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, Σ. 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 Σ. (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:

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

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

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

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 $f_y(v) = \dfrac{1}{\sqrt{2\pi}|\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)^t \Sigma^{-1}(v-\mu)}, v \in R^p$.

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

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, μ and Σ are the mean vector and the variance-covariance matrix of $y \sim N_p(\mu, \Sigma)$. Recall that we started with B nonsingular and hence Σ 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 μ and variance-covariance matrix Σ is called a nonsingular *p*-variate normal distribution, if Σ is positive definite. Later on, we shall also study the case where Σ 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 μ and Σ 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 μ is a fixed vector in $\mathbf{R}^p$ and Σ 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 Σ 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 μ_1 are of order $r \times 1 (1 \le r \le p)$ and Σ_{11} and Σ_{22} are of order $r \times r$ and $(p-r) \times (p-r)$, respectively. Such partitions of $\mathbf{v}, \mu$ and Σ 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 Σ is positive definite. Let $\mathbf{y}, \mathbf{v}, \mu$ and Σ 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 $f_y(\mathbf{v}) = \dfrac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)' \Sigma^{-1}(v-\mu)}$.

Since $\mathbf{\Sigma}_{12} = \mathbf{0}$, $|\mathbf{\Sigma}| = \left| \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix} \right| = |\mathbf{\Sigma}_{11}||\mathbf{\Sigma}_{22}|$.

Also $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix}$

Hence $(\mathbf{v}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{v}-\boldsymbol{\mu}) = \left(\mathbf{v}_1^t - \boldsymbol{\mu}_1^t : \mathbf{v}_2^t - \boldsymbol{\mu}_2^t\right)\begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{v}_1 - \boldsymbol{\mu}_1 \\ \mathbf{v}_2 - \boldsymbol{\mu}_2 \end{pmatrix}$

$$= (\mathbf{v}_1 - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{v}_1 - \boldsymbol{\mu}_1) + (\mathbf{v}_2 - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{v}_2 - \boldsymbol{\mu}_2).$$

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 $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where Σ is positive definite. Let $\mathbf{y}$, $\mathbf{v}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be partitioned as in Eqn. (1). Then the marginal distribution of $\mathbf{y}_1$ is $N_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

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

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^t & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{\Sigma}_{12}^t\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

Hence $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}$

Also $|\boldsymbol{\Sigma}| = \begin{vmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{vmatrix}\begin{vmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{vmatrix}\begin{vmatrix} \mathbf{I} & \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{0} & \mathbf{I} \end{vmatrix}$

$$= |\boldsymbol{\Sigma}_{11}|.|\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|$$

since $\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{B} & \mathbf{C} \end{vmatrix} = |\mathbf{A}|.|\mathbf{C}|$, where $\mathbf{A}$ and $\mathbf{C}$ are squares and $|\mathbf{I}| = 1$

$$\text{Now } f_y(\mathbf{v}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(v-\mu)'\Sigma^{-1}(v-\mu)} \, .$$

$$= \frac{e^{-\frac{1}{2}\left((v_1-\mu_1)^t : (v_2-\mu_2)^t\right)\begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ O & I \end{pmatrix}\begin{pmatrix} \Sigma_{11}^{-1} & O \\ O & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}\begin{pmatrix} I & O \\ -\Sigma_{12}^t\Sigma_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} v_1-\mu_1 \\ v_2-\mu_2 \end{pmatrix}}}{(2\pi)^{p/2}|\Sigma_{11}|^{1/2}|\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}$$

$$= \frac{e^{-\frac{1}{2}\left\{(\omega_1^t : \omega_2^t)\begin{pmatrix} \Sigma_{11}^{-1} & O \\ O & (\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}\right\}}}{(2\pi)^{p/2}|\Sigma_{11}|^{1/2}|\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}$$

where $\boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{v}_1-\boldsymbol{\mu}_1 \\ \mathbf{v}_2-\boldsymbol{\mu}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1-\boldsymbol{\mu}_1 \\ \mathbf{v}_2-\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{12}^t\boldsymbol{\Sigma}_{11}(\mathbf{v}_1-\boldsymbol{\mu}_1) \end{pmatrix}$

