# UNIT 14 SOME LINEAR ALGEBRA

**Structure** **Page No**



## 14.1 INTRODUCTION

In this unit, we discuss some concepts from linear algebra which will be useful in the study of multivariate statistical analysis. We look, in some detail, at idempotent matrices and quadratic forms. In the context of this course we start with the study of real symmetric matrices and the associated quadratic forms in Sec. 14.2. We define a classification for the quadratic forms and develop a method for determining the class to which a given quadratic form belongs.

In Sec.14.3, we study positive definite and semi definite matrices. In this Section we also obtain some characterizations of positive definite and semi definite matrices, and study some of their useful properties. Here we give a method of computing a square root of matrices; this plays an important role in transforming correlated random variables to uncorrelated random variables.

Idempotent matrices and Cochran's theorem play a key role in the distribution of quadratic forms in independent standard normal variables, particularly, in connection with the distribution of quadratic forms to become independent chi-squares. In Sections 14.4 and 14.5, respectively we study the properties of idempotent matrices, and prove the algebraic version of Cochran's theorem.

Singular value decompositions plays a very important role in developing the theory and studying the properties of canonical correlations between two random vectors. In Section 14.6, we study the singular value decomposition.

In this Unit, we shall be using the following notations. Matrices are denoted by capital letters like $A, B, C$. Vectors are denoted by boldface lower case letters like $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Scalars are denoted by lower case letters like $a, b, \alpha$. The transpose, rank and trace of a matrix $A$ are denoted by $A'$ or $A^t$, rank $A$ and $tr(A)$, respectively. $R^n$ denotes the n-dimensional Euclidean space.

### Objectives

After studying this unit, you should be able to

- determine the definiteness of a given quadratic form;
- apply the spectral decomposition in the study of principal components;
- compute a triangular square root of a positive definite matrix;
- apply the properties of positive and semi definite matrices to certain problems;

- apply Cochran's theorem to the distribution of quadratic forms in normal variables;
- apply the singular value decomposition in the development of canonical correlations.

Let us start our discussion with real symmetric matrices.

## 14.2 REAL SYMMETRIC MATRICES

Real symmetric matrices play a very important role in the study of multivariate statistical analysis. For example, the variance-covariance matrices are real symmetric matrices. They also play a crucial role in the distribution of quadratic forms in correlated normal random variables. We shall denote $(i, j)^{th}$ element of a matrix $A$ by $a_{ij}$. Then we write $A = (a_{ij})$.

**Definition 1:** A square matrix $A = (a_{ij})$ of order $n$ is called a **real symmetric matrix** if (i) all the elements of $A$ are real and (ii) $a_{ij} = a_{ji}$ for $i, j = 1, \ldots, n$.

**Definition 2:** A quadratic form in $n$ variables $x_1, x_2, \ldots, x_n$ is a homogeneous polynomial of degree 2 in these variable.

You also know that there is a unique real symmetric matrix $A$ associated with a given real quadratic form $\mathbf{Q(x)}$, in the sense that $\mathbf{Q(x) = x'Ax}$. This matrix $A$ is called the **matrix of the quadratic form $\mathbf{Q(x)}$**. (For reference, MTE-02, Sec. 14.3)

**Example 1:** Examine the following matrices for symmetric property

(i) $\begin{bmatrix} 2 & i \\ i & 3 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$ (iii) $\begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$ (iv) $\begin{bmatrix} -1 & 4 \\ 4 & 5 \end{bmatrix}$

**Solution:**

(i) is not a real symmetric matrix because all of its elements are not real.

(ii) is not a real symmetric matrix because it is not a square matrix.

(iii) is not a real symmetric matrix because $a_{12} = 3$ and $a_{21} = 2$ therefore $a_{12} \neq a_{21}$.

(iv) is a real symmetric matrix because (a) it is a square matrix (of order $2 \times 2$), (b) all of its elements are real and (c) $a_{12} = 4 = a_{21}$.

***

**Example 2:** Find the matrix of the quadratic form $\mathbf{Q(x)} = 2x_1x_2 + 5x_1x_3 + 3x_2^2 - x_3^2$.

**Solution:** Since there are three variables $x_1$, $x_2$ and $x_3$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. Let $A$ be the symmetric matrix such that $Q(\mathbf{x}) = \mathbf{x'Ax}$. Then $A$ is of order $3 \times 3$. Further $a_{21} = a_{12} = 1/2$ (coefficient of $x_1 x_2$ ) $= 1$. In general, whenever $i \neq j$, $a_{ij} = a_{ji} = 1/2$ (coefficient of $x_i x_j$). Also $a_{ii}$ = the coefficient of $x_i^2$, $i = 1, 2, 3$. Thus

$$A = \begin{bmatrix} 0 & 1 & 2.5 \\ 1 & 3 & 0 \\ 2.5 & 0 & -1 \end{bmatrix}$$

***

Let us now try an exercise.

E1) Find the matrices of the following quadratic forms:

(i) $x_1^2 - x_2^2$

(ii) $2x_1^2 + 3x_1x_2 + 5x_2^2$

(iii) $3x_1x_2 + 5x_2x_3 - 4x_1x_3$

(iv) $x_1^2 + x_2^2 + x_2x_4$ (in four variables $x_1$, $x_2$, $x_3$ and $x_4$)

Depending upon the range, every non null quadratic form $Q(\mathbf{x})$ in $n$ variables can be classified into one of the following mutually exclusive and collectively exhaustive classes:
(It is also said to be identification of the definiteness of the quadratic form.)

(a) positive definite (pd) if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbf{R}^n, \mathbf{x} \neq \mathbf{0}$,

(b) positive semidefinite (psd) if
(i) $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbf{R}^n$, and
(ii) $Q(\mathbf{x}) = 0$ for some $\mathbf{x} \neq \mathbf{0}$

(c) negative definite (nd) if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$ (i.e., if $-Q(\mathbf{x})$ is positive definite),

(d) negative semidefinite (nsd) if (i) $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ and (ii) $Q(\mathbf{x}) = 0$ for some $\mathbf{x} \neq \mathbf{0}$. (i.e., if $-Q(\mathbf{x})$ is positive semidefinite),

(e) indefinite, if it does not belong to any one of the above classes (a) – (d) (i.e., there exists $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^n$ such that $Q(\mathbf{x}) > 0$ and $Q(\mathbf{y}) < 0$)).

The quadratic form $Q(\mathbf{x}) \equiv 0$ can be classified into any one of the classes (b) and (d).

**Example 3**: Classify each of the following quadratic forms using the above classification. Also write down the matrices of the respective quadratic forms.

(i) $x_1^2 - x_2^2$ (ii) $x_1^2 + x_2^2$

(iii) $x_1^2 + x_2^2 + 2x_4^2$ (in four variables $x_1$, $x_2$, $x_3$ and $x_4$)

(iv) $-x_1^2 - x_2^2$

(v) $-x_1^2 - x_2^2 - 2x_4^2$ (again in four variables)

**Solution:**

(i) $x_1^2 - x_2^2 = 1$ if $x_1 = 1$ and $x_2 = 0$. Again $x_1^2 - x_2^2 = -1$ if $x_1 = 1$ and $x_2 = 1$. Thus it is indefinite. The matrix of the quadratic form is $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

(ii) $x_1^2 + x_2^2 > 0$ whenever at least one of $x_1$ and $x_2$ is not zero. Hence this quadratic form is positive definite. The matrix of the quadratic form is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, the identity matrix.

(iii) $Q(\mathbf{x}) = x_1^2 + x_2^2 + 2x_4^2 \geq 0$ for all values of $x_1$, $x_2$, $x_3$ and $x_4$. However, for $x_3 = 1$ and $x_1 = x_2 = x_4 = 0$, the value of $x_1^2 + x_2^2 + 2x_4^2 = 0$. Thus, there is a vector $\mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \neq \mathbf{0}$ such that $Q(\mathbf{x}) = 0$. Hence this quadratic form is

positive semi-definite. The matrix of the quadratic form is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

We leave it to you to show that the quadratic forms in (iv) and (v) are negative definite and negative semi-definite, respectively. (You can use the quadratic forms in (ii) and (iii) to arrive at this conclusion, and for writing down the matrices of the quadratic forms in (iv) and (v).)

***

In Example 3, we considered quadratic forms whose matrices are diagonal matrices. Here it is easy to identify the definiteness of the quadratic form. In fact, if $Q(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i x_i^2$ is a quadratic form in $n$ variables $x_1, \ldots, x_n$, then $Q(\mathbf{x})$ is p.d., p.s.d., n.d., n.s.d. or indefinite according as $\lambda_i > 0$ for all i; $\lambda_i \geq 0$ for all $i$ and $\lambda_j = 0$ for some j; $\lambda_i < 0$ for all i; $\lambda_i \leq 0$ for all $i$ and $\lambda_j = 0$ for some $j$; or if $\lambda_i < 0$ for some $i$ and $\lambda_j > 0$ for some $j$, respectively.

Now, what if we have a quadratic form $Q(\mathbf{x}) = 2x_1^2 - 3x_1x_2 + x_2^2$ or $Q(\mathbf{x}) = 2x_1^2 + x_2^2 + x_3^2 - 3x_1x_2 - 2x_1x_3 + 4x_2x_3$? (Notice that the matrices of these quadratic forms are $\begin{bmatrix} 2 & -1.5 \\ -1.5 & 1 \end{bmatrix}$ and $\begin{bmatrix} 2 & -1.5 & -1 \\ -1.5 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix}$, respectively, and are not diagonal matrices.)

In general, consider a quadratic form $Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$, where $A$ is not a diagonal matrix. How do we determine the definiteness of the quadratic form in such a case? The following results will be useful towards that end.

**Theorem 1:** Consider a quadratic form $Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$ where $A$ is symmetric. Make a nonsingular linear transformation of the variables: $\mathbf{y} = T\mathbf{x}$ (where $T$ is nonsingular). Call the transformed quadratic form as $\psi(\mathbf{y})(= \mathbf{y}'T^{-1'}AT^{-1}\mathbf{y})$. Then the ranges of $Q(\mathbf{x})$ and $\psi(\mathbf{y})$ are the same.

**Proof:** Let $\alpha$ belongs to the range of $Q(\mathbf{x})$. So, there is a vector $\mathbf{x}_0$ such the $\alpha = Q(\mathbf{x}_0) = \mathbf{x}_0'A\mathbf{x}_0$. Write $y_0 = T\mathbf{x}_0$. Now $\alpha = \mathbf{x}_0'A\mathbf{x}_0 = \mathbf{x}_0'T'T^{-1'}AT^{-1}T\mathbf{x}_0 = \mathbf{y}_0'T^{-1'}AT^{-1}\mathbf{y}_0 = \psi(\mathbf{y}_0)$. Hence $\alpha$ belongs to the range of $\psi(\mathbf{y})$. Thus, the range of $Q(\mathbf{x})$ is a subset of the range of $\psi(\mathbf{y})$. Since $T$ is nonsingular, by reversing the arguments, we can show that the range of $\psi(\mathbf{y})$ is a subset of the range of $Q(\mathbf{x})$. The proof is complete.

What we are saying through Theorem 1 is that the range of a quadratic form is invariant under nonsingular linear transformations. Thus, the definiteness of a quadratic form is invariant under nonsingular linear transformations. (Making a nonsingular linear transformation can also be interpreted as changing the basis.)

Recall that a real square matrix $S$ is called an **orthogonal matrix** if $S' = S^{-1}$. If $S$ and $T$ are orthogonal matrices of the same order, then so is $ST$. To see this we note that $T'S'ST = T'IT = I$. Similarly, $STT'S' = I$. Hence $(ST)' = T'S'$ is the inverse of $ST$.

Also, you should verify that $\begin{bmatrix} I & 0 \\ 0 & T \end{bmatrix}$ is an orthogonal matrix if $T$ is an orthogonal matrix.

So, now we want to determine the definiteness of a quadratic form $Q(\mathbf{x})$, the matrix of which is not necessarily diagonal. We shall now show that we can make an orthogonal transformation of the variables (i.e., we can make a transformation $\mathbf{y} = P\mathbf{x}$ where $P$ is an orthogonal matrix) such that under this transformation, the quadratic form is transformed into a quadratic form $\sum \lambda_i y_i^2$. Since we know how to determine the definiteness of $\sum \lambda_i y_i^2$, and since the definiteness of $\sum \lambda_i y_i^2$ is the same as that of $Q(\mathbf{x})$, we would then have the definiteness of $Q(\mathbf{x})$.

If $A$ is a real matrix, then it is not necessary that its eigenvalues are real. For example if $A = \begin{bmatrix} 0 & 1 \\ -1 & \end{bmatrix}$, then the eigenvalues are $i$ and $-i$. However, if $A$ is real, and symmetric then all its eigenvalues are real as shown below.

**Theorem 2:** Let $A$ be a real symmetric matrix. All the eigenvalues of $A$ are real and all the eigenvectors of $A$ can be chosen to be real.

**Proof:** Let $\lambda + i\mu$ be an eigenvalue of $A$ and let the corresponding eigen vector be $\mathbf{x} + i\mathbf{y}$, where $\lambda$, $\mu$ are real numbers and $\mathbf{x}, \mathbf{y}$ are real vectors. Clearly at least one of $\mathbf{x}$ and $\mathbf{y}$ is non-zero as $\mathbf{x} + i\mathbf{y}$, being an eigen vector is nonnull. Now,

$$A(\mathbf{x} + i\mathbf{y}) = (\lambda + i\mu)\,(\mathbf{x} + i\mathbf{y})$$

Equating the real and the imaginary parts on both sides of the above equality, we get,

$$A\mathbf{x} = \lambda\mathbf{x} - \mu\mathbf{y} \qquad (1)$$

$$A\mathbf{y} = \lambda\mathbf{y} + \mu\mathbf{x} \qquad (2)$$

Premultiplying (1) by $\mathbf{y}'$ and (2) by $\mathbf{x}'$, we get

$$\mathbf{y}'A\mathbf{x} = \lambda\mathbf{y}'\mathbf{x} - \mu\mathbf{y}'\mathbf{y} \qquad (3)$$

$$\mathbf{x}'A\mathbf{y} = \lambda\mathbf{x}'\mathbf{y} + \mu\mathbf{x}'\mathbf{x} \qquad (4)$$

Since $A$ is symmetric and $\mathbf{y}'A\mathbf{x}$ is a scalar we have $\mathbf{y}'A\mathbf{x} = (\mathbf{y}'A\mathbf{x})' = \mathbf{x}'A'\mathbf{y} = \mathbf{x}'A\mathbf{y}$. Similarly, $\mathbf{y}'\mathbf{x} = \mathbf{x}'\mathbf{y}$. Now subtracting Eqn.(3) from Eqn.(4) we get $\mu(\mathbf{x}'\mathbf{x} + \mathbf{yy}) = 0$. Since at least one of $x$ and $y$ is non-null, $\mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y} \neq 0$. So, $\mu = 0$.

Hence all the eigenvalues of $A$ are real. Further $A(\mathbf{x} + i\mathbf{y}) = \lambda(\mathbf{x} + i\mathbf{y})$, yields $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \lambda\mathbf{y}$. Since at least one of $\mathbf{x}$ and $\mathbf{y}$ is non-zero and $\mathbf{x}$, $\mathbf{y}$ are real, we can choose a non-zero vector among $\mathbf{x}$, $\mathbf{y}$ as an eigenvector of $A$ corresponding to $\lambda$. This complete the proof of the theorem.

Now we are ready to prove the result we had mentioned earlier in the following theorem.

**Theorem 3:** Let $A$ be a real symmetric matrix of order $n$. Then there exists a real orthogonal matrix $P$ of order $n$ such that $A = P\Lambda P'$, where $\Lambda$ is a real diagonal matrix.

**Proof:** We shall prove the theorem by induction on $n$. Let $A$ be a $1\times 1$ real symmetric matrix, i.e. $A = (\alpha)$, where $\alpha$ is a real number. Clearly $(1)'(\alpha)(1) = (1.\alpha.1) = (\alpha)$. Also $(1)$ is an orthogonal matrix of order $1\times 1$ since $(1)'.(1) = (1.1) = (1)$. So the theorem is true for $n = 1$. Let the theorem be true $n = m$ (a positive integer $> 1$). Let $A$ be a matrix of order $(m+1)\times(m+1)$. Let $x_1$ be a normalized eigenvector of $A$ corresponding to eigenvalue $\lambda_1$. Then $A\mathbf{x_1} = \lambda_1\mathbf{x_1}$. Now $\mathbf{x_1}$ can be extended to an orthonormal basis $\mathbf{x_1},\ldots,\mathbf{x_{m+1}}$ of $\mathbf{R}^{m+1}$. Write $R = (\mathbf{x_1} : \ldots : x_{m+1})$. Clearly $R$ is an orthogonal matrix. Now

$$AR = \mathbf{A}(\mathbf{x}_1 : \ldots : \mathbf{x}_{m+1}) = (\mathbf{x}_1 : \ldots : \mathbf{x}_{m+1})\begin{pmatrix} \lambda_1 & \mathbf{b}'_{12} \\ \mathbf{0} & B_{22} \end{pmatrix} = R\begin{pmatrix} \lambda_1 & \mathbf{b}'_{12} \\ \mathbf{0} & B_{22} \end{pmatrix},$$ where $\mathbf{0}$, $\mathbf{b}'_{12}$ and $B_{22}$ are of order $m\times 1$, $1\times m$ and $m\times m$, respectively. [This is because, $A\mathbf{x}_2,\ldots, A\mathbf{x}_{m+1}$ are vectors in $R^{m+1}$ and $\mathbf{x}_1,\ldots,\mathbf{x}_{m+1}$ form a basis of $R^{m+1}$, so $\mathbf{Ax}_i$ is a linear combination of $\mathbf{x}_1,\ldots,\mathbf{x}_{m+1}$.]

Therefore $R'AR = \begin{pmatrix} \lambda_1 & \mathbf{b}'_{12} \\ \mathbf{0} & B_{22} \end{pmatrix}$. Since $R'AR$ is real and symmetric it follows that

$\mathbf{b}_{12} = \mathbf{0}$ and $B_{22}$ is an $m\times m$ real symmetric matrix. Thus, $R'AR = \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & B_{22} \end{pmatrix}$. By induction hypothesis there exists an orthogonal matrix $S_1$ of order $m\times m$ that $B_{22} = S_1\Gamma S_1'$ where $\Gamma$ is a real diagonal matrix. Writing $S = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & S_1 \end{pmatrix}$, we notice that $S$ is an orthogonal matrix and

$$R'AR = S\begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \Gamma \end{pmatrix}S^t \text{ or } A = RS\begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \Gamma \end{pmatrix}S'R'$$

Write $P = RS$. Since $R$ and $S$ are orthogonal matrices so is $P$ as noticed earlier. Writing $D = \begin{pmatrix} \lambda & \mathbf{0} \\ \mathbf{0} & \Gamma \end{pmatrix}$, we observe that $D$ is a real diagonal matrix.

Thus, the theorem is true for $n = m+1$.

Hence the theorem follows by induction on $n$.

The beauty of Theorem 3 lies in its interpretation. Let $A$ be a real symmetric matrix and let $A = P\Lambda P'$ where $P$ is orthogonal and $\Lambda$ is a real diagonal matrix. We then have

$$AP = P\Lambda \quad \text{or} \quad A(\mathbf{p}_1 : \ldots : \mathbf{p}_n) = (\mathbf{p}_1 : \ldots : \mathbf{p}_n)\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \quad \text{or} \quad .$$

$AP_i = \lambda_i\mathbf{p_i}, i = 1,\ldots,n$.

Since $\mathbf{p}_1$ is a vector in an orthonormal basis, $\mathbf{p}_1$ is (a non-null vector) of unit norm. Hence $\lambda_1$ is an eigenvalue of $A$ and $\mathbf{p}_1$ is an eigenvector of $A$ corresponding to the eigenvalue $\lambda_1$. Thus, the diagonal elements of $\Lambda$ are the eigenvalues of $A$ and the columns of $P$ are the orthonormal eigenvectors of $A$. Further

$$A = P\Lambda P' = (\mathbf{p}_1 : \ldots : \mathbf{p}_n)\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}\begin{pmatrix} \mathbf{p}_1' \\ \vdots \\ \mathbf{p}_n' \end{pmatrix} = \lambda_1 \mathbf{p}_1 \mathbf{p}_1' + \cdots + \lambda_n \mathbf{p}_n \mathbf{p}_n'$$

Write $E_i = \mathbf{p}_i \mathbf{p}_i'$, $i = 1, \ldots, n$

Then $E_i E_j = \begin{cases} E_i \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}$

and rank $E_i$ = rank $\mathbf{p}_i \mathbf{p}_i'$ = rank $p_i = 1$.

Thus, we are able to write $A = \sum_{i=1}^{n} \lambda_i E_i$ where $E_1, \ldots, E_n$ are symmetric idempotent matrices of rank 1 such that $E_i E_j = 0$ whenever $i \neq j$. The set $\{\lambda_1, \ldots, \lambda_n\}$ is called the spectrum of $A$. Since the decomposition mentioned above involves the spectrum and the eigenvectors it is called a spectral decomposition of $A$.

**Example 4:** Obtain a spectral decomposition of the matrix $A = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$.

**Solution:** The characteristic equation of the matrix is $\begin{vmatrix} 4-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0$

or $(4-\lambda)(2-\lambda) - 1 = 0$

or $\lambda^2 - 6\lambda + 7 = 0$

The roots are $\dfrac{6 \pm \sqrt{36-28}}{2} = 3 \pm \sqrt{2}$

So the eigenvalues are $3+\sqrt{2}$ and $3-\sqrt{2}$.

Let $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ be the eigenvector of the given matrix corresponding to the eigenvalue $3+\sqrt{2}$.

Then $[A-(3+\sqrt{2})I]\mathbf{u} = 0$ or $\begin{bmatrix} 1-\sqrt{2} & 1 \\ 1 & -1-\sqrt{2} \end{bmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \mathbf{0}$

Notice that the second column of $\left(A-(3+\sqrt{2})I\right)$ is $\left(-1-\sqrt{2}\right)$ times the first column.

So, $u_1 = 1+\sqrt{2}$ and $u_2 = 1$ satisfy the equation $\left(A-(3+\sqrt{2})I\right)\mathbf{u} = 0$.

To normalize $\mathbf{u}$, we divide it by its norm namely $\sqrt{\left(1+\sqrt{2}\right)^2 + 1^2} = \sqrt{4+2\sqrt{2}}$. Thus, the normalized eigenvector corresponding to the eigenvalue $3+\sqrt{2}$ is

$\dfrac{1}{\sqrt{4+2\sqrt{2}}}\begin{pmatrix} 1+\sqrt{2} \\ 1 \end{pmatrix}$. It can be shown similarly that the normalized eigenvector

corresponding to $3-\sqrt{2}$ is $\dfrac{1}{\sqrt{4+2\sqrt{2}}}\begin{pmatrix} 1 \\ -\left(1+\sqrt{2}\right) \end{pmatrix}$. Hence $A = P\Lambda P'$

where $P = \dfrac{1}{\sqrt{4+2\sqrt{2}}}\begin{pmatrix} 1+\sqrt{2} & 1 \\ 1 & -\left(1+\sqrt{2}\right) \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 3+\sqrt{2} & 0 \\ 0 & 3-\sqrt{2} \end{pmatrix}$, is the required spectral decomposition.

***

Using Theorem 3, we can determine the definiteness of a quadratic form. Consider the quadratic form $Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$. Let $A = P\Lambda P'$ be a spectral decomposition of $A$. Then $Q(\mathbf{x}) = \mathbf{x}'P\Lambda P'\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y}$ where $\mathbf{y} = P'\mathbf{x}$. Since $P$ is nonsingular (in fact, orthogonal) the definiteness of $Q(\mathbf{x})$ is the same as the definiteness of $\mathbf{y}'\Lambda\mathbf{y}$. The definiteness of $\mathbf{y}'\Lambda\mathbf{y}$ is determined by the diagonal elements $\lambda_1,\ldots,\lambda_n$ in, $\Lambda$ the eigen values of $A$.

Thus

$$\mathbf{x}'A\mathbf{x} \text{ is } \begin{cases} \text{positive definite if } \lambda_i > 0 \text{ for all i} \\ \text{positive semidefinite if } \lambda_i \geq 0\ \forall i \text{ and } \lambda_j = 0 \text{ for some j} \\ \text{negative definite if } \lambda_i < 0 \text{ for all i} \\ \text{negative semidefinite if } \lambda_i \leq 0 \text{ for all i and } \lambda_j = 0 \text{ for some j} \\ \text{indefinite if } \lambda_i > 0 \text{ for some i and } \lambda_j = 0 \text{ for some j} \end{cases}$$

Because of the one-one correspondence between real symmetric matrices and the quadratic forms, we call a real symmetric matrix $A$ as positive definite, positive semidefinite, negative definite, negative semidefinite or indefinite accordingly as the corresponding quadratic form $\mathbf{x}'A\mathbf{x}$ is positive definite, positive semidefinite, negative definite, negative semidefinite or indefinite, respectively.

Now let us illustrate the following example as an application of this.

**Example 5:** Determine the definiteness of the quadratic forms (i) $2x_1^2 - x_1x_2 + x_2^2$ and (ii) $x_1^2 + x_2^2 + x_3^2 - 3x_1x_2 - 3x_1x_3 - 3x_2x_3$.

**Solution:** (i) The matrix of the quadratic form is $A = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 1 \end{pmatrix}$.

We see that its characteristic equation $|A - \lambda I| = 0$ is

$(2-\lambda)(1-\lambda) - 0.25 = 0$ or $\lambda^2 - 3\lambda + 1.75 = 0$

Hence the eigenvalues which are the roots of the above equation are $\frac{3 \pm \sqrt{9-7}}{2}$ or $\frac{3+\sqrt{2}}{2}$ and $\frac{3-\sqrt{2}}{2}$, which are positive.

Hence the quadratic form is positive definite.

(ii) The matrix of the quadratic form is $A = \begin{bmatrix} 1 & -1.5 & -1.5 \\ -1.5 & 1 & -1.5 \\ -1.5 & -1.5 & 1 \end{bmatrix}$.

It is easy to notice that the sum of each row in $A$ is $-2$.

Hence $A\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Thus, $-2$ is an eigenvalue of $A$. Further, the sum of the eigenvalues which is the same as the trace of $A$ is $3$. Hence there must be at least one positive eigenvalue of $A$. So the quadratic form is indefinite.

***

Now, try the following exercises.

---

E2) Let $A$ be a real symmetric matrix, a diagonal element of which is negative. Show that $A$ cannot be positive definite or positive semidefinite.

E3) Determine the definiteness of the following quadratic forms:

(i) $x_1^2 - 5x_1x_2 + 7x_2^2$, (ii) $x_1^2 - x_2^2 + x_3^2 - x_1x_2 + 10x_1x_3 - 2x_2x_3$,

(iii) $2x_1^2 + 3x_2^2 + 4x_3^2 + 6x_1x_2$

E4) Let $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Obtain the spectral decomposition of $A$. Hence write down $A^{100}$.

In next section, we shall discuss more about positive definite and non-negative definite matrices.

## 14.3 POSITIVE DEFINITE AND SEMIDEFINITE MATRICES

In the previous section, we noted that the definiteness of a quadratic form is also attributed to the matrix of the quadratic form. Thus, if $\mathbf{x}'A\mathbf{x}$ is positive definite where $A$ is a real symmetric matrix, then we call $A$ as a positive definite (pd) matrix. A real symmetrix matrix is called a Semidefinite matrix if it is either pd or psd i.e., if $\mathbf{x}'A\mathbf{x} \geq \mathbf{0}$ for all $\mathbf{x}$. Positive definite (pd) and nonnegative definite (nnd) matrices play a very important role in the multivariate analysis. *Unless stated otherwise, when we say a matrix is pd, psd, nnd, nd, nsd we mean that the matrix is real and symmetric. We may not state this fact explicitly each time.*

In this section, we study several properties of positive definite and nonnegative definite matrices. We shall also give an easy way to construct positive definite matrices and orthogonal matrices of order $n \times n$. Let us start with the following very useful theorem.

**Theorem 4:** a) A matrix $A$ is positive definite if and only if $A = BB'$ for some nonsingular matrix $B$.

b) A matrix $A$ is nonnegative definite if and only if $A = CC'$ for some matrix $C$.

**Proof:** a) If part. Let $A = BB'$ for some nonsingular matrix $B$. Let $\mathbf{x}$ be a nonnull vector. Then $\mathbf{x}'A\mathbf{x} = \mathbf{x}'BB'\mathbf{x} = \mathbf{y}'\mathbf{y} = y_1^2 + \cdots + y_n^2 \geq 0$ where $\mathbf{y} = B'\mathbf{x}$. Since $B$ is nonsingular, so is $B'$.

Let if possible $\mathbf{y} = B'\mathbf{x} = \mathbf{0}$. Then $\mathbf{x} = (B')^{-1}y = 0$.

Since $\mathbf{x} \neq \mathbf{0}$, there is a contradiction. So $\mathbf{y} \neq \mathbf{0}$.

Hence $\mathbf{x}'A\mathbf{x} = y'y > 0$. The choice of $\mathbf{x}$ being arbitrary, it follows, that $A$ is positive definite. Only if part. Let $A$ be positive definite. Then all its eigenvalues are strictly positive. Let $A = P\Lambda P'$ be a spectral decomposition of $A$. Let $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$ be positive square roots of $\lambda_1, \ldots, \lambda_n$, respectively. Write

$$\Lambda^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}$$

Then $B = P\Lambda^{1/2}P'$ is symmetric and $BB' = P\Lambda^{1/2}P'P\Lambda^{1/2}P' = P\Lambda^{1/2}\Lambda^{1/2}P' = P\Lambda P' = A$.

Further since $\lambda_1, \ldots, \lambda_n$ are strictly positive, so are $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$. Now $|B| = |P\Lambda^{1/2}P'| = |P||P'||\Lambda^{1/2}| = |PP'||\Lambda^{1/2}| = |I||\Lambda^{1/2}| = \sqrt{\lambda_1 \ldots \lambda_n} > 0$. Hence $B$ is nonsingular.

