Probability and Statistics: MA6.101

Tutorial 9 Solutions

Topics Covered: Random Vectors

Q1: Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be a 2-dimensional random vector. The components X_1 and X_2 are independent random variables with the following properties:

- $E[X_1] = 2$, $Var(X_1) = 4$
- $E[X_2] = 5$, $Var(X_2) = 1$

Now, consider a new random vector Y defined by the linear transformation $Y = AX$, where

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

Find the mean vector of Y and the covariance matrix of Y.

Solution:

First, we establish the mean vector (μ_X) and the covariance matrix (Σ_X) for the random vector X. The mean vector is simply the vector of the expected values of its components.

$$\mu_X = E[X] = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

The covariance matrix contains the variances of the components on its diagonal and the covariances between components on its off-diagonal. Since X_1 and X_2 are independent, their covariance is $Cov(X_1, X_2) = 0$.

$$\Sigma_X = C_X = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

(a) Find the mean of Y

For a linear transformation $Y = AX$, the mean of Y is given by $E[Y] = AE[X]$.

$$E[Y] = \mu_Y = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} (1)(2) + (2)(5) \\ (3)(2) + (-1)(5) \end{bmatrix} = \begin{bmatrix} 12 \\ 1 \end{bmatrix}$$

(b) Find the covariance matrix of Y

For a linear transformation $Y = AX$, the covariance matrix of Y is given by $C_Y = AC_XA^T$. First, we find the transpose of A:

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}$$

Now we compute the product AC_XA^T:

$$\begin{aligned}
C_Y = \Sigma_Y &= \left(\begin{bmatrix}1 & 2\\ 3 & -1\end{bmatrix}\begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix}\right)\begin{bmatrix}1 & 3\\ 2 & -1\end{bmatrix}\\
&= \begin{bmatrix}(1)(4)+(2)(0) & (1)(0)+(2)(1)\\ (3)(4)+(-1)(0) & (3)(0)+(-1)(1)\end{bmatrix}\begin{bmatrix}1 & 3\\ 2 & -1\end{bmatrix}\\
&= \begin{bmatrix}4 & 2\\ 12 & -1\end{bmatrix}\begin{bmatrix}1 & 3\\ 2 & -1\end{bmatrix}\\
&= \begin{bmatrix}(4)(1)+(2)(2) & (4)(3)+(2)(-1)\\ (12)(1)+(-1)(2) & (12)(3)+(-1)(-1)\end{bmatrix}\\
&= \begin{bmatrix}8 & 10\\ 10 & 37\end{bmatrix}
\end{aligned}$$

So, the mean vector of Y is $\mu_Y = \begin{bmatrix}12\\ 1\end{bmatrix}$ and its covariance matrix is $\Sigma_Y = \begin{bmatrix}8 & 10\\ 10 & 37\end{bmatrix}$. Notice that even though X_1 and X_2 were independent, the new components Y_1 and Y_2 are correlated ($Cov(Y_1, Y_2) = 10$).

Q2: Let X_1 be a uniform random variable with support $R_{X_1} = [1, 2]$ and probability density function

$$f_{X_1}(x_1) = \begin{cases}1 & \text{if } x_1 \in R_{X_1}\\ 0 & \text{if } x_1 \notin R_{X_1}\end{cases}$$

Let X_2 be a continuous random variable, independent of X_1, with support $R_{X_2} = [0, 2]$ and probability density function

$$f_{X_2}(x_2) = \begin{cases}\frac{3}{8}x_2^2 & \text{if } x_2 \in R_{X_2}\\ 0 & \text{if } x_2 \notin R_{X_2}\end{cases}$$

Let

$$\mathbf{Y} = \begin{bmatrix}Y_1\\ Y_2\end{bmatrix} = \begin{bmatrix}X_1^2\\ X_1 + X_2\end{bmatrix}$$

Find the joint probability density function of the random vector $\mathbf{Y}$.

Solution:

Since X_1 and X_2 are independent, their joint probability density function is the product of their marginals.