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 $\boldsymbol{\xi} = \boldsymbol{\xi} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{12}^t\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}(\mathbf{y}-\boldsymbol{\mu})$ is

$$\mathbf{f}_\xi(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{r/2}|\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(v_1-\mu_1)^t\Sigma_{11}^{-1}(v_1-\mu_1)} \times \frac{e^{-\frac{1}{2}\left(v_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(v_1-\mu_1)\right)^t\left(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\left(v_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(v_1-\mu_1)\right)}}{(2\pi)^{p-r/2}|\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}}$$

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 Σ 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 $\mathbf{y}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{y}_1$ and y_1 are independently distributed. Hence the conditional distribution of $\mathbf{y}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{y}_1$ given $\mathbf{y}_1$ is the same as the unconditional distribution of $\mathbf{y}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{y}_1$, which is $N_{p-r}\left(\mathbf{\mu}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\mu}_1,\ \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\right)$. 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

$$\mathbf{\mu}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\mu}_1 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{v}_1$$

since when $y_1 = v_1$, $\mathbf{y}_1 = \mathbf{v}_1$, $\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{y}_1 = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{v}_1$ is a fixed vector.

Also the conditional variance-covariance matrix is the same as that of $\mathbf{y}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{y}_1$ which is $\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$.

Thus, the conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1$ is
$N_{p-r}\left(\mathbf{\mu}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{v}_1 - \mathbf{\mu}_1),\ \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\right)$. This complete the proof.

Let $r = p-1$. Then $\mathbf{y}_2$ is univariate random variable $\mathbf{y}_p$. Also Σ_{21} is a row vector of order $1\times(p-1)$ and Σ_{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

$$\begin{aligned} E(y_p \mid v_1, \ldots, v_{p-1}) &= \mu_p + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{v}_1 - \mathbf{\mu}_1) \\ &= \beta_0 + \beta_1 v_1 + \cdots + \beta_{p-1}v_{p-1} \end{aligned}$$

where $\beta_0 = \mu_p - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\mu}_1$ 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 $\mathbf{\mu} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ and

$$\mathbf{\Sigma} = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix}.$$

(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 $\mathbf{\xi} + \mathbf{Ty} + \mathbf{c}$ so that ξ_1, ξ_2 and ξ_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 μ and Σ, 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 1×2 and $\boldsymbol{\Sigma}_{22}$ is of order 2×2.

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 ξ 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 Σ 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 $\mathbf{y} = \mathbf{B}^{-1}\mathbf{z}$, $\mathbf{v} = \mathbf{B}^{-1}\mathrm{u}$ 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 $f_z(\mathbf{u}) = \dfrac{1}{(2\pi)^{p/2}|\mathbf{B}\,\Sigma\,\mathbf{B}^t|^{1/2}} e^{-\frac{1}{2}(u-\theta)^t(\mathbf{B}\Sigma\mathbf{B}^t)^{-1}(u-\theta)}$

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 $N_p(\mathbf{B\mu}, \mathbf{B\Sigma B}^t)$.

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 $E(\mathbf{Cy}) = \mathbf{C\mu}$ and $D(\mathbf{Cy}) = \mathbf{C\Sigma C}^t$, it follows that

$$\mathbf{Cy} = N_r(\mathbf{C\mu}, \mathbf{C\Sigma C}^t).$$

If $\mathbf{\Sigma}$ 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 Σ 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 $\mathbf{y} \sim N_3(\mathbf{\mu}, \mathbf{\Sigma})$ where $\mathbf{\mu} = \begin{pmatrix} 4 \\ -2 \\ 6 \end{pmatrix}$ and $\mathbf{\Sigma} = \begin{pmatrix} 6 & 1 & 2 \\ 1 & 8 & 4 \\ 2 & 4 & 9 \end{pmatrix}$

(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 $\xi = \mathbf{l'y}$ of $\mathbf{y}$ such that ξ has a standard normal distribution.

