

A1

Define the probabilities:

$P(D)$ = The probability of having the disease,

$P(\text{Positive})$ = The probability of testing positive.

Using Bayes' Theorem:

$$P(D \mid \text{Positive}) = \frac{P(\text{Positive} \mid D)P(D)}{P(\text{Positive})}.$$

Known probabilities:

$$P(D) = 0.0001,$$

$$P(\neg D) = 0.9999,$$

$$P(\text{Positive} \mid D) = 0.99,$$

$$P(\text{Positive} \mid \neg D) = 0.01.$$

Calculate $P(\text{Positive})$:

$$P(\text{Positive}) = P(\text{Positive} \mid D)P(D) + P(\text{Positive} \mid \neg D)P(\neg D).$$

$$P(\text{Positive}) = (0.99)(0.0001) + (0.01)(0.9999).$$

$$P(\text{Positive}) = 0.000099 + 0.009999 = 0.010098.$$

Calculate $P(D \mid \text{Positive})$:

$$P(D \mid \text{Positive}) = \frac{(0.99)(0.0001)}{0.010098}.$$

$$P(D \mid \text{Positive}) = \frac{0.000099}{0.010098}.$$

$$P(D \mid \text{Positive}) \approx 0.0098.$$

Final result:

$$P(D \mid \text{Positive}) \approx 0.98\%.$$

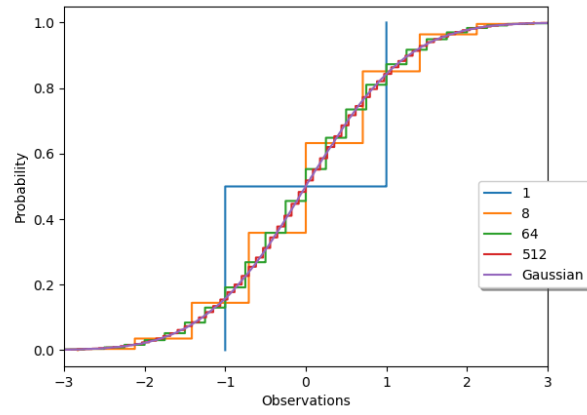


Figure 1: Empirical CDF of $Y^{(k)}$

A10

(a)

The value of n is calculated to be 40000.

(b)

As k increases, the empirical CDF of $Y^{(k)}$ becomes closer to the CDF of the standard normal distribution.

Question A2

(a)

Since $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$ (by the Law of Total Expectation), we know:

$$\mathbb{E}[Y \mid X] = X \implies \mathbb{E}[Y] = \mathbb{E}[X].$$

By the definition of covariance:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Expanding this out:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Substituting $\mathbb{E}[Y \mid X] = X$, we get:

$$\text{Cov}(X, Y) = \mathbb{E}[X \cdot X] - \mathbb{E}[X] \cdot \mathbb{E}[X].$$

This simplifies to:

$$\text{Cov}(X, Y) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Now, using the expansion:

$$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Thus:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

■

Question A2

(b)

We know that $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$. To show $\text{Cov}(X, Y) = 0$, we need to prove:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

By the definition of expectation of the joint PDF:

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy.$$

Using independence:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Substituting this into the expression for $\mathbb{E}[XY]$, we have:

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy.$$

The double integral can be separated:

$$\mathbb{E}[XY] = \left(\int_{-\infty}^{\infty} x f_X(x) dx \right) \cdot \left(\int_{-\infty}^{\infty} y f_Y(y) dy \right).$$

The first term is $\mathbb{E}[X]$, and the second term is $\mathbb{E}[Y]$. Thus:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Therefore:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0.$$

■

A3(a)

We know:

$$P(Z \leq z) = P(X + Y \leq z) = P(Y \leq z - X).$$

We can marginalize over X :

$$P(Z \leq z) = \int_{-\infty}^{\infty} P(Y \leq z - x) f(x) dx.$$

Now, let $G(y)$ denote the CDF of Y . Then:

$$P(Y \leq z - x) = G(z - x).$$

Substituting this into $P(Z \leq z)$:

$$P(Z \leq z) = \int_{-\infty}^{\infty} G(z - x) f(x) dx. \quad (1)$$

From the Fundamental Theorem of Calculus:

$$\frac{d}{dz} G(z - x) = g(z - x),$$

where $g(y)$ is the PDF of Y . This works because the derivative of the CDF $G(y)$ with respect to its variable yields the PDF $g(y)$, which represents the rate of change of the cumulative probability.

