Assignment 6

Course name: PROBABILITY AND STOCHASTIC PROCESSES (AI5030)

UGC NET DEC 2017 Ques. : 109

KSHITIZ KUMAR | AI22MTECH02002 | ai22mtech02002@iith.ac.in

Question 1

Let $\{X_n\}$ be a sequence of independent random variables where the distribution of X_n is normal with mean μ and variance σ^2 for $n = 1, 2, ...$
Define-

$$\bar{X_n} = \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

and,

$$S_n = \frac{\sum_{i=1}^{n}(\frac{X_i}{i})}{\sum_{i=1}^{n}(\frac{1}{i})}$$

Which of the following is true?

1. $E(\bar{X_n}) = E(S_n)$ for sufficiently large $'n'$.
2. $Var(S_n) < Var(\bar{X_n})$ for sufficiently large $'n'$.
3. $\bar{X_n}$ is consistent for μ.
4. $\bar{X_n}$ is sufficient for μ.

Answer:

Analysis of Option 1 :

$$\begin{aligned} E(\bar{X_n}) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}(n \times \mu) \\ &= \mu \end{aligned} \tag{1}$$

$$
\begin{aligned}
E(S_n) &= E\left(\frac{\sum_{i=1}^{n}(\frac{X_i}{i})}{\sum_{i=1}^{n}(\frac{1}{i})}\right) \\
&= \frac{\sum_{i=1}^{n}(\frac{E(X_i)}{i})}{\sum_{i=1}^{n}(\frac{1}{i})} \\
&= \frac{[\sum_{i=1}^{n}(\frac{1}{i})]\times\mu}{\sum_{i=1}^{n}(\frac{1}{i})} \\
&= \mu
\end{aligned} \tag{2}
$$

Hence, option 1 is correct for all values of n.

Analysis of Option 2 :

$$
\begin{aligned}
Var(\bar{X}_n) &= \frac{1}{n^2}\left(\sum_{i=1}^{n} Var(X_i)\right) \\
&= \frac{1}{n^2}\left(n\times\sigma^2\right) \\
&= \frac{\sigma^2}{n} \\
\implies \lim_{n\to\infty} Var(\bar{X}_n) &= \lim_{n\to\infty}\left(\frac{\sigma^2}{n}\right) \\
&= 0
\end{aligned} \tag{3}
$$

Lemma 1. *Partial sum of harmonic series,* $H_n = \sum_{i=1}^{n}\left(\frac{1}{i}\right)$ *approximates to* $H_n \to \log n$ *as* $n \to \infty$.

Theorem 2.

$$
\lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{1}{i}\right) \to \infty \tag{4}
$$

Theorem 3.

$$
\lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{1}{i^2}\right) = \frac{\pi^2}{6} \tag{5}
$$

Now, let's find the variance of S_n.

$$
\therefore Var(S_n) = Var\left(\frac{\sum_{i=1}^{n}(\frac{X_i}{i})}{\sum_{i=1}^{n}(\frac{1}{i})}\right)
$$

Applying the properties of variance and taking the constants out, we get- (6)

$$
= \left[\frac{\sum_{i=1}^{n}\left(\frac{1}{i^2}\right)}{\left(\sum_{i=1}^{n}\left(\frac{1}{i}\right)\right)^2}\right]\times Var(X_i)
$$

Applying Theorem(3) to numerator and Lemma(1) to denominator in equation 6, we get-

$$\begin{aligned} \therefore Var(S_n) &= \left(\frac{\frac{\pi^2}{6}}{\log n}\right) \times \sigma^2 \\ \implies \lim_{n\to\infty} Var(S_n) &= \lim_{n\to\infty}\left(\left(\frac{\frac{\pi^2}{6}}{\log n}\right) \times \sigma^2\right) \\ &= 0 \end{aligned} \tag{7}$$

Hence, option 2 is incorrect for sufficiently large values of n.