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \begin{cases}\frac{3}{8}x_2^2 & \text{if } x_1 \in [1, 2] \text{ and } x_2 \in [0, 2]\\ 0 & \text{otherwise}\end{cases}$$

The transformation is $g(x) = [X_1^2, X_1 + X_2]^T$. The inverse transformation $g^{-1}(y)$ is $x_1 = \sqrt{y_1}$ and $x_2 = y_2 - \sqrt{y_1}$. The Jacobian matrix of the inverse is:

$$J_{g^{-1}}(y) = \begin{bmatrix}\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2}\\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2}\end{bmatrix} = \begin{bmatrix}\frac{1}{2\sqrt{y_1}} & 0\\ -\frac{1}{2\sqrt{y_1}} & 1\end{bmatrix}$$

The determinant is $\det(J_{g^{-1}}(y)) = \frac{1}{2\sqrt{y_1}}$. The formula for the transformed PDF is $f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(g^{-1}(y))|\det(J_{g^{-1}}(y))|$.

$$f_{\mathbf{X}}(g^{-1}(y))|\det(J_{g^{-1}}(y))| = f_{X_1,X_2}(\sqrt{y_1}, y_2-\sqrt{y_1})\cdot\frac{1}{2\sqrt{y_1}} = \frac{3}{8}(y_2-\sqrt{y_1})^2\cdot\frac{1}{2\sqrt{y_1}} = \frac{3}{16\sqrt{y_1}}(y_2-\sqrt{y_1})^2$$

The support for Y is found from the support of X: $1 \le x_1 \le 2 \implies 1 \le y_1 \le 4$. And $0 \le x_2 \le 2 \implies 0 \le y_2 - \sqrt{y_1} \le 2 \implies \sqrt{y_1} \le y_2 \le \sqrt{y_1} + 2$. So the final PDF is:

$$f_{\mathbf{Y}}(y_1, y_2) = \begin{cases} \frac{3}{16\sqrt{y_1}}(y_2 - \sqrt{y_1})^2 & \text{if } 1 \le y_1 \le 4 \text{ and } \sqrt{y_1} \le y_2 \le \sqrt{y_1} + 2 \\ 0 & \text{otherwise} \end{cases}$$

Q3: Let X and Y be said to be bivariate normal if $aX + bY$ is normal for all a and b. If X and Y are bivariate normal with 0 mean, variance of 1, and ρ correlation, derive their joint PDF:

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(x^2 - 2\rho xy + y^2\right)\right)$$

Solution:
Let $\mathbf{X} = \begin{bmatrix} X \\ Y \end{bmatrix}$. The mean vector is $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The covariance matrix is $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. The determinant is $\det(\Sigma) = 1 - \rho^2$ and the inverse is $\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$. The multivariate normal PDF is $f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T\Sigma^{-1}(\mathbf{x} - \mu)\right)$. The quadratic form in the exponent is:

$$(\mathbf{x} - \mu)^T\Sigma^{-1}(\mathbf{x} - \mu) = \begin{bmatrix} x & y \end{bmatrix} \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{1-\rho^2}(x^2 - 2\rho xy + y^2)$$

Substituting this back gives the desired formula.

Q4: Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$. Now consider $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$. Then we have

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{A}E[\mathbf{X}] + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T).$$

Prove this using Moment-Generating Functions (MGFs).

A:
Our goal is to find the MGF of $\mathbf{Y}$, defined as $M_{\mathbf{Y}}(\mathbf{t}) = E[e^{\mathbf{t}^T\mathbf{Y}}]$, and show that it matches the MGF of the target normal distribution.
We are given that $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, where $E[\mathbf{X}] = \boldsymbol{\mu}$. The MGF of $\mathbf{X}$ is known to be:

$$M_{\mathbf{X}}(\mathbf{s}) = E[e^{\mathbf{s}^T\mathbf{X}}] = \exp\left(\mathbf{s}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}^T\mathbf{\Sigma}\mathbf{s}\right)$$

We begin with the definition of $M_{\mathbf{Y}}(\mathbf{t})$ and substitute $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$:

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E[e^{\mathbf{t}^T\mathbf{Y}}] \\ &= E[e^{\mathbf{t}^T(\mathbf{A}\mathbf{X}+\mathbf{b})}] \\ &= E[e^{\mathbf{t}^T\mathbf{A}\mathbf{X}+\mathbf{t}^T\mathbf{b}}] \\ &= E[e^{\mathbf{t}^T\mathbf{A}\mathbf{X}} \cdot e^{\mathbf{t}^T\mathbf{b}}] \end{aligned}$$

Since $e^{\mathbf{t}^T\mathbf{b}}$ is a scalar constant with respect to the expectation, we can factor it out:

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{b}} \cdot E[e^{\mathbf{t}^T\mathbf{A}\mathbf{X}}]$$

Now, we use the transpose property $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$, which implies $\mathbf{t}^T\mathbf{A} = (\mathbf{A}^T\mathbf{t})^T$.

