STAB52 Fall 2024 — Name (F, L): Zay Lin Aung

Quiz 6 — Student ID #: 1010388171

1. Let $n \geq 1$ and $X_{-1}, X_0, \ldots, X_n$ be i.i.d. random variables with $X_i \sim \text{Unif}(-1, 1)$. Define random variables $Y_1, \ldots, Y_n$ as follows:

$$Y_i = X_i - 2X_{i-1} + X_{i-2}, \quad i = 1, ..., n.$$

(a) (4 points) Compute directly, using the density of $\text{Unif}(-1, 1)$, the common mean and variance of X_i. Hence, find $\mathbb{E}[Y_i]$ and $\text{Var}(Y_i)$.
Note: The mean and variance of a uniform random variable are available in the notes/books. The point is to do this hands-on.

(b) (6 points) Using the general properties of covariance, compute $\text{Cov}(Y_i, Y_j)$ for $i < j$.
Hint: There are three cases: (i) $j = i + 1$, (ii) $j = i + 2$ and (iii) $j > i + 2$.

(c) (6 points) Using (a) and (b), find the expectation and variance of the random variable $\bar{Y}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$.
Hint: For the variance, the relevant formula can be found in Corollary 7.14 of the notes. Arrange the values $\text{Cov}(Y_i, Y_j)$ as the (i, j)-th entry of an $n \times n$ matrix. The matrix has a nice geometric pattern which can be exploited to facilitate the computation.

(d) (4 points) For $\epsilon > 0$, use (c) and Chebyshev's inequality to give a lower bound on the probability $\mathbb{P}(|\bar{Y}_n| < \epsilon)$.

Some perspective: In Figure 1 (left) we show (with the dots) a simulated sequence of $(X_{-1}, \ldots, X_n)$ with $n = 100$. In Figure 1 (right), we plot the corresponding series of $Y_i = X_i - 2X_{i-1} + X_{i-2}$. The graph of (X_i) shows no systematic patterns since the values are independent. But note that Y_i and Y_{i+1} tend to have opposite signs. You will see that this observation is consistent with the sign of $\text{Cov}(Y_i, Y_{i+1})$.

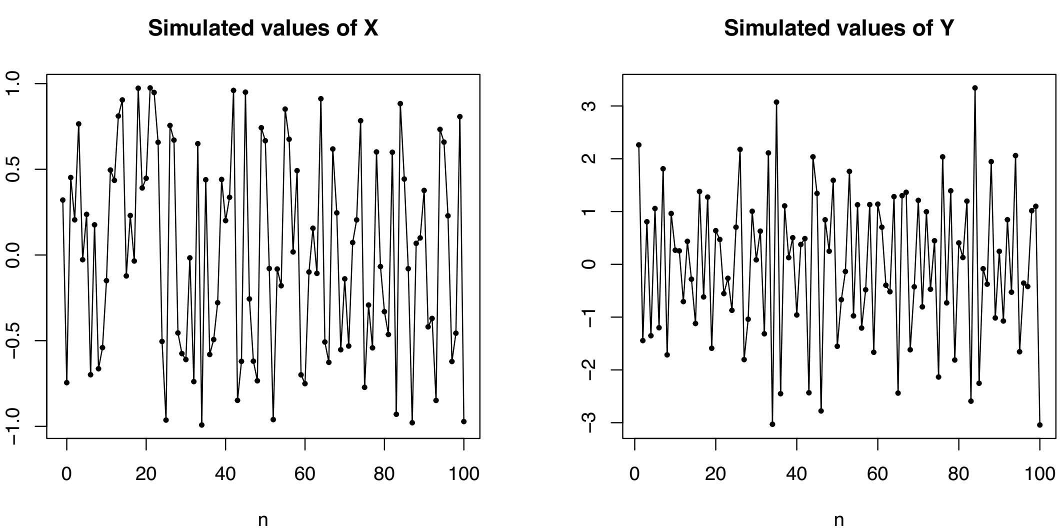

Figure 1: Simulated values of $(X_i)_{i=-1}^n$ (left) and $(Y_i)_{i=1}^n$ (right).