[Notice that $B = P\Lambda^{1/2}P'$ is a spectral decomposition of $B$. Now since diagonal elements of $\Lambda^{1/2}$ are strictly positive it follows that $B$ is pd. In fact, we proved that more stronger result that if $A$ is pd then $A = BB'$ for some symmetric pd matrix $B$.]

(b) It suffices to prove the statement for positive semidefinite matrices as an nnd matrix is either pd or psd. (For the pd matrices we already proved the statement in (a).)

If part. Let $A = CC'$. Then $\mathbf{x}'A\mathbf{x} = \mathbf{x}'CC'\mathbf{x} = \mathbf{u}'\mathbf{u} \geq 0$ where $\mathbf{u} = C'\mathbf{x}$. Hence $A$ is nnd.

Only if part. Notice that since $A$ is psd all its eigenvalues are nonnegative. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0$ be the eigenvalues of $A$. We can write a spectral decomposition of $A$ as

$$A = P\begin{pmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}P'$$

where $P$ is an orthogonal matrix and $\Lambda_1$ is a diagonal pd matrix of order $r \times r$. Write

$$\Lambda_1^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_r} \end{pmatrix}.$$

Then

$\mathbf{C} = P\begin{pmatrix} \Lambda_1^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}P'$ is symmetric and $CC^t = A$. This complete the proof.

A matrix $B$ such that $A = BB'$ is called a square root of $A$. Given $A$, $B$ is not unique since $A = BPP'B'$ where $P$ is any orthogonal matrix. In Theorem 4 we gave a method of computing a square root if we know the spectral decomposition of $A$. However, obtaining spectral decomposition is not easy in general. We shall now discuss a method of obtaining a square root of a positive definite matrix.

Let us start with an example.

**Example 6:** Obtain a square root of the positive definite matrix $A = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 3 & 1 \\ 2 & 1 & 5 \end{pmatrix}$

**Solution:** We shall obtain a lower triangular matrix $B$ such that $A = BB'$. Write

$B = \begin{pmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$. We shall solve for $b_{ij}$ $j = i, \ldots, 3$, $i = 1, 2, 3$ such that $A = BB'$.

Write

$$\begin{pmatrix} 4 & 1 & 2 \\ 1 & 3 & 1 \\ 2 & 1 & 5 \end{pmatrix} = \begin{pmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix}\begin{pmatrix} b_{11} & b_{21} & b_{31} \\ 0 & b_{22} & b_{32} \\ 0 & 0 & b_{33} \end{pmatrix}$$

Equating the elements on both sides, we get

$a_{11} = 4 = b_{11}^2$ or $b_{11} = 2$ (You can choose either $+2$ or $-2$ but choose and fix one of them.)

$a_{12} = 1 = b_{11}\, b_{21}$ so $b_{21} = \dfrac{1}{b_{11}} = 1/2$

$a_{13} = 2 = b_{11}\ b_{31}$ so $b_{31} = 1$

$a_{22} = 3 = b_{21}^2 + b_{22}^2$ or $b_{22}^2 = 3 - \frac{1}{4} = \frac{11}{4}$. So $b_{22} = \sqrt{\frac{11}{4}}$.

$a_{23} = 1 = b_{21}\ b_{31} + b_{22}\ b_{32}$ or $b_{22}\ b_{32} = 1 - 1/2.1 = 1/2$ or $b_{32} = \frac{1}{2}\sqrt{\frac{4}{11}} = \frac{1}{\sqrt{11}}$.

$a_{33} = 5 = b_{31}^2 + b_{32}^2 + b_{33}^2$ or $b_{33}^2 = 5 - 1^2 - \left(\frac{1}{\sqrt{11}}\right)^2 = 5 - 1 - \frac{1}{11} = \frac{43}{11}$ or $b_{33} = \sqrt{\frac{43}{11}}$.

Thus, $B = \begin{pmatrix} 2 & 0 & 0 \\ \frac{1}{2} & \sqrt{\frac{11}{4}} & 0 \\ 1 & \frac{1}{\sqrt{11}} & \sqrt{\frac{43}{11}} \end{pmatrix}$ is square root of $A$.

***

For the given matrix $A$ in Example 6, we could obtain a lower triangular matrix $B$ such that $A = BB'$. Can we always do this? Let us examine how we went about in solving for the elements of $B$. First we solved for the first column of $B$, then for the second column and so on. Also observe that each time we just had to solve one equation in one unknown to obtain the elements of $B$. Could there have been some hitch? What if the computed value for $b_{22}^2$ or later for $b_{33}^2$ turns out to be negative? If it happens to be so, we would not have been able to solve for $B$. It can be shown that if $A$ is positive definite then the above situation never arises.

Let us now try an exercise.

---

E 5) Compute a lower triangular square root in each of the following cases.

(i) $\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$ (ii) $\begin{pmatrix} 9 & 3 & 3 \\ 3 & 5 & 1 \\ 3 & 1 & 6 \end{pmatrix}$

---

Let us illustrate few more examples to understand the concept of definiteness.

**Example 7:** Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}$ be an $n \times n$ positive definite matrix where $A_{11}$ and $A_{22}$ are square matrices of order $r \times r$ and $(n-r)\times(n-r)$ respectively for some $r(1 \le r \le n-1)$. Show that $A_{11}$ is positive definite.

**Solution:** Let $\mathbf{x}$ be a $r \times 1$ nonnull vector.

Then $\mathbf{x}'A_{11}\mathbf{x} = (\mathbf{x}' : \mathbf{0})_{1\times n} \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}_{n\times n} \begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}_{n\times 1} = (\mathbf{x}' : \mathbf{0}) A \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix} > \mathbf{0}$, since $\begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}$ is a nonnull vector. Hence $A_{11}$ is pd.

***

**Example 8:** Let $A$ be a positive definite matrix. Then show that $|A| > \mathbf{0}$.

**Solution:** Since $A$ is positive definite, by theorem 4(a), $A = BB'$ for some nonsingular matrix $B$. So

$$|A| = |BB'| = |B|.|B'| = |B|^2 > 0$$

Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$ be a pd matrix.

Write $A_i = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{12} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1i} & a_{2i} & \cdots & a_{ii} \end{pmatrix}$, $i = 1, \ldots, n$.

The matrices $A_i, i = 1, \ldots, n$ are called leading principal submatrices of $A$. Combining Examples 6 and 7 we have $|A_i| > 0$ for $i = 1, \ldots, n$ if $A$ is p.d. Is the converse true? Next theorem has the answer of this which is stated below without proof.

***

**Theorem 5:** Let $A$ be a real symmetric matrix of order $n \times n$. Let $A_i, i = 1, \ldots, n$ be as defined above. Then $A$ is positive definite if and only if $|A_i| > 0$ for $A_i, i = 1, \ldots, n$.

**Example 9:** Let $A$ be a symmetric positive definite matrix. Show that $RAR'$ is pd where $R$ is any nonsingular matrix.

**Solution:** By Theorem 4(a) $A = BB'$, where $B$ is nonsingular. Then $RAR' = RBB'R' = CC'$ where $C = RB$. Further, $C$ is nonsingular since $R$ and $B$ are nonsingular. Hence $RAR'$ is pd.

***

**Example 10:** A symmetric matrix $A$ is positive definite if $RAR'$ is pd for some nonsingular matrix $R$.

**Solution:** Let $\mathbf{x} \neq \mathbf{0}$. Consider $\mathbf{x'Ax = x'R^{-1}RAR'R^{-1'}x = y'RAR'y}$ where $\mathbf{y} = R^{-1'}\mathbf{x} \neq \mathbf{0}$ since $\mathbf{x} \neq \mathbf{0}$. Hence $\mathbf{x'Ax = y'RAR'y > 0}$ since $\mathbf{y} \neq \mathbf{0}$ and $RAR'$ is pd. Hence $A$ is pd.

***

Now try the following exercises.

---

E6) Let $C$ and $D$ be symmetric matrices of order $r \times r$ and $(n-r) \times (n-r)$ respectively. Show that $\begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}$ is pd if and only if $C$ and $D$ are pd.

E7) Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{pmatrix}$ where $A_{11}$ and $A_{22}$ are square. Show that if $A$ is pd, then $A_{22}$ is pd.

E8) Let $A$ be as in E7. Show that if $A$ is nnd, then $A_{11}$ and $A_{22}$ are nnd.

---

Now let us take up some more theorems.

**Theorem 6:** Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A^t_{12} & A_{22} \end{pmatrix}$ be a partition of $A$, where $A_{11}$ and $A_{22}$ are square. $A$ is pd, if and only if $A_{11}$ and $A_{11} - A'_{12} A_{11}^{-1} A_{12}$ are pd.

**Proof:** For the 'if' part $A_{11}$ is pd. For the 'only if' part $A$ is pd and hence $A_{11}$ is pd by Theorem 5. Thus for both parts $A_{11}^{-1}$ exists. It is easy to see that

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^t & A_{22} \end{pmatrix} = \begin{pmatrix} I & 0 \\ A_{12}'A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{12}'A_{11}^{-1}A_{12} \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}$$

Hence $A = R\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{12}'A_{11}^{-1}A_{12} \end{pmatrix}R'$

where $R = \begin{pmatrix} I & 0 \\ A_{12}'A_{11}^{-1} & I \end{pmatrix}$. Notice that $|R| = |I|.|I| = 1$

Hence $R$ is nonsingular. By Examples 8 and 9, and E6 it follows that $A$ is pd if and only if $A_{11}$ and $A_{22} - A_{12}' A_{11}^{-1} A_{12}$ are pd. This completes the proof.

As we stated in the beginning that we shall give an easy way to construct pd matrices and orthogonal matrices. We shall do so now.

**Theorem 7:** Let $A$ be a symmetric matrix of order $n \times n$ with positive diagonal elements and let $a_{ii} > \sum_{j \neq i} |a_{ij}|$ $i = 1, \ldots, n$.

Then $A$ is positive definite.

The proof is beyond the scope of this notes.
Using the above theorem, it is easy to see that the matrix $A$ in Example 5 and those in E5 are pd.

**Theorem 8:** Let $\mathbf{u}$ be a vector with unit norm. Then $I - 2\mathbf{uu}'$ is a symmetric orthogonal matrix.

**Proof:** $(I - 2\mathbf{uu}')' = I' - 2(\mathbf{u}')'\mathbf{u}' = I - 2\mathbf{uu}'$
Hence $I - 2\mathbf{uu}'$ is symmetric.
Further, $(I - 2\mathbf{uu}')\,(I - 2\mathbf{uu}') = I - 2\mathbf{uu}' - 2\mathbf{uu}' + 4\mathbf{uu}'\mathbf{uu}' = I - 4\mathbf{uu}' + 4\mathbf{uu}' = I$ since $\mathbf{u}'\mathbf{u} = 1$. So $I - 2\mathbf{uu}'$ is orthogonal.

**Example 11:** Let $A$ be a positive semidefinite matrix of order $n \times n$ and let $a_{ii} = 0$. Then show that $a_{ij} = 0$ for $j = 1, \ldots, n$.

**Solution:** Consider $\mathbf{x} = \alpha\, \mathbf{e}_i + \mathbf{e}_j$ where $\mathbf{e}_i$ and $\mathbf{e}_j$ are $i^{th}$ and $j^{th}$ columns of the identity matrix of order $n \times n$ and $\alpha$ is a real number.

Now $$\mathbf{x}'A\mathbf{x} = \alpha^2 \mathbf{e}_i'A\mathbf{e}_i + \mathbf{e}_j'A\mathbf{e}_j + 2\alpha \mathbf{e}_i'A\mathbf{e}_j$$
$$= \alpha^2 \mathbf{a}_{ii} + \mathbf{a}_{jj} + 2\alpha \mathbf{a}_{ij} = \mathbf{a}_{jj} + 2\alpha \mathbf{a}_{ij}$$

Let if possible $a_{ij} \neq 0$. Choose $\alpha = \frac{-(a_{jj}+1)}{2a_{ij}}$. Then $\mathbf{x}'A\mathbf{x} = a_{jj} - (a_{jj}+1) = -1 < 0$ which is a contradiction since $A$ is psd. Hence $a_{ij} = 0$. Choice of $j$ being arbitrary, the result follows.

Now let us try the following exercises.

---

E9) Let $A$ be a nnd matrix. Show that $\mathbf{x}'A\mathbf{x} = 0$ if and only if $A\mathbf{x} = \mathbf{0}$.

E10) Show that $(1-\rho)I + \rho 1 1'$ is pd if and only if $-\frac{1}{n-1} < \rho < 1$ where $n$ is the order of the matrix and $1' = (1, \ldots, 1)$.

E11) Construct a $3\times3$ symmetric nondiagonal positive definite matrix $A$ such that $a_{11} = 2, a_{22} = 5, a_{33} = 4$.

E12) Let $A$ and $B$ be nnd matrices of the same order. Show that (i) $A+B$ is nnd; (ii) the column space of $A$ is a subspace of the column space of $A+B$.

So far we have discussed about various types of real symmetric matrices and other types of real symmetric matrices. Now let us discuss about idempotent matrices.

## 14.4 IDEMPOTENT MATRICES

A square matrix $A$ is said to be idempotent if $A^2 = A$. Can you quickly come up with some examples of idempotent matrices? Yes, you are right! $O$ and $I$ are idempotent matrices. In fact, the only nonsingular idempotent matrix is $I$. Why? This is so because $A^2 = A$ and $A$ is nonsingular implies $A = I$ (premultiply both sides of $A^2 = A$ by $A^{-1}$.) Similarly, the only rank 0 square matrix, namely $O$ is idempontent. What about idempotent matrices of order $n\times n$ of rank $r$ $(1\le r\le n-1)$? $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ is an example of an idempotent matrix of rank $r$.

Further, if $A$ is an idempotent matrix and $P$ is a nonsingular matrix of the same order, then $PAP^{-1}$. $PAP^{-1} = PA^2P^{-1} = PAP^{-1}$.

Thus, $PAP^{-1}$ is an idempotent matrix.

Hence $P\begin{pmatrix} I_r & 0 \\ \mathbf{0} & \mathbf{0} \end{pmatrix}P^{-1}$ is an idempotent matrix of rank $r$ for every nonsingular matrix $P$. We shall now show that every idempotent matrix of rank $r$ is of the form $P\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}P^{-1}$ for some nonsingular matrix $P$.

**Theorem 9:** Let $A$ be an $n\times n$ matrix of rank $r$ $(1\le r\le n-1)$. Then $A$ is idempotent if and only if $A = P\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}P^{-1}$ for some nonsingular matrix $P$.

**Proof:** 'If' part has already been proved above. For the 'only if' part, let $A$ be an : $n\times n$ idempotent matrix of rank $r$ $(1\le r\le n-1)$. Let $A = (\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_n)$. We have

$$(\mathbf{a}_1 : \ldots : \mathbf{a}_n) = A = A^2 = A(\mathbf{a}_1 : \ldots : \mathbf{a}_n)$$

Hence $A\mathbf{a}_i = \mathbf{a}_i, i = 1, \ldots, n$, Since rank $A = r$, there exist $r$ linearly independent columns $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}$ of $A$. Thus

$$A\,\mathbf{a}_{i_j} = \mathbf{a}_{i_j},\ j = 1, \ldots, r. \tag{1}$$

Again, $A(I-A) = \mathbf{0}$. Hence the set of columns of $(I-A)$ is a subspace of the null space $N(A)$ of $A$. We know that dimension of $N(A) = n - \text{rank } A = n - r$. So, rank of $I - A$ is at most $n - r$. On the other hand, since $I = A + (I - A)$, $n =$ rank $I \le$ rank $A +$ rank $(I - A)$.

Hence rank $(I-A) \ge n -$ rank $A$. Thus, rank $(I-A) = n -$ rank $A$. This, coupled with the fact the column space of $(I-A) \subset N(A)$, shows that the column space of

$(I-A)$ is the same as the null space of $A$. Let $\mathbf{e}_{l_1}-\mathbf{a}_{l_1},\ldots,\mathbf{e}_{l_{n-r}}-\mathbf{a}_{l_{n-r}}$ be linearly independent columns of $I-A$. Then

$$A\left(\mathbf{e}_{l_k}-\mathbf{a}_{l_k}\right)=0,\ k=1,\ldots,n-r \qquad (2)$$

Consider $P=\left(\mathbf{a}_{i_1}:\ldots:\mathbf{a}_{i_r}:\mathbf{e}_{l_1}-\mathbf{a}_{l_1}:\ldots:\mathbf{e}_{l_{n-r}}-\mathbf{a}_{l_{n-r}}\right)$. Clearly, $P$ is an $n\times n$ matrix. Let $P\mathbf{x}=\mathbf{0}$. Then, $x_1\mathbf{a}_{i_1}+x_2\mathbf{a}_{i_2}+\ldots+x_r\mathbf{a}_{i_r}+x_{r+1}\left(\mathbf{e}_{l_1}-\mathbf{a}_{l_1}\right)+\ldots+x_n\left(\mathbf{e}_{l_{n-r}}-\mathbf{a}_{l_{n-r}}\right)=\mathbf{0}$.

Now,

$$\mathbf{0}=AP\mathbf{x}=x_1A\mathbf{a}_{i_1}+x_2A\mathbf{a}_{i_2}+\ldots+x_rA\mathbf{a}_{i_r}+x_{r+1}A\left(\mathbf{e}_{l_1}-\mathbf{a}_{l_1}\right)+\ldots+x_nA\left(\mathbf{e}_{l_{n-r}}-\mathbf{a}_{l_{n-r}}\right)$$
$$=x_1\mathbf{a}_{i_1}+\ldots+x_r\mathbf{a}_{i_r} \quad \text{in view of Eqns. (1) and (2).}$$

This implies $x_1=x_2=x_r=0$ since $a_{i_1},\ldots,a_{i_r}$ are linearly independent. This in turn implies $x_{r+1}=\ldots,x_n=0$ since $\mathbf{e}_{l_i}-\mathbf{a}_{l_i},\ i=1,\ldots,n-r$ are linearly independently.

Thus, $P\mathbf{x}=\mathbf{0}$ implies $\mathbf{x}=\mathbf{0}$ or the columns of $P$ are linearly independent. So rank $P=n$ or in other words, $P$ is nonsingular.

Further, $AP=A\left(\mathbf{a}_{i_1}:\ldots:\mathbf{a}_{i_r}:\mathbf{e}_{l_1}-\mathbf{a}_{l_1}:\ldots:\mathbf{e}_{l_{n-r}}-\mathbf{a}_{l_{n-r}}\right)$

$$=(a_{i_1}:\ldots:a_{i_r}:0:\ldots:0)$$
$$=\left(\mathbf{a}_{i_1}:\ldots:\mathbf{a}_{i_r}:\mathbf{e}_{l_1}-\mathbf{a}_{l_1}:\ldots:\mathbf{e}_{l_{n-r}}-\mathbf{a}_{l_{n-r}}\right)\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}$$
$$=P\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}.$$

Thus, we have $A=P\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}P^{-1}$. This completes the proof.

Let $A$ be an idempotent matrix of order $n\times n$ with rank $r$. From Theorem 9, the following statements are clear.

(a) $A$ is similar to a diagonal matrix.

(b) $A$ has at most two distinct eigenvalues $1$ and $0$. Eigenvalue $1$ is with algebraic multiplicity $r$ and $0$ with algebraic multiplicity $n-r$.

Finding the rank of a matrix in general is not very easy. However, it is quite easy for idempotent matrices. We start with a definition.

**Definition:** The trace of a square matrix $A$ of order $n\times n$ is defined as the sum of its diagonal elements and is denoted by $\operatorname{tr}(A)$. Thus, $\operatorname{tr}(A)=\sum_{i=1}^{n} a_{ii}$

**Example 12**: Let $A$ and $B$ be are square matrices of the same order. Show that (i) $\operatorname{tr}(cA)=c.\operatorname{tr}(A)$ when $c$ is a real number; (ii) $\operatorname{tr}(A+B)=\operatorname{tr}(A)+\operatorname{tr}(B)$.

**Solution:** (i) $\operatorname{tr}(CA)=\sum_{i=1}^{n} caii=c\sum_{i=1}^{n} aii=c\operatorname{tr}(A)$

(ii) $\operatorname{tr}(A+B)=\sum_{i=1}^{n}(aii+bii)=\sum_{i=1}^{n} aii+\sum_{i=1}^{n} bii=\operatorname{tr}(A)+\operatorname{tr}(B)$

***

**Example 13:** Let $A$ and $B$ be matrices of order $m\times n$, and $n\times m$ respectively. Show that $tr(AB) = tr(BA)$.

**Solution:** The $(i, i)^{th}$ element of $AB$ is given by $\sum_{j=1}^{n} a_{ij}b_{ji}$. Hence

$tr(AB) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} = tr(BA)$, since $\sum_{i=1}^{m} b_{ji}a_{ij}$ is the $(j, j)^{th}$ element of $AB$.

***

We are now ready to prove the following theorems.

**Theorem 10:** Let $A$ be an idempotent matrix of any order $n\times n$. Then rank $A = tr(A)$.

**Proof:** If rank $A$ is 0 or $n$, then always $A = tr(A)$. Let rank $A = r$ when $1 \le r \le n-1$. Then by Theorem 9, there exists a nonsingular matrix $P$ such that $A = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}P^{-1}$.

Now $tr(A) = tr\left[P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}P^{-1}\right] = tr\left[\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}P^{-1}P\right] = tr\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = r =$ rank $A$.

We state below another result on idempotent matrices.

**Theorem 11:** A square matrix $A$ of order $n\times n$ is idempotent if and only if rank $(I - A) = n -$ rank $A$.

**Theorem 12:** Let $A$ be a real symmetric and idempotent matrix of rank $r$. Then there exists an orthogonal matrix $S$ such that $A = S\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}S'$. Hence $A$ is nonnegative definite.

**Proof:** Left as an exercise.

Let us now have an example.

**Example 14:** Let $A$ and $B$ be idempotent matrices of the same order. Then show that $A + B$ is idempotent if and only if $AB = BA = \mathbf{0}$.

**Solution:** 'If' part is trivial. For the 'only if' part, let $A, B$ and $A + B$ be idempotent. Then $A + B = (A + B)(A + B) = A^2 + B^2 + AB + BA = A + B + AB + BA$. So, $AB + BA = \mathbf{0}$. Pre-multiplying by $A$, we get $AB + ABA = \mathbf{0}$. Now post multiplying the previous equality by $A$ we get $ABA + ABA = \mathbf{0}$ or $ABA = \mathbf{0}$. Hence $AB = \mathbf{0}$ and as a consequence $BA = \mathbf{0}$.

***

Try the following Exercises.

---

E13) Let $A$ be a $2\times 2$ idempotent matrix. Can $a_{11}$ be equal to $2$?

E14) Let $A$ and $B$ be idempotent matrices. Then show that $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ is also idempotent.

E15) Show that if $A$ and $B$ are idempotent and the column space of $A$ is contained in the column space of $B$, the $BA = A$.



In this section, we shall discuss Cochran's theorem.

## 14.5 COCHRAN'S THEOREM

Cochran's theorem concerns with the probability distributions of quadratic forms in independent standard normal variables. This is a very important theorem which allows us to decompose sum of squares into several quadratic forms and identify their distributions and establish their independence.

Let $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ be a vector of $n$ independent standard normal variables. Let

$A_1, A_2, \ldots, A_k$ be real symmetric (nonrandom) matrices such that $\mathbf{x'x} = \mathbf{x}' A_1\mathbf{x} + \mathbf{x}'A_2\mathbf{x} + \cdots + \mathbf{x}' A_k\mathbf{x}$. We know that $\mathbf{x'x}$ is distributed as chi-square with $n$ degrees of freedom. Cochran's theorem asserts that $x'A_i\, x, i = 1, \ldots, k$ are distributed as independent chi-squares if and only if $\sum_{i=1}^{n}$ rank $A_i = n$. In this section, we prove an algebraic version of this result. In the next unit, we shall prove the statistical version.

**Theorem 13:** Let $A_1, A_2, \ldots, A_k$ be real symmetric matrices such that $A_1 + A_2 + \cdots + A_k = I$. The following are equivalent:

(a) $A_i$ is idempotent, $i = 1, \ldots, k$

(b) $\sum_{i=1}^{k}$ rank $A_i = n$

(c) $A_i A_j = \mathbf{0}$ whenever $i \neq j$

**Proof:** $(a) \Rightarrow (b): n = \text{rank } I = tr(I) = tr(A_1 + \cdots + A_k) = \sum_{i=1}^{k} tr(A_i) = \sum_{i=1}^{k} \text{rank } A_i$

(since $A_1, \ldots, A_k$ are idempotent rank $A_i = tr(A_i)$ by Theorem 10.)

$(b) \Rightarrow (c)$: Let rank $A_i = \mathbf{r}_i$. Then by (b), $\sum_{i=1}^{k} \mathbf{r}_i = n$. Since $A_i$ is a real symmetric matrix there exists a matrix $P_i$ of order $n \times r_i$ such that $A_i = P_i\Delta_i P_i'$, $P_i' P_i = I_{r_i}$, and $\Delta_i$ is a real nonsingular diagonal matrix, $i = 1, \ldots, k$. (Take the help of Theorem 3)

So, $I = A_1 + \cdots + A_k = \sum_{i=1}^{k} P_i\Delta_i P_i' = P\Delta P'$

where $P = (P_1 : P_2 : \ldots : P_k)$ and $\Delta = \begin{pmatrix} \Delta_1 & 0 & 0 & 0 \\ 0 & \Delta_2 & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \Delta_k \end{pmatrix}$ Notice that the number of

columns in $P$ is $\sum_{1}^{k} r_i$ which equals $n$ by hypothesis. Hence $P$ is a square matrix of order $n \times n$. So, $n \geq \text{rank } P \geq \text{rank } (P\Delta P') = \text{rank } I = n$. Hence $P$ is a nonsingular matrix. Similarly, $\Delta$ is also nonsingular. Since $P\Delta P' = I$ and $P\Delta$ is a square matrix,

$P' = (P\Delta)^{-1}$. In other words, $P'P\Delta = I$ or $P'P = \Delta^{-1}$ which is a diagonal matrix. So $P_i'P_j = \mathbf{0}$ whenever $i \neq j$. Hence $A_i\ A_j = P_i\Delta_i P_i' P_j \Delta_j P_j' = \mathbf{0}$ whenever $i \neq j$.

$(c) \Rightarrow (a)$: For each i, $A_i = A_i\ I = A_i(A_1 + \cdots + A_k) = A_i^2$ since $A_i\ A_j = \mathbf{0}$ whenever $i \neq j$. Thus, $A_i$ is idempotent for each $i$. Theorem is thus proved.

We now prove an algebraic version of another useful result in connection with the distribution of quadratic forms in normal variables.

**Theorem 14:** Let $A$ and $B$ be symmetric idempotent matrices and let $B - A$ be non negative definite. Then $B - A$ is also a symmetric idempotent matrix.

**Proof:** Since $A$ is symmetric idempotent, it is nnd. Since $B - A$ is nnd, the column space of $A$ is contained in the column space of $B$. So $BA = A$. Then $(B - A)A = 0 = A(B - A)$. Since $B(I - B) = (I - B)B = \mathbf{0}$, $A(I - B) = (I - B)A = \mathbf{0}$ and $(I - B)(B - A) = \mathbf{0} = (B - A)(I - B)$. Now $A + (B - A) + (I - B) = I$. By $(c) \Rightarrow (a)$ of Theorem 13 it follows that $B - A$ is idempotent.

After Cochran's theorem, we shall discuss the singular value decomposition in this section.

## 14.6 SINGULAR VALUE DECOMPOSITION

In Theorem 3, we showed that if $A$ is a real symmetric matrix, there exists an orthogonal matrix $P$ such that $A = P\Delta P'$. We also showed that the diagonal elements of $\Delta$ are the eigenvalues and the column of $P$ are the orthonormal eigenvectors of $A$. What about a real $m \times n$ matrix $A$? We know that every $m \times n$ matrix $A$ of rank $r(1 \le r \le \min\{m, n\})$ can be written as $A = R\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}S$ when $R$ and $S$ are nonsingular.

Can we replace the nonsingular matrices by orthogonal matrices if we can relax $I_r$ to a positive definite diagonal matrix? If so, what interpretation can we give to the orthogonal matrices and the diagonal elements of the diagonal matrix? We study these details in this section.

**Theorem 15:** Let $A$ be a real matrix of order $m \times n$ with rank $r(1 \le r \le \min\{m, n\})$. Then there exists orthogonal matrices U and $V$ of orders $m \times m$ and $n \times n$ respectively such that $A = U\begin{pmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}V'$ where $\Delta$ is a positive definite diagonal matrix.

**Proof:** Notice $AA'$ and $A'A$ are nonnegative definite matrices (Why? See Theorem 4). Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ be orthogonal eigenvectors of $AA'$ corresponding to the eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_r > \lambda_{r+1} = \ldots = \lambda_m = 0$. So $AA'\mathbf{u}_i = \lambda_i \mathbf{u}_i$, $i = 1, \ldots, m$. Write $\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}} A'u_i$, $i = 1, \ldots, r$, where $\sqrt{\lambda_i}$ is the positive square root of $\lambda_i$. Then for $i, j = 1, \ldots, r$

$$\mathbf{v}_i'\mathbf{v}_j = \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}} u_i' AA' u_j = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Thus, $\mathbf{v}_1,\ldots,\mathbf{v}_r$ are orthonormal vectors. Extend $\mathbf{v}_1,\ldots,\mathbf{v}_r$ to an orthonormal basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ of $R^n$. Write $U=(\mathbf{u}_1:\ldots:\mathbf{u}_m)$ and $V=(\mathbf{v}_1:\ldots:\mathbf{v}_n)$. Clearly U and V are orthogonal matrices.
Also $AA'\mathbf{u}_i=\mathbf{0}$ for $i=r+1,\ldots,m$. Hence $u_i'AA'u_i=0$ or $A'u_i=0$ for $i=r+1,\ldots,m$.

