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,..., let

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

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,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{2}$$

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

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

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

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

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

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

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

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xF_Xxdx \tag{6}$$
$$\because X \text{ is positive-valued} \tag{7}$$
$$= \int_{0}^{\infty} xF_Xxdx \tag{8}$$
$$\geq \int_{\epsilon^2}^{\infty} xF_Xxdx \tag{9}$$
$$\text{for some } \epsilon^2 > 0 \tag{10}$$
$$\geq \int_{\epsilon^2}^{\infty} kF_Xxdx \tag{11}$$
$$= \epsilon^2 \int_{\epsilon^2}^{\infty} F_Xxdx \tag{12}$$
$$\mathrm{E}(X) = \epsilon^2 \Pr\left(X \geq \epsilon^2\right) \tag{13}$$

This inequality is called Markov inequality
Now, we take

$$\sigma^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 \tag{14}$$
$$= \mathrm{E}(X^2) - \mu^2 \tag{15}$$
$$= \mathrm{E}(X^2) - \mathrm{E}(\mu^2) \tag{16}$$
$$\sigma^2 = \mathrm{E}(X^2 - \mu^2) \tag{17}$$

using (13),

$$\sigma^2 = \epsilon^2 \Pr\left(X^2 - \mu^2 \geq \epsilon^2\right) \tag{18}$$
$$\frac{\sigma^2}{\epsilon^2} = \Pr\left(X^2 - \mu^2 \geq \epsilon^2\right) \tag{19}$$
$$\Pr(|X - \mu| \geq \epsilon) = \frac{\sigma^2}{\epsilon^2} \tag{20}$$

This above inequality is called Chebyshev's inequaltiy
Now, as X_i is a sequence of independent and identically distributed random variables,
Let $Z_i = X_{2i-1}X_{2i}$.

$$p_{Z_iZ_j}(x) = p_{X_{2i-1}X_{2i}X_{2j-1}X_{2j}}(x) \tag{21}$$
$$= p_{X_{2i-1}}(x)p_{X_{2i}}(x)p_{X_{2j-1}}(x)p_{X_{2j}}(x) \tag{22}$$
$$= p_{X_{2i-1}X_{2i}}(x)p_{X_{2j-1}X_{2j}}(x) \tag{23}$$
$$= p_{Z_i}(x)p_{Z_j}(x) \tag{24}$$

Z_i is a independent R.V . Similarly,

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

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

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

Thus, Z_i is an identical distributed R.V.
Hence,

$$Z_n = \sum_{i=1}^{n} Z_i \tag{28}$$

is an i.i.d

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

$$= \mathrm{E}(X_{2i-1})\,\mathrm{E}(X_{2i}) = 0 \tag{30}$$

$$\mathrm{Var}(Z_i) = \mathrm{E}(Z_i^2) - \,[\mathrm{E}(Z_i)]^2 \tag{31}$$

$$= \mathrm{E}(Z_i^2) - 0 \tag{32}$$

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

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

Now,

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

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

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

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

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

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

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

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

$$= \frac{1}{n} \tag{42}$$

(A) by central limit theorem,

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

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

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

Thus,

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

where, $P \sim \mathcal{N}(0,1)$
So, option (A) is correct

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

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

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

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

$$\because [E(Y_n)]^2 = 0 \tag{50}$$

$$= \lim_{n\to\infty}\frac{1}{n} = 0 \tag{51}$$

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 $Y_n + \frac{1}{n} \xrightarrow{p} 0$

$$\lim_{n\to\infty}\Pr\left(\left|Y_n + \frac{1}{n} - Y\right| \geq \epsilon\right) = \lim_{n\to\infty}\Pr\left(\left|Y_n + \frac{1}{n}\right| \geq \epsilon\right) \tag{52}$$

$$\leq \lim_{n\to\infty}\frac{\mathrm{Var}(Y_n + \frac{1}{n})}{\epsilon^2} \tag{53}$$

$$\forall \epsilon > 0 \tag{54}$$

$$= \lim_{n\to\infty}\frac{\mathrm{Var}(Y_n) + \mathrm{Var}(\frac{1}{n})}{\epsilon^2} \tag{55}$$

$$= \lim_{n\to\infty}\frac{\frac{1}{n} + 0}{\epsilon^2} \tag{56}$$

$$= 0 \tag{57}$$

hence, option (C) is correct

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

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

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

$$= \sum_{i=1}^{n} 1 \tag{60}$$

$$= n \tag{61}$$

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

$$\leq 1 - \lim_{n\to\infty} \frac{Var(X_n)}{\epsilon^2} \tag{63}$$

$$\forall \epsilon > 0 \tag{64}$$

$$= \lim_{n\to\infty} 1 - \frac{Var(X_n)}{\epsilon^2} \tag{65}$$

$$= \lim_{n\to\infty} 1 - \frac{n}{\epsilon^2} = -\infty \tag{66}$$

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 intialize 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.

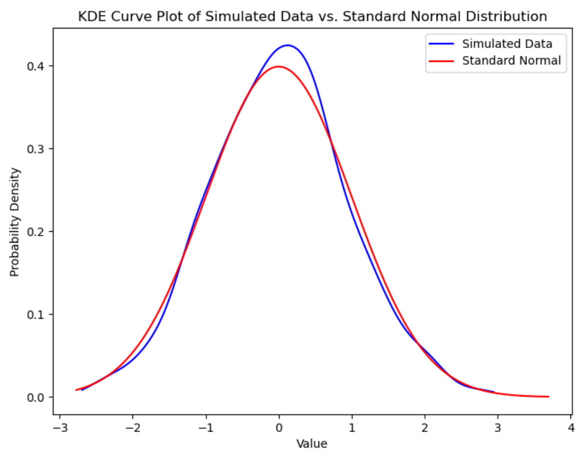

Fig. 6. simulated vs standard normal