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 ith 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

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

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 σ_{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 λ_{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 $\qquad D(\mathbf{x}) = E\left\{ (\mathbf{x} - E(\mathbf{x}))\,(\mathbf{x} - E(\mathbf{x}))^t \right\} \qquad (1)$

It can be shown similarly that

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

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 ρ_{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

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

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

$$\begin{aligned}\sigma_{13} &= \mathrm{Cov}(x_1, X_3) = \rho_{13}\sqrt{V(x_1).V(x_3)} \\ &= 0.3\times 2\times 4 = 2.4\end{aligned}$$

$$\begin{aligned}\sigma_{23} &= \mathrm{Cov}(x_2, x_3) = \rho_{23}\sqrt{V(x_2).\,V(x_3)} \\ &= -0.4\times 3\times 4 = -4.8\end{aligned}$$

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

Notice that $\mathbf{\Sigma}$ is symmetric in the above example. In fact, this is true for every variance-covariance matrix $\mathbf{\Sigma}$ because $\sigma_{ij} = \mathrm{Cov}(x_i, x_j) = \mathrm{Cov}(x_j, x_i) = \sigma_{ji}$ 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 \ 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 $E(\mathbf{x}) = \mu$ 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 $E(\mathbf{l}^t\mathbf{x}) = \mathbf{l}^t\ E(\mathbf{x}) = \mathbf{l}^t\mu$ and $V(\mathbf{l}^t\mathbf{x}) = \mathbf{l}^t\mathbf{\Sigma}\mathbf{l}$. Also $\mathrm{Cov}(\mathbf{l}^t\mathbf{x}, \mathbf{m}^t\mathbf{x}) = \mathbf{l}^t\mathbf{\Sigma}\mathbf{m}$ where $\mathbf{m}$ is a fixed vector.

Proof:
$$\begin{aligned}E(\mathbf{l}^t\mathbf{x}) &= E(l_1x_1 + l_2x_2 + \cdots + l_px_p) \\ &= l_1E(x_1) + l_2E(x_2) + \cdots + l_pE(x_p) \\ &= \mathbf{l}^tE(\mathbf{x}) = \mathbf{l}^t\mu\end{aligned}$$

$$V(\mathbf{l^t x}) = V(l_1x_1 + l_2x_2 + \cdots + l_px_p)$$

$$= Cov(l_1x_1 + l_2x_2 + \cdots l_px_p,\ l_1x_1 + l_2x_2 + \cdots + l_px_p)$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} l_i,\ l_j\sigma_{ij} = \mathbf{l^t\Sigma l}$$

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

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

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

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

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

$$= \frac{1}{3} \times -0.3 = -0.1$$

$$V(\mathbf{l^t x}) = \mathbf{l^t}\sum \mathbf{l} \qquad \text{(Usingh Theorem 1)}$$

$$= \frac{1}{3}(1\,1\,1)\begin{pmatrix}4.0 & 3.0 & 2.4\\ 3.0 & 9.0 & -4.8\\ 2.4 & -4.8 & 16\end{pmatrix}\frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$= \frac{1}{9}(9.4 \quad 7.2 \quad 13.6)\begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$= \frac{1}{9} \times 30.2 = 3.36$$

$$Cov(\mathbf{l^t x, m^t x}) = \mathbf{l^t}\sum \mathbf{m} \qquad \text{(using Theorem 1)}$$

$$= \frac{1}{3}(1\,1\,1)\begin{pmatrix}4.0 & 3.0 & 2.4\\ 3.0 & 9.0 & -4.8\\ 2.4 & -4.8 & 16\end{pmatrix}\begin{pmatrix}2\\-1\\-1\end{pmatrix}$$

$$= \frac{1}{3}(9.4 \quad 7.2 \quad 13.6)\begin{pmatrix}2\\-1\\-1\end{pmatrix} = \frac{1}{3} \times -2.0 = \frac{-2}{3}.$$

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 $\mathbf{B}$, $i = 1, \ldots, r$.

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

(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.

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

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

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

Example 3: Let Σ 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 $r_{ij} = \dfrac{\text{Cov}(x_i, x_j)}{\sqrt{V(x_i)V(x_j)}} = \dfrac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$

The $(i, j)^{th}$ element of $\mathbf{T}^{-1}\Sigma\mathbf{T}^{-1}$ is
$$\begin{pmatrix} 0 & \ldots & \sigma_{ii}^{-\frac{1}{2}} & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \sigma_{i1} & \ddots & \sigma_{ip} \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ \sigma_{jj}^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

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

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

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 Σ 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 Σ is nnd.

(b) Let Σ_{pxp} 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 Σ 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: Σ 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 Σ where $\mathbf{P}$ is orthogonal and Λ diagonal. Since Σ is pd, so is Λ. Write $\xi = \mathbf{P}^t\mathbf{x}$. Then $D(\xi) = \mathbf{P}^t\boldsymbol{\Sigma}\mathbf{P} = \mathbf{P}^t\mathbf{P}\Lambda\mathbf{P}^t\mathbf{P} = \Lambda$, since $\mathbf{P}$ is orthogonal. Since Λ is a diagonal matrix, the components of ξ are uncorrelated. Also $V(\xi_i) = \lambda_i$, the i^{th} diagonal element of Λ.

