P(X_1=x_1, X_2=x_2 \ldots X_n=x_n \cap T(x)=t)}{P(T(x)=t)}$$

In the numerator, the intersection is not ϕ

and probability is non zero

when $T = \sum_{i=1}^{n} x_i$

$$= \frac{P(X_1=x_1, \ldots, X_n=x_n)}{P(T(x)=t)}$$

$$= \frac{P(X_1=x_1) \cdot P(X_2=x_2) \ldots \cdot P(X_n=x_n)}{P(T(x)=\sum_{i=1}^{n} x_i)}$$

$$= \frac{\prod_{i=1}^{n} P(X_i=x_i)}{P(T(x)=\sum_{i=1}^{n} x_i)} \quad \text{— (1)}$$

Calculation of denominator :-

Since it is given to us that X_i follows a geometric distribution we know that

$\sum X_i$ will follow a negative binomial distribution

$$\therefore P\left(T(x) = \sum_{i=1}^{n} x_i\right) = \binom{\sum x_i - 1}{n-1} p^n (1-p)^{\sum x_i - n}$$

Proof for why negative binomial :-

$X \overset{iid}{\sim} Geometric(p)$

→ Success with probability p after $x-1$ failures

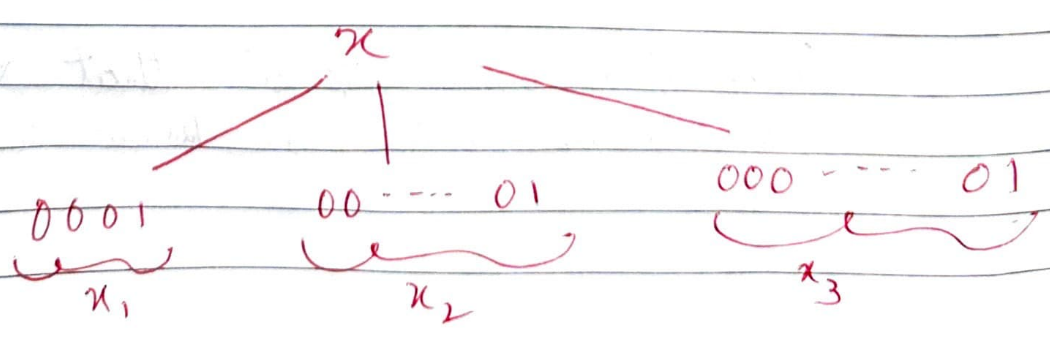
$$P(X = x) = (1-p)^{x-1} p$$

let $\sum x_i = t$

→ modeling t is similar to modeling x (i.e number of failures) in a negative binomial distribution.

$$P(X=x) = \binom{x-n-1}{n-1} p^n (1-p)^{x^*}$$ [negative binomial distribution]

Visualising an example :-



The overall count of trials needed is the sum of count of trials needed to reach n successes independently from a geometric distribution.

Let X' be negative binomial random variable and n is the fixed number of success needed.

$$\therefore \quad t = \sum x_i$$

$$x' + n = t$$

(number of failures) + (number of successes) = total trials

$$\therefore \quad x' = t - n$$
$$= \sum x_i - n$$

$$\therefore \quad P\left(T(x) = \sum_{i=1}^{n} x_i\right) = \binom{n + (\sum x_i - n) - 1}{n-1} p^n (1-p)^{\sum x_i - n}$$

$$= \binom{\sum x_i - 1}{n-1} p^n (1-p)^{\sum x_i - n}$$

Continuing :-

substituting in (1) :-

$$= \frac{\prod p (1-p)^{x_i - 1}}{\left(\sum_{i=1}^{n} x_i - 1\right) C (n-1) \cdot p^n (1-p)^{\sum x_i - n}}$$

$$= \frac{\cancel{p^n} \cancel{(1-p)^{\sum x_i - n}}}{\binom{\sum_{i=1}^{n} x_i - 1}{n-1} \cancel{p^n} \cancel{(1-p)^{\sum x_i - n}}}$$

$$= \frac{1}{\binom{\sum_{i=1}^{n} x_i - 1}{n-1}}$$

The above expression is independent of $\theta_{[p]}$ it doesn't have any term of p.