Since $\mathbf{u}_1\mathbf{u}_1'+\ldots+\mathbf{u}_m\mathbf{u}_m'=I$, we have

$$\begin{aligned}A&=(\mathbf{u}_1\mathbf{u}_1'+\ldots+\mathbf{u}_m\mathbf{u}_m')A\\&=(\mathbf{u}_1\mathbf{u}_1'+\ldots+\mathbf{u}_r\mathbf{u}_r')A \text{ since } \mathbf{u}_i'A=\mathbf{0} \text{ for } i=r+1,\ldots,m.\\&=\sum_{i=1}^{r}\sqrt{\lambda_i}\,\mathbf{u}_i\mathbf{v}_i'\end{aligned}$$

Denote $\delta_i=\sqrt{\lambda_i}$, $i=1,\ldots,r$ and $\Delta=\mathrm{diag}(\delta_1,\delta_2,\ldots,\delta_r$

It follows that $A=\sum_{i=1}^{r}\sqrt{\lambda_i}\,\mathbf{u}_i\mathbf{v}_i'=U\begin{pmatrix}\Delta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}V'$. The proof is complete.

We shall now interpret the columns of U and V and the diagonal elements of $\Delta$ in the above form.

Let $A=U\begin{pmatrix}\Delta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}V'$, where U and V are orthogonal and $\Delta$ is a positive definite diagonal matrix.

Then rank of A is the same as rank of $\Delta$ which in turn is the number of rows in $\Delta$.
Now

$$AA'=U\begin{pmatrix}\Delta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\Delta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}'V'=U\begin{pmatrix}\Delta^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}V'.$$

Thus, $U\begin{pmatrix}\Delta^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}V'$ is a spectral decomposition of $AA'$. Hence the diagonal elements of $\Delta^2$ are the nonzero eigenvalues of $AA'$ and the columns of U are the orthonormal eigenvectors of $\mathbf{AA'}$. To be more specific

$$AA'u_i=\begin{cases}\delta_i^2\mathbf{u}_i \text{ for } i=1,\ldots,r\\ \mathbf{0} \text{ for } i=r+1,\ldots,m\end{cases}$$

Again $A'A=V\begin{pmatrix}\Delta^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}U'$, which is a spectral decomposition of $AA'$. Hence

$$A'Av_i=\begin{cases}\delta_i^2\mathbf{v}_i \text{ for } i=1,\ldots,r\\ \mathbf{0} \text{ for } i=r+1,\ldots,n\end{cases}$$

Thus, the diagonal elements of $\Delta$ and the columns of U and V relate to the eigenvalues and eigenvectors of $AA'$ and $A'A$. *The diagonal elements of* $\Delta$ *are called the singular values and the columns of* U *and* V *are called the singular vectors of* A. *The decomposition* $A=U\begin{pmatrix}\Delta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}V'$ *is called the singular value decomposition of* A.

**Example 15:** Let $A=\frac{1}{3}\begin{pmatrix}1 & -2 & -2\\ -2 & 1 & -2\\ -2 & -2 & 1\end{pmatrix}\begin{pmatrix}2 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}\frac{1}{2}\begin{pmatrix}1 & -1 & -1 & -1\\ -1 & 1 & -1 & -1\\ -1 & -1 & 1 & -1\\ -1 & -1 & -1 & 1\end{pmatrix}$ be a singular value decomposition of $A$. What are the eigenvalues of $AA'$ and $A'A$? Identify the corresponding eigenvectors. What is the rank of $A$?

**Solution:** $A=U\begin{pmatrix}\Delta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}V'$

where $U=\frac{1}{3}\begin{pmatrix}1 & -2 & -2\\ -2 & 1 & -2\\ -2 & -2 & 1\end{pmatrix}$ and $V=\frac{1}{2}\begin{pmatrix}1 & -1 & -1 & -1\\ -1 & 1 & -1 & -1\\ -1 & -1 & 1 & -1\\ -1 & -1 & -1 & 1\end{pmatrix}$ are orthogonal matrices,

and $\Delta=\begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}$

The eigenvalues of $AA'$ are $\delta_1^2=4$, $\delta_2^2=1$ and $0$. The corresponding eigenvectors are the first, second and third columns respectively of $U$, namely $\frac{1}{3}\begin{pmatrix}1\\ -2\\ -2\end{pmatrix}$, $\frac{1}{3}\begin{pmatrix}-2\\ 1\\ -2\end{pmatrix}$, and $\frac{1}{3}\begin{pmatrix}-2\\ -2\\ 1\end{pmatrix}$ respectively.

***

Let us now sum up whatever we have studied in this unit.

## 14.7 SUMMARY

In this unit, we have covered the following points:

1. Definition of a real symmetric matrix
2. Classification of quadratic forms
3. Spectral decomposition of a real symmetric matrix
4. A method of determining the definiteness of a quadratic form
5. Properties of positive definite and nonnegative definite matrices
6. A method of finding a triangular square root of a pd matrix
7. Properties of idempotent matrices
8. Cochran's Theorem
9. Singular Value Decomposition.

## 14.8 SOLUTIONS/ANSWERS

E1. i) Coefficient of $x_1^2$, $x_2^2$, and $x_1x_2$ are respectively $1, -1$ and $0$. So the matrix of the quadratic form $x_1^2-x_2^2$ is $\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$.

ii) Coefficients of $x_1^2$, $x_2^2$, and $x_1x_2$ are respectively $2, 5$ and $3$. So the matrix of the quadratic form is $\begin{bmatrix} 2 & 1.5 \\ 1.5 & 5 \end{bmatrix}$.

iii) Coefficients of $x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3$ and $x_2x_3$ are respectively $0, 0, 0, 3, -4, 5$. So the matrix of the quadratic form $3x_1x_2 + 5x_2x_3 - 4x_1x_3$ is

$$\begin{bmatrix} 0 & 1.5 & -2 \\ 1.5 & 0 & 2.5 \\ -2 & 2.5 & 0 \end{bmatrix}.$$

iv) Coefficients of $x_1^2, x_2^2, x_3^2, x_4^2, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4$ and $x_3x_4$ are respectively $1, 1, 0, 0, 0, 0, 0, 0, 1, 0$. So the matrix of the quadratic form $x_1^2 + x_2^2 + x_2x_4$ is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

E2. Suppose $a_{ii} < 0$. Let $\mathbf{e}_i$ denotes the $i^{th}$ column of the identity matrix. Then $\mathbf{e}_i' \; A\mathbf{e}_i = a_{ii} < 0$. Hence $A$ cannot be pd or psd.

E3. i) The matrix of the quadratic form $x_1^2 - 5x_1x_2 + 7x_2^2$ is $\mathbf{A} = \begin{bmatrix} 1 & -2.5 \\ -2.5 & 7 \end{bmatrix}$.

The eigenvalues of $A$ are the roots of the characteristic equation $|A - \lambda I| = 0$ or $(1-\lambda)(7-\lambda) - 6.25 = 0$.

The Characteristic equation can be rewritten as $\lambda^2 - 8\lambda + 0.75 = 0$.

Hence the eigenvalues are $\dfrac{8+\sqrt{64-3}}{2}$ and $\dfrac{8-\sqrt{64-3}}{2}$ which are both positive. Hence the quadratic form $x_1^2 - 5x_1x_2 + 7x_2^2$ is positive definite.

ii) For $x_1 = 1$ and $x_2 = x_3 = 0$, the value of the quadratic form is $1$. Again for $x_2 = 1, x_1 = x_3 = 0$, the value of the quadratic form is $-1$. Hence the quadratic form is indefinite.

iii) The matrix of the quadratic form is $A = \begin{bmatrix} 2 & 3 & 0 \\ 3 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$. The characteristic equation of $A$ is $(4-\lambda)((2-\lambda)(3-\lambda) - 9) = 0$. So $4$ is a root of the above equation. The remaining two eigenvalues are the roots of the equation $(2-\lambda)(3-\lambda) - 9 = 0$ or $\lambda^2 - 5\lambda - 3 = 0$. So the eigenvalues are $\dfrac{5 \pm \sqrt{25-12}}{2}$ and $4$. Thus, all the three roots are positive. Hence the given quadratic form is positive definite.

E4. The eigenvalues of $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ are the roots of the characteristic equation $(2-\lambda)^2 - 1 = 0$ or $\lambda^2 - 4\lambda + 3 = 0$ or $(\lambda - 3)(\lambda - 1) = 0$. So the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = 1$. Let $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ be an eigenvector corresponding to $\lambda_1$. Then

$$2x_1 + x_2 = 3x_1$$
$$x_1 + 2x_2 = 3x_2$$
or $-x_1 + x_2 = 0$
$$x_1 - x_2 = 0$$

Thus, $x_1 = x_2$. So the normalized eigenvector corresponding to the eigenvalue 3 is $\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$.

It can similarly be shown that $\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$ is the normalized eigenvector corresponding to the eigenvalue 1 orthogonal to the first eigenvector. So the spectral decomposition of $A$ is

$$A = 3.\frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}(1 \quad 1) + 1.\frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}(1 \quad -1)$$

or $\frac{1}{2}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}3 & 0\\0 & 1\end{pmatrix}\left(\frac{1}{2}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\right)$

If $A_{n\times m} = P\Lambda P'$ is a spectral decomposition of $A$, then
$A^2 = P\Lambda P'P\Lambda P' = P\Lambda^2 P'$.
By induction it can be shown that $A^k = P\Lambda^k P'$ for $k = 1, 2, \ldots$
Thus, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A$, then $\lambda_1^k, \ldots, \lambda_n^k$ are the eigenvalues of $A^k$. The eigenvectors of $A^k$ can be taken to be the same as the eigenvectors of $A$.

Hence $A^{100} = \frac{1}{2}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}3^{100} & 0\\0 & 1\end{pmatrix}\left(\frac{1}{2}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\right)$.

E5. i) Let $\begin{pmatrix}4 & 1\\1 & 2\end{pmatrix} = \begin{pmatrix}b_{11} & 0\\b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}b_{11} & b_{21}\\0 & b_{22}\end{pmatrix}$

So $b_{11}^2 = 4$ or $b_{11} = 2$

$b_{21}\, b_{11} = 1$ or $b_{21} = \frac{1}{b_{11}} = \frac{1}{2}$

$b_{21}^2 + b_{22}^2 = 2$ or $b_{22}^2 = 2 - \frac{1}{4} = \frac{7}{4}$

so $b_{22} = \frac{\sqrt{7}}{2}$

Thus, the required lower triangular square root is $\begin{pmatrix}2 & 0\\ \frac{1}{2} & \frac{\sqrt{7}}{2}\end{pmatrix}$.

ii) Let $\begin{pmatrix}9 & 3 & 3\\3 & 5 & 1\\3 & 1 & 6\end{pmatrix} = \begin{pmatrix}b_{11} & 0 & 0\\b_{21} & b_{22} & 0\\b_{31} & b_{32} & b_{33}\end{pmatrix}\begin{pmatrix}b_{11} & b_{21} & b_{31}\\0 & b_{22} & b_{32}\\0 & 0 & b_{33}\end{pmatrix}$

so $b_{11}^2 = 9$ or $b_{11} = 3$
$b_{11}\, b_{21} = 3$ or $b_{21} = 1$
$b_{11}\, b_{31} = 3$ or $b_{31} = 1$
$b_{21}^2 + b_{22}^2 = 5$ or $b_{22}^2 = 5 - 1$ or $b_{22} = 2$

$$b_{21}\, b_{31} + b_{22}\, b_{32} = 1 \quad \text{or} \quad b_{22}\, b_{32} = 1 - 1 = 0 \quad \text{or} \quad b_{32} = 0$$
$$b_{31}^2 + b_{32}^2 + b_{33}^2 = 6 \quad \text{or} \quad b_{33}^2 = 6 - 0 - 1 = 5 \quad \text{or} \quad b_{33} = \sqrt{5}\,.$$

Thus, the required triangular square root is $\begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & \sqrt{5} \end{pmatrix}$.

E6. 'If' part: let C and D be positive definite. Then

$(\mathbf{x}' : \mathbf{y}')\begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{x}'C\mathbf{x} + \mathbf{y}'D\mathbf{y} > \mathbf{0}$ whenever at least one of $\mathbf{x}$ and $\mathbf{y}$ is non-null. Hence $\begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}$ is pd.

'Only if' part: Let $\begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}$ be pd consider $\mathbf{0} < (\mathbf{x}' : \mathbf{0})\begin{pmatrix} C & \mathbf{0} \\ \mathbf{0} & D \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix} = \mathbf{x}'C\mathbf{x}$ whenever $\mathbf{x} \neq \mathbf{0}$. Hence C is pd.

E7. We know that $\mathbf{0} < (\mathbf{0} : \mathbf{y}')\begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix}\begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix} = \mathbf{y}'A_{22}\mathbf{y}$ whenever $\mathbf{y} \neq \mathbf{0}$ ($\mathbf{y}$ is chosen so that $\mathbf{y}'A\mathbf{y}$ is conformable). Hence $A_{22}$ is pd.

E8. Use the same procedure as in E7.

E9. Let A be an nnd matrix. Then there exists B such that $A = BB'$. Hence $\mathbf{0} = \mathbf{x}'A\mathbf{x} = \mathbf{x}'BB'\mathbf{x}$ implies that $B'\mathbf{x} = \mathbf{0}$. So $A\mathbf{x} = BB'\mathbf{x} = \mathbf{0}$.

E10. Nonzero eigenvalues of $AA'$ and $A'A$ are the same. Hence nonzero eigenvalues of $\rho 1 1'$ are the same as the eigenvalues of the $1 \times 1$ matrix $\rho 1' 1 = n\rho$. The eigenvalues of $\rho 1' 1 = n\rho$. So, the eigenvalues of $\rho 1 1'$ are $n\rho$ and [0, 0,..., 0 $(n-1)$ times].

Let $\lambda$ be an eigenvalue of $\rho l l^t$. Let the corresponding eigenvector be $\mathbf{x}$. Then $(1-\rho)1 + \rho 1 1')x = (1-\rho)x + \lambda x = (1 - \rho + \lambda)x$.

Thus, the eigenvalues of $(1-\rho)1 + \rho 1 1'$ are $(1-\rho) + n\rho,\ (1-\rho), \ldots, (1-\rho)$ ($(n-1)$ times).

Hence $(1-\rho)1 + \rho 1 1'$ is pd if and only if $1 + (n-1)\,\rho > 0$ and $1 - \rho > 0$ or $-\dfrac{1}{n-1} < \rho < 1$.

E11. We use Theorem 7 for this purpose. Thus, $\begin{pmatrix} 2 & 1 & 0.5 \\ 1 & 5 & 3 \\ 0.5 & 3 & 4 \end{pmatrix}$ is a pd matrix with diagonal elements 2, 5 and 4, respectively.

E12. i) $\mathbf{x}'(A + B)\mathbf{x} = \mathbf{x}'A\mathbf{x} + \mathbf{x}'B\mathbf{x} \geq 0$ since $\mathbf{x}'A\mathbf{x}$ and $\mathbf{x}'B\mathbf{x}$ are nonnegative. Hence $A + B$ is nnd.

ii) Write $A = CC'$ and $B = DD'$. Hence $A + B = CC' + DD' = (C : D)\begin{pmatrix} C' \\ D' \end{pmatrix}$.

Hence column space of $A =$ Column space of $C \subseteq$ column space of $(C:D) =$ column space of $A+B$.

E13. Let $\begin{pmatrix} 2 & b \\ c & d \end{pmatrix}$ be an idempotent matrix. Then $\begin{pmatrix} 2 & b \\ c & d \end{pmatrix}\begin{pmatrix} 2 & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2 & b \\ c & d \end{pmatrix}$

$4+bc=2$. $b$ cannot be $0$ since $4+bc=2$. Also $bc=-2$

$2b+bd=b$. So $2+d=1$ or $d=-1$

$2c+cd=c$

$bc+d^2=d$

$bc+1=-1$

or $bc=-2$. Choose $b=-2$ and $c=1$

(in fact choose $b$ any nonzero number and $c=\frac{-2}{b}$)

Thus, $\begin{pmatrix} 2 & -2 \\ 1 & -1 \end{pmatrix}$ is idempotent. Hence there is an idempotent matrix with $a_{11}=2$.

E14. $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} A^2 & 0 \\ 0 & B^2 \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$

So $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ is idempotent if $A$ and $B$ are idempotent.

E15. Since the column space of $A$ is contained in the column space of $B$, we get $A = BD$ for some $D$.

Now $BA = B.BD = BD = A$.

—x—

# UNIT 15 DEFINITION AND PROPERTIES OF MVN-I

## Structure

Page No



## 15.1 INTRODUCTION

We begin the unit with an overview of the scope of multivariate analysis. We identify the extensions of problems of univariate analysis to higher dimensions and also outline the problems special to multivariate analysis which do not have univariate equivalence. We do this in Section 15.2. In Section 15.3, we study the properties of variance-covariance matrices in detail. We also identify the class of variance-covariance matrices with the class of nnd matrices. In Section 15.4, we present several examples of discrete and continuous multivariate distributions.

### Objectives:

After completing this unit, you should be able to

- define the scope and applications of multivariate analysis;
- distinguish between univariate and multivariate analysis;
- describe the bivariate normal distribution;
- compute with the mean vectors, variance-covariance matrices and covariance matrices of transformed variables;
- apply the concepts of marginal and conditional distributions and independence in multivariate probability distributions.

## 15.2 SCOPE OF MULTIVARIATE ANALYSIS

Let us start with an example. A software company wants to recruit 3 fresh engineering graduates. There are 20 applicants and their scores (each out of 100) in the aptitude test $(x_1)$, the test on software technology $(x_2)$ and the interview $(x_3)$ are recorded below in a matrix form

$$X = \begin{pmatrix} 52 & 61 & 35 \\ 64 & 75 & 72 \\ 68 & 60 & 58 \\ 50 & 65 & 59 \\ 49 & 58 & 65 \\ 70 & 64 & 81 \\ 50 & 54 & 47 \\ 70 & 71 & 54 \\ 61 & 65 & 63 \\ 62 & 65 & 78 \\ 58 & 70 & 64 \\ 50 & 48 & 31 \\ 47 & 59 & 56 \\ 55 & 61 & 60 \\ 46 & 48 & 42 \\ 60 & 62 & 68 \\ 52 & 53 & 64 \\ 58 & 61 & 50 \\ 78 & 82 & 65 \\ 42 & 47 & 49 \end{pmatrix}$$

In this matrix form, the vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ is called a random vector and the matrix **X** is called a **data matrix** related to $\mathbf{x}$. The (i, j)$^{th}$ element $x_{ij}$ of **X** denotes the score of the i$^{th}$ candidate in the aptitude test or the test on software technology or the interview according as $j = 1, 2, 3$ respectively. Thus, $x_{51} = 49$ is the score of the fifth candidate in the aptitude test. Each row of **X** corresponds to the scores of a candidate in three tests and each column of **X** corresponds to the scores of 20 candidates in a particular test/interview. Univariate analysis deals with the data on a single variable, say, those in the interview. Multivariate analysis deals with data on more than one variable (possibly correlated) collected on the same subjects. One of the major aims of statistics relates to dealing with variability in the data. By dealing with variability we mean (i) determining the extent of variability, (ii) identifying the sources of variability and (iii) either control the variability by taking suitable measures or taking advantage of the variability to select certain subject in an optimal manner or classifying the subjects or variables into different groups depending upon the variability. When we deal with a single variable, the variability is often quantified by the variance. When we deal with more than one variable, then the variability is often quantified by the matrix of variances and covariances. For example, in the case of the data mentioned above, the variability is quantified by the matrix

$$\sum_{3\times3} = (\sigma_{ij})$$

where $\sigma_{ij} = \mathrm{Cov}(x_i, x_j), i\,j = 1, 2, 3$.

Such a matrix $\sum$ is called the **variance covariance matrix of** $\mathbf{x}$ and is denoted by $D(\mathbf{x})$. We shall discuss about such a matrix in detail in the next section.

Let us turn our attention to the recruitment problem. If the recruitment is based on just the interview scores, then the candidates 6, 10 and 2 get selected. Again, if the recruitment is based on the aptitude test, the candidates 19, 6 and 8 get selected where as under the criterion of software technology scores, the candidates 19, 2 and 8 get selected. Notice that it is not realistic or optimal to base the judgment on the scores of just one of the three variables as we are ignoring useful information on the others. One possible way of using the scores on all the three is taking the average of the scores on the three (two tests and the interview) for each candidate and select the three candidates with the top three average scores. What should be the justification in choosing a criterion? We should choose a criterion which can distinguish among the candidates in the best possible manner. When we took the average, we took the linear combination $\mathbf{l}^t\mathbf{x}$ where $\mathbf{l}^t = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Why not look for a linear combination $\mathbf{p}^t\mathbf{x}$ over all linear combinations, which distinguishes among the candidates in the best manner or in other words which has the largest variability (variance) and use that as an index for selection criterion? This is precisely what is done in obtaining the first principal component. The first principal component $\mathbf{p}^t\mathbf{x}$ is a normalized linear combination of $\mathbf{x}$, which has the largest variance among all normalized linear combinations of $\mathbf{x}$.

Getting information on any variable is expensive in terms of time and or money. We may like to ask whether it is worthwhile conducting the test in software technology given that the aptitude test and interview are being conducted. Put in other words, does the test in software technology provide significant additional information in the presence of the aptitude test and the interview? This is called the assessment of additional information.

Based on the data $\mathbf{X}$ on $\mathbf{x}$ can we group the candidates into some well-defined classes? This may be a useful information if the company has jobs of different types – (a) requiring high skills and (b) requiring medium skills but intensive hard work. There may be a third group which is not of any use to the company. This is called the problem of discrimination.

How well can we predict the interview score of a candidate based on his two test scores? This problem is called the problem of multiple regression and correlation. Suppose the interview mentioned above is a technical interview. Assume that there is another HR interview and the score on HR interview be denoted by $x_4$. We may be interested in the association between the tests scores and the interviews scores, i.e., between $(x_1, x_2)$ and $(x_3, x_4)$. Such a problem is called the **problem of canonical correlations**. The problems mentioned above are some problems specific to multivariate analysis which do not occur in univariate analysis. In univariate analysis, we talk about inferences (estimation/testing) on the mean/proportion/variance of a variable. These problems can be extended to the inferences on mean vector/variance covariance matrix of a random vector. Univariate analysis of variance has an analogue in multivariate analysis of variance.

In this section, we shall learn to compute the variance covariance matrices.

## 15.3 VARIANCE COVARIANCE MATRICES

As we discussed in the previous section, the variance-covariance matrix of a random vector is a quantification of the joint variability of the components of the random vector. The variance-covariance matrices play a very important role in quantifying dependence structure in multivariate analysis. In this section, we formally define a random vector, its mean vector and variance-covariance matrix. We shall obtain formulae for the mean vector and variance-covariance matrix of linear compounds of a given random vector. We shall give a method of transforming correlated random variables to uncorrelated random variables. We shall show that every variance covariance matrix is nnd and that every nnd matrix is the variance-covariance matrix of a random vector.

**Definition:** A random vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ of order $p \times 1$ is a finite ordered p-triple sequence of random variables $x_1, x_2, \ldots, x_p$. $E(\mathbf{x}) = \begin{pmatrix} E(x_1) \\ \vdots \\ E(x_p) \end{pmatrix}$ is called the mean vector of $\mathbf{x}$, where $E(x_i)$ denotes the expected value of $x_i$. Let $\sigma_{ij}$ denotes the covariance between $x_i$ and $x_j$. Then the matrix $\sum = ((\sigma_{ij}))$ of order $p \times p$ is called the variance-covariance matrix of $\mathbf{x}$, denoted by $D(\mathbf{x})$. Let $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_q \end{pmatrix}$ be another random vector. Let $\lambda_{ij}$ denotes the covariance between $x_i$ and $y_j$, $i = 1, \ldots, p$, $j = 1, \ldots, q$. Then $\Lambda_{p \times q} = ((\lambda_{ij}))$ is called the covariance matrix between $\mathbf{x}$ and $\mathbf{y}$ and is denoted by $Cov(\mathbf{x}, \mathbf{y})$.

Clearly $D(\mathbf{x}) = Cov(\mathbf{x}, \mathbf{x})$.

We know that $V(x) = E(x - E(x))^2$ and $Cov(x, y) = E((x - E(x))(y - E(y))$. Is there a multivariate analogue to the above? Notice that

$$D(\mathbf{x}) = \sum = ((\sigma_{ij})) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & & & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

where $\sigma_{ij} = Cov(x_i, x_j) = E((x_i - E(x_i))(x_j - E(x_j))$.

Thus, $D(\mathbf{x}) = \left( E \begin{pmatrix} x_1 - E(x_1) \\ \vdots \\ x_p - E(x_p) \end{pmatrix} \left( x_1 - E(x_1) \cdots x_p - E(x_p) \right) \right)$

or $\quad D(\mathbf{x}) = E\left\{ (\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^t \right\}$ (1)

It can be shown similarly that

$$Cov(\mathbf{x}, \mathbf{y}) = E\left\{ (\mathbf{x} - E(\mathbf{x}))(\mathbf{y} - E(\mathbf{y}))^t \right\} \tag{2}$$

Let us illustrate this in the following example.

**Example 1:** Let $x_1$, $x_2$ and $x_3$ be random variables with means $2.3, -4.1$ and $1.5$ respectively and the variances $4, 9$ and $16$, respectively. Let $\rho_{ij}$ denotes the correlation coefficient between $x_i$ and $x_j$; $j = 1, 2, 3$ and $i = 1, 2, 3$. Let $\rho_{12} = 0.5$, $\rho_{13} = 0.3$ and $\rho_{23} = -0.4$. Write down the mean vector and the variance covariance matrix of $\mathbf{x} = (x_1\ x_2\ x_3)^t$.

**Solution:** The mean vector of $\mathbf{x}$ is $E(x) = \begin{pmatrix} E(x_1) \\ E(x_2) \\ E(x_3) \end{pmatrix} = \begin{bmatrix} 2.3 \\ -4.1 \\ 1.5 \end{bmatrix}$.

The variances and covariances are given by

$$\sigma_{11} = V(x_1) = 4,\ \sigma_{22} = V(x_2) = 9 \text{ and } \sigma_{33} = V(x_3) = 16$$

$$\sigma_{12} = \text{Cov}(x_1, x_2) = \rho_{12}\sqrt{V(x_1).\,V(x_2)} \quad \left[\text{we know that } \rho_{ij} = \frac{\text{cov}(x_i, x_j}{\sqrt{V(x_i).V(x_j)}}\right]$$
$$= 0.5 \times 2 \times 3 = 3.0$$

$$\sigma_{13} = \text{Cov}(x_1, X_3) = \rho_{13}\sqrt{V(x_1).V(x_3)}$$
$$= 0.3 \times 2 \times 4 = 2.4$$

$$\sigma_{23} = \text{Cov}(x_2, x_3) = \rho_{23}\sqrt{V(x_2).\,V(x_3)}$$
$$= -0.4 \times 3 \times 4 = -4.8$$

$$\text{Therefore } \Sigma = \begin{pmatrix} 4.0 & 3.0 & 2.4 \\ 3.0 & 9.0 & -4.8 \\ 2.4 & -4.8 & 16 \end{pmatrix}$$

Notice that $\mathbf{\Sigma}$ is symmetric in the above example. In fact, this is true for every variance-covariance matrix $\mathbf{\Sigma}$ because $\sigma_{ij} = \text{Cov}(x_i, x_j) = \text{Cov}(x_j, x_i) = \sigma_{ji}$ for all $i$ and $j$. Since the leading principal minors of $\mathbf{\Sigma}$ are 4.0, 27.0 and 218.88, respectively, it follows, from Theorem 5 of Unit 14 that $\mathbf{\Sigma}$ is positive definite.

Now try an exercise.

---

E1) Let $x_1$ and $x_2$ be two random variables with joint probability distribution given in the following table

| $x_1 \backslash x_2$ | 0 | 1 | $p_1(x_1)$ |
|---|---|---|---|
| −1 | 0.14 | 0.16 | 0.3 |
| 0 | 0.26 | 0.04 | 0.3 |
| 1 | 0.30 | 0.10 | 0.4 |
| $p_2(x_2)$ | 0.7 | 0.3 | |

Find the variance-covariance matrix for $x = (x_1, x_2)$.

---

Let $\mathbf{x}$ be a random vector. Consider $\mathbf{l}^t\mathbf{x}$ where $\mathbf{l}^t$ is a fixed vector (i.e., the components of $\mathbf{l}$ are not random variables). We shall now find the mean and variance of $\mathbf{l}^t\mathbf{x}$ in the following theorem.

**Theorem 1:** Let $\mathbf{x}_{p\times 1}$ be a random vector with $E(\mathbf{x}) = \boldsymbol{\mu}$ and its variance-covariance matrix equal to $\mathbf{\Sigma}$. Let $\mathbf{l}$ be a fixed vector and let $\mathbf{l}^t\mathbf{x} = l_1x_1 + l_2x_2 + \cdots + l_px_p$ be a linear combination of the components of $\mathbf{x}$. Then $E(\mathbf{l}^t\mathbf{x}) = \mathbf{l}^t\ E(\mathbf{x}) = \mathbf{l}^t\boldsymbol{\mu}$ and $V(\mathbf{l}^t\mathbf{x}) = \mathbf{l}^t\mathbf{\Sigma}\mathbf{l}$. Also $\text{Cov}(\mathbf{l}^t\mathbf{x}, \mathbf{m}^t\mathbf{x}) = \mathbf{l}^t\mathbf{\Sigma}\mathbf{m}$ where $\mathbf{m}$ is a fixed vector.

