Assignment 6

Gautham Bellamkonda - CS20BTECH11017

Download all latex-tikz codes from

https://github.com/GauthamBellamkonda/AI1103/tree/main/Assignment6

1 Problem

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n (≥ 2) from a distribution having the probability density function

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right) & x > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{1.0.1}$$

where $\theta \in (0, \infty)$. Let $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$ and $T = \sum_{i=1}^{n} X_i$. Then $E(X_{(1)}|T)$ equals ....

(A) $\frac{T}{n^2}$

(B) $\frac{T}{n}$

(C) $\frac{(n+1)T}{2n}$

(D) $\frac{(n+1)^2T}{4n^2}$

2 Prerequisites

Definition 1 (Complete Statistic)**.** *The statistic T is said to be complete for the distribution of X if, for every measurable function g, if*

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

Definition 2 (Sufficient Statistic)**.** *A statistic $t = T(X)$ is sufficient for underlying parameter θ precisely if the conditional probability distribution of the data X, given the statistic $t = T(X)$, does not depend on the parameter θ.*

Theorem 2.1 (Fischer-Neymann Factorisation theorem)**.** *If the probability density function is $f_\theta(x)$, then T is sufficient for θ if and only if nonnegative functions g and h can be found such that*

$$f_\theta(x) = h(x)g_\theta(T(x)) \tag{2.0.2}$$

Lemma 2.1.

$$T = \sum_{i=1}^{n} X_i \tag{2.0.3}$$

is a complete and sufficient statistic.

Proof.
(Sufficiency)

$$f_X(x_1, x_2, \ldots x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n) \tag{2.0.4}$$
$$= \frac{1}{\theta}\exp\left(-\frac{x_1}{\theta}\right)\cdots\frac{1}{\theta}\exp\left(-\frac{x_n}{\theta}\right) \tag{2.0.5}$$
$$= 1 \cdot \frac{1}{\theta^n}\exp\left(-\frac{T}{\theta}\right) \tag{2.0.6}$$
$$= h(X) \cdot g(T, \theta) \tag{2.0.7}$$

with

$$g(T, \theta) = \frac{1}{\theta^n}\exp\left(-\frac{T}{\theta}\right) \tag{2.0.8}$$
$$h(X) = 1 \tag{2.0.9}$$

Therefore, T is a sufficient statistic.
(Completeness)
X_i follow a gamma distribution, $X_i \sim \Gamma(1, \frac{1}{\theta})$. Hence, $T = \sum_{i=1}^{n} X_i$ follows a gamma distribution, $T \sim \Gamma(n, \frac{1}{\theta})$. Therefore, the expected value of a function g is

$$E(g(T)) = \int_0^\infty g(t)\frac{t^{n-1}e^{-\frac{t}{\theta}}}{\theta^n(n-1)!}dt \tag{2.0.10}$$
$$= \frac{1}{\theta^n(n-1)!}\int_0^\infty g(t)t^{n-1}e^{-\frac{t}{\theta}}dt \tag{2.0.11}$$
$$= \frac{1}{\theta^n(n-1)!}F\left(\frac{1}{\theta}\right) \tag{2.0.12}$$
$$= 0 \text{ for all } \theta > 0. \tag{2.0.13}$$

The integral in (2.0.11) can be interpreted as the Laplace transform $F(s) = \int_0^\infty f(t)e^{-st}dt$ of the function $f(t) = t^{n-1}g(t)$ evaluated at $\frac{1}{\theta}$. If this transform is 0 for all θ in $(0, \infty)$, then $f(t) = 0$ almost everywhere in $(0, \infty)$. Therefore, $g(t) = 0$ almost everywhere in

$(0, \infty)$ and hence,

$$\Pr(g(t) = 0) = 1 \tag{2.0.14}$$

Therefore, T is a complete statistic. □

Theorem 2.2 (Lehmann–Scheffé theorem)**.** *If T is a complete sufficient statistic for θ and*

$$E(g(T)) = \tau(\theta) \tag{2.0.15}$$

then $g(T)$ is the uniformly minimum-variance unbiased estimator (UMVUE) of $\tau(\theta)$.