Hence, we can say that T is a sufficient statistic ~~for~~ of θ.

2. Let $X_1, X_2, ..., X_n$ be i.i.d. with the following pdf

$$f_\theta(x_i) = \frac{2x_i}{\theta^2}, \quad 0 < x_i < \theta$$

Answer the following questions

(a) (2 points) Apart from the data itself, find sufficient statistic(s) for θ.

$$f_\theta(x_i) = \begin{cases} \frac{2x_i}{\theta^2} & 0 < x_i < \theta \\ 0 & otherwise \end{cases}$$

Let, $$I_{(0,\theta)}(x_i) = \begin{cases} 1 & 0 < x_i < \theta \\ 0 & otherwise \end{cases}$$

Therefore, $L(\theta) = \prod_{i=1}^{n} f_\theta(x_i)$

$$= \prod_{i=1}^{n} \frac{2x_i}{\theta^2} I_{(0,\theta)}(X_i)$$

$$= \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^{n} x_i I_{(0,\theta)}(X_i)$$

Let, $0 < X_{(1)} < X_{(2)} <... < X_{(n)} < \theta$

($L(\theta) = 0$, if for some x_i, the constraint: $0 < x_i < \theta$ does not hold)

Therefore, $L(\theta) = \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^{n} x_i I_{(0,\theta)}(X_i)$

$$= \left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}(X_{(n)}) \prod_{i=1}^{n} x_i$$

$$= g_\theta(T(X))h(x)$$

where, $g_\theta(T(X)) = \left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}(X_{(n)})$ and $h(x) = \prod_{i=1}^{n} x_i$

Thus, by the factorization theorem, $T(X) = X_{(n)}$ is a sufficient statistic for θ.

(b) (2 points) Find minimal sufficient statistic(s) for θ.

To get MSS (using Lehman-Scheffe' Theorem)

$$\frac{f_\theta(X)}{f_\theta(Y)} = \frac{\prod_{i=1}^{n} f_\theta(x_i)}{\prod_{i=1}^{n} f_\theta(y_i)} = \frac{\left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}\left(X_{(n)}\right)\prod_{i=1}^{n} x_i}{\left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}\left(Y_{(n)}\right)\prod_{i=1}^{n} y_i} = \frac{I_{(0,\theta)}\left(X_{(n)}\right)\prod_{i=1}^{n} x_i}{I_{(0,\theta)}\left(Y_{(n)}\right)\prod_{i=1}^{n} y_i}$$

Now, this ratio does not depend on θ iff $X_{(n)} = Y_{(n)}$

Thus, $T(X) = X_{(n)}$ is MSS for θ.

(c) (3 points) Is $T(X) = X_{(n)}$ a complete statistic?

The density of $T(X) = X_{(n)}$ is given by

$$f_T(t) = nf_X(t)\left[F_X(t)\right]^{n-1}$$

$$\Rightarrow f_T(t) = n\frac{2t}{\theta^2}\left[F_X(t)\right]^{n-1}$$

Now, let's find out the CDF of X.

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

$$= \int_{-\infty}^{0} f_X(u)du + \int_{0}^{x} f_X(u)du$$

$$= 0 + \int_{0}^{x} \frac{2u}{\theta^2}du \qquad \left\{x < \theta \; as \; X_{(n)} < \theta\right\}$$

$$= \frac{x^2}{\theta^2}$$

Now,

$$f_T(t) = n\frac{2t}{\theta^2}\left[F_X(t)\right]^{n-1}$$

$$\Rightarrow f_T(t) = n\frac{2t}{\theta^2}\left[\frac{t^2}{\theta^2}\right]^{n-1}$$

$$\Rightarrow f_T(t) = n\frac{2t^{2n-1}}{\theta^{2n}}$$

$$\Rightarrow f_T(t) = 2n \cdot t^{2n-1} \cdot \theta^{-2n}$$

To check $T(X) = X_{(n)}$ is a complete statistic or not, we will use the definition of completeness.