1.(a) For $i = 1, \dots, n$

$$Y_i = X_i - 2X_{i-1} + X_{i-2}$$

$$E[Y_i] = E[X_i - 2X_{i-1} + X_{i-2}]$$

$$= E[X_i] - 2E[X_{i-1}] + E[X_{i-2}] \quad \text{(by linearity of } E\text{)}$$

However, for each X_i for $i = -1, 0, 1, \dots$ $X_i \sim \text{Uniform}(-1, 1)$

Density $f_X(x) = \begin{cases} \frac{1}{1-(-1)} = \frac{1}{2}, & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

$$E[X_i] = \int x f_X(x) dx$$

$$= \int_{-1}^{1} \frac{x}{2} dx$$

$$= \frac{x^2}{4}\Big|_{-1}^{1}$$

$$= \frac{1^2}{4} - \left(\frac{(-1)^2}{4}\right)$$

$$= 0$$

$$\therefore E[Y_i] = 0 - 2(0) + 0 = 0$$

$$\text{Var}(Y_i) = \text{Var}(X_i - 2X_{i-1} + X_{i-2})$$

$$= \text{Var}(X_i) + (-2)^2 \text{Var}(X_{i-1}) + \text{Var}(X_{i-2}) \quad (X_i \text{ is i.i.d})$$

$$= \text{Var}(X_i) + 4\text{Var}(X_{i-1}) + \text{Var}(X_{i-2})$$

For each $i = -1, 0, 1, \dots$

$$\text{Var}(X_i) = E[X_i^2] - E[X_i]^2$$

$$= E[X_i^2] - 0$$

$$= \int x^2 f_X(x) dx$$

$$= \int_{-1}^{1} x^2 \frac{1}{2} dx$$

$$= \frac{x^3}{6}\Big|_{-1}^{1}$$

$$= \frac{1}{6} - \frac{(-1)}{6} = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$\therefore \text{Var}(Y_i) = \frac{1}{3} + 4\left(\frac{1}{3}\right) + \frac{1}{3}$$

$$= \frac{1+4+1}{3} = \frac{6}{3} = 2$$

(b) For $i < j$, $Cov(Y_i, Y_j) = E[Y_i Y_j] - E[Y_i]E[Y_j]$

$= E[Y_i Y_j] - 0$ (from a)

$= E[Y_i Y_j]$

Case (i) $j = i+1$

$$Cov(Y_i, Y_j) = E[Y_i \quad Y_{i+1}]$$

$$= E[(X_i - 2X_{i-1} + X_{i-2})(X_{i+1} - 2X_i + X_{i-1})]$$

$$= E[X_i X_{i+1} - 2X_i^2 + X_i X_{i-1} - 2X_{i-1}X_{i+1} + 4X_{i-1}X_i - 2X_{i-1}^2 + X_{i-2}X_{i+1}$$

$$- 2X_i X_{i-2} + X_{i-2}X_{i-1}]$$

$$= E[X_i X_{i+1}] - 2E[X_i^2] + E[X_i X_{i-1}] - 2E[X_{i-1}X_{i+1}] + 4E[X_{i-1}X_i] - 2E[X_{i-1}^2]$$

$$+ E[X_{i-2}X_{i+1}] - 2E[X_i X_{i-2}] + E[X_{i-2}X_{i-1}]$$

For each $i \neq j$, $E[X_i X_j] = E[X_i]E[X_j]$ ($\because X_i \perp\!\!\!\perp X_j$)

$= 0$ (from a)

Since $E[X_i^2] = Var(X_i) = \frac{1}{3}$ from a

Thus, $Cov(Y_i, Y_j) = (-2)(\frac{1}{3}) + (-2)(\frac{1}{3})$

$$= \frac{-2}{3} - \frac{2}{3} = \frac{-4}{3}$$

Case: $j = i+2$

$$Cov(Y_i, Y_j) = E[Y_i \quad Y_{i+2}]$$

$$= E[(X_i - 2X_{i-1} + X_{i-2})(X_{i+2} - 2X_{i+1} + X_i)]$$

Similarly, only same random variable pairs will be left

$\therefore Cov(Y_i, Y_j) = E[X_i^2] = Var(X_i) = \frac{1}{3}$

Case 3: $j > i+2$

$$Cov(Y_i, Y_j) = E[Y_i Y_j]$$

For $j > i+2$, Y_i and Y_{i+2} will not have same random variable X.

