

GATE: ST - 32.2023

EE22BTECH11039 - Pandrangi Aditya Sriram*

Question: Let $\{X_n\}_{n \geq 1}$ be a sequence of independent and identically distributed random variables each having a mean 4 and variance 9. If $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ for $n \geq 1$, then $\lim_{n \rightarrow \infty} E \left[\left(\frac{Y_n - 4}{\sqrt{n}} \right)^2 \right]$ (in integer) equals _____. (GATE ST 2023)

Solution:

- 1) **Theory:** For all X_i which are i.i.d's, mean $\mu = 4$ and variance $\sigma^2 = 9$,

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

The mean of a sum of i.i.d random variables is calculated as

$$E[Y_n] = E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] \quad (2)$$

$$= \frac{1}{n} \sum_{i=1}^n E[X_i] \quad (3)$$

$$= \frac{1}{n} (n\mu) \quad (4)$$

$$= \mu \quad (5)$$

The variance of a sum of i.i.d random variables is calculated as

$$\text{var}(Y_n) = E \left[\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \right] - \left(E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] \right)^2 \quad (6)$$

$$= \frac{1}{n^2} \left\{ E \left[\left(\sum_{i=1}^n X_i \right)^2 \right] - \left(E \left[\sum_{i=1}^n X_i \right] \right)^2 \right\} \quad (7)$$

But

$$E \left[\left(\sum_{i=1}^n X_i \right)^2 \right] = E \left[\sum_{i=1}^n \sum_{j=1}^n X_i X_j \right] \quad (8)$$

$$= \sum_{i=1}^n \sum_{j=1}^n E[X_i X_j] \quad (9)$$

and

$$\left(E \left[\sum_{i=1}^n X_i \right] \right)^2 = \left(\sum_{i=1}^n E[X_i] \right)^2 \quad (10)$$

$$= \sum_{i=1}^n \sum_{j=1}^n E[X_i] E[X_j] \quad (11)$$

Putting (9) and (11) in (7), and using the definition of covariance,

$$\text{var}(Y_n) = \frac{1}{n^2} \left\{ \sum_{i=1}^n \sum_{j=1}^n (E[X_i X_j] - E[X_i] E[X_j]) \right\} \quad (12)$$

$$= \frac{1}{n^2} \left\{ \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, X_j) \right\} \quad (13)$$

As all the variables are i.i.d's and are thus uncorrelated,

$$\text{cov}(X_i, X_j) = \begin{cases} 0 & \text{if } i \neq j \\ \text{var}(X_i) & \text{if } i = j \end{cases} \quad (14)$$

Putting (14) in (13),

$$\text{var}(Y_n) = \frac{1}{n^2} \left(\sum_{i=1}^n \text{cov}(X_i, X_i) \right) \quad (15)$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^n \text{var}(X_i) \right) \quad (16)$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^n \sigma^2 \right) \quad (17)$$

$$= \frac{\sigma^2}{n} \quad (18)$$

Consider the term $\left(\frac{Y_n - \mu}{\sqrt{n}} \right)^2$. Calculating its expectation,

$$E \left[\left(\frac{Y_n - \mu}{\sqrt{n}} \right)^2 \right] = \frac{1}{n} E[(Y_n - \mu)^2] \quad (19)$$

$$= \frac{1}{n} \text{var}(Y_n) \quad (20)$$

$$= \frac{\sigma^2}{n^2} \quad (21)$$

Substituting $\sigma^2 = 9$ and $\mu = 4$, we get

$$\lim_{n \rightarrow \infty} E \left[\left(\frac{Y_n - 4}{\sqrt{n}} \right)^2 \right] = \lim_{n \rightarrow \infty} \frac{9}{n^2} = 0 \quad (22)$$

- 2) **Simulation:** Any distribution with mean $\mu = 4$ and variance $\sigma^2 = 9$ can be used for the variable X_{ij} for all $i, j \in \mathbb{N}$; as shown in the Theory part, the limit is always zero regardless of the distribution. The most straightforward distribution that can be used for X_{ij} is:

$$p_{X_{ij}}(x) = \begin{cases} 0.5 & \text{if } x \in \{1, 7\} \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

This distribution has the following characteristics:

$$\mu = E[X_{ij}] = 0.5 \times 1 + 0.5 \times 7 = 4 \quad (24)$$

$$\sigma^2 = E[X_{ij}^2] - (E[X_{ij}])^2 \quad (25)$$

$$= (0.5 \times 1^2 + 0.5 \times 7^2) - 4^2 \quad (26)$$

$$= 9 \quad (27)$$

Alternatively, distributions such as $X_{ij} \sim \mathcal{N}(4, 9)$ could also be used. A matrix $X_{n \times m}$ is generated for all $i \leq n$ and $j \leq m$. Using this matrix, a set of m values for Y_j is generated as

$$Y_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad (28)$$

Now, the expression $\frac{(Y_j - 4)^2}{n}$ is calculated for all $j \leq m$ and their expectancy is calculated as follows:

$$E \left[\left(\frac{Y_n - 4}{\sqrt{n}} \right)^2 \right] = \frac{1}{m} \sum_{j=1}^m \frac{(Y_j - 4)^2}{n} \quad (29)$$

To calculate the limit $n \rightarrow \infty$, different values of n are taken, and the expected value is calculated (taking a fixed small value of m to reduce computational time) for each case. This output is plotted and is seen to be close to the curve $\frac{9}{n^2}$, as derived in (22). In both cases, we can observe the limit tends towards zero.

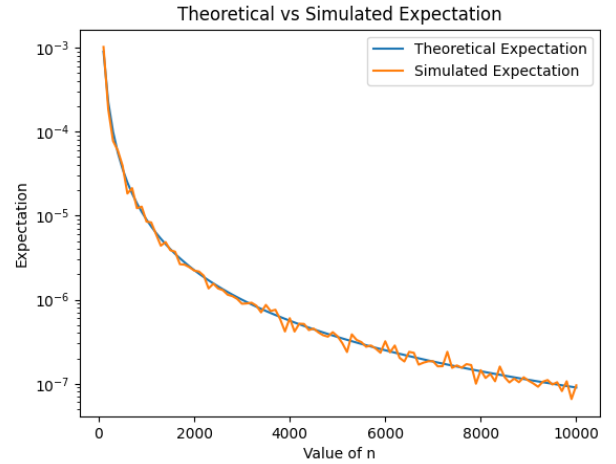


Fig. 2. Expectation vs n