Thus, we have to evaluate

$$E_\theta[g(t)] = 0 \qquad \forall\, \theta$$

Now, $T(X) = X_{(n)} \Rightarrow T(X) \sim f_\theta(x)$ $\qquad \left\{X_{(n)} \sim f_\theta(x)\right\}$

$$\Rightarrow E_\theta[g(t)] = 0 \qquad \forall\, \theta$$

$$\Rightarrow \int_0^\theta g(t) f_\theta(t) dt = 0 \qquad \forall\, \theta$$

$$\Rightarrow \int_0^\theta g(t)\left(2n \cdot t^{2n-1} \cdot \theta^{-2n}\right) dt = 0 \qquad \forall\, \theta$$

$$\Rightarrow \theta^{-2n} \int_0^\theta g(t)\left(2n \cdot t^{2n-1}\right) dt = 0 \qquad \forall\, \theta$$

$$\Rightarrow \int_0^\theta g(t)\left(2n \cdot t^{2n-1}\right) dt = 0 \qquad \forall\, \theta$$

$$\left\{\theta^{-2n} \neq 0 \; for \; \theta > 0\right\}$$

Differentiating both sides w.r.t. θ

$$\Rightarrow \frac{d}{d\theta}\left(\int_0^\theta g(t)\left(2n \cdot t^{2n-1}\right) dt\right) = 0 \qquad \forall\, \theta$$

$$\Rightarrow g(\theta)\left(2n \cdot \theta^{2n-1}\right) = 0 \qquad \forall\, \theta$$

$$\Rightarrow g(\theta) = 0 \qquad \forall\, \theta$$

$$\left\{Because, \; 2n \cdot \theta^{2n-1} \neq 0 \; for \; \theta > 0\right\}$$

Therefore, $g(\theta) = 0$ holds for every $\theta > 0$.

Hence, T is a complete statistic.

Q-3 Given : $Y_1 \ldots Y_n$ be independent random variables
$N(\beta x_i, \sigma^2)$

$\theta = (\beta, \sigma^2)$ are unknown and x_i's are known.

a) MME for β and σ^2

$\because \theta = (\beta, \sigma^2) = (\theta_1, \theta_2)$

$$E[Y_i] = \beta x_i = \frac{\Sigma Y_i}{n}$$

$$\Rightarrow \boxed{\hat{\beta} = \frac{\Sigma Y_i}{n x_i}}$$

$$E[Y_i^2] = Var[Y_i] + [E(Y_i)]^2$$

$$\Rightarrow \frac{\Sigma Y_i^2}{n} = \sigma^2 + (\beta x_i)^2$$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{\Sigma Y_i^2}{n} - \hat{\beta}^2 x_i^2}$$

b) MLE for β and σ^2

$$f_{(\beta,\sigma^2)}(Y_i) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-1}{2\sigma^2}(Y_i - \beta x_i)^2}$$

$$L(\theta) = \prod_{i=1}^{n} f_\theta(Y_i) \quad \text{where } \theta = (\beta, \sigma^2)$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\Sigma(Y_i - \beta x_i)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\Sigma(Y_i^2 + \beta^2 x_i^2 - 2\beta x_i Y_i)}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}(\Sigma Y_i^2 + \beta^2 \Sigma x_i^2 - 2\beta \Sigma x_i Y_i)}$$

Taking log of $L(\theta)$, we get

$$l(\theta) = \log(L(\theta))$$

$$= \frac{-n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\Sigma Y_i^2 + \beta^2 \Sigma x_i^2 - 2\beta \Sigma x_i Y_i)$$