Thus, $Cov(Y_i, Y_j) = 0$

$$\operatorname{Cov}(Y_i, Y_j) = \begin{cases} -4/3, & \text{if } j = i+1 \\ 1/3, & \text{if } j = i+2 \\ 0, & \text{if } j > i+2 \end{cases}$$

(c) $\bar{Y}_n = \frac{1}{n}(Y_1 + \dots + Y_n)$

$$E[\bar{Y}_n] = E\left[\frac{1}{n}(Y_1 + \dots + Y_n)\right]$$

$$= \frac{1}{n}(E[Y_1] + \dots + E[Y_n]) \quad \text{(by linearity of } E\text{)}$$

$$= \frac{1}{n}(0 + \dots + 0) \quad \text{(from (a))}$$

$$= 0$$

Case $n = 1$:

$$\operatorname{Var}(\bar{Y}_1) = \frac{1}{1^2}\operatorname{Var}(Y_1)$$

$$= 2$$

$$\operatorname{Var}(\bar{Y}_n) = \operatorname{Var}\left(\frac{1}{n}(Y_1 + \dots + Y_n)\right)$$

$$= \frac{1}{n^2}\operatorname{Var}(Y_1 + \dots + Y_n)$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n}\operatorname{Var}(Y_i) + 2\sum_{i<j}\operatorname{Cov}(Y_i, Y_j)\right)$$

Case: $n > 1$

In terms of covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & & \vdots \\ \vdots & & \ddots & \\ \sigma_{n1} & \cdots & & \sigma_{nn} \end{bmatrix}$$

Thus, $\operatorname{Cov}(Y_i, Y_j)$ can be observed in top half of matrix

# of cases $j = i+1$, $|\{\sigma_{12}, \sigma_{23}, \dots, \sigma_{n-1\,n}\}| = n-1$

# of cases $j = i+2$, $|\{\sigma_{13}, \sigma_{24}, \sigma_{35}, \dots, \sigma_{n-2\,n}\}| = n-2$

$$\therefore \operatorname{Var}(\bar{Y}_n) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\operatorname{Var}(Y_i) + 2\left(\sum_{j=i+1}\operatorname{Cov}(Y_i, Y_j) + \sum_{j=i+2}\operatorname{Cov}(Y_i, Y_j)\right)\right)$$

$$= \frac{1}{n^2}\left(2n + 2\left[(n-1)\left(-\frac{4}{3}\right) + (n-2)\left(\frac{1}{3}\right)\right]\right)$$

$$= \frac{1}{n^2}\left(2n + 2\,\frac{-4n+4+n-2}{3}\right)$$

$$= \frac{1}{n^2}\left(2n + \frac{-6n+4}{3}\right)$$

$$= \frac{1}{n^2}\cdot\frac{4}{3} = \frac{4}{3n^2}$$

(d) Let $\mu_{\bar{Y}}$ be expectation of $\bar{Y}_n$ and σ^2 be variance of $\bar{Y}_n$

by Chebyshev's inequality,

$$P(|\bar{Y}_n - \mu_{\bar{Y}}| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$P(|\bar{Y}_n - 0| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$P(|\bar{Y}_n| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Since both $P(|\bar{Y}_n| \geq \varepsilon) > 0$ and $\frac{\sigma^2}{\varepsilon^2} > 0$

$$-P(|\bar{Y}_n| \geq \varepsilon) \geq \frac{-\sigma^2}{\varepsilon^2}$$

$$1 - P(|\bar{Y}_n| \geq \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

$$P(|\bar{Y}_n| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

$$P(|\bar{Y}_n| < \varepsilon) \geq 1 - \frac{2}{\varepsilon^2}, \quad n = 1$$

$$P(|\bar{Y}_n| < \varepsilon) \geq 1 - \frac{4}{3n^2\varepsilon^2}, \quad n > 1$$