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{b}} \cdot E[e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{X}}]$$

We recognize the expectation term $E[e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{X}}]$ as the MGF of $\mathbf{X}$, $M_{\mathbf{X}}(\mathbf{s})$, evaluated at the vector $\mathbf{s} = \mathbf{A}^T\mathbf{t}$.

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{b}} \cdot M_{\mathbf{X}}(\mathbf{A}^T\mathbf{t})$$

Next, we substitute the known formula for $M_{\mathbf{X}}$:

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{b}} \cdot \exp\left((\mathbf{A}^T\mathbf{t})^T\boldsymbol{\mu} + \frac{1}{2}(\mathbf{A}^T\mathbf{t})^T\boldsymbol{\Sigma}(\mathbf{A}^T\mathbf{t})\right)$$

We simplify the terms in the exponent using $(\mathbf{A}^T\mathbf{t})^T = \mathbf{t}^T(\mathbf{A}^T)^T = \mathbf{t}^T\mathbf{A}$:

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= e^{\mathbf{t}^T\mathbf{b}} \cdot \exp\left((\mathbf{t}^T\mathbf{A})\boldsymbol{\mu} + \frac{1}{2}(\mathbf{t}^T\mathbf{A})\boldsymbol{\Sigma}(\mathbf{A}^T\mathbf{t})\right) \\ &= e^{\mathbf{t}^T\mathbf{b}} \cdot \exp\left(\mathbf{t}^T(\mathbf{A}\boldsymbol{\mu}) + \frac{1}{2}\mathbf{t}^T(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)\mathbf{t}\right) \end{aligned}$$

Finally, we combine the terms into a single exponent:

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= \exp\left(\mathbf{t}^T\mathbf{b} + \mathbf{t}^T(\mathbf{A}\boldsymbol{\mu}) + \frac{1}{2}\mathbf{t}^T(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)\mathbf{t}\right) \\ &= \exp\left(\mathbf{t}^T(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) + \frac{1}{2}\mathbf{t}^T(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)\mathbf{t}\right) \end{aligned}$$

This final expression is the MGF of a multivariate normal distribution. By the uniqueness property of MGFs, we can identify the mean and covariance matrix by comparing our result to the general form $M(\mathbf{t}) = \exp(\mathbf{t}^T\boldsymbol{\mu}_{\text{new}} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}_{\text{new}}\mathbf{t})$.
We find:

- New Mean: $\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{A}\boldsymbol{\mu} + \mathbf{b} = \mathbf{A}E[\mathbf{X}] + \mathbf{b}$
- New Covariance: $\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$

Therefore, we have shown that $\mathbf{Y} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.

Q5: Let $\mathbf{X} = [X_1, X_2]^T$ be a bivariate normal random vector with mean $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and covariance $\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$. Find a constant a such that the random variables $Y_1 = X_1 + aX_2$ and $Y_2 = X_2$ are independent.

Solution:

For jointly normal random variables, independence is equivalent to being uncorrelated. So, we need to find the value of a that makes Y_1 and Y_2 uncorrelated, which means their covariance must be zero: $Cov(Y_1, Y_2) = 0$.

We can compute the covariance:

$$\begin{aligned} Cov(Y_1, Y_2) &= Cov(X_1 + aX_2, X_2) \\ &= Cov(X_1, X_2) + Cov(aX_2, X_2) \quad \text{(by bilinearity of covariance)} \\ &= Cov(X_1, X_2) + a \cdot Cov(X_2, X_2) \\ &= Cov(X_1, X_2) + a \cdot Var(X_2) \end{aligned}$$

From the given covariance matrix Σ, we have $Cov(X_1, X_2) = 2$ and $Var(X_2) = 1$. Substituting these values into the equation:

$$Cov(Y_1, Y_2) = 2 + a(1)$$

To make them uncorrelated, we set the covariance to zero:

$$2 + a = 0 \implies a = -2$$

Thus, when $a = -2$, the random variables $Y_1 = X_1 - 2X_2$ and $Y_2 = X_2$ are uncorrelated, and because they are linear combinations of jointly normal variables, they are also independent.