Solution: **(a)** $\mathbf{C}$ is a 2×3 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 $\mathbf{x} = \mathbf{Cy} \sim N_2(\mathbf{C\mu}, \mathbf{C\Sigma C}^t)$.

Now $\mathbf{C\mu} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 5 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ -2 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\mathbf{C\Sigma C}^t = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 5 & 1 \end{pmatrix} \begin{pmatrix} 6 & 1 & 2 \\ 1 & 8 & 4 \\ 2 & 4 & 9 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 5 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 25 & 63 \\ 63 & 269 \end{pmatrix}$$

(b) Let $\mathbf{m} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$. By (c) of Theorem 5 and part (a) given above, we have

$\mathbf{m'y} =\sim N(\mathbf{m'\mu}, \mathbf{m'\Sigma m})$, where $\mathbf{m'\mu} = 0$ 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 $\xi = \mathbf{l'y} \sim N(0, 1)$.

Now, try the following exercise.

E2) (a) Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\bar{y} = \frac{1}{p}(y_1 + \cdots + y_p)$. 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 $\mathbf{x} = \mathbf{Cy}$ and $\mathbf{w} = \mathbf{Ty}$, where $\mathbf{x}$ and $\mathbf{w}$ are vectors of order 2×1 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 $p \times 1$, $\Phi_y(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{y}})$ 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 $\Phi(\mathbf{t})$ of a random vector $\mathbf{y}$ uniquely determines its distribution.

(b) If $\Phi(\mathbf{t}) = \psi_1(t_1) \ldots \psi_p(t_p)$, 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

$$\Phi(\mathbf{t}) = \psi_1(\mathbf{t}_1) \ldots \psi_k(\mathbf{t}_k)$$

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 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

$\Phi(t) = E(e^{it'y}) = \Phi(\mathbf{t}) = E(e^{it^ty}) = e^{it^t\mu - \frac{1}{2}t^t\Sigma t}$

Proof: Recall that the characteristic function of univariate random variable ξ having a normal distribution with mean θ and variance σ^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 $\Psi(s, t) = E(e^{ist'y}) =$

$\psi(s, \mathbf{t}) = E(e^{ist'y}) = e^{ist'\mu - \frac{1}{2}s^2t'\Sigma t}$, for every $\mathbf{t}$.

Now $\Phi(\mathbf{t}) = E(e^{it^ty}) = \Phi(1, \mathbf{t}) = e^{it'\mu - \frac{1}{2}t'\Sigma t}$.

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

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_{kk} \end{pmatrix}.$$

For $i \neq j$, $\begin{pmatrix} \mathbf{y}_i \\ \mathbf{y}_j \end{pmatrix}$ has a multivariate normal distribution with mean vector $\begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}$ and variance-covariance matrix $\begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj} \end{pmatrix}$ (why?). If $\mathbf{y}_i$ and $\mathbf{y}_j$ are independent, then $\Sigma_{ij} = \mathbf{0}$.

$$\text{Thus, } \Sigma = \begin{pmatrix} \Sigma_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \Sigma_{22} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \Sigma_{kk} \end{pmatrix}$$

Now the characteristic function of $\mathbf{y}$ is

$$\Phi(\mathbf{t}) = E(e^{it^t y}) = e^{it^t\mu - \frac{1}{2}t^t\Sigma t} = e^{it_1^t\mu_1 - \frac{1}{2}t_1^t\Sigma_{11}t_1} \cdots e^{it_k^t\mu_k - \frac{1}{2}t_k^t\Sigma_{kk}t_k}.$$

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 μ and variance covariance matrix Σ. Then $\mathbf{y} - \mu$ belongs to the column space of Σ with probability 1.