Differentiating $P(Z \leq z)$ using (1):

$$h(z) = \frac{d}{dz} P(Z \leq z) = \frac{d}{dz} \int_{-\infty}^{\infty} G(z - x) f(x) dx.$$

By the chain rule and the Fundamental Theorem of Calculus:

$$h(z) = \int_{-\infty}^{\infty} g(z - x) f(x) dx.$$

Thus:

$$h(z) = \int_{-\infty}^{\infty} f(x) g(z - x) dx.$$

■

A3(b)

We proved in (a) that:

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x) dx.$$

Since f and g are nonzero only on $[0, 1]$, we note that $g(z-x)$ is:

$$g(z-x) = \begin{cases} 1 & \text{for } 0 \leq z-x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Plugging in values:

$$h(z) = \int_0^1 g(z-x) dx.$$

For $0 \leq z \leq 1$, $g(z-x) = 1$ for $x \in [0, z]$:

$$h(z) = \int_0^z 1 dx = z.$$

For $1 < z \leq 2$, $g(z-x) = 1$ for $x \in [z-1, 1]$:

$$h(z) = \int_{z-1}^1 1 dx = 1 - (z-1) = 2 - z.$$

Thus:

$$h(z) = \begin{cases} z & \text{for } 0 \leq z \leq 1, \\ 2 - z & \text{for } 1 < z \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

■

A4(a)

Let $Y = aX_1 + b$. Since $Y \sim \mathcal{N}(0, 1)$, we need to determine a and b such that Y has mean 0 and variance 1.

Mean

The mean of Y is:

$$\mathbb{E}[Y] = \mathbb{E}[aX_1 + b] = a\mathbb{E}[X_1] + b.$$

Since $\mathbb{E}[X_1] = \mu$ and $\mathbb{E}[Y] = 0$ (as given), this becomes:

$$a\mu + b = 0.$$

Solve for b :

$$b = -a\mu. \tag{1}$$

Variance

The variance of Y is:

$$\text{Var}(Y) = \text{Var}(aX_1 + b) = a^2\text{Var}(X_1).$$

Since $\text{Var}(X_1) = \sigma^2$ and $\text{Var}(Y) = 1$ (as given), we have:

$$a^2\sigma^2 = 1.$$

Solve for a :

$$a = \frac{1}{\sigma}. \tag{2}$$

Substituting Values

Substitute $a = \frac{1}{\sigma}$ from (2) into (1):

$$b = -\frac{\mu}{\sigma}.$$

Final Results

The values of a and b are:

$$a = \frac{1}{\sigma}, \quad b = -\frac{\mu}{\sigma}.$$

■

A4(b)

Mean

Let $Z = X_1 + 2X_2$. Then, by the linearity of expectation:

$$\mathbb{E}[Z] = \mathbb{E}[X_1 + 2X_2] = \mathbb{E}[X_1] + 2\mathbb{E}[X_2].$$

Since $\mathbb{E}[X_1] = \mu$ and $\mathbb{E}[X_2] = \mu$, we have:

$$\mathbb{E}[Z] = \mu + 2\mu = 3\mu.$$

Variance

The variance of Z is:

$$\text{Var}(Z) = \text{Var}(X_1 + 2X_2).$$

Using the variance properties for linear combinations:

$$\text{Var}(Z) = \text{Var}(X_1) + 2^2\text{Var}(X_2).$$

Since $\text{Var}(X_1) = \sigma^2$ and $\text{Var}(X_2) = \sigma^2$, this becomes:

$$\text{Var}(Z) = \sigma^2 + 4\sigma^2 = 5\sigma^2.$$

Final Results

- **Mean:**

$$\mathbb{E}[Z] = 3\mu.$$

- **Variance:**

$$\text{Var}(Z) = 5\sigma^2.$$

■

A4(c)