**Proof:** $E(\mathbf{l}^t\mathbf{x}) = E(l_1x_1 + l_2x_2 + \cdots + l_px_p)$

$$= l_1E(x_1) + l_2E(x_2) + \cdots + l_pE(x_p)$$
$$= \mathbf{l}^tE(\mathbf{x}) = \mathbf{l}^t\boldsymbol{\mu}$$

$$V(\mathbf{l}^t\mathbf{x}) = V(l_1x_1 + l_2x_2 + \cdots + l_px_p)$$
$$= Cov(l_1x_1 + l_2x_2 + \cdots l_px_p, l_1x_1 + l_2x_2 + \cdots + l_px_p)$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{p} l_i, l_j \sigma_{ij} = \mathbf{l}^t\Sigma\mathbf{l}$$

$$Cov(\mathbf{l}^t\mathbf{x}, \mathbf{m}^t\mathbf{x}) = \sum_{i=1}^{p}\sum_{j=1}^{p} l_i, m_j \sigma_{ij} = \mathbf{l}^t\Sigma\mathbf{m}$$

Now let us illustrate the above theorem in the following example.

**Example 2:** Find the mean and variance of $\mathbf{l}^t\mathbf{x}$ in Example 1, where $\mathbf{l} = \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}$. Also find the covariance between $\mathbf{l}^t\mathbf{x}$ and $\mathbf{m}^t\mathbf{x}$, where $\mathbf{m} = \begin{pmatrix}2\\-1\\-1\end{pmatrix}$.

**Solution:** Mean of $\mathbf{l}^t\mathbf{x} = \mathbf{l}^tE(\mathbf{x}) = l_1E(x_1) + l_2E(x_2) + l_3E(x_3)$

$$= \frac{1}{3}(E(x_1) + E(x_2) + E(x_3)) = \frac{1}{3}(2.3 - 4.1 + 1.5)$$
$$= \frac{1}{3} \times -0.3 = -0.1$$

$$V(\mathbf{l}^t\mathbf{x}) = \mathbf{l}^t\sum\mathbf{l} \qquad \text{(Usingh Theorem 1)}$$
$$= \frac{1}{3}(1\,1\,1)\begin{pmatrix}4.0 & 3.0 & 2.4\\3.0 & 9.0 & -4.8\\2.4 & -4.8 & 16\end{pmatrix}\frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}$$
$$= \frac{1}{9}(9.4 \quad 7.2 \quad 13.6)\begin{pmatrix}1\\1\\1\end{pmatrix}$$
$$= \frac{1}{9} \times 30.2 = 3.36$$

$$Cov(\mathbf{l}^t\mathbf{x}, \mathbf{m}^t\mathbf{x}) = \mathbf{l}^t\sum\mathbf{m} \qquad \text{(using Theorem 1)}$$
$$= \frac{1}{3}(1\,1\,1)\begin{pmatrix}4.0 & 3.0 & 2.4\\3.0 & 9.0 & -4.8\\2.4 & -4.8 & 16\end{pmatrix}\begin{pmatrix}2\\-1\\-1\end{pmatrix}$$
$$= \frac{1}{3}(9.4 \quad 7.2 \quad 13.6)\begin{pmatrix}2\\-1\\-1\end{pmatrix} = \frac{1}{3} \times -2.0 = \frac{-2}{3}.$$

Now try an exercise.

---

E2) Let $\mathbf{x} = (x_1 \; x_2 \; x_3 \; x_4)^t$ be an random vector with mean vector $\mu = (2\,1 -1 -2)^t$ and variance covariance matrix $\sum = \begin{pmatrix}1 & 0.2 & 0.2 & 0.2\\0.2 & 1 & 0.2 & 0.2\\0.2 & 0.2 & 1 & 0.2\\0.2 & 0.2 & 0.2 & 1\end{pmatrix}$.

Find the mean and variance of $\mathbf{l}^t\mathbf{x}$ and $\mathrm{Cov}(\mathbf{l}^t\mathbf{x}, \mathbf{m}^t\mathbf{x})$, where $\mathbf{l}^t = (1\,1\,1\,1)$ and $\mathbf{m}^t = (1\,1\,-1\,-1)$.

---

We shall now extend the results of Theorem 1 to several linear functions.

**Theorem 2:** Let $\mathbf{x}$ and $\mathbf{y}$ be random vectors of orders $p\times 1$ and $q\times 1$ respectively. Let $E(\mathbf{x}) = \mu$, $E(\mathbf{y}) = \theta$, $D(\mathbf{x}) = \Sigma$, $D(\mathbf{y}) = \Gamma$ and $\mathrm{Cov}(\mathbf{x}, \mathbf{y}) = \Delta$.

Let $\mathbf{B}_{r\times p}$ and $\mathbf{C}_{s\times q}$ be fixed matrices (non-random). Then the following hold.

(a) $E(\mathbf{Bx}) = \mathbf{B}\mu$

(b) $D(\mathbf{Bx}) = \mathbf{B\Sigma B}^t$

(c) $\mathrm{Cov}(\mathbf{Bx}, \mathbf{Cy}) = \mathbf{B\Delta C}^t$

**Proof:** (a) $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_r \end{pmatrix}$ where $\mathbf{b}_i$ is the $i^{th}$ row of $\mathbf{B}$, $i = 1, \ldots, r$.

Now $E(\mathbf{Bx}) = E\begin{pmatrix} \mathbf{b}_1\mathbf{x} \\ \vdots \\ \mathbf{b}_r\mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 E(\mathbf{x}) \\ \vdots \\ \mathbf{b}_r E(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{M} \\ \mathbf{b}_r \end{pmatrix} E(\mathbf{x}) = \mathbf{B}\mu$

(b) $D(\mathbf{Bx}) = \mathrm{Cov}(\mathbf{Bx}, \mathbf{Bx}) = \mathrm{Cov}\left[\begin{pmatrix} \mathbf{b}_1\mathbf{x} \\ \vdots \\ \mathbf{b}_r\mathbf{x} \end{pmatrix}, \begin{pmatrix} \mathbf{b}_1\mathbf{x} \\ \vdots \\ \mathbf{b}_r\mathbf{x} \end{pmatrix}\right]$

Thus, the $(i, j)^{th}$ element of $D(\mathbf{Bx})$ is $\mathrm{Cov}(\mathbf{b}_i\mathbf{x}, \mathbf{b}_j\mathbf{x}) = \mathbf{b}_i\Sigma\mathbf{b}_j^t$ for $i, j = 1, \ldots, p$ from Theorem 1.

Thus, $D(\mathbf{Bx}) = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_r \end{pmatrix} \sum (\mathbf{b}_1^t : \ldots : \mathbf{b}_r^t) = \mathbf{B\Sigma B}^t$.

(c) $\mathrm{Cov}(\mathbf{Bx}, \mathbf{Cy}) = \mathrm{Cov}\left[\begin{pmatrix} \mathbf{b}_1\mathbf{x} \\ \vdots \\ \mathbf{b}_r\mathbf{x} \end{pmatrix}, \begin{pmatrix} \mathbf{c}_1\mathbf{y} \\ \vdots \\ \mathbf{c}_s\mathbf{y} \end{pmatrix}\right]$ where $\mathbf{C}_j$ is the $j^{th}$ row of $\mathbf{C}$, $j = 1, \ldots, s$.

The $(i, j)^{th}$ element $\mathrm{Cov}(\mathbf{Bx}, \mathbf{Cy}) = \mathbf{b}_i\Delta\mathbf{c}_j^t$.

Hence $\mathrm{Cov}(\mathbf{Bx}, \mathbf{Cy}) = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_r \end{pmatrix} \Delta(\mathbf{c}_1^t : \ldots : \mathbf{c}_s^t) = \mathbf{B\Delta C}^t$

**Example 3:** Let $\Sigma$ be the variance-covariance matrix of a random vector $\mathbf{x}$ of order $p\times 1$. Let $r_{ij}$ denotes the correlation coefficient between $x_i$ and $x_j$, $i, j = 1, \ldots, p$. Write $\mathbf{R} = ((r_{ij}))$. $\mathbf{R}$ is called the correlation matrix of $\mathbf{x}$.

$$\text{Write} = \mathbf{T} = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\sigma_{pp}} \end{pmatrix}.$$

Thus, $\mathbf{T}$ is a diagonal matrix with $i^{th}$ diagonal element equal to the standard deviation of $x_i, i = 1, \ldots, p$. Assume that $\sigma_{11}, \ldots, \sigma_{pp}$ are strictly positive. Show that $\mathbf{R} = \mathbf{T}^{-1}\Sigma\mathbf{T}^{-1}$.

**Solution:** Notice that $r_{ij} = \dfrac{\mathrm{Cov}(x_i, x_j)}{\sqrt{V(x_i)V(x_j)}} = \dfrac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$

The $(i, j)^{th}$ element of $\mathbf{T}^{-1}\Sigma\mathbf{T}^{-1}$ is

$$\begin{pmatrix} 0 & \ldots & \sigma_{ii}^{-\frac{1}{2}} & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \sigma_{i1} & \ddots & \sigma_{ip} \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ \sigma_{jj}^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{ii}^{-\frac{1}{2}}\sigma_{i1} & \ldots & \underbrace{\sigma_{ii}^{-\frac{1}{2}}\sigma_{ij}}_{j^{th}\text{ element}} & \ldots & \sigma_{ii}^{-\frac{1}{2}}\sigma_{ip} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ \sigma_{jj}^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ j^{th}\text{ element} \\ \\ \\ \\ \end{array}$$

$$= \sigma_{ii}^{-\frac{1}{2}}\sigma_{ij}\sigma_{jj}^{-\frac{1}{2}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} = r_{ij}.$$

Thus, $\mathbf{R} = \mathbf{T}^{-1}\Sigma\mathbf{T}^{-1}$.

This establishes a relationship between the variance-covariance matrix and the correlation matrix of a random vector.

Now, try the following exercise.

---

E3) (a) Give an alternative matrix proof of Theorem 2(b) and (c).

(b) Obtain the correlation matrix of a random vector $\boldsymbol{x}$ with the following variance-covariance matrix.

$$\Sigma = \begin{pmatrix} 4.0 & 3.0 & 2.4 \\ 3.0 & 9.0 & -4.8 \\ 2.4 & -4.8 & 16 \end{pmatrix}$$

(Compare your results with the matrices in Example 1.)

---

We are ready now to show the equality of the class of all nnd matrices with the class of all variance-covariance matrices.

**Theorem 3:** (a) Every variance-covariance matrix is nonnegative definite (nnd).

(b) Every nnd matrix is the variance-covariance matrix of a random vector.

**Proof:** (a) Let $\Sigma$ be the variance-covariance matrix of a random vector $\mathbf{x}$. Then for each fixed $\mathbf{l}$, $\mathbf{l}^t\Sigma\,\mathbf{l} = V(\mathbf{l}^t\mathbf{x}) \geq 0$ (Since variance of a random variable is nonnegative.) Hence $\Sigma$ is nnd.

(b) Let $\Sigma_{p\times p}$ be an nnd matrix. Then by Theorem 4(b) of Unit 14, there exists a matrix $\mathbf{C}$ of order $p\times r$ for some positive integer $r$ such that $\Sigma = \mathbf{CC}^t$. Let $x_1, x_2, \ldots, x_r$ be independent random variables each with variance 1. Write $\mathbf{x}^t = (x_1, \ldots, x_r)$. Then $D(\mathbf{x}) = \mathbf{I}_{r\times r}$ ($\mathbf{I}_{r\times r}$ is the identity matrix of order $r\times r$). Write $\mathbf{y} = \mathbf{Cx}$. Then by theorem 2, $D(\mathbf{y}) = \mathbf{CIC}^t = \mathbf{CC}^t = \Sigma$.

**Corollary:** The variance-covariance matrix $\Sigma$ of a random vector $\mathbf{x}$ is positive semi-definite if and only if there exists a fixed non-null vector $\mathbf{l}$ such that $\mathbf{l}^t\mathbf{x}$ is a constant with probability 1.

**Proof:** $\Sigma$ is positive semi-definite

$\Leftrightarrow$ There exists fixed $\mathbf{l} \neq \mathbf{0}$, such that $\mathbf{l}^t\Sigma\,\mathbf{l} = 0$

$\Leftrightarrow$ There exists fixed $\mathbf{l} \neq \mathbf{0}$ such that $V(\mathbf{l}^t\mathbf{x}) = 0$

$\Leftrightarrow$ There exists a fixed vector $\mathbf{l} \neq \mathbf{0}$ such that $\mathbf{l}^t\mathbf{x}$ is a constant with probability 1.

In general, independent/uncorrelated random variables are easier to handle statistically than the correlated random variables. We shall now give methods of transforming a random vector (the components of which are correlated) with positive definite variance covariance matrix to a random vector the components of which are uncorrelated, by a suitable linear transformation.

Let $\mathbf{x}$ be a random vector of order $p\times 1$ with $D(\mathbf{x}) = \Sigma$ is a non-diagonal pd matrix.

**Method 1:** Let $\Sigma = \mathbf{P}\Lambda\mathbf{P}^t$ be a spectral decomposition of $\Sigma$ where $\mathbf{P}$ is orthogonal and $\Lambda$ diagonal. Since $\Sigma$ is pd, so is $\Lambda$. Write $\xi = \mathbf{P}^t\mathbf{x}$. Then $D(\xi) = \mathbf{P}^t\mathbf{\Sigma}\mathbf{P} = \mathbf{P}^t\mathbf{P}\Lambda\mathbf{P}^t\mathbf{P} = \Lambda$, since $\mathbf{P}$ is orthogonal. Since $\Lambda$ is a diagonal matrix, the components of $\xi$ are uncorrelated. Also $V(\xi_i) = \lambda_i$, the $i^{th}$ diagonal element of $\Lambda$.

If we write $\mathbf{y} = \Lambda^{-1/2}\mathbf{P}^t\mathbf{x}$, then $D(\mathbf{y}) = \Lambda^{-1/2}\mathbf{P}^t\mathbf{P}\Lambda\mathbf{P}^t\mathbf{P}\Lambda^{-1/2} = \mathbf{I}$. Here

$$\Lambda^{-1/2} = \mathrm{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_p}}\right).$$

Thus, the components of $\mathbf{y}$ are uncorrelated each with variance 1.

Before proceeding further, let us recall that if $\Sigma$ is pd, then there exists a nonsingular matrix $\mathbf{B}$ such that $\Sigma = \mathbf{BB}^t$. Write $\mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$. Then $D(\mathbf{y}) = \mathbf{B}^{-1}\Sigma\mathbf{B}^{-1^t} = \mathbf{B}^{-1}\mathbf{BB}^t\,\mathbf{B}^{-1^t} = \mathbf{I}$. Hence the components of $\mathbf{y}$ are uncorrelated, each with variance 1. Notice that in method 1, $\mathbf{P}\Lambda^{1/2}$ is a choice for the matrix $\mathbf{B}$.

In Unit 14, we gave a method of computing a lower triangular square root of a pd matrix. Before giving Method 2, we shall give another algorithm of obtaining a lower triangular square root of a pd matrix. This algorithm also helps us in getting the inverse of the triangular square root as a bonus whereby we can write down $\mathbf{y}$ immediately.

Let $\Sigma$ be a pd matrix of order $p \times p$ and $I$ be $p \times p$ unit matrix.

Algorithm:

Step 1: Form the matrix $\mathbf{T} = (\Sigma : \mathbf{I})$

Step 2: Set $i = 1$ ($i$ is the sweep out number)

Step 3: Replace the $i^{th}$ row of $\mathbf{T}$ by ($i^{th}$ row of $\mathbf{T}$) $\div \sqrt{t_{ii}}$

Step 4: Is $i = p$? If yes go to Step 9. If no go to Step 5.

Step 5: Set $j = i + 1$

Step 6: Replace $j^{th}$ row of $\mathbf{T}$ by ($j^{th}$ row of $\mathbf{T}$) $- \dfrac{t_{ji}}{t_{ii}}$ ($i^{th}$ row of $\mathbf{T}$)

Step 7: Is $j = p$? If yes go to Step 9. If no go to Step 8.

Step 8: Replace $j$ by $j+1$ and go to Step 6.

Step 9: Is $i = p$? If yes, go to Step 11. If no go to Step 10.

Step 10: Replace $i$ by $i+1$ and go to Step 3.

Step 11: The first $p$ columns of $\mathbf{T}$ form $\mathbf{B}^t$ and the last $p$ columns of $\mathbf{T}$ form $\mathbf{B}^{-1}$ where $\Sigma = \mathbf{BB}^t$ with $\mathbf{B}$ lower triangular.

**Method 2:** Obtain a lower triangular square root of $\Sigma$ and $\mathbf{B}^{-1}$ by the above algorithm. Then write down $\mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$. Since $\mathbf{B}$ is lower triangular, so is $\mathbf{B}^{-1}$.

$$\text{So } \mathbf{y} = \mathbf{y} = \begin{pmatrix} b^{11} & 0 & \cdots & 0 \\ b^{21} & b^{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b^{p1} & 0 & 0 & b^{pp} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \text{ where } \mathbf{B}^{-1} = ((b^{ij})).$$

Observe that

$$\begin{aligned} y_1 &= b^{11}x_1 \\ y_2 &= b^{21}x_1 + b^{22}x_2 \\ &\vdots \\ y_i &= b^{i1}x_1 + \cdots + b^{ii}x_i \\ y_p &= b^{p1}x_1 + \cdots + b^{pp}x_p. \end{aligned}$$

Thus, the first component of $\mathbf{y}$ is a scalar multiple of the first component of $\mathbf{x}$. The second component of $\mathbf{y}$ is a linear combination of the first two components of $\mathbf{x}$, and so on.

We now illustrate the above methods with examples.

**Example 4:** Let $\mathbf{x}$ be a random vector with $D(\mathbf{x}) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

(a) Find an orthogonal transformation $\xi = \mathbf{Px}$ (where $\mathbf{P}$ is an orthogonal matrix) such that the components of $\xi$ are uncorrelated. Obtain the variances of $\xi_1$ and $\xi_2$.

(b) Find a nonsingular linear transformation $\mathbf{y} = \mathbf{Bx}$, so that the components of $\mathbf{y}$ are uncorrelated each with variance 1 by both the methods described above.

**Solution:**(a) Using the solution of E4 of Unit 1, we have the spectral decomposition of

$$D(\mathbf{x}) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3\mathbf{p}_1\mathbf{p}_1^t + 1\mathbf{p}_2\mathbf{p}_2^t = (\mathbf{p}_1 : \mathbf{p}_2) \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{p}_1^t \\ \mathbf{p}_2^t \end{pmatrix} = \mathbf{P} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{P}^t$$

where $\mathbf{p}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$ and $\mathbf{p}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$ and $\mathbf{P} = (\mathbf{p}_1 : \mathbf{p}_2)$.

Hence the required orthogonal matrix is $\mathbf{P}' = \mathbf{P}^t = \begin{pmatrix}\mathbf{p}_1^t\\\mathbf{p}_2^t\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}$. The transformation is $\xi = \mathbf{P}^t\mathbf{x}$.

Thus, $\xi_1 = \frac{1}{\sqrt{2}}(x_1 + x_2)$ and $\xi_2 = \frac{1}{\sqrt{2}}(x_1 - x_2)$

Now $D(\xi) = \mathbf{P}^t\mathbf{P}\begin{pmatrix}3 & 0\\0 & 1\end{pmatrix}\mathbf{P}^t\mathbf{P} = \begin{pmatrix}3 & 0\\0 & 1\end{pmatrix}$

So $V(\xi_1) = 3$ and $V(\xi_2) = 1$

**(b)** Using Method 2 write $y_1 = \frac{1}{\sqrt{3}}\xi_1$ and $y_2 = \xi_2$

or $\mathbf{y} = \begin{pmatrix}\frac{1}{\sqrt{3}} & 0\\0 & 1\end{pmatrix}\xi = \begin{pmatrix}\frac{1}{\sqrt{3}} & 0\\0 & 1\end{pmatrix}\mathbf{P}^t\mathbf{x}$

Then $y_1$ and $y_2$ are uncorrelated and $V(y_1)$ and $V(y_2) = 1$.

Using Method 1 for the matrix $\begin{pmatrix}2 & 1 & 1 & 0\\1 & 2 & 0 & 1\end{pmatrix}$ and proceed as below.

$$\begin{array}{cccc l}
2 & 1 & 1 & 0 & \ldots\ldots\ldots\ldots (1)\\
1 & 2 & 0 & 1 & \ldots\ldots\ldots\ldots (2)\\
\hline
\sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \ldots\ldots\ldots\ldots (3) = (1) \div \sqrt{2}\\
0 & \frac{3}{2} & -\frac{1}{2} & 1 & \ldots\ldots\ldots\ldots (4) = (2) - \frac{1}{\sqrt{2}}(3)\\
\hline
\sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \ldots\ldots\ldots\ldots (5) = (3)\\
0 & \sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{6}} & \frac{\sqrt{2}}{\sqrt{3}} & \ldots\ldots\ldots\ldots (6) = (4) \div \sqrt{\frac{3}{2}}.
\end{array}$$

Thus, $\mathbf{B} = \begin{pmatrix}\sqrt{2} & 0\\\frac{1}{\sqrt{2}} & \frac{\sqrt{3}}{\sqrt{2}}\end{pmatrix}$ and $\mathbf{B}^{-1} = \begin{pmatrix}\frac{1}{\sqrt{2}} & 0\\-\frac{1}{\sqrt{6}} & \frac{\sqrt{2}}{\sqrt{3}}\end{pmatrix}$

The required transformation is $\mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$.

Hence $y_1 = \frac{1}{\sqrt{2}}x_1$

$y_2 = -\frac{1}{\sqrt{6}}x_1 + \frac{\sqrt{2}}{\sqrt{3}}x_2$.

Let us try the following exercises.

E4) Consider a random vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$ with $E(\mathbf{x}) = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 3 \\ -2 \end{pmatrix}$ and

$D(\mathbf{x}) = \begin{pmatrix} 5 & 1 & 1 & 1 & 1 \\ 1 & 4 & 0 & 2 & 0 \\ 1 & 0 & 6 & 1 & 2 \\ 1 & 2 & 1 & 8 & 3 \\ 1 & 0 & 2 & 3 & 9 \end{pmatrix}$. Write $\mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and $\xi = \begin{pmatrix} x_4 \\ x_5 \end{pmatrix}$. Thus, $\mathbf{x} = \begin{pmatrix} \mathbf{y} \\ \xi \end{pmatrix}$.

Let $\mathbf{B} = \begin{pmatrix} 2 & -1 & 1 \\ 4 & 5 & 2 \end{pmatrix}$ and $\mathbf{C} = \begin{pmatrix} 4 & 2 \\ -1 & 3 \\ 0 & 1 \end{pmatrix}$. Write $\mathbf{u} = \mathbf{B}\mathbf{y}$ and $\mathbf{v} = \mathbf{C}\xi$.

Compute the following:

(a) $E(\mathbf{u}), D(\mathbf{u})$

(b) $E(\mathbf{v}), D(\mathbf{v})$ and

(c) $\mathrm{Cov}(\mathbf{u}, \mathbf{v})$.

E5) Let $\mathbf{x}$ be a random vector with $E(\mathbf{x}) = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ and $D(\mathbf{x}) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$. Notice that each row sum of $D(\mathbf{x})$ is $0$. Obtain a linear combination $\mathbf{l}^t x$ which is constant with probability $1$. What is the value of this constant?

E6) Let $\mathbf{x}$ be a random vector with $D(\mathbf{x}) = \begin{pmatrix} 4 & 2 & 6 \\ 2 & 17 & 27 \\ 6 & 27 & 70 \end{pmatrix}$. Use Method 2 to obtain a lower triangular square root matrix $\mathbf{B}$ and its inverse such that the components of $\mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$ are uncorrelated each with variance $1$.

E7) Let $\Sigma$ be the variance covariance matrix of a random vector $\mathbf{x}$. Let $\sigma_{ii} = 0$. Show that the covariance of $x_i$ with all other components is zero.

In the next section, we shall discuss the multivariate distributions.

## 15.4 MULTIVARIATE DISTRIBUTIONS

In this section, we give an introduction to multivariate distributions. By a multivariate distribution, we mean the joint distribution of more than one random variables. We give examples of discrete and continuous multivariate distributions. From the joint distribution of random variables $x_1, \ldots, x_p$ we obtain the individual distribution (which we call the marginal distribution) of each $x$. We define the concept of conditional distribution. We briefly study the concept of independence of random variables and its relation to uncorrelatedness. First, let us consider a few examples.

**Example 5:** A college has 2 specialists in long distance running, 4 specialists in Tennis and 6 top level cricketers among its students. The college plans to send 3 sportsmen from the above for participating in the University sports and games. The three sportsmen are selected randomly from among the above 12. Let $x_1$ and $x_2$ denote respectively the number of long distance specialists and the number of tennis specialists chosen. The joint probability mass function of $x_1$ and $x_2$ is defined as $P\{x_1 = i, x_2 = j\}$ for $i = 0, 1, 2$ and $j = 0, 1, 2, 3$. Obtain the joint probability mass function of $x_1, x_2$.

**Solution:** The joint probability mass function of $x_1, x_2$ is given by

$$p(i,\ j) = P(x_1 = i,\ x_2 = j)$$

$$= \begin{cases} \dfrac{{}^{2}C_i \ \ {}^{4}C_j \ \ {}^{6}C_{3-i-j}}{{}^{12}C_3}; & 0 \le i + j \le, \ \ i = 0, 1, 2, \quad j = 0, 1, 2, 3 \\ 0 \ \ \text{otherwise} & \end{cases}$$

The marginal probability mass function of $x_1$ is given by

$$p_{x_1}(i) = P(x_1 = i) = \sum_{j=0}^{3-i} P(x_1 = i,\ x_2 = j)$$

$$= \frac{\left({}^{2}C_i\right)}{\left({}^{12}C_i\right)} \sum_{j=0}^{3-i} \left({}^{4}C_j\right)\left({}^{6}C_{3-i-j}\right)$$

$$= \begin{cases} \dfrac{\left({}^{2}C_i\right)\left({}^{10}C_{3-i}\right)}{\left({}^{12}C_3\right)}, & i = 0, 1, 2 \\ 0, \ \ \text{otherwise} & \end{cases}$$

Similarly, the marginal probability mass function of $x_2$ is computed as

$$p_{x_2}(j) = \frac{{}^{4}C_j \ \ {}^{8}C_{3-j}}{{}^{12}C_3}$$

$$j = 0, 1, 2, 3$$

$$= 0, \ \ \text{otherwise}$$

The values taken by $x_1$ and $x_2$ and the corresponding probabilities constitute *the joint distribution* of $x_1$ and $x_2$ and can be represented in a tabular form as follows:

**Table 15.1: Joint distribution of $x_1$ and $x_2$**

| Value taken by | $\to x_2$ | 0 | 1 | 2 | 3 | Row sum |
|---|---|---|---|---|---|---|
| $\downarrow x_1$ | | Joint probability | | | | |
| 0 | Joint probability | $\frac{20}{220}$ | $\frac{60}{220}$ | $\frac{36}{220}$ | $\frac{4}{220}$ | $\frac{120}{220}$ |
| 1 | | $\frac{30}{220}$ | $\frac{48}{220}$ | $\frac{12}{220}$ | 0 | $\frac{90}{220}$ |
| 2 | | $\frac{6}{220}$ | $\frac{4}{220}$ | 0 | 0 | $\frac{10}{220}$ |
| Column sum | | $\frac{56}{220}$ | $\frac{112}{220}$ | $\frac{48}{220}$ | $\frac{4}{220}$ | 1 |

For the joint distribution table, it is easy to write down the distributions of $x_1$ and $x_2$ which we call the marginal distributions of $x_1$ and $x_2$, respectively.

$$P\{x_1 = 0\} = P\{x_1 = 0,\ x_2 = 0\} + P\{x_1 = 0,\ x_2 = 1\} + P\{x_1 = 0,\ x_2 = 2\} + P\{x_1 = 0,\ x_2 = 3\}$$
$$= \frac{20}{220} + \frac{60}{220} + \frac{36}{220} + \frac{4}{220}.$$

Notice that $P\{x_1 = 0\}$ is the row-sum corresponding to $x_1 = 0$ in the Table 15.1. Accordingly this is recorded as row-sum corresponding to $x_1 = 0$. Similarly, the second and third row-sums are the probabilities of $x_1 = 1$ and $x_1 = 2$, respectively. Thus, the marginal distribution of $x_1$ is

**Table 15.2: Marginal distribution of $x_1$**

| Value | 0 | 1 | 2 |
|---|---|---|---|
| Probability | $\frac{120}{220}$ | $\frac{90}{220}$ | $\frac{10}{220}$ |

Similarly, the marginal distribution of $x_2$ is obtained using the column sums in Table 15.1. Thus, the marginal distribution of $x_2$ is

**Table 15.3: Marginal distribution of $x_2$**

| Value | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Probability | $\frac{56}{220}$ | $\frac{112}{220}$ | $\frac{48}{220}$ | $\frac{4}{220}$ |

Suppose we are given additional information that no long distance running specialist is selected, or in other words, we know that $x_1 = 0$. Then what are the probabilities for $x_2 = 0, 1, 2, 3$ given this additional information? Notice that we are looking for the conditional probabilities: $P\{x_2 = j \mid x_1 = 0\}$, for $j = 0, 1, 2, 3$. We can compute them as

$$p_{x_2/x_1}(j/0)$$
$$= P(x_2 = j/x_1 = 0) = \frac{P(x_1 = 0,\ x_2 = j)}{P(x_1 = 0)}$$
$$= \frac{\left({}^2C_0\right)\left({}^4C_j\right)\left({}^6C_{3-j}\right)}{\left({}^2C_0\right)\left({}^{10}C_3\right)}$$
$$= \frac{\left({}^4C_j\right)\left({}^6C_{3-j}\right)}{\left({}^{10}C_3\right)},\quad j = 0, 1, 2, 3$$

The above distribution of $x_2$ given $x_1 = 0$ is called the *conditional distribution* of $x_2$ given $x_1 = 0$ and can be expressed neatly in the following table.

**Table 15.4: Conditional distribution of $x_1 | x_2 = 0$**

| Value | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Probability | $\frac{1}{6}$ | $\frac{1}{2}$ | $\frac{3}{10}$ | $\frac{1}{30}$ |

Now, try the following exercises.

E8) In Example 5, let $x_3$ = number of cricketers chosen. Write down the joint distribution of $x_2$ and $x_3$. Obtain the marginal distributions of $x_2$ and $x_3$. Obtain the conditional distribution of $x_3$ given $x_2 = 1$.