Proof: We shall show that if a vector $\boldsymbol{l}$ is orthogonal to the columns of Σ, then $\boldsymbol{l}^t(\mathbf{y} - \mu) = 0$ with probability 1.

$\boldsymbol{l}^t\Sigma = 0 \Rightarrow \boldsymbol{l}^t\Sigma\boldsymbol{l} = 0 \Rightarrow V(\boldsymbol{l}^t(\mathbf{y} - \mu)) = 0 \Rightarrow \boldsymbol{l}^t(\mathbf{y} - \mu)$ is a constant with probability $1 \Rightarrow \boldsymbol{l}^t(\mathbf{y} - \mu) = E(\boldsymbol{l}^t(\mathbf{y} - \mu)) = 0$ with probability 1.

Theorem 14: $\mathbf{y} \sim N_p(\mu, \Sigma)$ if and only if there exists a random vector $\mathbf{x}$ of independent standard normal variables such that $\mathbf{y} = \mu + \mathbf{B}\mathbf{x}$ with probability 1 for some matrix $\mathbf{B}$ of full column rank such that $\mathbf{B}\mathbf{B}^t = \Sigma$.

Proof: If part: Let $\boldsymbol{l}$ be a fixed vector. Then $\boldsymbol{l}^t\mathbf{y} = \boldsymbol{l}^t\mu + \boldsymbol{l}^t\mathbf{B}\mathbf{x}$. Now $\boldsymbol{l}^t\mathbf{B}\mathbf{x}$ is a linear combination of independent standard normal variables and hence has a univariate normal distribution. Hence $\boldsymbol{l}^t\mathbf{y} = \boldsymbol{l}^t\mu + \boldsymbol{l}^t\mathbf{B}\mathbf{x}$ has a univariate normal distribution. The choice of $\boldsymbol{l}$ being arbitrary, it follows that $\mathbf{y}$ has a multivariate normal distribution. Now $E(\mathbf{y}) = E(\mu + \mathbf{B}\mathbf{x}) = \mu + \mathbf{B}E(\mathbf{x}) = \mu$, since $E(\mathbf{x}) = \mathbf{0}$

$$D(\mathbf{y}) = D(\mu + \mathbf{B}\mathbf{x}) = D(\mathbf{B}\mathbf{x}) = \mathbf{B}\mathbf{I}\mathbf{B}^t = \Sigma.$$

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

'Only if' part: Let the rank of Σ be equal to r. Let a spectral decomposition of Σ be

$$\mathbf{\Sigma} = (\mathbf{P}_1 : \mathbf{P}_2)\begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{P}_1^t \\ \mathbf{P}_2^t \end{pmatrix} = \mathbf{P}_1\mathbf{\Lambda}\mathbf{P}_1^t$$

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 $\mathbf{\Lambda}$ is a pd diagonal matrix of order $r \times r$. Thus, $\mathbf{\Sigma} = \mathbf{BB}^t$ where $\mathbf{B} = \mathbf{P}_1\mathbf{\Lambda}^{1/2}$ and $\mathbf{\Lambda}^{1/2}$ is the pd square root of $\mathbf{\Lambda}$. 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 $(\mathbf{\Lambda}^{1/2})\mathbf{P}_1^t\mathbf{P}_1\mathbf{\Lambda}^{1/2} = \mathbf{I}$.

Now write $\mathbf{x} = (\mathbf{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t(\mathbf{y} - \boldsymbol{\mu})$. Clearly $\mathbf{x}$ is a random vector of order $r \times 1$. $\mathbf{x}$ has an *r*-variate normal distribution. Now, $E(\mathbf{x}) = (\mathbf{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t E(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$.