Mean

First, consider the expectation of $\hat{\mu}_n$:

$$\mathbb{E}[\hat{\mu}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i].$$

Since $\mathbb{E}[X_i] = \mu$ for all i , this simplifies to:

$$\mathbb{E}[\hat{\mu}_n] = \frac{1}{n} \cdot n \cdot \mu = \mu.$$

Now, compute the expectation of $\sqrt{n}(\hat{\mu}_n - \mu)$:

$$\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] = \sqrt{n} \cdot \mathbb{E}[\hat{\mu}_n - \mu].$$

Substituting $\mathbb{E}[\hat{\mu}_n] = \mu$, we get:

$$\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] = \sqrt{n} \cdot (\mathbb{E}[\hat{\mu}_n] - \mu) = \sqrt{n} \cdot (\mu - \mu) = 0.$$

Variance

The variance of $\sqrt{n}(\hat{\mu}_n - \mu)$ is:

$$\text{Var}(\sqrt{n}(\hat{\mu}_n - \mu)) = \text{Var}\left(\sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right).$$

Simplify the scaling factor $\sqrt{n} \cdot \frac{1}{n} = \frac{1}{\sqrt{n}}$:

$$\text{Var}(\sqrt{n}(\hat{\mu}_n - \mu)) = \frac{1}{n} \cdot \text{Var}\left(\sum_{i=1}^n (X_i - \mu)\right).$$

Since X_1, X_2, \dots, X_n are independent, the variance of the sum is the sum of variances:

$$\text{Var}\left(\sum_{i=1}^n (X_i - \mu)\right) = \sum_{i=1}^n \text{Var}(X_i - \mu).$$

Since $\text{Var}(X_i - \mu) = \text{Var}(X_i) = \sigma^2$ for all i , this becomes:

$$\text{Var}\left(\sum_{i=1}^n (X_i - \mu)\right) = n \cdot \sigma^2.$$

Substitute this back:

$$\text{Var}(\sqrt{n}(\hat{\mu}_n - \mu)) = \frac{1}{n} \cdot n \cdot \sigma^2 = \sigma^2.$$

Final Results

- **Mean:**

$$\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] = 0.$$

- **Variance:**

$$\text{Var}[\sqrt{n}(\hat{\mu}_n - \mu)] = \sigma^2.$$

■

A5(a)

Matrix A

We have

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}.$$

We perform row operations to find its rank:

$$r_2 \leftarrow r_2 - r_3 \implies \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

$$r_3 \leftarrow r_3 - r_1 \implies \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 1 \end{bmatrix},$$

$$r_3 \leftarrow r_3 - r_2 \implies \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the last row is all zeroes, the rank of A is 2.

Matrix B

We have

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

We perform the following operations:

$$r_1 \leftrightarrow r_3 \implies \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix},$$

$$r_2 \leftrightarrow r_3 \implies \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 0 & 1 \end{bmatrix},$$

$$r_2 \leftarrow r_2 - r_1 \implies \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

$$r_3 \leftarrow r_3 - r_1 \implies \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix},$$

$$r_3 \leftarrow r_3 + r_2 \quad \implies \quad \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, the rank of B is also 2.

A5(b)

For both matrices A and B , the pivot columns are the first two columns. Hence a minimal basis for the column space (of either matrix) is given by their first two original columns:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}.$$

■

A6(a)

We are given the matrix

$$A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \quad \text{and the vector} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

We want to compute $A\mathbf{c}$.

$$\begin{aligned} A\mathbf{c} &= \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (0 \cdot 1) + (2 \cdot 1) + (4 \cdot 1) \\ (2 \cdot 1) + (4 \cdot 1) + (2 \cdot 1) \\ (3 \cdot 1) + (3 \cdot 1) + (1 \cdot 1) \end{bmatrix} \\ &= \begin{bmatrix} 0 + 2 + 4 \\ 2 + 4 + 2 \\ 3 + 3 + 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}. \end{aligned}$$

Thus,

$$A\mathbf{c} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}.$$

A6(b)

We want to solve the system $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix}.$$