For β:

Taking $\frac{\partial l(\theta)}{\partial \beta}$, we get

$$\frac{\partial l(\theta)}{\partial \beta} = 0 - \frac{1}{2\sigma^2}(0 + 2\beta \Sigma x_i^2 - 2\Sigma x_i Y_i)$$

for maximum, equating $\frac{\partial l(\theta)}{\partial \beta} = 0$

$$\Rightarrow \quad \frac{-\cancel{2}\beta \sum x_i^2}{\cancel{2}\sigma^2} + \frac{\cancel{2} \sum x_i y_i}{\cancel{2}\sigma^2} = 0$$

$$\Rightarrow \quad \beta \frac{\sum x_i^2}{\cancel{\sigma^2}} = \frac{\sum x_i y_i}{\cancel{\sigma^2}}$$

$$\Rightarrow \quad \boxed{\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}}$$

Taking $\frac{\partial^2 l(\theta)}{\partial \beta^2}$ to ensure maxima

$$\frac{\partial^2 l(\theta)}{\partial \beta^2} = -\frac{\sum x_i^2}{\sigma^2} < 0 \quad \because \frac{\sum x_i^2}{\sigma^2} > 0$$

Therefore maxima exists

for σ^2:

Taking $\frac{\partial l(\theta)}{\partial \sigma^2}$, we get

$$\frac{\partial l(\theta)}{\partial \sigma^2} = \frac{-n}{2 \times 2\pi\sigma^2} \times 2\pi + \frac{1}{2\sigma^4}\left(\Sigma y_i^2 + \hat{\beta}^2 \Sigma x_i^2 - 2\hat{\beta}\Sigma x_i y_i\right)$$

Equating $\frac{\partial l(\theta)}{\partial \sigma^2} = 0$

$$\Rightarrow \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \Sigma (y_i - \hat{\beta} x_i)^2 = 0$$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{1}{n} \Sigma (y_i - \hat{\beta} x_i)^2}$$

Taking $\frac{\partial^2 l(\theta)}{\partial \sigma^4}$ to ensure maxima

$$= \frac{+n}{2\sigma^4} - \frac{\cancel{2}}{\cancel{2}\sigma^6} \Sigma (y_i - \beta x_i)^2$$

$$= \frac{1}{\sigma^4}\left[\frac{n}{2} - \frac{\Sigma (y_i - \beta x_i)^2}{\sigma^2}\right]$$

$$= \frac{1}{\sigma^4}\left[\frac{n}{2} - n\right] = \frac{-n}{2\sigma^4} < 0$$

maxima exists

(c) unbiased estimator of β

Let $\hat{\beta}_{MLE} = \frac{\Sigma x_i Y_i}{\Sigma x_i^2}$ (from part (b))

$$\Rightarrow E[\hat{\beta}_{MLE}] = E\left[\frac{\Sigma x_i Y_i}{\Sigma x_i^2}\right]$$

$$= \frac{1}{\Sigma x_i^2} E[\Sigma x_i Y_i]$$

$$= \frac{1}{\Sigma x_i^2} E[x_1 Y_1 + x_2 Y_2 + \dots x_n Y_n]$$

$$= \frac{\Sigma x_i \beta x_i}{\Sigma x_i^2}$$

$$= \frac{\Sigma \beta x_i^2}{\Sigma x_i^2}$$

$$= \frac{\beta \Sigma x_i^2}{\Sigma x_i^2}$$

$$= \beta$$

$E[\hat{\beta}_{MLE}] - \beta = 0$

Hence, $\hat{\beta}_{MLE}$ is an unbiased estimator of β.

Date
DELTA Pg No.