Q6: Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ belong to a bivariate normal distribution $\mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$.

Show that $x_1|x_2 \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$ where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

A: The marginal distribution of x_2 is given by $\mathcal{N}(\mu_2, \Sigma_2)$. Using conditional probability

$$p(x_1|x_2) = \frac{\mathcal{N}(x; \mu, \Sigma)}{\mathcal{N}(x_2; \mu_2, \Sigma_{22})}$$

$$p(x; \mu, \Sigma) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right]$$

$$
\begin{aligned}
&= p(x_1|x_2) = \frac{1}{\sqrt{(2\pi)\Sigma_{22}}} \exp\left[-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\Sigma_{22}}\right] \\
&= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}\frac{(x_2-\mu_2)^2}{\Sigma_{22}}\right] \\
&= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} + \frac{1}{2}\frac{(x_2-\mu_2)^2}{\Sigma_{22}}\right]
\end{aligned}
$$

For a 2x2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ the inverse is given by

$$
A^{-1} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}
$$

Using this and expanding

$$
\begin{aligned}
p(x_1|x_2) &= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\Bigg[-\frac{1}{2|\Sigma|}(\Sigma_{11}(x_2-\mu_2)^2 \\
&\quad +2\Sigma_{12}(x_2-\mu_2)(x_1-\mu_1) + \Sigma_{22}(x_1-\mu_1)^2) + \frac{1}{2}\frac{(x_2-\mu_2)^2}{\Sigma_{22}}\Bigg] \\
&= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\Bigg[-\frac{1}{2|\Sigma|\Sigma_{22}}(\Sigma_{12}^2(x_2-\mu_2)^2 \\
&\quad -2\Sigma_{12}\Sigma_{22}(x_2-\mu_2)(x_1-\mu_1) + \Sigma_{22}^2(x_1-\mu_1)^2)\Bigg] \\
&= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2|\Sigma|\Sigma_{22}}(\Sigma_{22}(x_1-\mu_1) - \Sigma_{12}(x_2-\mu_2))^2\right] \\
&= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\left[-\frac{\Sigma_{22}}{2|\Sigma|}\left(x_1-\mu_1-\frac{\Sigma_{12}}{\Sigma_{22}}(x_2-\mu_2)\right)^2\right] \\
&= \frac{1}{\sqrt{(2\pi)|\Sigma_{11}|}} \cdot \frac{\sqrt{|\Sigma_{22}|}}{\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}\left(\frac{x_1-(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))}{\sqrt{\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}}}\right)^2\right]
\end{aligned}
$$

Thus the mean and variance are

$$
\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)
$$

$$
\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
$$

Q7: Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector with mean $\mu = E[\mathbf{X}]$ and covariance matrix $\Sigma = \mathrm{Cov}(\mathbf{X})$. Let $a, b \in \mathbb{R}^n$ be fixed (non-random) vectors.

(a) Show that

$$
\mathrm{Var}(a^T\mathbf{X}) = a^T\Sigma a, \qquad \text{and} \qquad \mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X}) = a^T\Sigma b.
$$

(b) Using

$$
\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}, \quad a = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 3 \end{bmatrix},
$$

compute both $\mathrm{Var}(a^T\mathbf{X})$ and $\mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X})$ numerically.