Analysis of Option 3 :

Definition 4 (Consistent Estimator)**.** *An estimator T of parameter θ is said to be **consistent**, if it converges in probability to the true value of the parameter; i.e.,*

$$\lim_{n\to\infty} T = \theta$$

Now, we already know that Sample Mean($\bar{X}_n$) is an unbiased estimator of Population Mean(μ). So, we only need to show that the variance of Sample Mean in this case goes to zero as $n \to \infty$ for proving that Sample Mean is also a consistent estimator of Population Mean.
From equation 3, we can infer the same. Hence, option 3 is correct.

Analysis of Option 4 :

Definition 5 (Sufficient Statistic)**.** *A statistic $T(X_1, X_2, ..., X_n)$ is a **sufficient statistic**, if for each 't' the conditional distribution of $X_1, X_2, ..., X_n$, given $T = t$ and θ, doesn't depend on θ.*
*Since, finding conditional distribution is complex for large number of random variables, we use **Neyman's Factorization Theorem**.*

Theorem 6 (Neyman's Factorization Theorem)**.** *Suppose, we have random sample $X = (X_1, X_2, ..., X_n)$ from some function $f(x|\theta)$ and that function is the joint probability distribution of X.*
A statistic is sufficient if we can write joint pdf as -

$$f(x|\theta) = g(T(X)|\theta)h(x) \tag{8}$$

where,
θ : unknown parameter
and, pdf's exists for all values of x and θ.

Here, Joint PDFs can be expressed as product of marginals, becoz they are independent.

$$\begin{aligned}
\therefore JointPDF &= f(x_1;\mu,\sigma^2)\times f(x_2;\mu,\sigma^2)\times f(x_3;\mu,\sigma^2)\times\ldots\times .f(x_n;\mu,\sigma^2)\\
&= \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2}\times\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2}\times\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_3-\mu}{\sigma}\right)^2}\times\ldots\times\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}\\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\times e^{-\frac{1}{2\sigma^2}\left((x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+\ldots+(x_n-\mu)^2\right)}\\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\times e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n(x_i-\mu)^2\right)} \qquad (9)\\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\times e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2-2\mu\sum_{i=1}^n x_i+n\mu^2\right)}\\
&= \left(\left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\times e^{-\frac{1}{2\sigma^2}(n\mu^2)}\right)\times\left(e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2\right)}\right)\times\left(e^{\frac{\mu}{\sigma^2}\left(\sum_{i=1}^n x_i\right)}\right)\right)\times(1)
\end{aligned}$$

Comparing equations (8) and (9), we get-

$$\begin{aligned}
\therefore JointPDF &= \left(\left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\times e^{-\frac{1}{2\sigma^2}(n\mu^2)}\right)\times\left(e^{-\frac{1}{2\sigma^2}(t_1(x))}\right)\times\left(e^{\frac{\mu}{\sigma^2}(t_2(x))}\right)\right)\times(1)\\
&= \left(g_{(\mu,\sigma^2)}(t_1(x),t_2(x))\right)\times(h(x))
\end{aligned} \qquad (10)$$

where,

$$g_{(\mu,\sigma^2)}(t_1(x),t_2(x)) = \left(\left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\times e^{-\frac{1}{2\sigma^2}(n\mu^2)}\right)\times\left(e^{-\frac{1}{2\sigma^2}(t_1(x))}\right)\times\left(e^{\frac{\mu}{\sigma^2}(t_2(x))}\right)\right)$$

$$t_1(x) = \sum_{i=1}^n x_i^2$$

$$t_2(x) = \sum_{i=1}^n x_i$$

$$h(x) = 1$$

Thus, we have successfully factorized the Joint PDF of given distribution and proved that there is an existence of sufficient statistic for the given distribution. Now, let's move on to identifying those estimators and their corresponding parameters.

Here, $t_1(x)$ is a sufficient estimator of Population Mean (μ) in a normal population, and $t_2(x)$ is a sufficient estimator for population variance (σ^2).
Hence, option 4 is correct.

Hence, the options 1, 3 and 4 are correct.