E9) In Example 5, let $p_{ijk}$ denotes $P\{x_1 = i, x_2 = j, x_3 = k\}$. Obtain $p_{ijk}$ for $i = 0, 1, 2,\ j = 0, 1, 2, 3$ and $k = 0, 1, 2, \ldots, 6,\ i + j + k = 3$. The values of $x_1$, $x_2$ and $x_3$ and the corresponding $p_{ijk}$ constitute the joint distribution of $x_1$, $x_2$ and $x_3$.

E10) In Example 5, obtain the variance-covariance matrix of $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$.

The random variables $x_1, x_2$ and $x_3$ in Example 5, E9 and E10 are discrete. Then $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ is called a **discrete random vector** and the distribution of $\mathbf{x}$ (the joint distribution of $x_1$, $x_2$ and $x_3$) in such a case is called discrete multivariate distribution. More generally, we can define discrete multivariate distribution for $x = (x_1, x_2, \ldots, x_p)^t$ by joint probability.

$P(x_1 = i_1, x_2 = i_2, \ldots, x_p = i_p) = p_{i_1 i_2 \cdots i_p}$.

Then for a set A of p-tuples $P(x \in A) = \sum p_{i_1, i_2, \ldots, i_p}\ (i_1, i_2, \ldots, i_p) \in A$.

On the other hand we say that $x_1, \ldots, x_p$ are jointly continuous ($\mathbf{x} = (x_1 \ldots x_p)^t$ is continuous) if there exists a function $f(u_1, \ldots, u_p)$ defined for all $u_1, \ldots, u_p$ having the property that for a set A of p-tuples, i.e. $A \subseteq R^p$.

$P((x_1, \ldots, x_p) \in A) = \int \ldots \int f(u_1, \ldots, u_2)\, du_1, \ldots, du_p$, where the integration is over $(u_1, \ldots, u_p) \in A$. Thus, $P(x \in R^p) = \int \ldots \int_{R^p} f(u_1, \ldots, u_p)\, du_1, \ldots du_p = 1$.

The function $f(u_1, \ldots, u_p)$ is called the probability density function of $\mathbf{x}$ (or joint probability density function of $(x_1, \ldots, x_p)$). If $A_1, \ldots, A_p$ are sets of real numbers such that $\mathbf{A} = \{(u_1, \ldots, u_p), u_i \in A_i, i = 1, \ldots, p\}$ we can write

$$P\{(x_1, \ldots, x_p) \in A\} = P\{x_i \in A_i, i = 1, \ldots, p\} = \int_{A_1} \ldots \int_{A_p} f(u_1 \ldots u_p)\, du_1 \ldots du_p$$

The distribution function of $\mathbf{x} = (x_1 \ldots x_p)^t$ is defined as

$$F(a_1, a_2, \ldots, a_p) = P(x_1 \le a_1, \ldots, x_p \le a_p\}$$

$$= \int_{-\infty}^{a_1} \ldots \int_{-\infty}^{a_p} f(u_1 \ldots u_p)\, du_1 \ldots du_p,\ (a_1, a_2, \ldots, a_p) \in R^p$$

It follows upon differentiation, that $f(a_1, \ldots, a_p) = \frac{\partial^p}{\partial a_1 \ldots \partial a_p} F(a_1,\ldots,a_p)$ provided the partial derivatives are defined. At a countable set where partial derivatives are not defined we set it equal to zero.

Another interpretation of the density function of $\mathbf{x}$ can be given using the following:

$$P\{a_i < x_i < a_i + \delta a_i, i = 1, \ldots, p\} = \int_{a_1}^{a_1+\delta a_1} \ldots \int_{a_1}^{a_1+\delta a_p} f(u_1,\ldots,u_p)du_1,\ldots,du_p$$
$$\approx f(a_1, \ldots, a_p)\delta a_1, \ldots \delta a_p .$$

when $\delta a_1, i = 1, \ldots, p$ are infinitesimally small and $f$ is continuous. Thus, $f(a_1, \ldots, a_p)\delta a_1, \delta a_2, \ldots, \delta a_p$ is a measure of the chance that the random vector $\mathbf{x}$ is in a small neighborhood of $(a_1, \ldots, a_p)$.

Let $f_x(u_1, u_2, \ldots, u_p)$ be the density function of random vector $\mathbf{x}$ of order $p \times 1$.

Then the marginal density of $x_i$, denoted by $f_{x_i}(u_i)$ is defined as

$$f_{x_i}(u_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(u_1, u_2, \ldots u_p) du_1 \ldots du_{i-1} du_{i+1} \ldots du_p$$

Let $\mathbf{x}_1 = (x_1, \ldots, x_r)^t$ and $\mathbf{x}_2 = (x_{r+1}, \ldots, x_p)^t$. Then the joint marginal density of $\mathbf{x}_1$ is defined as $f_{x_1}(\mathbf{u}_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots u_p) du_{r+1} \ldots du_p$ where $\mathbf{u}_1 = (u_1, \ldots, u_r)^t$.

Likewise we can define $f_{\mathbf{x}_2}(\mathbf{u}_2)$, where $\mathbf{u}_2 = (u_{r+1}, \ldots, u_p)^t$.

The conditional density of $\mathbf{x}_1$ given $\mathbf{x}_2 = (u_{r+1}, \ldots, u_p)^t$ is defined as

$$f_{x_1|x_2=u_2}(\mathbf{u}_1 | \mathbf{u}_2) = \frac{f_x(u_1, \ldots u_p)}{f_{x_2}(\mathbf{u}_2)}.$$

Why is the conditional density defined thus? To see this let us multiply the both sides of the above equality by $du_1, \ldots, du_r$.

Then $f_{x_1|x_2=u_2}(\mathbf{u}_1 | \mathbf{u}_2) du_1 \ldots du_r = \frac{f_x(u_1, \ldots u_p) du_1 \ldots du_p}{f_{u_2}(u_{r+1}, \ldots, u_p) du_{r+1} \ldots du_p}$.

$$\approx \frac{P\left(u_1 \le x_1 \le u_1 + du_1, \ldots, u_p \le x_p \le u_p + du_p\right)}{P\left(u_{r+1} \le x_{r+1} \le u_{r+1} + du_{r+1}, \ldots, u_p \le x_p \le u_p + du_p\right)}$$

$$= P\left\{u_1 \le x_1 \le u_1 + du_1, \ldots, u_r \le x_r \le u_r + du_r \mid u_{r+1} \le x_{r+1} \le u_{r+1} + du_{r+1}, \ldots, u_p \le x_p \le u_p + du\right.$$

Thus, $f_{x_1|x_2=u_2}(\mathbf{u}_1 | \mathbf{u}_2), du_1, \ldots, du_p$ represents the conditional probability that $x_i$ lies between $u_i$ and $u_1 + du_i, i = 1, \ldots, r$ given that $x_j$ lies in small neighbourhood of $u_j$ that is between $u_j$ and $u_j + du_j, j = r+1, \ldots, p$.

Let us now consider a few examples.

**Example 6:** Let $\mathbf{x} = (x_1, x_2)^t$ has the joint density

$$f_x(u_1, u_2) = \begin{cases} c(2-u_1-u_2) & 0<u_1<1,\ 0<u_2<1 \\ 0 & \text{otherwise} \end{cases}$$

where $c$ is a constant.

(a) Obtain the value of $c$

(b) Find the marginal density of $x_2$

(c) Find the conditional density of $\mathbf{x}_1$ given $x_2 = u_2$, where $0 < u_2 < 1$.

(d) Find the probability that $\mathbf{x}_1 > 1/4$ given $\mathbf{x}_2 = u_2, 0 < u_2 < 1$.

**Solution:** **(a)** Since $f_x(u_1, u_2)$ is a density function, $\int\limits_{-\infty}^{\infty}\dots\int\limits_{-\infty}^{\infty} f_x(u_1,u_2)du_1,du_2 = 1$

Now $1 = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_x(u_1,u_2)du_1du_2 = \int\limits_0^1\int\limits_0^1 c.(2-u_1-u_2)du_1du_2$

$$= c.\int\limits_0^1 (2-u_1)du_1 - \int\limits_0^1 u_2du_2 \quad \text{since} \int\limits_0^1 du_1 = \int\limits_0^1 du_2 = 1$$

$$= c.\left(2u_1 - \frac{u_1^2}{2}\right)\Bigg]_0^1 - \left(\frac{u_2^2}{2}\right)\Bigg]_0^1$$

$$= \left[2 - \frac{1}{2} - \frac{1}{2}\right] = c.1 = c \text{ or } c = 1.$$

**(b)** The marginal density of $x_1$ in the range (0, 1) is

$$f_{x_1}(u_1) = \int\limits_0^1 f_x(u_1,u_2)du_2$$

$$= \int\limits_0^1 (2-u_1-u_2)du_2$$

$$= (2-u_1) - \int\limits_0^1 u_2du_2$$

$$= 2 - u_1 - \frac{1}{2} = \frac{3}{2} - u_1.$$

Thus, $f_{x_1}(u_1) = \begin{cases} \frac{3}{2} - u_1, & \text{for } 0<u_1<1 \\ 0, & \text{elsewhere} \end{cases}$

**(c)** By symmetry, the marginal density of $x_2$ is $f_{x_2}(u_2) = \begin{cases} \frac{3}{2} - u_2, & \text{for } 0<u_2<1 \\ 0, & \text{elsewhere} \end{cases}$

So, the conditional density of $x_1$ given $x_2 = u_2, 0 < u_2 < 1$ is

$f_{x_1|x_2=u_2}(u_1 | u_2) = \dfrac{f_x(u_1,u_2)}{f_{x_2}(u_2)} = \dfrac{2-u_1-u_2}{\frac{3}{2}-u_2}$ where $0 < u_1 < 1$. Since $f_x(u_1, u_2) = 0$,

whenever $u_1 \notin (0, 1)$ whatever be given $u_2$.

we have $f_{x_1|x_2=u_2}(u_1 \mid u_2) = \begin{cases} \dfrac{2-u_1-u_2}{\frac{3}{2}-u_2}, & \text{whenever } 0 < u_1 < 1 \\ 0 & , \quad \text{otherwise} \end{cases}$

for any given $u_2 \in (0, 1)$.

**(d)** $$P\left\langle x_1 > \frac{1}{4} \middle| x_2 = u_2 \right\rangle = \int_{\frac{1}{4}}^{1} f_{x_1/x_2=u_2}(u_1 \mid u_2)\,du_1$$

$$= \int_{\frac{1}{4}}^{1} \frac{2-u_1-u_2}{\frac{3}{2}-u_2}\,du_1$$

$$= \frac{1}{\frac{3}{2}-u_2} \int_{\frac{1}{4}}^{1} (2-u_1-u_2)\,du_1$$

$$= \frac{1}{\frac{3}{2}-u_2}\left(2u_1 - \frac{u_1^2}{2} - u_1u_2\right)\Bigg]_{\frac{1}{4}}^{1}$$

$$= \frac{1}{\frac{3}{2}-u_2}\left(2-\frac{1}{2}-u_2-\left(\frac{2}{4}-\frac{1}{32}-\frac{u_2}{4}\right)\right)$$

$$= \frac{1}{\frac{3}{2}-u_2}\left(\frac{31}{32}-\frac{3u_2}{4}\right) = \frac{31-24u_2}{16(3-2u_2)}$$

**Example 7:** Consider a random vector $\mathbf{x}$ with joint density

$$f_x(u_1, u_2) = \begin{cases} 2e^{-u_1}e^{-2u_2}, & 0 < u_1 < \infty, \quad 0 < u_2 < \infty \\ 0 & , \quad \text{elsewhere} \end{cases}$$

(a) Obtain the marginal density of $x_2$

(b) Obtain the conditional density of $\mathbf{x}_1$ given $\mathbf{x}_2 = u_2, u_2 > 0$.

**Solution: (a)** Clearly $f_{x_2}(u_2) = 0$ whenever $-\infty < x_2 \le 0$. Let $0 < x_2 < \infty$, then

$$f_{x_2}(u_2) = \int_0^\infty 2e^{-u_1}e^{-2u_2}\,du_1 = 2e^{-2u_2}\int_0^\infty e^{-u_1}\,du_1$$

$$= 2e^{-2u_2}\, e^{-u_1}\Big]_0^\infty$$

$$= 2e^{-2u_2}, \text{ whenever } 0 < u_2 \le \infty.$$

which is an exponential distribution with parameter 2.

**(b)** Given $u_2 > 0$ $f_{x_1|x_2=u_2}(u_1 \mid u_2) = \dfrac{f_x(u_1, u_2)}{f_{x_2}(u_2)} = \dfrac{2e^{-u_1}e^{-u_2}}{2e^{-u_2}} = e^{-u_1}$, whenever $0 < u_1 \le \infty$.

Again $f_{x_1|x_2=u_2}(u_1 \mid u_2) = 0, u_1 \notin (0, \infty)$ for any given $u_2$.

Thus, $f_{x_1|x_2=u_2}(u_1 \mid u_2) = \begin{cases} e^{-u_1} & \text{if } u_1 \in (0,\infty), u_2 \in (0,\infty) \\ 0 & \text{otherwise} \end{cases}$

It can be easily checked that this is the same as the marginal distribution of $x_1$. Thus, in this example the joint density of $x_1$ and $x_2$ is the product of the marginal densities of $x_1$ and $x_2$.

Let a random vector $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$, where $\mathbf{x}_1$ is of order $r \times 1$ and $\mathbf{x}_2$ is of order $(p-r) \times 1$ have joint density $f_x(\mathbf{u})$, where $\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$ is partitioned according to the partition of $\mathbf{x}$. We say that $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent if
$P\{\mathbf{x}_1 \in A, \mathbf{x}_2 \in B\} = P(\mathbf{x}_1 \in A\}P\{\mathbf{x}_2 \in B\}$ all subsets A and B of $R^r$ and $R^{(p-r)}$ respectively.

It can be shown that $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent if and only if the joint density of $\mathbf{x}_1$ and $\mathbf{x}_2$ (i.e., the density of $\mathbf{x}$) is equal to the product of the marginal density of $\mathbf{x}_1$ and $\mathbf{x}_2$ or in other words

$f_x(\mathbf{u}) = f_{x_1}(\mathbf{u}_1).f_{x_2}(\mathbf{u}_2)$, for all $\mathbf{u}_1$ and $\mathbf{u}_2$.

We give below a relationship between uncorrelatedness and independence.

**Theorem 4:** Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be independent vectors. Then the matrix $\mathrm{Cov}(\mathbf{x}_1, \mathbf{x}_2) = 0$.

**Proof:** Let $f_{x_1}(\mathbf{u}_1)$ and $f_{x_2}(\mathbf{u}_2)$ be the densities of $\mathbf{x}_1$ and $\mathbf{x}_2$. Then the joint density of $\mathbf{x}_1$ and $\mathbf{x}_2$ (i.e., the density of $x = \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$) is $f_x(\mathbf{u}) = f_{x_1}(\mathbf{u}_1).f_{x_2}(\mathbf{u}_2)$ where $\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$.

Let $\mathbf{u}_1 = (u_1, \ldots, u_r)^t$ and $\mathbf{u}_2 = (u_{r+1}, \ldots, u_p)^t$.

$$\mathrm{Cov}(\mathbf{u}_1, \mathbf{u}_2) = E(\mathbf{u}_1 \mathbf{u}_2^t) - E(\mathbf{u}_1) E(\mathbf{u}_2^t)$$

$$= \int \ldots \int \mathbf{u}_1 \mathbf{u}_2^t f_{x_1}(\mathbf{u}_1) f_{x_2}(\mathbf{u}_2) du_1 \ldots du_p$$
$$- \left[ \int \ldots \int \mathbf{u}_1 f_{x_1}(\mathbf{u}_1) du_1 \ldots du_r \right] \left[ \int \ldots \int \mathbf{u}_2^t f_{x_2}(\mathbf{u}_2) du_{r+1} \ldots du_p \right] = 0$$

since the first integral in the previous expression splits into the product of the two later integrals. Also for a matrix $A = ((a_{ij}))$, we define

$$\int \ldots \int A \, dx_1, \ldots, dx_p = \left( \left( \int \ldots \int a_{ij} dx_1, \ldots, dx_p \right) \right).$$

However, the converse is not true as shown through the following exercise E11.

Now, try these exercises.

---

E11) Let x have the following probability distribution

| Value | –3 | –1 | 1 | 3 |
|---|---|---|---|---|
| Probability | ¼ | ¼ | ¼ | ¼ |

(a) Show that the probability distribution $x^2$ is

| Value | 1 | 9 |
|---|---|---|
| Probability | ½ | ½ |

(b) Write down the joint distribution of $x$ and $x^2$
(c) Show that $x$ and $x^2$ are uncorrelated.
(d) Show that $x$ and $x^2$ are not independent.

E12) Consider a random vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ with the density

$$f_x(u_1, u_2) = \begin{cases} cu_1(2-u_1-u_2) & 0<u_1<1,\ 0<u_2<1 \\ 0 & \text{otherwise} \end{cases}$$

where $c$ is a constant.

(a) Obtain the value of $c$.
(b) Find the marginal densities of $x_1$ and $x_2$.
(c) Find the conditional density of $x_2$ given $x_1 = u_1$.
(d) Are $x_1$ and $x_2$ independent?

---

Let us consider the bivariate normal distribution. This is a special case of the multivariate normal distribution which we shall study in detail in the next few sections.

**Example 8 (Bivariate normal distribution):** Let $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ have joint density

$$f_x(u_1,u_2) = \frac{e^{\left\{-\frac{1}{2(1-\rho^2_{x_1x_2})}\left[\left(\frac{u_1-\mu_{x_1}}{\sigma_{x_1}}\right)^2+\left(\frac{u_2-\mu_{x_2}}{\sigma_{x_2}}\right)^2-2\rho_{x_1x_2}\frac{(u_1-\mu_x)(u_2-\mu_y)}{\sigma_{x_1}\sigma_{x_2}}\right]\right\}}}{2\pi\sigma_{x_1}\sigma_{x_2}\sqrt{(1-\rho^2_{x_1x_2})}}, -\infty < u_1 < \infty, -\infty < u_2 < \infty.$$

This distribution will be denoted by $N_2\left(\mu_{x_1}, \mu_{x_2}, \sigma^2_{x_1}, \sigma^2_{x_2}, \rho_{x_1x_2}\right)$.

(a) Rewrite the above density in terms of variance-covariance matrix $\Sigma$ of $\mathbf{x}$.

(b) Obtain a lower triangular square root $\mathbf{B}$ of $\Sigma$.

(c) Write down the density of $\mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$.

(d) Hence write down the marginal density of $y_1$ and $x_1$.

(e) Show that $y_1$ and $y_2$ are independent.

(f) Obtain the conditional distribution of $x_1$ given $x_2 = u_2$.

**Solution: (a)** Let $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ be the variance covariance matrix of $\mathbf{x}$. Then $\sigma_{11} = \sigma^2_{x_1}$, $\sigma_{22} = \sigma^2_{x_2}$ and $\sigma_{12} = \rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2}$.

Thus, $|\Sigma| = \sigma_{11}\sigma_{22} - \sigma^2_{12} = \sigma^2_{x_1}\sigma^2_{x_2} - \rho_{x_1x_2}\sigma^2_{x_1}\sigma^2_{x_2} = \sigma^2_{x_1}\sigma^2_{x_2}\left(1-\rho^2_{x_1x_2}\right)$

Also $\left(\frac{1}{1-\rho^2_{x_1x_2}}\right)\left[\left(\frac{u_1-\mu_{x_1}}{\sigma_{x_1}}\right)^2+\left(\frac{u_2-\mu_{x_2}}{\sigma_{x_2}}\right)^2-2\rho_{x_1x_2}\frac{(u_1-\mu_{x_1})(u_2-\mu_{x_2})}{\sigma_{x_1}\sigma_{x_2}}\right]$

$$=\frac{1}{1-\rho^2_{x_1x_2}}\left[(u_1-\mu_{x_1})\ (u_2-\mu_{x_2})\right]\begin{bmatrix}\frac{1}{\sigma^2_{x_1}} & -\frac{\rho_{x_1x_2}}{\sigma_{x_1}\sigma_{x_2}}\\ -\frac{\rho_{x_1x_2}}{\sigma_{x_1}\sigma_{x_2}} & \frac{1}{\sigma^2_{x_2}}\end{bmatrix}\begin{pmatrix}u_1-\mu_{x_1}\\ u_2-\mu_{x_2}\end{pmatrix}$$

$$=\left[(u_1-\mu_{x_1}):(u_2-\mu_{x_2})\right]\left[\frac{1}{(1-\rho^2_{x_1x_2})(\sigma^2_{x_1}\sigma^2_{x_2})}\begin{pmatrix}\sigma^2_{x_2} & -\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2}\\ -\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} & \sigma^2_{x_1}\end{pmatrix}\right]\begin{pmatrix}u_1-\mu_{x_1}\\ u_2-\mu_{x_2}\end{pmatrix}$$

$$=\begin{pmatrix}u_1-\mu_{x_1} & u_2-\mu_{x_2}\end{pmatrix}\left[\frac{1}{|\Sigma|}\begin{pmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{pmatrix}\right]\begin{pmatrix}u_1-\mu_{x_1}\\ u_2-\mu_{x_2}\end{pmatrix}$$

$$=(\mathbf{u}-\boldsymbol{\mu}_x)^t\,\Sigma^{-1}(\mathbf{u}-\boldsymbol{\mu}_x),$$

where $\mathbf{u}=\begin{pmatrix}u_1\\ u_2\end{pmatrix}$ and $\boldsymbol{\mu}_x=\begin{pmatrix}\mu_{x_1}\\ \mu_{x_2}\end{pmatrix}$.

Thus, the density of $\mathbf{x}$ can be rewritten in terms of $\Sigma$ as

$$f_x(\mathbf{u})=\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}e^{\frac{1}{2}(\mathbf{u}-\boldsymbol{\mu}_x)^t\Sigma^{-1}(\mathbf{u}-\boldsymbol{\mu}_x)},u\in R^2.$$

Hence we can denote this distribution as $N_2(\boldsymbol{\mu}_x,\Sigma)$.

**(b)** Let $\Sigma=\mathbf{BB}'$, where $\mathbf{B}$ is lower triangular.

Writing $\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}b_{11} & 0\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}b_{11} & b_{21}\\ 0 & b_{22}\end{pmatrix}$

We have $b_{11}=\sqrt{\sigma_{11}}$

$$b_{21}=\frac{\sigma_{12}}{\sqrt{\sigma_{11}}}$$

and $b_{22}=\sqrt{\sigma_{22}-\frac{\sigma^2_{12}}{\sigma_{11}}}$

**(c)** Consider $\mathbf{y}=\mathbf{B}^{-1}\mathbf{x}$. Write $\mathbf{v}=\mathbf{B}^{-1}\mathbf{u}$ and $\boldsymbol{\mu}_y=\mathbf{B}^{-1}\boldsymbol{\mu}_x$.

Then the density of $y$ is given by (density of $\mathbf{x}$ written in terms of $y$).

$$f_{\mathbf{y}}(\mathbf{v})=f_{\mathbf{x}}(\mathbf{Bv})\,|J(u_1,u_2)|^{-1}$$

Now $\mathbf{B}^{-1}=\begin{pmatrix}\frac{1}{b_{11}} & 0\\ -\frac{b_{21}}{b_{11}b_{22}} & \frac{1}{b_{22}}\end{pmatrix}$

So $y_1=\frac{1}{b_{11}}x_1;\quad v_1=\frac{1}{b_{11}}u_1$

and $\quad y_2 = \frac{1}{b_{22}}\left(x_2 - \frac{b_{21}}{b_{11}}x_1\right); \quad v_2 = \frac{1}{b_{22}}\left(u_2 - \frac{b_{21}}{b_{11}}u_1\right)$

Then $\quad J(u_1,u_2) = \begin{vmatrix} \frac{\partial v_1}{\partial u_1} & \frac{\partial v_1}{\partial u_2} \\ \frac{\partial v_2}{\partial u_1} & \frac{\partial v_2}{\partial u_2} \end{vmatrix}$

$$= \begin{vmatrix} \frac{1}{b_{11}} & 0 \\ \frac{b_{21}}{b_{11}b_{22}} & \frac{1}{b_{22}} \end{vmatrix} = \frac{1}{b_{11}b_{22}} = \left[\sqrt{\sigma_{11}}\sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\right]^{-1} = \left[\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\right]^{-1} = |\Sigma|^{-\frac{1}{2}}$$

Also $\boldsymbol{\Sigma}^{-1} = (\mathbf{BB}^t) = \mathbf{B}^{t^{-1}}\mathbf{B}^{-1} = \mathbf{B}^{-1^t}\mathbf{B}^{-1}$

Hence $\quad (\mathbf{u} - \boldsymbol{\mu}_x)^t\boldsymbol{\Sigma}^{-1}(\mathbf{u} - \boldsymbol{\mu}_x)$

$$= (\mathbf{u} - \boldsymbol{\mu}_x)^t\mathbf{B}^{-1^t}\mathbf{B}^{-1}(\mathbf{u} - \boldsymbol{\mu}_x)$$

$$= (\mathbf{v} - \boldsymbol{\mu}_y)^t(\mathbf{v} - \boldsymbol{\mu}_y)$$

Thus, the density of $\mathbf{y}$ is $f_y(\mathbf{v}) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{v}-\boldsymbol{\mu}_y)^t(\mathbf{v}-\boldsymbol{\mu}_y)} . |\Sigma|^{\frac{1}{2}} = \frac{1}{2\pi} e^{-\frac{1}{2}(\mathbf{v}-\boldsymbol{\mu}_y)^t(\mathbf{v}-\boldsymbol{\mu}_y)}$.

The range of values for $y_1$ and $y_2$ are clearly the same as the range of values of $x_1$ and $x_2$, namely, $-\infty < y_1 < \infty, -\infty < y_2 < \infty$.

Hence $\mathbf{y} \sim N_2(\boldsymbol{\mu}_y, \mathbf{I})$

(**d**) The joint density of $y_1$ and $y_2$ is $\frac{1}{2\pi} e^{-\frac{1}{2}\left\{(v_1-\mu_{y_1})^2 + (v_2-\mu_{y_2})^2\right\}}, -\infty < v_1, v_2 < \infty$.

Hence the marginal density of $y_1$ is $\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(v_1-\mu_{y_1})^2} . e^{-\frac{1}{2}(v_2-\mu_{y_2})^2} dv_2$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v_1-\mu_{y_1})^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v_2-\mu_{y_2})^2} dv_2$$

$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v_1-\mu_{y_1})^2}$, since the integrand above as the density of a normal distribution.

Thus, the marginal density of $y_1 = \frac{1}{b_{11}}x_1$ is $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v_1-\mu_{y_1})^2}, -\infty < v_1 < \infty$ which is the density of $N(\mu_{y_1}, 1)$.

So, $x_1$ has a normal distribution with mean $= b_{11}\mu_{y_1} = b_{11}\frac{1}{b_{11}}\mu_{x_1} = \mu_{x_1}$ and variance $= b_{11}^2 V(y_1) = b_{11}^2 = \sigma_{11}$.

Thus, the marginal distribution of $x_1$ is $N(\mu_{x_1}, \sigma_{11})$.

**(e)** The joint density of $y_1$ and $y_2$ is (i.e. the density of $\mathbf{y}$) is

$$\frac{1}{2\pi}e^{-\frac{1}{2}\left(v_1-\mu_{y_1}\right)^2}.e^{-\frac{1}{2}\left(v_2-\mu_{y_2}\right)^2}=h_1(v_1).h_2(v_2) \text{ for all } v_1,\ v_2 \text{ where}$$

$$h_1(v_1)=\frac{1}{2\pi}e^{-\frac{1}{2}\left(v_1-\mu_{y_1}\right)^2},\ -\infty<v_1<\infty \text{ and } h_2(v_2)=\frac{1}{2\pi}e^{-\frac{1}{2}\left(v_2-\mu_{y_2}\right)^2},\ -\infty<v_2<\infty.$$

Notice that $h_1$ depends only on $v_1$ and $h_1$ depends only on $v_2$. Hence $y_1$ and $y_2$ are independent.