Solution:

(a) Detailed Proof:

We start with the definitions of variance and covariance. For any random variables Y_1, Y_2,

$$\mathrm{Cov}(Y_1, Y_2) = E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])].$$

Let

$$Y_1 = a^T\mathbf{X}, \qquad Y_2 = b^T\mathbf{X}.$$

Then

$$E[Y_1] = a^T E[\mathbf{X}] = a^T\mu, \qquad E[Y_2] = b^T E[\mathbf{X}] = b^T\mu.$$

Now,

$$\mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X}) = E[(a^T\mathbf{X} - a^T\mu)(b^T\mathbf{X} - b^T\mu)].$$

Simplifying,

$$\mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X}) = E[a^T(\mathbf{X} - \mu)(b^T(\mathbf{X} - \mu))].$$

Since $(b^T(\mathbf{X} - \mu))$ is a scalar, we can rewrite it as

$$(a^T(\mathbf{X} - \mu))(b^T(\mathbf{X} - \mu)) = a^T(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T b.$$

Therefore,

$$\mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X}) = E[a^T(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T b] = a^T E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]b.$$

By definition of covariance matrix, $E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] = \Sigma$. Hence,

$$\boxed{\mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X}) = a^T\Sigma b.}$$

If we take $a = b$, then this reduces to

$$\mathrm{Var}(a^T\mathbf{X}) = a^T\Sigma a.$$

Alternative viewpoint: By the bilinearity of covariance,

$$\mathrm{Var}(a^T\mathbf{X} + b^T\mathbf{X}) = \mathrm{Var}(a^T\mathbf{X}) + \mathrm{Var}(b^T\mathbf{X}) + 2\,\mathrm{Cov}(a^T\mathbf{X}, b^T\mathbf{X}),$$

and since $\mathrm{Var}(c^T\mathbf{X}) = c^T\Sigma c$ for any c, comparing terms yields $\mathrm{Cov}(a^T\mathbf{X}, b^T\mathbf{X}) = a^T\Sigma b$ as before.

(b) Numerical Computation:

Given

$$\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}, \quad a = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Step 1: Compute Σa and Σb.

$$\Sigma a = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 8 - 1 \\ 2 - 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \end{bmatrix}.$$

$$\Sigma b = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 + 3 \\ 1 + 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}.$$

Step 2: Compute $\mathrm{Var}(a^T\mathbf{X})$ **and** $\mathrm{Cov}(a^T\mathbf{X}, b^T\mathbf{X})$**.**

$$\mathrm{Var}(a^T\mathbf{X}) = a^T\Sigma a = \begin{bmatrix}2 & -1\end{bmatrix}\begin{bmatrix}7\\0\end{bmatrix} = 14.$$

$$\mathrm{Cov}(a^T\mathbf{X}, b^T\mathbf{X}) = a^T\Sigma b = \begin{bmatrix}2 & -1\end{bmatrix}\begin{bmatrix}7\\7\end{bmatrix} = 14 - 7 = 7.$$

Final Answers:

$$\boxed{\mathrm{Var}(a^T\mathbf{X}) = 14, \qquad \mathrm{Cov}(a^T\mathbf{X},\, b^T\mathbf{X}) = 7.}$$

Q8: Let $(\mathbf{X}, \mathbf{Y})$ be a jointly distributed random vector in $\mathbb{R}^{n+m}$, where

$$\mathbf{X} = \begin{bmatrix}X_1\\ \vdots \\ X_n\end{bmatrix} \in \mathbb{R}^n, \qquad \mathbf{Y} = \begin{bmatrix}Y_1\\ \vdots \\ Y_m\end{bmatrix} \in \mathbb{R}^m,$$

and assume all components have finite second moments. Define the conditional mean and conditional covariance of $\mathbf{Y}$ given $\mathbf{X}$ as

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} := E[\mathbf{Y} \mid \mathbf{X}], \qquad \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}} := Cov(\mathbf{Y} \mid \mathbf{X}) = E\big[(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})^T \mid \mathbf{X}\big].$$

(a) Show that the unconditional mean of $\mathbf{Y}$ can be expressed as

$$E[\mathbf{Y}] = E[\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}] = E[E[\mathbf{Y} \mid \mathbf{X}]].$$

(b) Prove that the covariance matrix of $\mathbf{Y}$ satisfies

$$Cov(\mathbf{Y}) = E[\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}] + Cov(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}).$$

(c) Suppose $\mathbf{Y}$ is given by the linear model

$$\mathbf{Y} = A\mathbf{X} + \mathbf{Z},$$

where A is an $(m \times n)$ constant matrix, and $\mathbf{X}$ and $\mathbf{Z}$ are independent random vectors satisfying

$$E[\mathbf{X}] = \mathbf{0}, \quad Cov(\mathbf{X}) = \boldsymbol{\Sigma}_{\mathbf{X}}, \qquad E[\mathbf{Z}] = \mathbf{0}, \quad Cov(\mathbf{Z}) = \boldsymbol{\Sigma}_{\mathbf{Z}}.$$