Form the augmented matrix:

$$\left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right].$$

Row Operations

$$\begin{aligned} r_1 &\longleftrightarrow r_2 &\implies &\left[\begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 2 & 4 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right], \\ r_1 &\leftarrow \frac{1}{2}r_1, \quad r_2 \leftarrow \frac{1}{2}r_2 &\implies &\left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 3 & 3 & 1 & -4 \end{array} \right], \\ r_3 &\leftarrow r_3 - 3r_1 &\implies &\left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & -3 & -2 & -1 \end{array} \right], \\ r_3 &\leftarrow r_3 + 3r_2 &\implies &\left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 4 & -4 \end{array} \right], \\ r_3 &\leftarrow \frac{1}{4}r_3 &\implies &\left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array} \right]. \end{aligned}$$

Back-Substitution and Solution

From the final augmented matrix, we read off the equations:

$$\begin{cases} x_1 + 2x_2 + x_3 = -1, \\ x_2 + 2x_3 = -1, \\ x_3 = -1. \end{cases}$$

Starting from the bottom row, $x_3 = -1$. Plugging into the second row:

$$x_2 + 2(-1) = -1 \implies x_2 - 2 = -1 \implies x_2 = 1.$$

Finally, from the first row:

$$x_1 + 2(1) + (-1) = -1 \implies x_1 + 1 = -1 \implies x_1 = -2.$$

Hence the solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}.$$

A7

(a)

We expand $f(x, y)$ as:

$$f(x, y) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n B_{ij} y_i x_j + c.$$

(b)

The gradient of $f(x, y)$ with respect to x is given by:

$$\nabla_x f(x, y) = (A + A^\top) x + B^\top y.$$

In summation form, the k -th component of the gradient is:

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^n (A_{kj} + A_{jk}) x_j + \sum_{i=1}^n B_{ik} y_i.$$

This arises because A is not necessarily symmetric; thus the partial derivatives from both $A_{kj} x_k x_j$ and $A_{ik} x_i x_k$ add up to $\sum_j (A_{kj} + A_{jk}) x_j$, rather than $2 \sum_j A_{kj} x_j$.

This is derived as follows:

1. **For the term** $x^\top A x = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$:

When we take $\frac{\partial}{\partial x_k}$, only terms where $i = k$ or $j = k$ contribute.

$$\text{If } i = k, A_{kj} x_k x_j \implies \frac{\partial}{\partial x_k} (A_{kj} x_k x_j) = A_{kj} x_j.$$

$$\text{If } j = k, A_{ik} x_i x_k \implies \frac{\partial}{\partial x_k} (A_{ik} x_i x_k) = A_{ik} x_i.$$

Adding these contributions:

$$\sum_{j=1}^n A_{kj} x_j + \sum_{i=1}^n A_{ik} x_i = \sum_{j=1}^n (A_{kj} + A_{jk}) x_j,$$

$$\text{i.e. } [(A + A^\top) x]_k.$$

2. **For the term** $y^\top B x = \sum_{i=1}^n \sum_{j=1}^n B_{ij} y_i x_j$:

When we take $\frac{\partial}{\partial x_k}$, only terms where $j = k$ contribute.

$$\text{If } j = k, B_{ik} y_i x_k \implies \frac{\partial}{\partial x_k} (B_{ik} y_i x_k) = B_{ik} y_i.$$

Summing over i yields: $\sum_{i=1}^n B_{ik} y_i$, i.e. $[B^\top y]_k$.

3. **For the constant term** c :

$$\frac{\partial}{\partial x_k} c = 0.$$

Combining these results for each component k :

$$\frac{\partial f}{\partial x_k} = [(A + A^\top) x]_k + [B^\top y]_k.$$

Hence in vector form:

$$\nabla_x f(x, y) = (A + A^\top) x + B^\top y.$$

(c)

The gradient of $f(x, y)$ with respect to y is:

$$\nabla_y f(x, y) = Bx.$$

In summation form, the k -th component of the gradient is:

$$\frac{\partial f}{\partial y_k} = \sum_{j=1}^n B_{kj} x_j.$$

This follows because $x^\top Ax$ does not involve y . Taking partials of $y^\top Bx$ with respect to y_k isolates exactly those terms $B_{kj}x_j$.

