

GATE-EC2023

Jay Vikrant EE22BTECH11025

Q65ST.2023: Let $\{X_n\}_{n \geq 1}$ be a sequence of independent and identically distributed random variables with mean 0 and variance 1, all of them defined on the same probability space. For $n=1,2,3,\dots$, let

$$Y_n = \frac{1}{n}(X_1X_2 + X_3X_4 + \dots + X_{2n-1}X_{2n}) \quad (1)$$

Then which of the following statements is/are true?

- (A) $\{\sqrt{n}Y_n\}_{n \geq 1}$ converges in distribution to a standard normal random variable.
- (B) $\{Y_n\}_{n \geq 1}$ converges in 2nd mean to 0.
- (C) $\{Y_n + \frac{1}{n}\}_{n \geq 1}$ converges in probability to 0.
- (D) $\{X_n\}_{n \geq 1}$ converges almost surely to 0.

Solution:

- 1) X_n converges in distribution to a standard normal random variable X , $X_n \xrightarrow{d} X$, then for all x ,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (2)$$

where, $X \sim \mathcal{N}(0, 1)$

- 2) X_n converges in p^{th} mean to X , then we have

$$\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0 \quad (3)$$

- 3) X_n converges in probability to X , $X_n \xrightarrow{p} X$, then for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| \geq \epsilon) = 0 \quad (4)$$

- 4) X_n converges almost surely to X , $X_n \xrightarrow{a.s} X$, then for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| < \epsilon) = 1 \quad (5)$$

$$E(X) = \int_{-\infty}^{\infty} xF_X x dx \quad (6)$$

$$\because X \text{ is positive-valued} \quad (7)$$

$$= \int_0^{\infty} xF_X x dx \quad (8)$$

$$\geq \int_{\epsilon^2}^{\infty} xF_X x dx \quad (9)$$

$$\text{for some } \epsilon^2 > 0 \quad (10)$$

$$\geq \int_{\epsilon^2}^{\infty} kF_X x dx \quad (11)$$

$$= \epsilon^2 \int_{\epsilon^2}^{\infty} F_X x dx \quad (12)$$

$$E(X) = \epsilon^2 \Pr(X \geq \epsilon^2) \quad (13)$$

This inequality is called Markov inequality
Now, we take

$$\sigma^2 = E(X^2) - [E(X)]^2 \quad (14)$$

$$= E(X^2) - \mu^2 \quad (15)$$

$$= E(X^2) - E(\mu^2) \quad (16)$$

$$\sigma^2 = E(X^2 - \mu^2) \quad (17)$$

using (13),

$$\sigma^2 = \epsilon^2 \Pr(X^2 - \mu^2 \geq \epsilon^2) \quad (18)$$

$$\frac{\sigma^2}{\epsilon^2} = \Pr(X^2 - \mu^2 \geq \epsilon^2) \quad (19)$$

$$\Pr(|X - \mu| \geq \epsilon) = \frac{\sigma^2}{\epsilon^2} \quad (20)$$

This above inequality is called Chebyshev's inequality

Now, as X_i is a sequence of independent and identically distributed random variables,

Let $Z_i = X_{2i-1}X_{2i}$.

$$p_{Z_i Z_j}(x) = p_{X_{2i-1}X_{2i}X_{2j-1}X_{2j}}(x) \quad (21)$$

$$= p_{X_{2i-1}}(x)p_{X_{2i}}(x)p_{X_{2j-1}}(x)p_{X_{2j}}(x) \quad (22)$$

$$= p_{X_{2i-1}X_{2i}}(x)p_{X_{2j-1}X_{2j}}(x) \quad (23)$$

$$= p_{Z_i}(x)p_{Z_j}(x) \quad (24)$$

Z_i is a independent R.V . Similarly,

$$F_{X_{2i-1}}(x) = F_{X_{2i}}(x) = F_{X_{2j-1}}(x) = F_{X_{2j}}(x) \quad (25)$$

$$F_{X_{2i-1}}(x)F_{X_{2i}}(x) = F_{X_{2j-1}}(x)F_{X_{2j}}(x) \quad (26)$$

$$F_{Z_i}(x) = F_{Z_j}(x) \quad (27)$$

Thus, Z_i is an identical distributed R.V.

Hence,

$$Z_n = \sum_{i=1}^n Z_i \quad (28)$$

is an i.i.d

$$E(Z_i) = E(X_{2i-1}X_{2i}) \quad (29)$$

$$= E(X_{2i-1})E(X_{2i}) = 0 \quad (30)$$

$$\text{Var}(Z_i) = E(Z_i^2) - [E(Z_i)]^2 \quad (31)$$

$$= E(Z_i^2) - 0 \quad (32)$$

$$= E(X_{2i-1}^2 X_{2i}^2) \quad (33)$$

$$= E(X_{2i-1}^2)E(X_{2i}^2) = 1 \quad (34)$$

Now,

$$Y_n = \frac{1}{n} \left(\sum_{i=1}^n Z_i \right) \quad (35)$$

$$E(Y_n) = E\left(\frac{1}{n} \sum_{i=1}^n Z_i\right) \quad (36)$$

$$= \frac{1}{n} \sum_{i=1}^n E(Z_i) = 0 \quad (37)$$

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

$$\text{cov}(X_i, X_j) = \begin{cases} 0 & \text{if } i \neq j \\ \text{Var}(X_i) & \text{if } i = j \end{cases} \quad (38)$$