Using parts (a) and (b), derive expressions for $E[\mathbf{Y}]$ and $Cov(\mathbf{Y})$. Then compute $Cov(\mathbf{Y})$ numerically for

$$A = \begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}, \qquad \boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix}4 & 1\\1 & 3\end{bmatrix}, \qquad \boldsymbol{\Sigma}_{\mathbf{Z}} = \begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}.$$

Solution:

(a) By definition of conditional expectation,

$$E[\mathbf{Y}] = \int \mathbf{y}\, f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y} = \int E[\mathbf{Y} \mid \mathbf{X} = \mathbf{x}]\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = E[E[\mathbf{Y} \mid \mathbf{X}]] = E[\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}].$$

Hence, the overall mean equals the expected conditional mean:

$$\boxed{E[\mathbf{Y}] = E[E[\mathbf{Y} \mid \mathbf{X}]]}.$$

(b) Start from the definition of covariance:

$$Cov(\mathbf{Y}) = E\big[(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])^T\big].$$

Add and subtract $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}$:

$$\mathbf{Y} - E[\mathbf{Y}] = (\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}) + (\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}]).$$

Then expand the product:

$$\begin{aligned} Cov(\mathbf{Y}) = {} & E\big[(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})^T\big] \\ & + E\big[(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}])^T\big] \\ & + E\big[(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}])(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})^T\big] \\ & + E\big[(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}])(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}])^T\big]. \end{aligned}$$

Since $E[\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} \mid \mathbf{X}] = 0$, the second and third terms vanish. Thus,

$$Cov(\mathbf{Y}) = E\big[E[(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})^T \mid \mathbf{X}]\big] + E\big[(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}])(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} - E[\mathbf{Y}])^T\big].$$

The first term equals $E[\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}]$ and the second equals $Cov(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})$. Therefore,

$$\boxed{Cov(\mathbf{Y}) = E[\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}] + Cov(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}}).}$$

(c) For $\mathbf{Y} = A\mathbf{X} + \mathbf{Z}$ with $\mathbf{X}$ and $\mathbf{Z}$ independent and zero mean:

$$E[\mathbf{Y} \mid \mathbf{X}] = A\mathbf{X} + E[\mathbf{Z} \mid \mathbf{X}] = A\mathbf{X}, \qquad Cov(\mathbf{Y} \mid \mathbf{X}) = Cov(\mathbf{Z}) = \boldsymbol{\Sigma}_{\mathbf{Z}}.$$

By part (a),

$$E[\mathbf{Y}] = E[E[\mathbf{Y} \mid \mathbf{X}]] = AE[\mathbf{X}] = \mathbf{0}.$$

By part (b),

$$Cov(\mathbf{Y}) = E[Cov(\mathbf{Y} \mid \mathbf{X})] + Cov(E[\mathbf{Y} \mid \mathbf{X}]) = \boldsymbol{\Sigma}_{\mathbf{Z}} + Cov(A\mathbf{X}) = \boldsymbol{\Sigma}_{\mathbf{Z}} + A\boldsymbol{\Sigma}_{\mathbf{X}}A^T.$$

Numerical computation:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{Z}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Compute

$$A\boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 6 & 7 \\ 1 & 3 \end{bmatrix}, \qquad A\boldsymbol{\Sigma}_{\mathbf{X}}A^T = \begin{bmatrix} 20 & 7 \\ 7 & 3 \end{bmatrix}.$$

Thus,

$$Cov(\mathbf{Y}) = A\boldsymbol{\Sigma}_{\mathbf{X}}A^T + \boldsymbol{\Sigma}_{\mathbf{Z}} = \begin{bmatrix} 21 & 7 \\ 7 & 5 \end{bmatrix}.$$

Final Results:

$$E[\mathbf{Y}] = \mathbf{0}, \qquad Cov(\mathbf{Y}) = \boxed{\begin{bmatrix} 21 & 7 \\ 7 & 5 \end{bmatrix}}.$$