A8(a)

We are given that:

$$\text{diag}(v) = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n \end{bmatrix} \quad \text{and} \quad \text{diag}(w) = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}.$$

If we assume that $\text{diag}(v)^{-1} = \text{diag}(w)$, it follows that:

$$w_i = \frac{1}{v_i}, \quad \forall i \in \{1, 2, \dots, n\}.$$

We are also given that $g(v_i) = w_i$, so substituting the relationship between v_i and w_i , we find:

$$g(v_i) = \frac{1}{v_i}.$$

■

A8(b)

By definition, the squared Euclidean norm of Ax is:

$$\|Ax\|_2^2 = (Ax)^\top (Ax).$$

Using the associative property of matrix multiplication:

$$(Ax)^\top (Ax) = x^\top (A^\top A)x.$$

Since A is orthonormal, it satisfies $A^\top A = I$, where I is the identity matrix:

$$x^\top (A^\top A)x = x^\top Ix = x^\top x.$$

The expression $x^\top x$ is exactly the squared Euclidean norm of x , so:

$$\|Ax\|_2^2 = \|x\|_2^2.$$

Thus, we have shown that:

$$\|Ax\|_2^2 = \|x\|_2^2.$$

■

A8(c)

We start with the following property of an invertible matrix:

$$BB^{-1} = I,$$

where I is the identity matrix.

Taking the transpose of both sides:

$$(BB^{-1})^{\top} = I^{\top}.$$

Using the property of transposes for matrix multiplication, $(AB)^{\top} = B^{\top}A^{\top}$, this becomes:

$$(B^{-1})^{\top}B^{\top} = I^{\top}.$$

Since the transpose of the identity matrix is itself ($I^{\top} = I$), we have:

$$(B^{-1})^{\top}B^{\top} = I.$$

Because B is symmetric, $B^{\top} = B$, so:

$$(B^{-1})^{\top}B = I.$$

By the definition of the inverse, this implies:

$$(B^{-1})^{\top} = B^{-1}.$$

Thus, B^{-1} is symmetric. ■

A8(d)

Let λ be an eigenvalue of C , and let v be the corresponding eigenvector. By definition of eigenvalues and eigenvectors, we have:

$$Cv = \lambda v.$$

Taking the quadratic form $v^\top Cv$, we substitute $Cv = \lambda v$:

$$v^\top Cv = v^\top (\lambda v).$$

Since λ is a scalar, we can factor it out:

$$v^\top Cv = \lambda(v^\top v).$$

Since C is PSD, it satisfies:

$$v^\top Cv \geq 0 \quad \text{for all vectors } v.$$

Substituting $v^\top Cv = \lambda(v^\top v)$, we get:

$$\lambda(v^\top v) \geq 0.$$

The term $v^\top v$ is the squared norm of v , which is strictly positive since $v \neq 0$ (by definition of an eigenvector). Thus:

$$\lambda \geq 0.$$

Thus, all eigenvalues of C are non-negative. ■

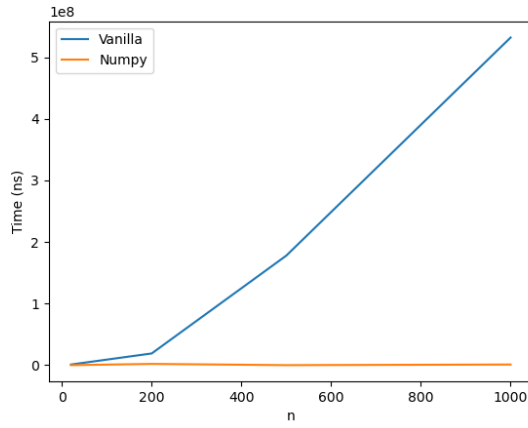


Figure 1: Runtime comparison of vanilla vs. numpy implementations for varying sizes of n .

A9

(c)

The runtime differences between the vanilla and numpy implementations for varying sizes of n are summarized below:

n	Vanilla Implementation (ms)	Numpy Implementation (ms)
20	0.999	0.0
200	18.978	1.9981
500	177.8502	0.0
1000	531.933	1.0056

Explanation: The vanilla implementation is significantly slower compared to the numpy implementation because it relies on Python loops for matrix multiplication and vector operations, which are less efficient. Numpy, on the other hand, uses highly optimized C and Fortran libraries for linear algebra computations, which execute operations in a much faster and parallelized manner.