Also
$$\begin{aligned} D(\mathbf{x}) &= (\mathbf{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t\mathbf{\Sigma}\mathbf{P}_1(\mathbf{\Lambda}^{1/2})^{-1} \\ &= (\mathbf{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t\mathbf{P}_1\mathbf{\Lambda}\mathbf{P}_1^t\mathbf{P}_1(\mathbf{\Lambda}^{1/2})^{-1} \\ &= (\mathbf{\Lambda}^{1/2})^{-1}\mathbf{\Lambda}(\mathbf{\Lambda}^{1/2})^{-1} = \mathbf{I}. \end{aligned}$$

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

$$\mathbf{x} = (\mathbf{\Lambda}^{1/2})^{-1}\mathbf{P}_1^t(\mathbf{y} - \boldsymbol{\mu})$$
or
$$\mathbf{P}_1\mathbf{\Lambda}^{1/2}\mathbf{x} = \mathbf{P}_1\mathbf{P}_1^t(\mathbf{y} - \boldsymbol{\mu})$$

By Lemma 1, $(\mathbf{y} - \boldsymbol{\mu})$ belongs to the column space of $\mathbf{\Sigma}$ (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, $\mathbf{P}_1\mathbf{\Lambda}^{1/2}\mathbf{x} = \mathbf{P}_1\mathbf{P}_1^t\ \mathbf{P}_1\mathbf{v} = \mathbf{P}_1\mathbf{v} = \mathbf{y} - \boldsymbol{\mu}$ with probability 1

or $\mathbf{y} = \boldsymbol{\mu} + \mathbf{Bx}$ where $\mathbf{B} = \mathbf{P}_1\mathbf{\Lambda}^{1/2}$

Also $\mathbf{BB}^t = \mathbf{P}_1\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}_1^t = \mathbf{P}_1\mathbf{\Lambda}\mathbf{P}_1^t\mathbf{\Sigma}$.

Thus, we have established the existence of a random vector $\mathbf{y}$ having *p*-variate normal distribution via linear functions.

Example 4: Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ as defined via the linear functions. Let $\mathbf{\Sigma}$ be pd. Then show that the density of $\mathbf{y}$ is

$$f_y(\mathbf{v}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}}e^{-\frac{1}{2}(v-\mu)^t\Sigma^{-1}(v-\mu)}$$

Solution: Since $\mathbf{\Sigma}$ 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 $\mathbf{y} = \boldsymbol{\mu} + \mathbf{Bx}$ with probability 1. where $\mathbf{B}$ is a $p \times p$ matrix such that $\mathbf{\Sigma} = \mathbf{BB}^t$.

Since $\mathbf{\Sigma}$ 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 $\boldsymbol{\Sigma}$ 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

$$y_3 = \frac{\mathrm{Cov}(y_1, y_3)}{\sqrt{V(y_1)}\sqrt{V(y_3)}} = \frac{2}{3\sqrt{6}} = \frac{1}{3}\sqrt{\frac{2}{3}}$$

E2) (a) $\bar{y} = \frac{1}{p}\mathbf{1}^t\mathbf{y}$ where $\mathbf{1}^t = (1, \ldots, 1)$

By Theorem 9(c), we have $\bar{y} \sim N\left(\frac{1}{p}\mathbf{1}^t\mu, \frac{1}{p^2}\mathbf{1}^t\mathbf{\Sigma}\mathbf{1}\right)$

Now $\mathbf{1}^t\mu = (\mu_1 + \ldots + \mu_p)$. So $\frac{1}{p}\mathbf{1}^t\mu = \bar{\mu}$.

$$\frac{1}{p^2}\mathbf{1}^t\mathbf{\Sigma}\mathbf{1} = \frac{1}{p^2}\sum_1^p\sum_1^p \sigma_{ij}.$$

(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 $\mathrm{Cov}(\mathbf{x}, \mathbf{w}) = \mathbf{C\Sigma T}^t = 0$. 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.