(d) unbiased estimator for σ^2

$\because \hat{\sigma}^2_{MLE} = \frac{1}{n} \Sigma (\cancel{y} y_i - \hat{\beta} x_i)^2$

and $\hat{\beta}_{MLE} = \frac{\Sigma y_i x_i}{\Sigma x_i^2}$

$E[\hat{\sigma}^2_{MLE}] = E\left[\frac{1}{n} \Sigma (y_i - \hat{\beta} x_i)^2\right]$

$= \frac{1}{n} \Sigma \left(E(y_i^2) + E(\hat{\beta}^2 x_i^2) - 2x_i E[\hat{\beta} \cancel{y} y_i] \right)$

$\because E[y_i^2] = (E[y_i])^2 + Var[y_i]$ $\left[y_i \sim N(\beta x_i, \sigma^2) \right]$
$= (\beta x_i)^2 + \sigma^2$

$= \frac{1}{n} \Sigma \left[\sigma^2 + \beta^2 x_i^2 + x_i^2 E(\hat{\beta}^2) - 2x_i E\left(\frac{\Sigma y_i x_i y_i}{\Sigma x_i^2} \right) \right]$

$\because Var(\hat{\beta}) = Var\left(\frac{\Sigma y_i x_i}{\Sigma x_i^2} \right)$

$= \frac{\Sigma x_i^2 Var(y_i)}{(\Sigma x_i^2)^2}$

$= \frac{\sigma^2}{(\Sigma x_i^2)}$ $\because Var(y_i) = \sigma^2$

$\Rightarrow E[\hat{\beta}^2] = [E(\hat{\beta})]^2 + Var(\hat{\beta})$

$= \beta^2 + \frac{\sigma^2}{(\Sigma x_i^2)}$ $[\because E(\hat{\beta}) = \beta \text{ part (c)}]$

$$= \frac{1}{n} \Sigma \left[\sigma^2 + \beta^2 x_i^2 + x_i^2 \left(\beta^2 + \frac{\sigma^2}{\Sigma x_i^2} \right) - \frac{2x_i}{\Sigma x_i^2} \Sigma E(Y_i^2 x_i) \right]$$

$$= \frac{1}{n} \Sigma \left[\sigma^2 + \beta^2 x_i^2 + x_i^2 \left(\beta^2 + \frac{\sigma^2}{\Sigma x_i^2} \right) - \frac{2x_i \Sigma x_i}{\Sigma x_i^2} E(Y_i^2) \right]$$

$$= \frac{1}{n} \Sigma \left[\sigma^2 + \beta^2 x_i^2 + x_i^2 \left(\beta^2 + \frac{\sigma^2}{\Sigma x_i^2} \right) - \frac{2x_i \Sigma x_i}{\Sigma x_i^2} (\sigma^2 + \beta^2 x_i^2) \right.$$

$$= \frac{1}{n} \left(n\sigma^2 + \cancel{\beta^2 \Sigma x_i^2} + \cancel{\beta^2 \Sigma x_i^2} + \frac{\sigma^2 \cancel{\Sigma x_i^2}}{\cancel{\Sigma x_i^2}} - 2\frac{\Sigma x_i^2}{\Sigma x_i^2} \sigma^2 - \cancel{2\beta^2 \frac{(\Sigma x_i^2}{\Sigma x_i^2}} \right.$$

$$= \frac{1}{n} \left(n\sigma^2 - \sigma^2 \right)$$

$$= \frac{(n-1)}{n} \sigma^2$$

$$E\left[\frac{n}{(n-1)} \hat{\sigma}^2_{MLE} \right] = \sigma^2$$

$\Rightarrow$ $\frac{n}{n-1} \hat{\sigma}^2_{MLE}$ is an unbiased estimator of σ^2

Date
DELTA Pg No

(i) For a distribution to belong to exponential family of distributions, the likelihood fn should be of the form

$$L(\theta) = \prod_{i=1} \{h(x_i) [c(\theta)]^n \exp[w_1(\theta)\sum_{i=1}^{n} t_1(x_i) + w_2(\theta)\sum_{i=1}^{n} t_2(x_i)$$

$$+ \dots w_k(\theta)\sum_{i=1}^{n} t_n(x_i)]$$

Given $f_\theta(Y_i) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-1}{2\sigma^2}(Y_i - \beta x_i)^2}$

$$L(\theta) = \prod_{i=1}^{n} f_\theta(Y_i)$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\Sigma(Y_i - \beta x_i)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}(\Sigma Y_i^2 + \beta^2 \Sigma x_i^2 - 2\beta \Sigma x_i Y_i)}$$