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

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

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

Before proceeding further, let us recall that if Σ 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 Σ 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 $\mathbf{T}$) $-\frac{t_{ji}}{t_{ii}}$ (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 Σ 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 $\mathbf{B}^{-1}$.

So $\mathbf{y}=\mathbf{y}=\begin{pmatrix} b^{11} & 0 & \cdots & 0 \\ b^{21} & b^{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b^{p1} & 0 & 0 & b^{pp} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$, 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 ξ are uncorrelated. Obtain the variances of ξ_1 and ξ_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.

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

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 $\mathbf{u} = \mathbf{By}$ 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) $\text{Cov}(\mathbf{u}, \mathbf{v})$.

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 Σ 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

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

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

The marginal probability mass function of x_1 is given by

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

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

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

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

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

$j = 0, 1, 2, 3$

$= 0$, 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 $x_1 \mid x_2 = 0$

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:

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

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

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

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

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

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

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

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

Thus, $f_{x_1|x_2=u_2}(\mathbf{u}_1 \mid \mathbf{u}_2), du_1, \ldots, du_p$ 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, $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_x(u_1, u_2) du_1, du_2 = 1$

Now $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_x(u_1, u_2) du_1 du_2 = \int_0^1\int_0^1 c.(2-u_1-u_2) du_1 du_2$

$$= c.\int_0^1 (2-u_1)du_1 - \int_0^1 u_2 du_2 \quad \text{since} \quad \int_0^1 du_1 = \int_0^1 du_2 = 1$$

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

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

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

$$f_{x_1}(u_1) = \int_0^1 f_x(u_1, u_2) du_2$$

$$= \int_0^1 (2-u_1-u_2) du_2$$

$$= (2-u_1) - \int_0^1 u_2 du_2$$

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

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 $-\infty < x_2 \leq 0$. Let $0 < x_2 < \infty$, then

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

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

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

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 $0 < u_1 \leq \infty$.

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 $\text{Cov}(\mathbf{x}_1, \mathbf{x}_2) = 0$.

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

$$\begin{aligned}
\text{Cov}(\mathbf{u}_1, \mathbf{u}_2) &= E(\mathbf{u}_1 \mathbf{u}_2^t) - E(\mathbf{u}_1)E(\mathbf{u}_2^t) \\
&= \int \ldots \int \mathbf{u}_1 \mathbf{u}_2^t f_{x_1}(\mathbf{u}_1) f_{x_2}(\mathbf{u}_2) du_1 \ldots du_p \\
&\quad - \left[\int \ldots \int \mathbf{u}_1 f_{x_1}(\mathbf{u}_1) du_1 \ldots du_r\right]\left[\int \ldots \int \mathbf{u}_2^t f_{x_2}(\mathbf{u}_2) du_{r+1} \ldots du_p\right] = 0
\end{aligned}$$

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

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

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

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

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

(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 $\sigma_{11} = \sigma_{x_1}^2$, $\sigma_{22} = \sigma_{x_2}^2$ and $\sigma_{12} = \rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2}$.

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

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

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

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

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

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

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 Σ 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 $\quad b_{11}=\sqrt{\sigma_{11}}$

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

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

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

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

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

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

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

Hence $(\mathbf{u} - \mu_x)^t \Sigma^{-1} (\mathbf{u} - \mu_x)$

$= (\mathbf{u} - \mu_x)^t \mathbf{B}^{-1^t}\mathbf{B}^{-1}(\mathbf{u} - \mu_x)$

$= (\mathbf{v} - \mu_y)^t (\mathbf{v} - \mu_y)$

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

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 $\mathbf{y} \sim N_2(\mu_y, \mathbf{I})$

(**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

$$\frac{1}{2\pi}e^{-\frac{1}{2}\left(v_1-\mu_{y_1}\right)^2}.e^{-\frac{1}{2}\left(v_2-\mu_{y2}\right)^2}=h_1(v_1).h_2(v_2) \text{ for all } v_1,\ v_2 \text{ where}$$

$$h_1(v_1)=\frac{1}{2\pi}e^{-\frac{1}{2}\left(v_1-\mu_{y1}\right)^2},\ -\infty<v_1<\infty \text{ and } h_2(v_2)=\frac{1}{2\pi}e^{-\frac{1}{2}\left(v_2-\mu_{y2}\right)^2},\ -\infty<v_2<\infty.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\text{Cov}(\mathbf{1^t x, m^t x}) = \mathbf{1^t \Sigma m} = 0.8\mathbf{1^t m} + 0.2\mathbf{1^t} ll^t \mathbf{m} = 0$$
$$= 0.8 \times 0 + 0.2 \times 4 \times 0 - 0$$

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

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

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

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

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

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

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

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

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

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

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

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

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 $\mathrm{Cov}(x_i, x_j) = 0$ for all j.

$\mathbf{\Sigma}$ 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 $\mathrm{Cov}(x_i, x_j) = \sigma_{ij} = 0$ 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.

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

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

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

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

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

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

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

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 $x_3 \mid x_2 = 1$ is obtained using the row corresponding to $x_2 = 1$ in the joint probability table as follows:

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

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

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

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

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

Conditional Distribution of $x_3 \mid x_2 = 1$

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 $i+j+k=3$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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–