$$\text{Var}(Y_n) = \frac{1}{n^2} \left(\sum_{i=1}^n \text{cov}(Z_i, Z_i) \right) \quad (39)$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^n \text{var}(Z_i) \right) \quad (40)$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^n 1 \right) \quad (41)$$

$$= \frac{1}{n} \quad (42)$$

(A) by central limit theorem,

$$\frac{Y_n - \mu}{\sigma} \sim \mathcal{N}(0, 1) \quad (43)$$

$$= \frac{Y_n - 0}{\sqrt{\frac{1}{n}}} \sim \mathcal{N}(0, 1) \quad (44)$$

$$= \sqrt{n}Y_n \sim \mathcal{N}(0, 1) \quad (45)$$

Thus,

$$\lim_{n \rightarrow \infty} F_{\sqrt{n}Y_n} = F_P(x) \quad (46)$$

where, $P \sim \mathcal{N}(0, 1)$

So, option (A) is correct

(B) For 2nd mean to be converging to 0,

$$\lim_{n \rightarrow \infty} E(|Y_n - 0|^2) = \lim_{n \rightarrow \infty} E(Y_n^2) \quad (47)$$

$$\because \frac{1}{n} = E(Y_n^2) - [E(Y_n)]^2 \quad (48)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} + [E(Y_n)]^2 \quad (49)$$

$$\because [E(Y_n)]^2 = 0 \quad (50)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} = 0 \quad (51)$$

Thus, $\{Y_n\}_{n \geq 1}$ converges in 2nd mean to 0

Hence, option (B) is correct.

(C) For $\{Y_n + \frac{1}{n}\}_{n \geq 1}$ to be converging $Y_n + \frac{1}{n} \xrightarrow{P} 0$

$$\lim_{n \rightarrow \infty} \Pr\left(\left|Y_n + \frac{1}{n} - 0\right| \geq \epsilon\right) = \lim_{n \rightarrow \infty} \Pr\left(\left|Y_n + \frac{1}{n}\right| \geq \epsilon\right) \quad (52)$$

$$\leq \lim_{n \rightarrow \infty} \frac{\text{Var}(Y_n + \frac{1}{n})}{\epsilon^2} \quad (53)$$

$$\forall \epsilon > 0 \quad (54)$$

$$= \lim_{n \rightarrow \infty} \frac{\text{Var}(Y_n) + \text{Var}(\frac{1}{n})}{\epsilon^2} \quad (55)$$

$$= \lim_{n \rightarrow \infty} \frac{\frac{1}{n} + 0}{\epsilon^2} \quad (56)$$

$$= 0 \quad (57)$$

hence, option (C) is correct

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

$$\text{Var}(X_n) = \sum_{i=1}^n \text{cov}(X_i, X_i) \quad (58)$$

$$= \sum_{i=1}^n \text{var}(X_i) \quad (59)$$

$$= \sum_{i=1}^n 1 \quad (60)$$

$$= n \quad (61)$$

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| < \epsilon) = 1 - \lim_{n \rightarrow \infty} \Pr(|X_n - 0| \geq \epsilon) \quad (62)$$

$$\leq 1 - \lim_{n \rightarrow \infty} \frac{\text{Var}(X_n)}{\epsilon^2} \quad (63)$$

$$\forall \epsilon > 0 \quad (64)$$

$$= \lim_{n \rightarrow \infty} 1 - \frac{\text{Var}(X_n)}{\epsilon^2} \quad (65)$$

$$= \lim_{n \rightarrow \infty} 1 - \frac{n}{\epsilon^2} = -\infty \quad (66)$$

Hence, option (D) is incorrect

Simulation:

- 1) we take X_i as uniform R.V and generate it by using **generateUniformRandom()**
- 2) for A, we initialize Y_n in the **main()** and standardize Y_n and save it to file and plot it in python. Also, compare it to standard normal distribution.
- 3) for B, we define a function **calculate2ndmean** to compute 2nd mean to see its convergence
- 4) for C, we define a **calculateYnplus1overn** to define the sequence $Y_n + \frac{1}{n}$ and compute the proportion of simulation i.e $\Pr\left(\left|Y_n + \frac{1}{n} - Y\right| \geq \epsilon\right)$
- 5) for D, we define a **simulateXn** to define the sequence X_n .
- 6) To check convergence almost surely by a function **checkconvergencealmostsurely** that computes $\Pr(|X_n| < \epsilon)$ and if ≥ 1 then it does not converges almost surely.

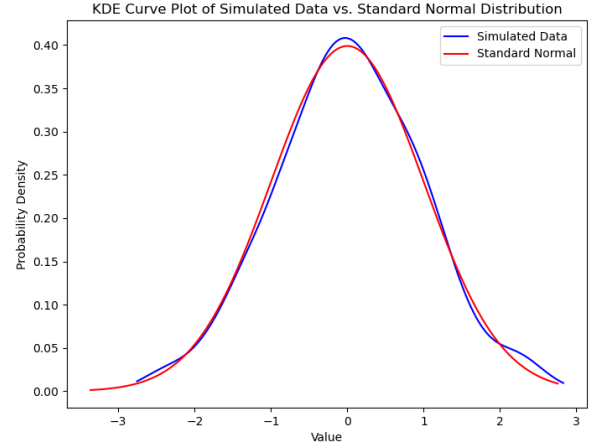


Fig. 6. simulated vs standard normal