Comparing it to exponential family of dist. $L(\theta)$

$h(y_i) = 1$ $\quad c(\theta) = \frac{1}{(2\pi\sigma^2)^{1/2}}$

$w_1(\theta) = \frac{-1}{2\sigma^2}$ $\quad t_1(Y_i) = Y_i^2$

$w_2(\theta) = \frac{-\beta^2}{2\sigma^2}$ $\quad t_2(Y_i) = \Sigma x_i^2$

$w_3(\theta) = \frac{-2\beta}{2\sigma^2}$ $\quad t_3(Y_i) = x_i Y_i$

Hence, Yes it belongs to exponential family of distribution.

Date

DELTA Pg No.

(f) $Z_i = a + bY_i$ where $a \in (-\infty, \infty)$

$b \in (0, \infty)$

a is a location parameter and b is a scale parameter

$$Y_i = \frac{Z_i - a}{b} \Rightarrow \frac{dY}{dz} = \frac{1}{b}$$

$$f_0(Z_i) = \frac{1}{b} f\left(\frac{Z_i - a}{b}\right)$$

$$= \frac{1}{b}\left[\frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-1}{2\sigma^2}\left(\frac{Z_i - a}{b} - \beta x_i\right)^2}\right]$$

$$= \frac{1}{b}\left[\frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-1}{2\sigma^2 b^2}\left[Z_i - (a + \beta b x_i)\right]^2}\right]$$

$$= \frac{1}{(2\pi\sigma^2 b^2)^{1/2}} e^{\frac{-1}{2\sigma^2 b^2}\left[Z_i - (a + \beta b x_i)\right]^2}$$

$$\sim N(a + \beta b x_i, \sigma^2 b^2)$$

Yes, it belongs to the location scale family.

Q4.

ans) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Uniform}(0, \theta)$, $\theta > 0$

(a) The distribution of $R \sim \text{Beta}(n-1, 2)$

We need to see if R is an ancillary statistic

Now, a statistic is ancillary if its pdf does not depend on θ

So,

$$f_R(x) = \frac{x^{n-2}(1-x)^{2-1}}{B(n-1, 2)}$$

[pdf of Beta dist^n $\rightarrow \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$]

$$= \frac{x^{n-2}(1-x)}{B(n-1, 2)}$$

As we can see the distribution of R is independent of θ

So, R is an ancillary statistic.

(b) We need to find an unbiased estimator of θ

An estimator $\hat{\theta}$ is unbiased if

$$E[\hat{\theta}] = \theta \quad \text{for all } \theta$$

Initially we will start with MLE

$$\hat{\theta} = X_{(n)}$$

The pdf of $T = X_{(n)}$ is

$$f_T(t) = n f_X(t) [F_X(t)]^{n-1} \quad \text{(given)}$$

$$= n \cdot \frac{1}{\theta} \cdot \left(\frac{t}{\theta}\right)^{n-1}$$

CDF of Uniform distr^n

$$= \frac{n t^{n-1}}{\theta^n}$$

So,

$$E[\hat{\theta}] = \int_0^\theta t \, p_{\hat{\theta}}(t) \, dt$$

$$= \int_0^\theta t \frac{n t^{n-1}}{\theta^n} dt$$

$$= \frac{n}{\theta^n} \int_0^\theta t^n dx$$

$$= \left(\frac{n}{n+1}\right) \theta$$

Since this is not equal to θ, we see that $\hat{\theta}$ is biased

So, to obtain the unbiased estimator, we will rescale this as follows:

$$\hat{\theta}_{new} = \frac{n+1}{n} \cdot \hat{\theta}$$

$$= \frac{n+1}{n} \cdot X_{(n)}$$

Verification

Again,

$$E[\theta_{new}] = E\left[\frac{n+1}{n} \cdot X_{(n)}\right]$$

$$= \frac{n+1}{n} E[X_{(n)}]$$

$$= \left(\frac{n+1}{n}\right)\left(\frac{n}{n+1}\right)\theta$$ [$E(X_{(n)})$ calc before]

$$= \theta$$

$\therefore$,

$$\hat{\theta}_{unbiased} = \frac{n+1}{n} \cdot X_{(n)}$$

(c) We need to find MSE of $T = X_{(n)}$

We know,

$$MSE(T) = Var(T) + (Bias)^2$$

Bias for $T = X_{(n)}$ has been calculated in (b)

$$Bias(T) = E[T] - \theta$$
$$= \left(\frac{n}{n+1}\right)\theta - \theta$$
$$= \left(\frac{-1}{n+1}\right)\theta$$

Now, $Var(X_{(n)}) = E[X_{(n)}^2] - (E[X_{(n)}])^2$

$$E[X_{(n)}^2] = \int_0^\theta t^2 p_\theta(t)\, dt$$
$$= \int_0^\theta t^2 \cdot \frac{n t^{n-1}}{\theta^n}\, dt$$
$$= \frac{n}{\theta^n}\int_0^\theta t^{n+1}\, dt$$
$$= \left(\frac{n}{n+2}\right)\theta^2$$

So, $Var(a_{(n)}) = \left(\frac{n}{n+2}\right)\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2$

$$= \frac{n}{(n+1)^2(n+2)} \cdot \theta^2$$

So,

$$\text{mse} = \text{Var} + (\text{Bias})^2$$

$$= \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{\theta^2}{(n+1)^2}$$

$$= \frac{\theta^2}{(n+1)^2}\left(\frac{n}{n+2} + 1\right)$$

$$\boxed{\text{mse} = \frac{2 \cdot \theta^2}{(n+1)(n+2)}}$$

(d) In part (b) we had identified the unbiased estimator of θ as: $\frac{n+1}{n} \cdot X_{(n)}$

Now we need to check if above sufficient and complete

→ Sufficiency

Given $X_i \stackrel{iid}{\sim} \text{Uniform}(0, \theta) \quad \theta > 0$

So,

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x)$$

$$= \prod_{i=1}^{n} \frac{1}{\theta} I_{(0,\theta)}(X_{(i)})$$

$$= \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta)}(X_{(n)})$$

here, $h(x) = 1$

$$g_\theta(T(x)) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta)}(x_{(n)})$$

using factorization theorm,

$T = X_{(n)}$ is a sufficient statistic

→ Completeness

A statistic T is complete if $E_\theta[g(T)] = 0 \ \forall \theta$ implies $g(T) = 0$

where $g(T)$ is a func^n of T

So,

$$E_\theta[g(T)] = 0 \quad \forall \theta$$

$$\Rightarrow \int_0^\theta g(t) \cdot f_T(t) \cdot dt = 0$$

$$\Rightarrow \int_0^\theta g(t) \cdot \frac{n t^{n-1}}{\theta^n} dt = 0$$

[Recall pdf of $X_{(n)}$ $f_T(t) = \frac{n t^{n-1}}{\theta^n}$]

$$\Rightarrow \int_0^\theta g(t) \cdot t^{n-1} dt = 0$$

Differentiating both sides w.r.t t we get

$$g(t) = 0$$

Hence, T is complete

Now that we have proved sufficiency and completeness applying Lehmann - Scheffé theorem, the unbiased estimator calculated earlier is the UMVUE $\forall \theta$

$\therefore$, $\frac{n+1}{n} \cdot X_{(n)}$ is UMVUE for θ