**(f)** We can write

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}1 & 0\\ \frac{\sigma_{12}}{\sigma_{11}} & 1\end{pmatrix}\begin{pmatrix}\sigma_{11} & 0\\ 0 & \sigma_{22}-\sigma_{21}\sigma_{11}^{-1}\sigma_{12}\end{pmatrix}\begin{pmatrix}1 & \frac{\sigma_{12}}{\sigma_{11}}\\ 0 & 1\end{pmatrix}$$

$$\text{Hence } \Sigma^{-1}=\begin{pmatrix}1 & -\frac{\sigma_{12}}{\sigma_{11}}\\ 0 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_{11}} & 0\\ 0 & \frac{1}{\sigma_{22}-\sigma_{12}\sigma_{11}^{-1}\sigma_{12}}\end{pmatrix}\begin{pmatrix}1 & 0\\ -\frac{\sigma_{12}}{\sigma_{11}} & 1\end{pmatrix}.$$

Hence the density of $\mathbf{x}=\begin{pmatrix}x_1\\ x_2\end{pmatrix}$ can be rewritten as

$$\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}.e^{-\frac{1}{2}\left(u_1-\mu_{x_1};u_2-\mu_{x_2}\right)\begin{pmatrix}1 & -\frac{\sigma_{12}}{\sigma_1}\\ 0 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_{11}} & 0\\ 0 & \frac{1}{\sigma_{22}-\sigma_{12}\sigma_{11}^{-1}\sigma_{12}}\end{pmatrix}\begin{pmatrix}1 & 0\\ -\frac{\sigma_{12}}{\sigma_{11}} & 1\end{pmatrix}\begin{pmatrix}u_1-\mu_{x_1}\\ u_2-\mu_{x_2}\end{pmatrix}}$$

$$\text{Writing } v=\begin{pmatrix}1 & 0\\ -\frac{\sigma_{12}}{\sigma_{11}} & 1\end{pmatrix}\begin{pmatrix}u_1-\mu_{x_1}\\ u_2-\mu_{x_2}\end{pmatrix}=\begin{pmatrix}u_1-\mu_{x_1}\\ \left(u_2-\mu_{x_2}\right)-\frac{\sigma_{12}}{\sigma_{11}}\left(u_1-\mu_{x_1}\right)\end{pmatrix}$$

We have the density of $\mathbf{x}$ as

$$\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}.e^{-\frac{1}{2}\frac{\left(u_1-\mu_{x_1}\right)^2}{\sigma_{11}}}e^{-\frac{1}{2}\frac{\left(u_2-\mu_{x_2}-\frac{\sigma_{12}}{\sigma_{11}}\left(u_1-\mu_{x_1}\right)\right)^2}{\sigma_{22}-\sigma_{12}\sigma_{11}^{-1}\sigma_{12}}},\ -\infty<v_1,\ v_2<\infty.$$

Also notice that $|\Sigma|=\sigma_{11}\sigma_{22}-\sigma_{12}^2=\sigma_{11}\left(\sigma_{22}-\sigma_{12}\sigma_{11}^{-1}\sigma_{12}\right)$

Hence the conditional density of $x_2$ given $x_1=u_1$ is

$$f_{x_2|x_1}(u_2\mid u_1)=\frac{\frac{1}{2\pi\left[\sigma_{11}\left(\sigma_{22}-\sigma_{12}\sigma_{11}^{-1}\sigma_{12}\right)\right]^{\frac{1}{2}}}.e^{-\frac{1}{2}\frac{\left(u_1-\mu_{x_1}\right)^2}{\sigma_{11}}}.e^{-\frac{1}{2}\frac{\left(\left(u_2-\mu_{x_2}\right)-\frac{\sigma_{12}}{\sigma_{11}}\left(u_1-\mu_{x_1}\right)\right)^2}{\sigma_{22}-\sigma_{12}\sigma_{11}^{-1}\sigma_{12}}}}{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}}.e^{-\frac{1}{2}\frac{\left(u_1-\mu_{x_1}\right)^2}{\sigma_{11}}}}$$

$$= \frac{1}{\sqrt{2\pi}\left(\sigma_{22} - \sigma_{12}\sigma_{11}^{-1}\sigma_{12}\right)^{\frac{1}{2}}} . e^{-\frac{1}{2}\frac{\left(\left(u_2 - \mu_{x_2}\right) - \frac{\sigma_{12}}{\sigma_{11}}\left(u_1 - \mu_{x_1}\right)\right)^2}{\sigma_{22} - \sigma_{12}\sigma_{11}^{-1}\sigma_{12}}}, \quad -\infty < u_2 < \infty,$$

which is the density function of $N\left(\mu_{x_2} + \frac{\sigma_{12}}{\sigma_{11}}\left(u_1 - \mu_{x_1}\right), \sigma 11\left(\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)\right)$

or $N\left(\mu_{x_2} + \rho_{x_1x_2}\frac{\sigma_{x_2}}{\sigma_{x_1}}\left(u - \mu_{x_1}\right), \sigma_{x_2}^2\left(1 - \rho_{x_1x_2}^2\right)\right)$.

Try the following exercise.

---

E13) Let **x** have a bivariate normal distribution $N_2\left(\mu_{x_1}, \mu_{x_2}, \sigma_{x_1}, \sigma_{x_2}, \rho_{x_1x_2}\right)$. Show that $x_1$ and $x_2$ are independent if and only if $\rho_{x_1x_2} = 0$. (Recall that, in general, uncorrelatedness does not imply in dependence. However if $x_1$ and $x_2$ have a bivariate normal distribution then $x_1$ and $x_2$ are independent if and only if they are uncorrelated.)

---

Now, we shall summarize the unit.

## 15.5 SUMMARY

In this unit, we have covered the following points:

- Nature of multivariate problems
- Computation of the mean vector and variance-covariance matrix of a linear transformation of a random vector
- An algorithm to compute a lower triangular square root of a positive definite matrix and its inverse simultaneously
- Discrete and continuous multivariate distributions
- Uncorrelatedness and independence
- Bivariate normal distribution.

## 15.6 SOLUTIONS TO EXERCISES

E1) $\mu_1 = E(X_1) = (-1)\ (0.3) + (0)\ (0.3) + (1)\ (0.4) = 0.1$

$\mu_2 = E(X_2) = (0)\ (0.7) + (1)\ (0.3) = 0.3$

$\sigma_{11} = E(X_1 - \mu_1)^2$

$= (-1-0.1)^2\ (0.3) + (0-0.1)^2(0.3) + (1-0.1)^2(0.4) = 0.69$

$\sigma_{22} = E(X_2 - \mu_2)^2$

$= (0-0.3)^2(0.7) + (1-0.3)^2(0.3) = 0.78$

$\sigma_{12} = E(X_1 - \mu_1)\ (X_2 - \mu_2)$

$= (-1-0.1)\ (0-0.3)\ (0.21) + (-1-0.1)\ (1-0.3)\ (0.09)$

$+ (0-0.1)\ (0-0.3)\ (0.21) + (0-0.1)\ (1-0.3)\ (0.09)$

$+ (1-0.1)\ (0-0.3)\ (0.28) + (1-0.1)\ (1-0.3)\ (0.12) = 0$

$\sigma_{21} = 0$

Therefore, mean $= \mu = \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix}$

and $\sum = \begin{bmatrix} 0.69 & 0 \\ 0 & 0.78 \end{bmatrix}$.

E2) The mean of $\mathbf{1^t x} = \mathbf{l}\mu = (1 \quad 1 \quad 1 \quad 1)\begin{pmatrix} 2 \\ 1 \\ -1 \\ -2 \end{pmatrix} = 0 = 0$

We can write $\mathbf{\Sigma} = 0.8\,\mathbf{I} + 0.2 ll^t$ [where I is identity matrix of order $4 \times 4$]

where $l^t = (1 \quad 1 \quad 1 \quad 1)$ $(= \mathbf{1}^t$ of the present exercise)

Now $\quad \mathbf{V(1^t x) = 1^t \Sigma 1 = 0.81^t 1 + 0.21^t} ll^t \mathbf{1)}$

$$= 0.8 \times 4 + 0.2 \times 4 \times 4 = 3.2 + 3.2 = 6.4$$

$$\mathrm{Cov}(\mathbf{1^t x, m^t x}) = \mathbf{1^t \Sigma m} = 0.8\mathbf{1^t m} + 0.2\mathbf{1^t} ll^t \mathbf{m} = 0$$

$$= 0.8 \times 0 + 0.2 \times 4 \times 0 - 0$$

E3) (a) Let $z = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$

The $\quad D(z) = \begin{pmatrix} D(\mathbf{x}) & \mathrm{Cov}(\mathbf{x}, \mathbf{y}) \\ \mathrm{Cov}(\mathbf{y}, \mathbf{x}) & D(\mathbf{y}) \end{pmatrix}$

$$= \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Delta} \\ \mathbf{\Delta}^t & \mathbf{\Gamma} \end{pmatrix}$$

Let $\mathbf{u} = \mathbf{Bx}$ and $\mathbf{v} = \mathbf{Cy}$

Write $\mathbf{w} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$

Then $\mathbf{w} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{Bx} \\ \mathbf{Cy} \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} z$

Thus $\quad D(\mathbf{w}) = D\left(\begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}\mathbf{z}\right) = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} D(z) \begin{pmatrix} \mathbf{B}^t & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^t \end{pmatrix}$

$$= \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}\begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Delta} \\ \mathbf{\Delta}^t & \mathbf{\Gamma} \end{pmatrix}\begin{pmatrix} \mathbf{B}^t & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^t \end{pmatrix} = \begin{pmatrix} \mathbf{B\Sigma B}^t & \mathbf{B\Delta C}^t \\ \mathbf{C\Delta^t B}^t & \mathbf{C\Gamma C}^t \end{pmatrix}$$

whence it follows that $D(\mathbf{Bx}) = D(\mathbf{u}) = \mathbf{B\Sigma B}^t$ and

$\mathrm{Cov}(\mathbf{Bx}, \mathbf{Cy}) = \mathrm{Cov}(\mathbf{u}, v) = \mathbf{B\Delta C}^t$.

(b) The standard deviations of $x_1, x_2$ and $x_3$ are 2.0, 3.0 and 4.0, respectively. Hence the correlation matrix $\mathbf{R}$ of the random vector $\mathbf{x}$ is

$$\mathbf{R} = \mathbf{B}^{-1}\mathbf{\Sigma}\mathbf{B}^{-1'} \text{ where } \mathbf{B} = \begin{pmatrix} 2.0 & 0 & 0 \\ 0 & 3.0 & 0 \\ 0 & 0 & 4.0 \end{pmatrix}.$$

You can easily check that $\mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{2.0} & 0 & 0 \\ 0 & \frac{1}{3.0} & 0 \\ 0 & 0 & \frac{1}{4.0} \end{bmatrix}$

Therefore,

$$\mathbf{R} = \begin{pmatrix} \frac{1}{2.0} & 0 & 0 \\ 0 & \frac{1}{3.0} & 0 \\ 0 & 0 & \frac{1}{4.0} \end{pmatrix} \begin{pmatrix} 4.0 & 3.0 & 2.4 \\ 3.0 & 9.0 & -4.8 \\ 2.4 & -4.8 & 16.0 \end{pmatrix} \begin{pmatrix} \frac{1}{2.0} & 0 & 0 \\ 0 & \frac{1}{3.0} & 0 \\ 0 & 0 & \frac{1}{4.0} \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & -0.4 \\ 0.3 & -0.4 & 1 \end{pmatrix}.$$

which is exactly the same as in Example 1 (as expected).

E4) From the given information,

$$E(\mathbf{y}) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, E(\mathbf{z}) = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

$$D(\mathbf{y}) = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 6 \end{pmatrix}, D(\mathbf{z}) = \begin{pmatrix} 8 & 3 \\ 3 & 9 \end{pmatrix} \text{ and } \mathrm{Cov}(\mathbf{y}, \mathbf{z}) \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 1 & 2 \end{pmatrix}$$

(a) $E(\mathbf{u}) = E(\mathbf{B}\,\mathbf{y}) = \mathbf{B}E(\mathbf{y}) = \begin{pmatrix} 2 & -1 & 1 \\ 4 & 5 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 14 \end{pmatrix}$

$$D(\mathbf{u}) = \mathbf{B}D(\mathbf{y})\mathbf{B}^t = \begin{pmatrix} 2 & -1 & 1 \\ 4 & 5 & 2 \end{pmatrix}\begin{pmatrix} 5 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 6 \end{pmatrix}\begin{pmatrix} 2 & 4 \\ -1 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 30 & 46 \\ 46 & 260 \end{pmatrix}$$

(b) $E(\boldsymbol{\nu}) = E(\mathbf{C}\,\mathbf{z}) = \mathbf{C}\,E(\mathbf{z}) = E(\mathbf{v}) = E(\mathbf{C}\,\mathbf{z}) = \mathbf{C}E(\mathbf{z}) = \begin{pmatrix} 4 & 2 \\ -1 & 3 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} -8 \\ 9 \\ 2 \end{pmatrix}$

$$D(\mathbf{v}) = D(\mathbf{C}\,\mathbf{z}) = \mathbf{C}D(\mathbf{z})\mathbf{C}^t = \begin{pmatrix} 4 & 2 \\ -1 & 3 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 8 & 3 \\ 3 & 9 \end{pmatrix}\begin{pmatrix} 4 & -1 & 0 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 212 & 52 & 30 \\ 52 & 71 & 24 \\ 30 & 24 & 9 \end{pmatrix}$$

(c) $\mathrm{Cov}(\mathbf{u}, \mathbf{v}) = \mathrm{Cov}(\mathbf{By}, \mathbf{Cz}) = \mathbf{B}\mathrm{Cov}(\mathbf{y}, \mathbf{z})\,\mathbf{C}^t$

$$= \begin{pmatrix} 2 & -1 & 1 \\ 4 & 5 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 4 & -1 & 0 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 12 & 11 & 4 \\ 80 & 8 & 8 \end{pmatrix}$$

**Note:** Since $D(\mathbf{v}) = D(C\,\mathbf{z}) = CD(\mathbf{z})C^t$, the rank of $D(\mathbf{v})$ is at most equal to the rank of $D(\mathbf{z})$ which is equal to 2 (since the determinant of $D(\mathbf{z}) = 63 \neq 0$). So $D(\mathbf{v})$ is a positive semidefinite matrix. It can be shown that $\mathbf{g} = \begin{pmatrix} 1 \\ 4 \\ -14 \end{pmatrix}$ is orthogonal to both $\begin{pmatrix} 4 \\ -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$ [use Gram-Schmidt orthogonalization process on $\begin{pmatrix} 4 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and hence $(1 \quad 4 \quad -14)C = 0$.

Thus, $0 = \mathbf{g}^t C D(z)\, C^t \mathbf{g} = V(\mathbf{g}^t\,\mathbf{v})$. Hence $\mathbf{g}^t\mathbf{v}$ is a constant with probability 1. Let us find out the constant now.

$E(\mathbf{g}^t\mathbf{v}) = \mathbf{g}^t E(\mathbf{v}) = (1 \quad 4 \quad -14)\begin{pmatrix} -8 \\ 9 \\ 2 \end{pmatrix} = 0$. Since $\mathbf{g}^t\mathbf{v}$ is a constant with probability 1, $\mathbf{g}^t\mathbf{v} = E(\mathbf{g}^t\mathbf{v}) = 0$ with probability 1.

E5) This is similar to the note in E3. Since each row sum (same as the column sum) of $D(\mathbf{x})$ is 0. We have $\mathbf{1}^t D(\mathbf{x})\mathbf{1} = 0$ where $\mathbf{1}^t = (1 \quad 1 \quad 1)$.

Hence $\mathbf{1}^t\mathbf{x} = x_1 + x_2 + x_3$ is a constant with probability 1. The constant is $E(\mathbf{1}^t\mathbf{x}) = E(x_1) + E(x_2) + E(x_3) = 1 + 2 - 1 = 2$.

E6) We form $D(\mathbf{x}) : \mathbf{I}$ and follow the procedure in Method 2. Thus,

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | 2 | 6 | 1 | 0 | 0 | (1) |
| 2 | 17 | 27 | 0 | 1 | 0 | (2) |
| 6 | 27 | 70 | 0 | 0 | 1 | (3) |
| 2 | 1 | 3 | ½ | 0 | 0 | (4) = (1) ÷ 2 |
| 0 | 16 | 24 | $-\frac{1}{2}$ | 1 | 0 | (5) = (2) − (4) |
| 0 | 24 | 61 | $-\frac{3}{2}$ | 0 | 1 | (6) = (3) − 3 ×(4) |
| 2 | 1 | 3 | ½ | 0 | 0 | (7) = (4) |
| 0 | 4 | 6 | $-\frac{1}{8}$ | ¼ | 0 | (8) = (5) ÷ 4 |
| 0 | 0 | 25 | $-\frac{3}{4}$ | $-\frac{3}{2}$ | 1 | (9) = (6) - 6 x (8) |
| 2 | 1 | 3 | ½ | 0 | 0 | (10) = (7) |
| 0 | 4 | 6 | $-\frac{1}{8}$ | ¼ | 0 | (11) = (8) |
| 0 | 0 | 5 | $-\frac{3}{20}$ | $-\frac{3}{10}$ | $\frac{1}{5}$ | (12) = (9) ÷ 5 |

Hence $\mathbf{B} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 4 & 0 \\ 3 & 6 & 5 \end{pmatrix}$ is a lower triangular square root of $D(\mathbf{x})$ (i.e.

$D(\mathbf{x}) = \mathbf{BB}^t$ ) and $\mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{8} & \frac{1}{4} & 0 \\ \frac{3}{20} & -\frac{3}{10} & \frac{1}{5} \end{pmatrix}$. The components of $\mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$ are

uncorrelated each with variance 1 as $D(\mathbf{y}) = \mathbf{B}^{-1}D(\mathbf{x})\mathbf{B}^{-1^t} = \mathbf{B}^{-1}\mathbf{BB}^t = \mathbf{I}$.

E7) $V(x_i) = \sigma_{ii} = 0$. So $x_i$ is a constant with probability 1. Hence $Cov(x_i, x_j) = 0$ for all $j$.

$\Sigma$ being a variance-covariance matrix is nnd by Theorem 3(a). Now by Example 11 of Unit 14, $\sigma_{ii} = 0 \Rightarrow \sigma_{ij} = 0$ for all $j$. Hence $Cov(x_i, x_j) = \sigma_{ij} = 0$ for all $j$.

E8) $x_2$ = No. of tennis specialists chosen
$x_3$ = No. of cricketers chosen

Let $p_{ij}$ denotes the probability that $x_2 = i$ and $x_3 = j, i = 1, \ldots, 4$ and $j = 1, \ldots, 6$
Clearly $p_{ij} = 0$ whenever $1 + j \geq 4$ as only 3 sportsmen were chosen.

$p_{00} = 0$ since there are only 2 specialists in long distance running and 3 sportsmen are selected.

$$p_{01} = \left({}^2C_2\right)\left({}^6C_1\right)/\left({}^{12}C_3\right) = \frac{6}{220}$$
$$p_{02} = \left({}^2C_1\right)\left({}^6C_2\right)/\left({}^{12}C_3\right) = \frac{30}{220}$$
$$p_{03} = \left({}^6C_3\right)/\left({}^{12}C_3\right) = \frac{20}{220}$$

$$p_{04} = p_{05} = p_{06} = 0$$

$$p_{10} = \left({}^4C_{10}\right)\left({}^2C_2\right)/\left({}^{12}C_3\right) = \frac{4}{220}$$
$$p_{11} = \left({}^4C_1\right)\left({}^6C_1\right)\left({}^2C_1\right)/\left({}^{12}C_3\right) = \frac{48}{220}$$
$$p_{12} = \left({}^4C_1\right)\left({}^6C_2\right)\left({}^{12}C_3\right)/\left({}^{12}C_3\right) = \frac{60}{220}$$

$$p_{13} = p_{14} = p_{15} = p_{16} = 0$$

$$p_{20} = \left({}^4C_2\right)\left({}^2C_1\right)/\left({}^{12}C_3\right) = \frac{12}{220}$$
$$p_{21} = \left({}^4C_2\right)\left({}^6C_1\right)/\left({}^{12}C_3\right) = \frac{36}{220}$$

$$p_{22} = p_{23} = p_{24} = p_{26} = p_{26} = 0$$

$$p_{30} = \left({}^4C_3\right)/\left({}^{12}C_3\right) = \frac{4}{220} = \frac{4}{220}$$
$$p_{31} = p_{32} = p_{33} = p_{34} = p_{35} = p_{36} = 0$$

Also $p_{ij} = 0$ for $i = 4$ or $j \geq 4$.

Thus, the joint distribution of $x_2$ and $x_3$ is given by

Joint Distribution of $x_2$ and $x_3$

| Value taken by | → $x_3$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Row sum |
|---|---|---|---|---|---|---|---|---|---|
| ↓ $x_2$ | | Joint Probability | | | | | | | |
| 0 | Joint Probability | 0 | $\frac{6}{220}$ | $\frac{30}{220}$ | $\frac{20}{220}$ | 0 | 0 | 0 | $\frac{56}{220}$ |
| 1 | | $\frac{4}{220}$ | $\frac{48}{220}$ | $\frac{60}{220}$ | 0 | 0 | 0 | 0 | $\frac{112}{220}$ |
| 2 | | $\frac{12}{220}$ | $\frac{36}{220}$ | 0 | 0 | 0 | 0 | 0 | $\frac{48}{220}$ |
| 3 | | $\frac{4}{220}$ | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{4}{220}$ |
| 4 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Column sum | | $\frac{20}{220}$ | $\frac{90}{220}$ | $\frac{90}{220}$ | $\frac{20}{220}$ | 0 | 0 | 0 | 1 |

The marginal distribution of $x_2$ is as follows:

Marginal Distribution of $x_2$

| Value | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Probability | $\frac{56}{220}$ | $\frac{112}{220}$ | $\frac{48}{220}$ | $\frac{4}{220}$ | 0 |

The marginal distribution of $x_2$ is as follows:

Marginal Distribution of $x_3$

| Value | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Probability | $\frac{20}{220}$ | $\frac{90}{220}$ | $\frac{90}{220}$ | $\frac{20}{220}$ | 0 | 0 | 0 |

The conditional distribution of $x_3 | x_2 = 1$ is obtained using the row corresponding to $x_2 = 1$ in the joint probability table as follows:

$$P(x_3 = 0 | x_2 = 1) = \frac{4}{220} \div \frac{112}{220} = \frac{4}{112}$$

$$P(x_3 = 1 | x_2 = 1) = \frac{48}{220} \div \frac{112}{220} = \frac{48}{112}$$

$$P(x_3 = 2 | x_2 = 1) = \frac{60}{220} \div \frac{112}{220} = \frac{60}{112}$$

$$P(x_3 = j | x_2 = 1) = 0 \text{ for } j \geq 3$$

Thus, the conditional distribution of $x_3 | x_2 = 1$ is as follows:

Conditional Distribution of $x_3 | x_2 = 1$

| Value | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Probability | $\frac{4}{112}$ | $\frac{48}{112}$ | $\frac{60}{112}$ | 0 | 0 | 0 | 0 |

E9) Clearly $p_{ijk} = 0$ whenever $i + j + k \neq 3$. Hence we shall consider only those combinations of i, j and k such that $i + j + k = 3$.

$$p_{003} = \left({}^6C_3\right) \Big/ \left({}^{12}C_3\right) = \frac{20}{220}$$

$$p_{012} = \left({}^4C_1\right)\left({}^6C_2\right) \Big/ \left({}^{12}C_3\right) = \frac{60}{220}$$

$$p_{021} = \left({}^4C_2\right)\left({}^6C_1\right) \Big/ \left({}^{12}C_3\right) = \frac{36}{220}$$

$$p_{030} = \left({}^4C_3\right) \Big/ \left({}^{12}C_3\right) = \frac{4}{220}$$

$$p_{102} = \left({}^2C_1\right)\left({}^6C_2\right) \Big/ \left({}^{12}C_3\right) = \frac{30}{220}$$

$$p_{111} = \left({}^2C_1\right)\left({}^4C_1\right)\left({}^6C_1\right) \Big/ \left({}^{12}C_3\right) = \frac{48}{220}$$

$$p_{120} = \left({}^2C_1\right)\left({}^4C_2\right) \Big/ \left({}^{12}C_3\right) = \frac{12}{220}$$

$$p_{201} = \left({}^2C_2\right)\left({}^6C_1\right) \Big/ \left({}^{12}C_3\right) = \frac{6}{220}$$

$$p_{210} = \left({}^2C_2\right)\left({}^4C_1\right) \Big/ \left({}^{12}C_3\right) = \frac{4}{220}.$$

E10) From Table 15.2 of Example 5,

$$E(x_1) = 1.\frac{90}{220} + 2.\frac{10}{220} = \frac{110}{220} = \frac{1}{2}$$

$$E(x_1^2) = 1.\frac{90}{220} + 4.\frac{10}{220} = \frac{130}{220}$$

$$V(x_1) = E(x_1^2) - (E(x_1)^2) = \frac{130}{220} - \frac{1}{4} = \frac{75}{220}$$

From Table 15.3 of Example 5,

$$E(x_2) = 1.\frac{112}{220} + 2.\frac{48}{220} + 3.\frac{4}{220} = \frac{220}{220} = 1$$

$$E(x_2^2) = 1.\frac{112}{220} + 4.\frac{48}{220} + 9.\frac{4}{220} = \frac{340}{220}$$

So $V(x_2) = E(x_2^2) - (E(x_2)^2) = \frac{340}{220} - 1 = \frac{120}{220}$

From E8), we have

$$E(x_3) = 1.\frac{90}{220} + 2.\frac{90}{220} + 3.\frac{20}{220} = \frac{330}{220} = \frac{3}{2}$$

$$E(x_3^2) = 1.\frac{90}{220} + 4.\frac{90}{220} + 9.\frac{20}{220} = \frac{630}{220}$$

$$V(x_3) = \frac{630}{220} - \frac{9}{4} = \frac{135}{220}$$

From Table 15.1 of Example 5

$$E(x_1x_2) = 1.1.\frac{48}{220} + 1.2.\frac{12}{220} + 2.1.\frac{4}{220} = \frac{80}{220}$$

$$Cov(x_1x_2) = E(x_1x_2) - E(x_1).E(x_2) = \frac{80}{220} - \frac{1}{2} = -\frac{30}{220}$$

From the table of joint distribution of $x_2$ and $x_3$ in E8,

$$E(x_2x_3) = 1.1.\frac{48}{220} + 1.2.\frac{60}{220} + 2.1.\frac{36}{220} = \frac{240}{220}$$

$$Cov(x_2x_3) = \frac{240}{220} - 1.\frac{3}{2} = -\frac{90}{220}$$

From the computations of E8,

$$Prob\{x_1 = 1, x_3 = 1\} = p_{111} = \frac{48}{220}$$

$$Prob\{x_1 = 1, x_3 = 2\} = p_{102} = \frac{30}{220}$$

$$Prob\{x_1 = 2, x_3 = 1\} = p_{201} = \frac{6}{220}$$

Now $E(x_1x_3) = 1.1.\frac{48}{220} + 1.2.\frac{30}{220} + 2.1.\frac{6}{220} = \frac{120}{220}$

So $Cov(x_1 x_3) = \frac{120}{220} - \frac{1}{2}.\frac{3}{2} = -\frac{45}{220}$.

Hence the variance covariance matrix of $\mathbf{x} = (x_1, x_2, x_3)^t$ is

$$\frac{1}{220}\begin{bmatrix} 75 & -30 & -45 \\ -30 & 120 & -90 \\ -45 & -90 & 135 \end{bmatrix}$$

Remark: Notice that each row sum of the above dispersion matrix is $0$. Thus, it is a positive semidefinite matrix. Is it surprising? No! Why not? That is because we know that $x_1 + x_2 + x_3 = 3$ (a constant). Thus, $V(x_1 + x_2 + x_3) = 0$.

E11) (a) Notice that $x^2$ can take only two values 1 and 9.

$$P(x^2 = 1) = P(x = 1 \text{ or } x = -1) = P(x = 1) + (p(x = -1) = \frac{1}{2}$$

Similarly, $P(x^2 = 9) = P(x = 3) + P(x = -3) = \frac{1}{2}$

(b) The joint distribution of $x$ and $x^2$ is given as under:

Joint Distribution of $x$ and $x^2$

| Value taken by | $\rightarrow$ x | −3 | −1 | 1 | 3 | Row sum |
|---|---|---|---|---|---|---|
| $\downarrow x^2$ | | Joint Probability | | | | |
| 1 | Joint Probability | 0 | ¼ | ¼ | 0 | ½ |
| 9 | | ¼ | 0 | 0 | ¼ | ½ |
| Column sum | | ¼ | ¼ | ¼ | ¼ | 1 |

(c) Clearly, $E(x) = -\frac{1}{4} + \frac{1}{4} = 0$

So $\text{Cov}(x, x^2) = E(x.x^2) - E(x).E(x^2) = E(x.x^2)$

$$= (-3)^3.\frac{1}{4} + (-1)^3.\frac{1}{4} + 1^3.\frac{1}{4} + 3^3.\frac{1}{4} = 0.$$

(d) $P(x = 3 \text{ and } x^2 = 1) = 0$

But $P(x = 3) = \frac{1}{4}$ and $P(x^2 = 1) = \frac{1}{2}$

Thus, $P(x = 3 \text{ and } x^2 = 1) \neq P(x = 3).P(x^2 = 1)$ .

Hence $x$ and $x^2$ are not independent.

E12) (a) $f_x(u_1, u_2) = \begin{cases} c.u_1(2 - u_1 - u_2) & \text{when } 0 < u_1 < 1 \text{ and } 0 < u_2 < 1 \\ 0 & \text{otherwise} \end{cases}$

So $1 = \int_0^1\int_0^1 f_x(u_1, u_2)du_1du_2$

$$= \int_0^1\int_0^1 c\left(2u_1 - u_1^2 - u_1u_2\right)du_1du_2$$

$$= c\left[\int_0^1\left[(2u_1 - u_1^2) - u_1\left(\int_0^1 u_2 du_2\right)\right]du_1\right]$$

$$= c\int_0^1\left(2u_1 - u_1^2 - \frac{1}{2}u_1\right)du_1$$

$$= c\left[\frac{3}{2}.\frac{u_1^2}{2} - \frac{u_1^3}{3}\right]_0^1$$

$$= c\left[\frac{3}{4}. - \frac{1}{3}\right] = c.\frac{5}{12} = \frac{5}{12}.c$$

Hence $c = \frac{12}{5}$

(b) The marginal density of $x_1$ is

$$\frac{12}{5}\int_0^1 u_1(2 - u_1 - u_2)du_2$$

$$= \frac{12}{5}u_1\left(2 - u_1 - \frac{1}{2}\right) = \frac{12}{5}u_1\left(\frac{3}{2} - u_1\right) \text{ if } 0 < u_1 < 1 \text{ and } 0 \text{ otherwise.}$$

Similarly, the marginal density of $x_2$ is

$$\frac{12}{5}\int_0^1 u_1(2 - u_1 - u_2)du_1 = \frac{12}{5}\left(\frac{2}{3} - \frac{u_2}{2}\right)$$

(c) The conditional density of $x_2$ given $x_1 = u_1, 0 < u_1 < 1$, is given by

$$\frac{f_x(u_1, u_2)}{f_{x_1}(u_1)} = \frac{\frac{12}{5}(2 - u_1 - u_2).u_1}{\frac{12}{5}u_1\left(\frac{3}{2} - u_1\right)} = \frac{2 - u_1 - u_2}{\frac{3}{2} - u_1}, \quad 0 < u_2 < 1$$

(d) Since the conditional density of $x_2$ given $x_1$ is different from the marginal distribution of $x_2$, we conclude that $x_1$ and $x_2$ are not independent.