3 Solution

By the law of total expectation,

$$E\left(E(X_{(1)}|T)\right) = E(X_{(1)}) \tag{3.0.1}$$

By Lehmann–Scheffé theorem, with

$$\theta = X_{(1)}, \tag{3.0.2}$$
$$\tau(x) = E(x), \tag{3.0.3}$$
$$g(T) = E(X_{(1)}|T). \tag{3.0.4}$$

it follows from (2.0.16) that $E(X_{(1)}|T)$ is the UMVUE of $E(X_{(1)})$.

$$\Pr(X_{(1)} > x) = \Pr(X_1 > x)\dots\Pr(X_n > x) \tag{3.0.5}$$
$$= (1 - F_{X_1}(x))\dots(1 - F_{X_n}(x)) \tag{3.0.6}$$
$$= (1 - F_{X_1}(x))^n \tag{3.0.7}$$
$$= \exp\left(-\frac{nx}{\theta}\right) \tag{3.0.8}$$
$$F_{X_{(1)}}(x) = 1 - \exp\left(-\frac{nx}{\theta}\right) \tag{3.0.9}$$
$$f_{X_{(1)}}(x) = \frac{n}{\theta}\exp\left(-\frac{nx}{\theta}\right) \tag{3.0.10}$$

Therefore, $X_{(1)}$ follows an exponential distribution with mean $\frac{\theta}{n}$.

$$E(X_{(1)}) = \frac{\theta}{n} \tag{3.0.11}$$

Note that,

$$E\left(\frac{T}{n^2}\right) = E\left(\frac{\sum_{i=1}^{n} X_i}{n^2}\right) \tag{3.0.12}$$
$$= \frac{E(\sum_{i=1}^{n} X_i)}{n^2} \tag{3.0.13}$$
$$= \sum_{i=1}^{n} \frac{E(X_i)}{n^2} \tag{3.0.14}$$
$$= \sum_{i=1}^{n} \frac{\theta}{n^2} \tag{3.0.15}$$
$$= \frac{\theta}{n} \tag{3.0.16}$$
$$= E(X_{(1)}) \tag{3.0.17}$$

Therefore, by Lehmann–Scheffé theorem, with

$$\theta = X_{(1)}, \tag{3.0.18}$$
$$\tau(x) = E(x), \tag{3.0.19}$$
$$g(T) = \frac{T}{n^2}, \tag{3.0.20}$$

it follows that $\frac{T}{n^2}$ is UMVUE of $E(X_{(1)})$.

Since there exists a unique UMVUE for $E(X_{(1)})$, it follows that

$$E(X_{(1)}|T) = \frac{T}{n^2} \tag{3.0.21}$$

Hence, option A is correct.

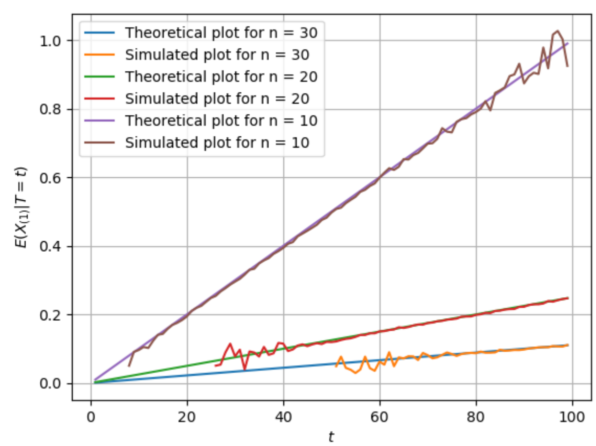

Fig. 4: Theory vs Simulated plot of $E(X_{(1)}|T)$