E13) From Example 8, if $\rho_{x_1x_2} = 0$, the density of $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is

$$f_x(u_1, u_2) = \frac{e^{\left\{-\frac{1}{2}\left(\frac{u_1 - \mu_{x_1}}{\sigma_{x_1}}\right)^2 + \left(\frac{u_2 - \mu_{x_2}}{\sigma_{x_2}}\right)^2\right\}}}{2\pi\sigma_{x_1}\sigma_{x_2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_{x_1}}e^{-\frac{1}{2}\left(\frac{u_1 - \mu_{x_1}}{\sigma_{x_1}}\right)^2}\frac{1}{\sqrt{2\pi}\sigma_{x_2}}e^{-\frac{1}{2}\left(\frac{u_2 - \mu_{x_2}}{\sigma_{x_2}}\right)^2}$$

$$= f_{x_1}(u_1).f_{x_2}(u_2)$$

Hence $x_1$ and $x_2$ are independent. Conversely, if $x_1$ and $x_2$ are any two independent random variables, then we know that $\rho_{x_1x_2} = 0$ and hence it holds in this particular case.

–x–

# UNIT 16 DEFINITION AND PROPERTIES OF MVN-II

## Structure

Page No



## 16.1 INTRODUCTION

In this unit, the concepts of marginal and conditional distributions and the important concept of independence with the help of examples will be continued. We also give a method of obtaining the density of a transformed random vector. In Sec. 16.2, we introduce multivariate normal distribution using its density function. In Sec. 16.3, we study the multivariate normal distribution defined via linear zero functions and obtain several properties. We also show that the two definitions of multivariate normal coincide if the variance-covariance matrix is positive definite.

### Objectives

After completing this unit, you should be able to

- apply the properties of multivariate normal distribution to the problems of multivariate analysis;
- appreciate the beauty of the density-free approach to multivariate normal distribution.

## 16.2 NONSINGULAR MULTIVARIATE NORMAL DISTRIBUTION

The multivariate normal distribution is a generalization of the univariate normal distribution to higher dimensions. This distribution plays a fundamental role in multivariate analysis. While it is true that the real data virtually never follow multivariate normal, the multivariate normal distribution is often a good approximation to the population distribution. Also the sampling distributions of many multivariate statistics are approximately normal, regardless of the form of the parent distribution (discrete or continuous) in view of the multivariate central limit theorem which can be stated as follows:

**Theorem 1 (Multivariate Central Limit Theorem):** Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be independent observations from a population with mean vector $\boldsymbol{\mu}$ and finite variance-covariance matrix, $\Sigma$. Let $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$. Then for large n, $\sqrt{n}\,(\overline{\mathbf{x}} - \boldsymbol{\mu})$ has an approximate multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\Sigma$. ($n$ should be large relative to the number of components in $\overline{\mathbf{x}}$.)

Moreover, the multivariate normal theory is easily tractable mathematically and nice and elegant results can be obtained. Thus, the study of multivariate normal distribution serves the dual purpose of usefulness in practice and mathematical elegance.

Let us start with the simplest extension of a univariate standard normal distribution. Let $x_1, \ldots, x_p$ be independent standard normal variables. Then their joint density which we have identified as the density of $\mathbf{x} = (x_1, \ldots, x_p)^t$ is given by

$$f_x(u_1, \ldots, u_p) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u_2^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{u_p^2}{2}}$$

$$= \frac{1}{(2\pi)^{p/2}} e^{-\frac{u^t u}{2}}, \ u \in R^p, \text{ where } \mathbf{u} = (u_1, \ldots, u_p)^t.$$

Let us make a nonsingular linear transformation $\mathbf{y} = \mathbf{Bx} + \boldsymbol{\mu}$, where $\mathbf{B}$ is a fixed nonsingular matrix and $\boldsymbol{\mu}$ is a fixed vector. Accordingly let us set $\mathbf{v} = \mathbf{Bx} + \boldsymbol{\mu}$. Then the density of $\mathbf{y}$ is obtained as follows.

First we obtain the Jacobian of the transformation:

$$J(u_1, \ldots, u_p) = \text{Absolute value of} \begin{vmatrix} \frac{\partial v_1}{\partial u_1} & \cdots & \frac{\partial v_1}{\partial u_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial v_p}{\partial u_1} & \cdots & \frac{\partial v_p}{\partial u_p} \end{vmatrix} = \begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ \cdots & \ddots & \cdots & \cdots \\ \cdots & \cdots & \ddots & \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{vmatrix} = |\mathbf{B}|$$

Also $\mathbf{u} = \mathbf{B}^{-1}(v - \mu)$.

So the density of $\mathbf{y}$ is $\frac{1}{(2\pi)^{p/2}} . e^{-\frac{1}{2}(v-\mu)^t \mathbf{B}^{-1^t} \mathbf{B}^{-1}(v-\mu)} . \frac{1}{[\text{absolute value of } |\mathbf{B}|]}$

Let us write $\Sigma = \mathbf{BB}^t$. Then $|\Sigma| = |\mathbf{B}||\mathbf{B}^t| = |\mathbf{B}|^2$. So the absolute value of $|\mathbf{B}|$ is $|\Sigma|^{1/2}$, the positive square root of $|\Sigma|$.

Thus, the density of $\mathbf{y}$ can be rewritten as $f_y(v) = \frac{1}{\sqrt{2\pi}|\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)^t \Sigma^{-1}(v-\mu)}, v \in R^p$.

Notice that the density of $\mathbf{y}$ depends on the parameters $\mu$ and $\Sigma$. The probability distribution with this density is called *p*-variate normal distribution with parameters $\mu$ and $\Sigma$ and we denote the distribution as

$$\mathbf{y} \sim N_p(\mu, \Sigma)$$

In the same notation, the vector $\mathbf{x}$ that we considered originally has the distribution $N_p(\mathbf{0}, \mathbf{I})$.

Let us now identify the parameters $\mu$ and $\Sigma$.

We know that $E(\mathbf{x}) = \mathbf{0}$ and the variance-covariance matrix $D(\mathbf{x}) = \mathbf{I}$, since $x_1, \ldots, x_p$ are independent standard normal variates.

Since $\mathbf{y} = \mathbf{Bx} + \boldsymbol{\mu}$, we have

$E(\mathbf{y}) = E(\mathbf{Bx} + \boldsymbol{\mu}) = \mathbf{B}E(\mathbf{x}) + \boldsymbol{\mu} = \boldsymbol{\mu}$

$D(\mathbf{y}) = D(\mathbf{Bx} + \boldsymbol{\mu}) = D(\mathbf{Bx}) = \mathbf{BIB}^t = \mathbf{BB}^t = \Sigma$

Thus, $\mu$ and $\Sigma$ are the mean vector and the variance-covariance matrix of $\mathbf{y} \sim N_p(\mu, \Sigma)$. Recall that we started with B nonsingular and hence $\Sigma$ is nonsingular (in fact, positive definite). The fact that B is nonsingular was crucial in obtaining the density of y as above (Notice that the jacobian not being 0 was an assumption while obtaining the density of the transformed random vector in the form mentioned above.) Thus, the distribution $N_p(\mu, \Sigma)$, i.e., the *p*-variate normal distribution with mean vector $\mu$ and variance-covariance matrix $\Sigma$ is called a nonsingular *p*-variate normal distribution, if $\Sigma$ is positive definite. Later on, we shall also study the case where $\Sigma$ need not necessarily be positive definite. Let us now summarize the above discussion.

**Definition 1:** A random vector $\mathbf{y}$ of order $p \times 1$ is said to have a nonsingular *p*-variate normal distribution with parameters $\mu$ and $\Sigma$ if it has the density

$$f_y(\mathbf{v}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)' \Sigma^{-1}(v-\mu)}$$

where $\mu$ is a fixed vector in $\mathbf{R}^p$ and $\Sigma$ is a $p \times p$ positive definite matrix. Also then, we use the notation $\mathbf{y} \sim N_p(\mu, \Sigma)$.

**Remark:** If $\mathbf{y} \sim N_p(\mu, \Sigma)$ where $\Sigma$ is a positive definite, then $E(\mathbf{y}) = \mu$ and $D(\mathbf{y}) = \Sigma$.

We shall now turn our attention to the marginal distributions. Let $\mathbf{y} \sim N_p(\mu, \Sigma)$.

$$\text{Partition } \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix} \tag{1}$$

where $\mathbf{y}_1$, $\mathbf{v}_1$ and $\mu_1$ are of order $r \times 1 (1 \le r \le p)$ and $\Sigma_{11}$ and $\Sigma_{22}$ are of order $r \times r$ and $(p-r) \times (p-r)$, respectively. Such partitions of $\mathbf{v}$, $\mu$ and $\Sigma$ are called conformable partitions to that of $\mathbf{y}$.

We first show that if $\Sigma_{12} = 0$, then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent. Notice that if $\Sigma_{12} = 0$, then the covariance between $y_i$ and $y_j$, $i = 1, \ldots, r$ and $i = r+1, \ldots, p$ is 0 and hence each $y_i, \ldots, y_r$ is uncorrelated with each of $y_{r+1}, \ldots, y_p$.

**Theorem 2:** Let $\mathbf{y} \sim N_p(\mu, \Sigma)$ where $\Sigma$ is positive definite. Let $\mathbf{y}$, $\mathbf{v}$, $\mu$ and $\Sigma$ be partitioned as given in Eqn. (1). Then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent if and only if $\Sigma_{12} = 0$.

**Proof:** We need to prove only the 'if' part as the only if part has already proved in Theorem 4 of Unit 15.

The density of $\mathbf{y}$ is $f_y(\mathbf{v}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)' \Sigma^{-1}(v-\mu)}$.

Since $\mathbf{\Sigma}_{12} = \mathbf{0}$, $|\Sigma| = \left| \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix} \right| = |\mathbf{\Sigma}_{11}||\mathbf{\Sigma}_{22}|$.

Also $\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$

Hence $(\mathbf{v}-\boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{v}-\boldsymbol{\mu}) = \left(\mathbf{v}_1^t - \boldsymbol{\mu}_1^t : \mathbf{v}_2^t - \boldsymbol{\mu}_2^t\right) \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 - \boldsymbol{\mu}_1 \\ \mathbf{v}_2 - \boldsymbol{\mu}_2 \end{pmatrix}$

$$= (\mathbf{v}_1 - \boldsymbol{\mu}_1)^t \Sigma_{11}^{-1} (\mathbf{v}_1 - \boldsymbol{\mu}_1) + (\mathbf{v}_2 - \boldsymbol{\mu}_2)^t \Sigma_{22}^{-1} (\mathbf{v}_2 - \boldsymbol{\mu}_2).$$

Thus, $f_y(\mathbf{v})$ can be rewritten as

$$f_y(\mathbf{v}) = \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2} |\Sigma_{22}|^{1/2}} e^{-\frac{1}{2}\left\{(v_1-\mu_1)^t \Sigma_{11}^{-1}(v_1-\mu_1) + (v_2-\mu_2)^t \Sigma_{22}^{-1}(v_2-\mu_2)\right\}}$$

$$= \frac{1}{(2\pi)^{\frac{r}{2}} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(v_1-\mu_1)^t \Sigma_{11}^{-1}(v_1-\mu_1)} \frac{1}{(2\pi)^{\frac{p-r}{2}} |\Sigma_{22}|^{1/2}} e^{-\frac{1}{2}(v_2-\mu_2)^t \Sigma_{22}^{-1}(v_2-\mu_2)}$$

$$= f_{y_1}(\mathbf{v}_1).f_{y_2}(\mathbf{v}_2).$$

Thus, joint density of $\mathbf{y}$ factors into product of r-variate and $(p-r)$-variate normal density involving only $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively.

Hence $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent.

We now obtain the marginal distribution of $\mathbf{y}_1$ in general case.

**Theorem 3:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ where $\Sigma$ is positive definite. Let $\mathbf{y}$, $\mathbf{v}$, $\boldsymbol{\mu}$ and $\Sigma$ be partitioned as in Eqn. (1). Then the marginal distribution of $\mathbf{y}_1$ is $N_r(\boldsymbol{\mu}_1, \Sigma_{11})$.

**Proof:** The technique of the proof lies in transforming from $\mathbf{y}$ to $\boldsymbol{\xi}$ by a nonsingular linear transformation such that $\mathbf{y}_1 = \boldsymbol{\xi}_1$, and $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are independent. Then the marginal distribution of $\boldsymbol{\xi}_1$ and hence $\mathbf{y}_1$ can be easily obtained using Theorem 6 of Unit 15. Towards this end, write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 \\ \Sigma_{12}^t \Sigma_{11}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \Sigma_{11}^{-1}\Sigma_{12} \\ 0 & \mathbf{I} \end{pmatrix}$$

Hence $\Sigma^{-1} = \begin{pmatrix} \mathbf{I} & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & 0 \\ -\Sigma_{12}^t \Sigma_{11}^{-1} & \mathbf{I} \end{pmatrix}$

Also $|\Sigma| = \begin{vmatrix} \mathbf{I} & 0 \\ -\Sigma_{12}^t \Sigma_{11}^{-1} & \mathbf{I} \end{vmatrix} \begin{vmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{vmatrix} \begin{vmatrix} \mathbf{I} & \Sigma_{11}^{-1}\Sigma_{12} \\ 0 & \mathbf{I} \end{vmatrix}$

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

since $\begin{vmatrix} \mathbf{A} & 0 \\ \mathbf{B} & \mathbf{C} \end{vmatrix} = |\mathbf{A}|.|\mathbf{C}|$, where $\mathbf{A}$ and $\mathbf{C}$ are squares and $|\mathbf{I}| = 1$

Now $f_y(v) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)'\Sigma^{-1}(v-\mu)}$

$$= \frac{e^{-\frac{1}{2}\left((v_1-\mu_1)' : (v_2-\mu_2)'\right)\begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ O & I \end{pmatrix}\begin{pmatrix} \Sigma_{11}^{-1} & O \\ O & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}\begin{pmatrix} I & O \\ -\Sigma_{12}^t\Sigma_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} v_1-\mu_1 \\ v_2-\mu_2 \end{pmatrix}}}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2} |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}$$

$$= \frac{e^{-\frac{1}{2}\left\{(\omega_1^t : \omega_2^t)\begin{pmatrix} \Sigma_{11}^{-1} & O \\ O & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}\right\}}}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2} |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}$$

where $\boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 - \boldsymbol{\mu}_1 \\ \mathbf{v}_2 - \boldsymbol{\mu}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 - \boldsymbol{\mu}_1 \\ \mathbf{v}_2 - \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{12}^t \boldsymbol{\Sigma}_{11}(\mathbf{v}_1 - \boldsymbol{\mu}_1) \end{pmatrix}$

The Jacobian of the transformation is $J = \begin{vmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{vmatrix} = 1$.

Hence the density of $\xi = \boldsymbol{\xi} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix} (\mathbf{y} - \boldsymbol{\mu})$ is

$$\mathbf{f}_\xi(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{r/2} |\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(v_1-\mu_1)'\Sigma_{11}^{-1}(v_1-\mu_1)} \times \frac{e^{-\frac{1}{2}\left(v_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(v_1-\mu_1)\right)'\left(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\left(v_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(v_1-\mu_1)\right)}}{(2\pi)^{p-r/2} |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}$$

From the above it is clear that $(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ and $\left(\mathbf{y}_2 - \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)\right)$ are independent with densities $N_r(0, \Sigma_{11})$ and $N_{p-1}(0, \Sigma_{22} - \Sigma_{21}, \Sigma_{11}^{-1}\Sigma_{12})$, respectively. Hence the marginal distribution of $\mathbf{y}_1$ from the above factorization is $N_r(\mu_1, \Sigma_{11})$. Also the marginal distribution of $\left(\mathbf{y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_1\right)$ is

$N_{p-r}\left(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)$. (Notice that $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^t$ ).

This complete the proof.

**Corollary:** If $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\Sigma$ is positive definite, then $\mathbf{y}_1 \sim N(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_{ii})$ for $i = 1, \ldots, p$.

We shall now obtain the conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1 = \mathbf{v}_1$, where $\mathbf{y}(\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$ is partitioned as in Eqn. (1). Once again we assume that $\boldsymbol{\Sigma}$ is p.d.

**Theorem 4:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. Let $\mathbf{y}$, $\mathbf{v}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be partitioned as in Eqn. (1). The conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1$ is

$N_{p-r}\left(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{v}_1 - \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)$.

**Proof:** In the proof of Theorem 7 of Unit 15, we showed that $\mathbf{y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_1$ and $y_1$ are independently distributed. Hence the conditional distribution of $\mathbf{y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_1$ given $\mathbf{y}_1$ is the same as the unconditional distribution of $\mathbf{y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_1$, which is $N_{p-r}\left(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1,\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)$. Thus, the conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1 = \mathbf{v}_1$ is a *p-r* variate normal with mean equal to

$$\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{v}_1$$

since when $y_1 = v_1$, $\mathbf{y}_1 = \mathbf{v}_1$, $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_1 = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{v}_1$ is a fixed vector.

Also the conditional variance-covariance matrix is the same as that of $\mathbf{y}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{y}_1$ which is $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Thus, the conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1$ is $N_{p-r}\left(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{v}_1 - \boldsymbol{\mu}_1),\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)$. This complete the proof.

Let $r = p-1$. Then $\mathbf{y}_2$ is univariate random variable $\mathbf{y}_p$. Also $\Sigma_{21}$ is a row vector of order $1 \times (p-1)$ and $\Sigma_{11}$ is of order $(p-1) \times (p-1)$. Then conditional expectation of $\mathbf{y}_p$ given $y_1 = v_1, \ldots, y_{p-1} = v_{p-1}$ obtained from the above theorem is

$$\begin{aligned} E(y_p \mid v_1, \ldots, v_{p-1}) &= \mu_p + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{v}_1 - \boldsymbol{\mu}_1) \\ &= \beta_0 + \beta_1 v_1 + \cdots + \beta_{p-1} v_{p-1} \end{aligned}$$

where $\beta_0 = \mu_p - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$ and $\beta = (\beta_1, \ldots, \beta_{p-1})^t - \Sigma_{11}^{-1}\Sigma_{12}$.

This conditional expectation is called regression of $y_p$ on $y_1, \ldots, y_{p-1}$.

So it is clear that if $y_1, \ldots, y_p$ have a joint *p*-variate normal distribution, then the regression of $y_p$ on $y_1, \ldots, y_{p-1}$ is linear in $y_1, \ldots, y_{p-1}$. $\beta_1, \ldots, \beta_{p-1}$ are called the regression coefficients.

**Example 1:** Let $\mathbf{y} = (y_1, y_2, y_3)^t$ have $N_3(\mu, \Sigma)$, where $\boldsymbol{\mu} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix}.$$

(a) Write down the marginal distributions of $y_1$, $y_2$, and $y_3$.

(b) Write down the marginal distribution of $\begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$.

(c) Write down the conditional distribution of $y_1$ given $y_2 = -0.5$ and $y_3 = 0.2$.

(d) Make a nonsingular linear transformation $\boldsymbol{\xi} + \mathbf{Ty} + \mathbf{c}$ so that $\xi_1$, $\xi_2$ and $\xi_3$ are independent standard normal variables.

**Solution:** (**a**) In view of the corollary to Theorem 3 of Unit 15,
$y_1 \sim N(1, 4)$, $y_2 \sim N(-1, 4)$ and $y_3 \sim N(0, 4)$

(**b**) It is easy to see that the marginal distribution of $(y_1, y_3)^t$ is a bivariate normal. From the given information on $\mu$ and $\Sigma$, we have

$$E(y_1) = 1,\ E(y_3) = 0,\ V(y_1) = V(y_3) = 4,\ Cov(y_1, y_3) = 2$$

Hence $\begin{pmatrix} y_1 \\ y_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}\right)$.

(**c**) The conditional distribution of $y_1$ given $y_2 = -0.5$ and $y_3 = 0.2$ is normal by Theorem 4.

Partition $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}^t & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ where $\boldsymbol{\sigma}_{12}$ is of order $1\times 2$ and $\boldsymbol{\Sigma}_{22}$ is of order $2\times 2$.

Then by Theorem 4, the conditional mean of $y_1$ is $\mu_1 + \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\begin{pmatrix} -0.5 - \mu_2 \\ 0.2 - \mu_3 \end{pmatrix}$

$$= 1 + (1\ \ 2)\frac{1}{12}\begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} -0.5-(-1) \\ 0.2-0 \end{pmatrix}$$

$$= 1 + \frac{1}{12}(0\ \ 6)\begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} = 1 + 0.1 = 1.1$$

The conditional variance is

$$\boldsymbol{\sigma}_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}^t = 4 - \frac{1}{12}(0\ 6)\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 4 - 1 = 3.$$

So the conditional distribution of $y_1$ given $y_2 = -0.5$ and $y_3 = 0.2$ is $N(1.1, 3)$.

(**d**) We have $y - \mu \sim N_3(0, \Sigma)$

We shall obtain a lower triangular matrix $\mathbf{B}$ such that $\mathbf{BB}^t = \Sigma$. Then $\xi = \mathbf{B}^{-1}(\mathbf{y} - \mu)$ has a trivariate normal distribution with independent components. Further,

$$E(\xi) = E(\mathbf{B}^{-1}(\mathbf{y} - \mu)) = \mathbf{B}^{-1}(E(\mathbf{y} - \mu)) = \mathbf{B}^{-1}(\mu - \mu) = 0.\ D(\xi) = \mathbf{B}^{-1}D(y)\mathbf{B}^{-1^t} = \mathbf{B}^{-1}\Sigma\mathbf{B}^{-1^t}$$
$$= \mathbf{B}^{-1}\mathbf{BB}^t\mathbf{B}^{-1^t} = \mathbf{I}$$

We shall use the algorithm given in Sec. 15.4 to get as required $\mathbf{B}$ and $\mathbf{B}^{-1}$. Thus, we form

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | 1 | 2 | 1 | 0 | 0 | (1) |
| 1 | 4 | 2 | 0 | 1 | 0 | (2) |
| 2 | 2 | 4 | 0 | 0 | 1 | (3) |
| 2 | ½ | 1 | ½ | 0 | 0 | (4) = (1) ÷ $\sqrt{4}$ |
| 0 | $\frac{15}{4}$ | $\frac{3}{2}$ | $-\frac{1}{4}$ | 1 | 0 | (5) = (2) – (4)x½ |
| 0 | $\frac{3}{2}$ | 3 | $-\frac{1}{2}$ | 0 | 1 | (6) = (3) – (4) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | ½ | 1 | ½ | 0 | 0 | (7) = (4) |
| 0 | $\frac{\sqrt{15}}{2}$ | $\frac{\sqrt{3}}{\sqrt{5}}$ | $-\frac{1}{\sqrt{60}}$ | $\frac{2}{\sqrt{15}}$ | 0 | (8) = (5) ÷ $\sqrt{\frac{15}{4}}$ |
| 0 | 0 | $\frac{12}{5}$ | $-\frac{2}{5}$ | $-\frac{2}{5}$ | 1 | (9) = (6) - $\frac{\sqrt{3}}{\sqrt{5}}$ X(8) |
| 2 | ½ | 1 | ½ | 0 | 0 | (10) = (4) |
| 0 | $\frac{\sqrt{15}}{2}$ | $\frac{\sqrt{3}}{\sqrt{5}}$ | $-\frac{1}{\sqrt{60}}$ | $\frac{2}{\sqrt{15}}$ | 0 | (11) = (8) |
| 0 | 0 | $\sqrt{\frac{12}{5}}$ | $-\frac{1}{\sqrt{15}}$ | $-\frac{1}{\sqrt{15}}$ | $\frac{\sqrt{5}}{\sqrt{12}}$ | (12) = (9) ÷ $\sqrt{\frac{12}{5}}$ |

Hence $\mathbf{B} = \begin{pmatrix} 2 & 0 & 0 \\ \frac{1}{2} & \frac{\sqrt{15}}{2} & 0 \\ 1 & \frac{\sqrt{3}}{\sqrt{5}} & \frac{\sqrt{12}}{\sqrt{5}} \end{pmatrix}$ and $\mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{\sqrt{60}} & \frac{2}{\sqrt{15}} & 0 \\ -\frac{1}{\sqrt{15}} & -\frac{1}{\sqrt{15}} & \frac{\sqrt{5}}{\sqrt{12}} \end{pmatrix}$

Hence if $\xi = \mathbf{T}\mathbf{y} + \mathbf{c}$, where $\mathbf{T} = \mathbf{B}^{-1}$ and $c = \mathbf{T}\,\mathbf{c} = \mathbf{T}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, then $\xi$ is a vector of independent N(0, 1) variables.

Now, try an exercise.

---

E1) Let $N_4(\mu, \Sigma)$, where $\mu = \begin{pmatrix} 2 \\ 4 \\ 1 \\ -3 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 9 & 0 & 2 & 0 \\ 0 & 4 & 0 & 1 \\ 2 & 0 & 6 & 0 \\ 0 & 1 & 0 & 9 \end{pmatrix}$.

(a) Obtain the marginal distribution of $\begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$.

(b) Show that $\begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$ and $\begin{pmatrix} y_2 \\ y_4 \end{pmatrix}$ are independent.

(c) Obtain the conditional distribution of $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ given $\begin{pmatrix} y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1.2 \\ -2.6 \end{pmatrix}$

(d) Write down the correlation coefficient between $y_1$ and $y_3$.

---

So far, we have discussed multivariate normal distribution. Let us now discuss characterization of multivariate normal via linear functions.

## 16.3 CHARACTERIZATION OF MULTIVARIATE NORMAL VIA LINEAR FUNCTIONS

Let $\mathbf{y}$ have a nonsingular $p$-variate normal distribution. In the previous section, we saw that each component of y has a univariate normal distribution. Notice that $\mathbf{y}_i = \mathbf{e}_i^t\mathbf{y}$, where $\mathbf{e}_i$ is the $i^{th}$ column of the identity matrix. Thus, $\mathbf{y}_i$ is a linear combination of the components of $\mathbf{y}$. In this section, we show that every fixed linear combination of the components of $\mathbf{y}$ (where $\mathbf{y}$ is as specified above) has a univariate normal distribution. We then go on to show that the distribution of a $p$-variate random vector is completely determined by the class of distributions of all fixed linear combinations of its components. Hence we show that a $p$-variate random vector $\mathbf{y}$ has a $p$-variate normal distribution if and only if every fixed linear combination of the components of $\mathbf{y}$ has a univariate normal distribution. Using this characterization, we derive several properties of multivariate normal distribution, some of which we studied in the previous section.

**Theorem 5:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\Sigma$ is positive definite. Then the following hold.

(a) Let $\mathbf{z} = \mathbf{B}\mathbf{y}$ where $\mathbf{B}$ is a fixed nonsingular matrix. Then $\mathbf{z} \sim N_p(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t)$.

(b) Let $\mathbf{x} = \mathbf{C}\mathbf{y}$, where $\mathbf{C}$ is a fixed $r \times p$ matrix of rank $r (1 \le r \le p)$. Then $\mathbf{x} \sim N_r(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^t)$.

(c) Let $\mathbf{w} = \mathbf{l}'\mathbf{y}$ where $\mathbf{l}$ is a fixed nonnull vector. Then $\mathbf{w} \sim N(\mathbf{l}'\boldsymbol{\mu}, \mathbf{l}'\boldsymbol{\Sigma}\mathbf{l})$.

Further the distributions of $\mathbf{z}$ and $\mathbf{x}$ are nonsingular multivariate normal.

**Proof:** **(a)** The density of $\mathbf{y}$ is $f_y(\mathbf{v}) = \dfrac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)^t \Sigma^{-1}(v-\mu)}$

Since for the transformation $z = By$, $B$ is non-singular, the Jacobian of the transformation is the absolute value of $|\mathbf{B}|$ as we saw in the beginning of Sec. 15.4.

Let $\mathbf{u} = \mathbf{B}\mathbf{v}$ and $\boldsymbol{\theta} = \mathbf{B}\boldsymbol{\mu}$. We have $\mathbf{y} = \mathbf{B}^{-1}\mathbf{z}$, $\mathbf{v} = \mathbf{B}^{-1}u$ and $\boldsymbol{\mu} = \mathbf{B}^{-1}\boldsymbol{\theta}$. We can now write down the density of $\mathbf{z}$ as

$$f_z(\mathbf{u}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(B^{-1}u - B^{-1}\theta)^t \Sigma^{-1}(B^{-1}u - B^{-1}\theta)} \times \frac{1}{\text{abs. value of } |\mathbf{B}|}.$$

Notice that $|\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t| = |\mathbf{B}^2|.|\Sigma|$. Hence $|\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t|^{1/2} = |\Sigma|^{1/2}$ (absolute value of $|\mathbf{B}|$).

Further $(\mathbf{B}^{-1}\mathbf{u} - \mathbf{B}^{-1}\boldsymbol{\theta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{B}^{-1}\mathbf{u} - \mathbf{B}^{-1}\boldsymbol{\theta})$

$$= (\mathbf{u} - \boldsymbol{\theta})^t \mathbf{B}^{-1^t}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{-1}(\mathbf{u} - \boldsymbol{\theta})$$

$$= (\mathbf{u} - \boldsymbol{\theta})^t (\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t)^{-1}(\mathbf{u} - \boldsymbol{\theta})$$

Hence $f_z(\mathbf{u}) = \dfrac{1}{(2\pi)^{p/2}|\mathbf{B}\,\Sigma\,\mathbf{B}^t|^{1/2}} e^{-\frac{1}{2}(u-\theta)^t (\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t)^{-1}(u-\theta)}$

which is the density of $N_p(\boldsymbol{\theta}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t)$ or $N_p(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t)$

**(b)** Since $\mathbf{C}$ is an $r \times p$ matrix of rank $r$, the rows of $\mathbf{C}$ are linearly independent. Hence there exists a matrix $\mathbf{T}$ of order $(p-r) \times p$ of rank $(p-r)$ such that $\mathbf{B} = \begin{pmatrix} \mathbf{C} \\ \mathbf{T} \end{pmatrix}$ is nonsingular. [We can extend the rows of $\mathbf{C}$ to a basis of $\mathbf{R}^p$. The rows of $\mathbf{T}$ are additional vectors in the extended basis of $\mathbf{R}^p$.]

From (a) we have $\mathbf{z} = \mathbf{By} = \begin{pmatrix} \mathbf{Cy} \\ \mathbf{Ty} \end{pmatrix}$ follows $N_p(\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}^t)$.

Observe that the components of $\mathbf{Cy}$ are the first $r$ components of $\mathbf{z}$. By Theorem 3, the marginal distribution of the first $r$ components of $\mathbf{z}$, i.e., distribution of $\mathbf{Cy}$ is $r$-variate normal. Since $E(\mathbf{Cy}) = \mathbf{C}\boldsymbol{\mu}$ and $D(\mathbf{Cy}) = \mathbf{C\Sigma C}^t$, it follows that

$$\mathbf{Cy} = N_r(\mathbf{C}\boldsymbol{\mu}, \mathbf{C\Sigma C}^t).$$

If $\boldsymbol{\Sigma}$ is positive definite then so in the matrix $\mathbf{C\Sigma C}^t$, (Why?) hence distribution of $\mathbf{Cy}$ is non-singular multivariate normal.

(c) This is a special case of (b) where $r = 1$.

Thus, we have proved that if $\mathbf{y} \sim N_p(\mu, \Sigma)$, where $\Sigma$ is positive definite, then every nonnull fixed linear combination $\mathbf{l'y}$ of $\mathbf{y}$ has a univariate normal distribution. If $\mathbf{l} = 0$, then $\mathbf{l'y} = 0$ with probability 1. It has mean 0 and variance 0. It can be thought of as $N(0, 0)$

**Example 2:** Let $\mathbf{y} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = \begin{pmatrix} 4 \\ -2 \\ 6 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 6 & 1 & 2 \\ 1 & 8 & 4 \\ 2 & 4 & 9 \end{pmatrix}$

(a) Obtain the distribution of $\mathbf{x} = \mathbf{Cy}$ where $\mathbf{C} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 5 & 1 \end{pmatrix}$

(b) Obtain a linear combination $\boldsymbol{\xi} = \mathbf{l'y}$ of $\mathbf{y}$ such that $\boldsymbol{\xi}$ has a standard normal distribution.

**Solution:** **(a)** $\mathbf{C}$ is a $2\times3$ matrix of rank 2. (Both the rows of $\mathbf{C}$ are non-null and neither is a scalar multiple of the other.) Hence by Theorem 5(b), we have $\mathbf{x} = \mathbf{Cy} \sim N_2(\mathbf{C}\boldsymbol{\mu}, \mathbf{C\Sigma C}^t)$.

Now $\mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 5 & 1 \end{pmatrix}\begin{pmatrix} 4 \\ -2 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\mathbf{C\Sigma C}^t = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 5 & 1 \end{pmatrix}\begin{pmatrix} 6 & 1 & 2 \\ 1 & 8 & 4 \\ 2 & 4 & 9 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ 1 & 5 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 25 & 63 \\ 63 & 269 \end{pmatrix}$$

**(b)** Let $\mathbf{m} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$. By (c) of Theorem 5 and part (a) given above, we have

$\mathbf{m'y} =\sim N(\mathbf{m'}\boldsymbol{\mu}, \mathbf{m'\Sigma m})$, where $\mathbf{m'}\boldsymbol{\mu} = 0$ and $\mathbf{m'\Sigma m} = 25$.

So $\mathbf{m'y} \sim N(0, 25)$.

Taking $\mathbf{l} = \frac{1}{5}\mathbf{m} = \begin{pmatrix} -\frac{1}{5} \\ \frac{1}{5} \\ \frac{1}{5} \end{pmatrix}$,, we have $\boldsymbol{\xi} = \mathbf{l'y} \sim N(0, 1)$.

Now, try the following exercise.

---

E2) (a) Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\bar{y} = \frac{1}{p}\left(y_1 + \cdots + y_p\right)$. Obtain the distribution of $\bar{y}$.

(b) Let $\mathbf{y} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = \begin{pmatrix} 2 & 1 & 3 & -4 \end{pmatrix}$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & -2 & -1 \\ 1 & -2 & 9 & -1 \\ 1 & -1 & -1 & 16 \end{pmatrix}$$

Find $\mathbf{x} = \mathbf{C}\mathbf{y}$ and $\mathbf{w} = \mathbf{T}\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{w}$ are vectors of order $2 \times 1$ such that $\mathbf{x}$ and $\mathbf{w}$ have independent nonsingular bivariate normal distributions. Compute the parameters of these distributions.

---

We now embark on the issue of characterizing the multivariate normal distribution via linear functions.

**Definition 2 (Characteristic Function):** For a univariate random variable $x$, $\Phi_x(t) = E(e^{itx})$ is called the characteristic function of $x$. For a multivariate random variable $\mathbf{y}$ of order $p \times 1$, $\Phi_{\mathbf{y}}(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{y}})$ is called the characteristic function of $\mathbf{y}$.

The characteristic function of a univariate random variable is a function of a real variable $t$, where as that of a *p*-variate random variable is a function of *p* real variables $t_1, \ldots, t_p$ or the vector $\mathbf{t} = (t_1, \ldots, t_p)^t$ in $R^p$.

We state the following results on characteristic functions the proof of which are beyond the scope of this notes.

**Theorem 6:** Every random vector has a characteristic function.

**Theorem 7:** (a) The characteristic function $\boldsymbol{\Phi}(\mathbf{t})$ of a random vector $\mathbf{y}$ uniquely determines its distribution.

(b) If $\boldsymbol{\Phi}(\mathbf{t}) = \psi_1(t_1)\ldots\psi_p(t_p)$, then the components of $\mathbf{y}$ are independent.

Let $y$ = $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix}$ be a partition of $\mathbf{y}$. Partition $\mathbf{t}$ accordingly. If

$$\boldsymbol{\Phi}(\mathbf{t}) = \psi_1(\mathbf{t}_1)\ldots\psi_k(\mathbf{t}_k)$$

then rectors $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are mutually independent.

We now prove a theorem due to Cramér and Word that connects the distribution of a *p*-variate random vector with the distributions of the linear combinations of its components.

**Theorem 8:** Let $\mathbf{y}$ be a $p$-variate random vector. Then the distribution of $\mathbf{y}$ is completely determined by the class of univariate distributions of all linear functions $\boldsymbol{l}'\mathbf{y}, \boldsymbol{l} \in \mathbf{R}^p$. ($\boldsymbol{l}$ fixed).

**Proof:** Let the characteristic function of $\boldsymbol{l}'\mathbf{y}$ be $\Phi(t, \boldsymbol{l}) = E(e^{it\boldsymbol{l}'y})$.

Now $\Phi(1, \boldsymbol{l}) = E(e^{i\boldsymbol{l}'y}) = \psi(\boldsymbol{l})$ is the characteristic function of $\mathbf{y}$ as a function of $\boldsymbol{l} = (l_1, \ldots, l_p)^t$. By Theorem 7(a) above, the distribution of $\mathbf{y}$ is completely specified by the characteristic function of $\mathbf{y}$.

Motivated by Theorem 8 together with the fact that every linear function of a random vector having a (nonsingular) multivariate normal, has a univariate normal distribution, we define multivariate normal distribution as follows.

**Definition 3:** A $p$-dimensional random vector $\mathbf{y}$ is said to have a $p$-variate normal distribution if every linear function $\boldsymbol{l}'\mathbf{y}$ (with constant $\boldsymbol{l}$) of component of $\mathbf{y}$ has a univariate normal distribution.

From now on, in this section, we use Definition 3 and obtain several important properties of multivariate normal distribution. We shall also show that this definition coincides with the earlier definition through density, whenever the density exists. This approach is called a density-free approach.

**Theorem 9:** Let $\mathbf{y}$ be a $p$-dimensional random vector having a $p$-variate normal distribution. Then $E(\mathbf{y})$ and $D(\mathbf{y})$ exist.

**Proof:** Since every linear function has univariate normal and $y_i$ is a linear function of $\mathbf{y}$, hence by definition, $E(y_i) = \mu_i$ and $V(y_i) = \sigma_{ii}$ exist and are finite for $i = 1, \ldots, p$. Again, since $y_i + y_j$ is a linear function of

$\mathbf{y}$, $V(y_i + y_j) = V(y_i) + 2Cov(y_i, y_j) + V(y_j)$ exists and is finite. Hence

$Cov(y_i, y_j) = \sigma_{ii}$ exists and is finite. Hence $E(\mathbf{y}) = (\mu_1, \ldots, \mu_p)^t$ and

$D(\mathbf{y}) = \mathbf{\Sigma} = ((\sigma_{ij}))$ exist and are finite.

We now obtain the characteristic function of $\mathbf{y} \sim N_p(\mu, \Sigma)$.

**Theorem 10:** Let $\mathbf{y}$ have a $p$-variate normal distribution with $E(\mathbf{y}) = \boldsymbol{\mu}$ and variance-covariance matrix $D(\mathbf{y}) = \mathbf{\Sigma}$. Then the characteristic function of $\mathbf{y}$ is

$\Phi(t) = E(e^{it'y}) = \Phi(\mathbf{t}) = E(e^{it^t y}) = e^{it^t\mu - \frac{1}{2}t^t\Sigma t}$

**Proof:** Recall that the characteristic function of univariate random variable $\xi$ having a normal distribution with mean $\theta$ and variance $\sigma^2$ is $\psi(x) = E(e^{is\xi}) = e^{i\theta s - \frac{s^2\sigma^2}{2}}$.

Also for each fixed $\mathbf{t}$, $\mathbf{t}^t\mathbf{y} \sim N(\mathbf{t}^t\boldsymbol{\mu}, \mathbf{t}^t\mathbf{\Sigma}\mathbf{t})$.

Hence the characteristic function of $\mathbf{t}^t\mathbf{y}$ denoted by $\Psi(s, \boldsymbol{t}) = E(e^{ist'y}) =$

$\psi(s, \mathbf{t}) = E(e^{ist'y}) = e^{ist'\mu - \frac{1}{2}s^2t'\Sigma t}$, for every $\mathbf{t}$.

Now $\Phi(\mathbf{t}) = E(e^{it^t y}) = \Phi(1, \mathbf{t}) = e^{it'\mu - \frac{1}{2}t'\Sigma t}$.

Notice that the characteristic function of $\mathbf{y}$ depends on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, its mean vector and variance-covariance matrix respectively. Also by Theorem 7, the distribution of $\mathbf{y}$ is completely specified by its characteristic function. Hence the distribution of $\mathbf{y}$ is completely specified by its mean vector and variance-covariance matrix. Henceforth, we shall use $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote that y has a *p*-variate normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Where we do not insist that $\boldsymbol{\Sigma}$ is positive definite. However, in view of Theorem 3 of Section 15.3, $\boldsymbol{\Sigma}$ is nnd.

**Theorem 11:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then every linear function $\boldsymbol{l}'\mathbf{y} \sim N_p(\boldsymbol{l}^t\boldsymbol{\mu}, \boldsymbol{l}^t\boldsymbol{\Sigma}\boldsymbol{l})$.

**Proof:** By definition, $\boldsymbol{l}^t\mathbf{y}$ has a univariate normal distribution. Further $E(\boldsymbol{l}^t\mathbf{y}) = \boldsymbol{l}^t\boldsymbol{\mu}$ and $V(\boldsymbol{l}^t\mathbf{y}) = \boldsymbol{l}^t\boldsymbol{\Sigma}\boldsymbol{l}$.

Try an exercise.

---

E3) Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{B}$ be a fixed $r \times p$ matrix. Show that $\mathbf{By} \sim N_r(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t)$.

---

Here, we shall discuss few more theorems.

**Theorem 12:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\boldsymbol{\Sigma}$ is a diagonal matrix, then the components of $\mathbf{y}$ are independent.

**Proof:** The characteristic function of

$$\boldsymbol{\Phi}(\mathbf{t}) = E(e^{it^t y}) = e^{it^t\mu - \frac{1}{2}t^t\Sigma t} = e^{it_1\mu_1 - \frac{1}{2}t_1^2\sigma_{11}} \ldots e^{it_p\mu_p - \frac{1}{2}t_p^2\sigma_{pp}} \text{ (since } t^t\Sigma t = t_1^2\sigma_{11} + \ldots + t_p^2\sigma_{pp})$$

So, by Theorem 7(b) $y_1, \ldots, y_p$ are independent.

If $\boldsymbol{\Sigma}$ is a diagonal matrix, then $y_1, \ldots, y_p$ are uncorrelated. Thus, we have shown in Theorem 7 that uncorrelatedness implies independence if $y_1, \ldots, y_p$ have a multivariate normal distribution.

**Theorem 13:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Write $y = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ when $\mathbf{y}_1$ has r components.

Partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ conformably as $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. Then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independently distributed if and only if $\Sigma_{12} = 0$.

The proof is similar to that of Theorem 12.

**Example 3:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Partition $\mathbf{y}$ as $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix}$. Show that if $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are independent pair-wise, then they are mutually independent. (In general, pair-wise independence does not imply mutually independence, but it holds if $\mathbf{y}_1, \ldots, \mathbf{y}_k$ have a multivariate normal distribution.)

**Solution:** Partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ conformably to that of $\mathbf{y}$. Thus, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_2 \end{pmatrix}$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{k1} & \boldsymbol{\Sigma}_{k2} & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix}.$$

For $i \neq j$, $\begin{pmatrix} \mathbf{y}_i \\ \mathbf{y}_j \end{pmatrix}$ has a multivariate normal distribution with mean vector $\begin{pmatrix} \boldsymbol{\mu}_i \\ \boldsymbol{\mu}_j \end{pmatrix}$ and variance-covariance matrix $\begin{pmatrix} \boldsymbol{\Sigma}_{ii} & \boldsymbol{\Sigma}_{ij} \\ \boldsymbol{\Sigma}_{ji} & \boldsymbol{\Sigma}_{jj} \end{pmatrix}$ (why?). If $\mathbf{y}_i$ and $\mathbf{y}_j$ are independent, then $\boldsymbol{\Sigma}_{ij} = \mathbf{0}$:

Thus, $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \boldsymbol{\Sigma}_{22} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix}$

Now the characteristic function of $\mathbf{y}$ is

$$\Phi(\mathbf{t}) = E(e^{it'\mathbf{y}}) = e^{it'\mu - \frac{1}{2}t'\Sigma t} = e^{it_1'\mu_1 - \frac{1}{2}t_1'\Sigma_{11}t_1} \cdots e^{it_k'\mu_k - \frac{1}{2}t_k'\Sigma_{kk}t_k}.$$

Thus, $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are mutually independent. This complete the proof.

So far we have not shown the existence of a random vector defined as in the definition of multivariate normal via linear functions. We shall now do this. First we shall prove a Lemma.

**Lemma 1:** Let $\mathbf{y}$ have a multivariate distribution with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$. Then $\mathbf{y} - \boldsymbol{\mu}$ belongs to the column space of $\boldsymbol{\Sigma}$ with probability 1.

**Proof:** We shall show that if a vector $\boldsymbol{l}$ is orthogonal to the columns of $\boldsymbol{\Sigma}$, then $\boldsymbol{l}^t(\mathbf{y} - \boldsymbol{\mu}) = 0$ with probability 1.

$\boldsymbol{l}^t\boldsymbol{\Sigma} = 0 \Rightarrow \boldsymbol{l}^t\boldsymbol{\Sigma}\boldsymbol{l} = 0 \Rightarrow V(\boldsymbol{l}^t(\mathbf{y} - \boldsymbol{\mu})) = 0 \Rightarrow \boldsymbol{l}^t(\mathbf{y} - \boldsymbol{\mu})$ is a constant with probability $1 \Rightarrow \boldsymbol{l}^t(\mathbf{y} - \boldsymbol{\mu}) = E(\boldsymbol{l}^t(\mathbf{y} - \boldsymbol{\mu})) = 0$ with probability 1.

**Theorem 14:** $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if there exists a random vector $\mathbf{x}$ of independent standard normal variables such that $\mathbf{y} = \boldsymbol{\mu} + \mathbf{Bx}$ with probability 1 for some matrix $\mathbf{B}$ of full column rank such that $\mathbf{BB}^t = \boldsymbol{\Sigma}$.

**Proof:** If part: Let $\boldsymbol{l}$ be a fixed vector. Then $\boldsymbol{l}^t\mathbf{y} = \boldsymbol{l}^t\boldsymbol{\mu} + \boldsymbol{l}^t\mathbf{Bx}$. Now $\boldsymbol{l}^t\mathbf{Bx}$ is a linear combination of independent standard normal variables and hence has a univariate normal distribution. Hence $\boldsymbol{l}^t\mathbf{y} = \boldsymbol{l}^t\boldsymbol{\mu} + \boldsymbol{l}^t\mathbf{Bx}$ has a univariate normal distribution. The choice of $\boldsymbol{l}$ being arbitrary, it follows that $\mathbf{y}$ has a multivariate normal distribution. Now $E(\mathbf{y}) = E(\boldsymbol{\mu} + \mathbf{Bx}) = \boldsymbol{\mu} + \mathbf{B}E(\mathbf{x}) = \boldsymbol{\mu}$, since $E(\mathbf{x}) = \mathbf{0}$

$$D(\mathbf{y}) = D(\boldsymbol{\mu} + \mathbf{Bx}) = D(\mathbf{Bx}) = \mathbf{BIB}^t = \boldsymbol{\Sigma}.$$

Hence $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

'Only if' part: Let the rank of $\boldsymbol{\Sigma}$ be equal to *r*. Let a spectral decomposition of $\boldsymbol{\Sigma}$ be

$$\boldsymbol{\Sigma} = (\mathbf{P}_1 : \mathbf{P}_2)\begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{P}_1^t \\ \mathbf{P}_2^t \end{pmatrix} = \mathbf{P}_1\boldsymbol{\Lambda}\mathbf{P}_1^t$$

where $\mathbf{P}_1$ is a matrix of order $P \times r$ such that $\mathbf{P}_1^t\mathbf{P}_1 = \mathbf{I}_r$ and $\boldsymbol{\Lambda}$ is a pd diagonal matrix of order $r \times r$. Thus, $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^t$ where $\mathbf{B} = \mathbf{P}_1\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{\Lambda}^{1/2}$ is the pd square root of $\boldsymbol{\Lambda}$. Notice that the rank of $\mathbf{B}$ is the same as rank of $\mathbf{P}_1$ which is equal to r since $\mathbf{P}_1^t\mathbf{P}_1 = \mathbf{I}_r$. Also $(\boldsymbol{\Lambda}^{1/2})\mathbf{P}_1^t\mathbf{P}_1\boldsymbol{\Lambda}^{1/2} = \mathbf{I}$.

Now write $\mathbf{x} = (\boldsymbol{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t(\mathbf{y} - \boldsymbol{\mu})$. Clearly $\mathbf{x}$ is a random vector of order $r \times 1$. $\mathbf{x}$ has an *r*-variate normal distribution. Now, $E(\mathbf{x}) = (\boldsymbol{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t E(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$.

Also
$$\begin{aligned} D(\mathbf{x}) &= (\boldsymbol{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t\boldsymbol{\Sigma}\mathbf{P}_1(\boldsymbol{\Lambda}^{1/2})^{-1} \\ &= (\boldsymbol{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t\mathbf{P}_1\boldsymbol{\Lambda}\mathbf{P}_1^t\mathbf{P}_1(\boldsymbol{\Lambda}^{1/2})^{-1} \\ &= (\boldsymbol{\Lambda}^{1/2})^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^{1/2})^{-1} = \mathbf{I}. \end{aligned}$$

Thus, we have manufactured a random vector $\mathbf{x} \sim N_r(\mathbf{0}, \mathbf{I})$ (i.e., $\mathbf{x}$ is a vector of r independent standard normal variables) given by

$$\mathbf{x} = (\boldsymbol{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t(\mathbf{y} - \boldsymbol{\mu})$$

or
$$\mathbf{P}_1\boldsymbol{\Lambda}^{1/2}\mathbf{x} = \mathbf{P}_1\mathbf{P}_1^t(\mathbf{y} - \boldsymbol{\mu})$$

By Lemma 1, $(\mathbf{y} - \boldsymbol{\mu})$ belongs to the column space of $\boldsymbol{\Sigma}$ (with probability 1) which is the same as the column space of $\mathbf{P}_1$.

Thus, $(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{P}_1\mathbf{v}$ for some $\mathbf{v}$ with probability 1.

Thus, $\mathbf{P}_1\boldsymbol{\Lambda}^{1/2}\mathbf{x} = \mathbf{P}_1\mathbf{P}_1^t\ \mathbf{P}_1\mathbf{v} = \mathbf{P}_1\mathbf{v} = \mathbf{y} - \boldsymbol{\mu}$ with probability 1

or $\mathbf{y} = \boldsymbol{\mu} + \mathbf{B}\mathbf{x}$ where $\mathbf{B} = \mathbf{P}_1\boldsymbol{\Lambda}^{1/2}$

Also $\mathbf{B}\mathbf{B}^t = \mathbf{P}_1\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{P}_1^t = \mathbf{P}_1\boldsymbol{\Lambda}\mathbf{P}_1^t\boldsymbol{\Sigma}$.

Thus, we have established the existence of a random vector $\mathbf{y}$ having *p*-variate normal distribution via linear functions.

**Example 4:** Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as defined via the linear functions. Let $\boldsymbol{\Sigma}$ be pd. Then show that the density of $\mathbf{y}$ is

$$f_y(\mathbf{v}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(v-\mu)^t\Sigma^{-1}(v-\mu)}$$

**Solution:** Since $\boldsymbol{\Sigma}$ is pd, by Theorem 14, there exists a random vector $\mathbf{x}$ of order $p \times 1$ which is a vector of independence $N(0, 1)$ variables such that $\mathbf{y} = \boldsymbol{\mu} + \mathbf{B}\mathbf{x}$ with probability 1. where $\mathbf{B}$ is a $p \times p$ matrix such that $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^t$.

Since $\boldsymbol{\Sigma}$ is nonsingular, it follows that $\mathbf{B}$ is nonsingular.

The density of $\mathbf{y}$ is

$$f_y(v) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)'\Sigma^{-1}(v-\mu)}$$

Thus, the two definitions of multivariate normal are equivalent if $\Sigma$ is pd.

Now, we shall summarize the unit.

## 16.4 SUMMARY

In this unit, we have covered the following points.

- Marginal and conditional distributions.
- Density of a nonsingular multivariate normal distribution.
- Marginal and conditional distributions in multivariate normal.
- Definition of multivariate normal via linear functions.
- Characteristic function of multivariate normal.

## 16.5 SOLUTIONS/ANSWERS

E1) (a) Clearly, $y_1$ and $y_3$ have a joint bivariate normal distribution. $E(y_1)=2, E(y_3)=1$, $V(y_1)=9, V(y_3)=6$ and $Cov(y_1, y_3)=2$. Thus

$$\begin{pmatrix} y_1 \\ y_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 9 & 2 \\ 2 & 6 \end{pmatrix}\right).$$

(b) $$Cov\left(\begin{pmatrix} y_1 \\ y_3 \end{pmatrix}, \begin{pmatrix} y_2 \\ y_4 \end{pmatrix}\right) = \begin{pmatrix} Cov(y_1, y_2) & Cov(y_1, y_4) \\ Cov(y_3, y_2) & Cov(y_3, y_4) \end{pmatrix} = 0$$

Since $y_1, y_3, y_2$ and $y_4$ have a joint (multivariate) normal distribution, uncorrelatedness implies independence as shown in Theorem 6.

(c) Partitioning $\mathbf{y}, \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as in Eqn. (1) where $r = 2$, we get

$$\mathbf{y}_1 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \mathbf{y}_2 = \begin{pmatrix} y_3 \\ y_4 \end{pmatrix}, \boldsymbol{\mu}_1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{11} = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}, \boldsymbol{\Sigma}_{22} = \begin{pmatrix} 6 & 0 \\ 0 & 9 \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_{12} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Along the lines of Theorem 8, we can show that the distribution of $\mathbf{y}_1 | \mathbf{y}_2 = \mathbf{v}_2$ is $N_2\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{v}_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$

Here $\mathbf{v}_2 = \begin{pmatrix} 1.2 \\ -2.6 \end{pmatrix}$. So

$$\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{v}_2 - \boldsymbol{\mu}_2) = \begin{pmatrix} 2 \\ 4 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{9} \end{pmatrix}\left(\begin{pmatrix} 1.2 \\ -2.6 \end{pmatrix} - \begin{pmatrix} 1 \\ -3 \end{pmatrix}\right)$$

$$= \begin{pmatrix} \frac{31}{15} \\ \frac{182}{42} \end{pmatrix} = \frac{1}{45}\begin{pmatrix} 93 \\ 182 \end{pmatrix}$$

$$\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} = \begin{pmatrix} \frac{25}{3} & 0 \\ 0 & 3 \end{pmatrix}$$

So, the conditional distribution of $\mathbf{y}_1 \mid \mathbf{y}_2 = \begin{pmatrix} 1.2 \\ -2.6 \end{pmatrix}$ is

$$N_2\left(\frac{1}{45}\begin{pmatrix} 93 \\ 182 \end{pmatrix}, \begin{pmatrix} \frac{25}{3} & 0 \\ 0 & 3 \end{pmatrix}\right).$$

(d) The correlation coefficient between $y_1$ and

$$y_3 = \frac{\text{Cov}(y_1, y_3)}{\sqrt{V(y_1)}\sqrt{V(y_3)}} = \frac{2}{3\sqrt{6}} = \frac{1}{3}\sqrt{\frac{2}{3}}$$

E2) (a) $\bar{\mathbf{y}} = \frac{1}{p}\mathbf{1}^t\mathbf{y}$ where $\mathbf{1}^t = (1, \ldots, 1)$

By Theorem 9(c), we have $\bar{\mathbf{y}} \sim N\left(\frac{1}{p}\mathbf{1}^t\boldsymbol{\mu}, \frac{1}{p^2}\mathbf{1}^t\mathbf{\Sigma}\mathbf{1}\right)$

Now $\mathbf{1}^t\boldsymbol{\mu} = (\mu_1 + \ldots + \mu_p)$. So $\frac{1}{p}\mathbf{1}^t\boldsymbol{\mu} = \bar{\mu}$.

$$\frac{1}{p^2}\mathbf{1}^t\mathbf{\Sigma}\mathbf{1} = \frac{1}{p^2}\sum_1^p\sum_1^p \sigma_{ij}.$$

(b) Since $\begin{pmatrix} \mathbf{x} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{C} \\ \mathbf{T} \end{pmatrix}\mathbf{y}$, by Theorem 9(b) (since $\mathbf{\Sigma}$ is pd) we have $\begin{pmatrix} \mathbf{x} \\ \mathbf{w} \end{pmatrix}$ have a joint multivariate normal distribution. So $\mathbf{x}$ and $\mathbf{w}$ have independent bivariate normal distributions if and only if Cov($\mathbf{x}$, $\mathbf{w}$) $= \mathbf{C\Sigma T}^t = 0$. Let us choose $\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. Then $\mathbf{x} = \mathbf{Cy} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$.

Also $\mathbf{C\Sigma} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & -2 & -1 \end{pmatrix}$. We need $\mathbf{C\Sigma T}^t = 0$, or in other words we need two vectors (non-null) each with 4 components which are orthogonal to both the rows of $\mathbf{C\Sigma}$. This can be achieved using Gram-Schmidt orthogonalization process (as described earlier) on

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 2 \\ -2 \\ -1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \mathbf{u}_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$ As a result, we get

$\mathbf{T} = \begin{pmatrix} 13 & -9 & -1 & -3 \\ 0 & 1 & 3 & -4 \end{pmatrix}$ so that $\mathbf{C\Sigma T}^t = 0$. Now $\mathbf{x} = \mathbf{Cy} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ and

$\mathbf{w} = \mathbf{Ty} = \begin{pmatrix} 13y_1 - 9y_2 - y_3 - 3y_4 \\ y_2 + 3y_3 - 4y_4 \end{pmatrix}$. So $\mathbf{x}$ and $\mathbf{w}$ being linear compounds of $\mathbf{y}$ have bivariate normal distributions.

$$E(\mathbf{x}) = \begin{pmatrix} E(y_1) \\ E(y_2) \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$D(\mathbf{x}) = D\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$E(\mathbf{w}) = E(\mathbf{Ty}) = \begin{pmatrix} 13.2 - 9.1 - 1.3 - 3(-4) \\ 1 + 3.3 - 4(-4) \end{pmatrix} = \begin{pmatrix} 26 \\ 26 \end{pmatrix}$$

$$D(\mathbf{w}) = \mathbf{T\Sigma T}^t = \begin{pmatrix} 69 & 144 \\ 144 & 374 \end{pmatrix}$$

Notice that $D(\mathbf{x})$ and $D(\mathbf{w})$ are nonsingular. Hence $\mathbf{x}$ and $\mathbf{w}$ have nonsingular bivariate normal distributions.

E3) Consider $\boldsymbol{l}^t\mathbf{By}$ where $\boldsymbol{l}$ is a fixed vector. Since it is a linear combination of components of $\mathbf{y}$,

$$\boldsymbol{l}^t\mathbf{By} \sim N(\boldsymbol{l}^t\mathbf{B}\boldsymbol{\mu}, \boldsymbol{l}^t\mathbf{B\Sigma B}^t\boldsymbol{l})$$

The choice of $\boldsymbol{l}$ being arbitrary, it follows that $\mathbf{By} \sim N(\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}^t)$ by definition.

–x–

# PRACTICAL ASSIGNMENT

## Session I

1. Write a program in C-language to find the lower triangular square root of a pd matrix. Also test your program on E6) of Unit 15.

2. Consider $y = (y_1, y_2, y_3)^t \sim N_3(\mu, \Sigma)$. Write a program in C-language to find the marginal distributions of $y_1$, $y_2$ and $y_3$. Also extend it to find the conditional distribution of $y_1$ given $y_2$ and $y_3$. Also test your program on Example 1 of Unit 16.