CHAPTER

# 5 some special multivariate distributions

## 5.1 INTRODUCTION

In this chapter we shall present some of the basic properties of six special multivariate distributions that will appear in later chapters of this book. These distributions are the multinomial, the Dirichlet, the multivariate normal, the Wishart, the multivariate $t$, and what is called here the bilateral bivariate Pareto. The format of this chapter is essentially the same as that of the preceding chapter on univariate distributions. However, since it is anticipated that many readers will not be sufficiently familiar with some of the multivariate distributions, many properties of these distributions will be given here in relatively great detail.

## 5.2 THE MULTINOMIAL DISTRIBUTION

Consider an experiment whose outcome must belong to one of $k$ ($k \geq 2$) mutually exclusive and exhaustive categories, and let $p_i$ ($0 < p_i < 1$) be the probability that the outcome belongs to the $i$th category ($i = 1, \ldots, k$). Here $\Sigma_{i=1}^k p_i = 1$. Suppose that the experiment is performed $n$ times and that the $n$ outcomes are independent. Furthermore, let $X_i$ denote the number of these outcomes that belong to category $i$ ($i = 1, \ldots, k$). Then the random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a *multinomial distribution with parameters $n$ and $\mathbf{p} = (p_1, \ldots, p_k)'$*. Let $\mathbf{x} = (x_1, \ldots, x_k)'$ be any point in $R^k$ each of whose components $x_i$ is a nonnegative integer ($i = 1, \ldots, k$) such that $\Sigma_{i=1}^k x_i = n$. Then at the point $\mathbf{x}$, the value of the p.f. $f(\cdot|n, \mathbf{p})$ of $\mathbf{X}$ is specified by the equation





$$f(\mathbf{x}|n, \mathbf{p}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}. \tag{1}$$

It follows from the definition of the random vector $\mathbf{X}$ that $f(\mathbf{x}|n, \mathbf{p}) = 0$ at any other point $\mathbf{x} \in R^k$.

Since the probability that $\Sigma_{i=1}^k X_i = n$ is 1, one of the $k$ random variables $X_1, \ldots, X_k$ can be eliminated, and the multinomial p.f. defined by Eq. (1) can be written as a $(k - 1)$-dimensional distribution. However, this reduction introduces an asymmetry among the $k$ categories that was not originally present. Indeed, a reduction similar to this was made when the binomial distribution was defined in Sec. 4.3. Hence, if $X$ has a binomial distribution with parameters $n$ and $p$, then the two-dimensional random vector $(X, n - X)'$ has a multinomial distribution with parameters $n$ and $(p, 1 - p)'$.

If a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a multinomial distribution with parameters $n$ and $\mathbf{p}$, then the mean vector of $\mathbf{X}$ is $E(\mathbf{X}) = n\mathbf{p}$ and the elements of the covariance matrix of $\mathbf{X}$ are as follows (see Exercise 2):

$$\begin{aligned} \operatorname{Var}(X_i) &= np_i(1 - p_i) \qquad & i &= 1, \ldots, k, \\ \operatorname{Cov}(X_i, X_j) &= -np_ip_j & i, j &= 1, \ldots, k;\ i \neq j. \end{aligned} \tag{2}$$

The following properties result from the description of the multinomial distribution given at the beginning of this section. If $\mathbf{X}$ has the multinomial distribution defined by Eq. (1), then the marginal distribution of any component $X_i$ is binomial with parameters $n$ and $p_i$ ($i = 1, \ldots, k$). Also, if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_r$ are independent $k$-dimensional random vectors and if $\mathbf{X}_i$ has a multinomial distribution with parameters $n_i$ and $\mathbf{p}$ ($i = 1, \ldots, r$), then the sum $\mathbf{X}_1 + \cdots + \mathbf{X}_r$ has a multinomial distribution with parameters $n_1 + \cdots + n_r$ and $\mathbf{p}$ (see Exercise 3).

## 5.3 THE DIRICHLET DISTRIBUTION

The conventions which we shall use in introducing the p.d.f. of the Dirichlet distribution are slightly different from those which we have been using. The appropriate explanation will be given after Eq. (1) has been presented.

A random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a *Dirichlet distribution with parametric vector* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$ ($\alpha_i > 0$; $i = 1, \ldots, k$) if the p.d.f. $f(\cdot\,|\boldsymbol{\alpha})$ of $\mathbf{X}$ satisfies the following properties: Let $\mathbf{x} = (x_1, \ldots, x_k)'$ be any point in $R^k$ such that $x_i > 0$ for $i = 1, \ldots, k$ and $\Sigma_{i=1}^k x_i = 1$. Then

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} x_1^{\alpha_1 - 1} \cdots x_k^{\alpha_k - 1}. \tag{1}$$

Also, $f(\mathbf{x}|\boldsymbol{\alpha}) = 0$ at any other point $\mathbf{x} \in R^k$.

Equation (1) requires further explanation. The function $f(\cdot\,|\boldsymbol{\alpha})$ is positive only on the $(k-1)$-dimensional simplex of points $\mathbf{x} = (x_1, \ldots, x_k)'$ such that $x_i > 0$ for $i = 1, \ldots, k$ and $\Sigma_{i=1}^k x_i = 1$. Therefore,

$$\Pr\Big(\sum_{i=1}^{k} X_i = 1\Big) = 1. \tag{2}$$

It follows from this linear relation that the $k$ random variables $X_1, \ldots, X_k$ cannot have a joint $k$-dimensional p.d.f. Thus, despite its appearance, $f(\cdot\,|\boldsymbol{\alpha})$ is not a $k$-dimensional p.d.f. Rather, it gives the joint p.d.f. of any subcollection of $k - 1$ of the random variables $X_1, \ldots, X_k$ after the other one, say $X_j$, has been eliminated through use of the relation $X_1 + \cdots + X_k = 1$. This joint $(k-1)$-dimensional p.d.f. is obtained by eliminating $x_j$ from the right side of Eq. (1) through use of the relation $x_1 + \cdots + x_k = 1$. The resulting p.d.f. is positive at those points in $R^{k-1}$ such that each of the $k - 1$ coordinates is positive and their sum is less than 1.

A comparison of the expression in Eq. (1) with the p.d.f. of the beta distribution in Eq. (1) of Sec. 4.9 reveals why the Dirichlet distribution is sometimes referred to as a multivariate beta distribution. In fact, if a random variable $X$ has a beta distribution with parameters $\alpha$ and $\beta$, then the random vector $\mathbf{Y} = (X, 1 - X)'$ has a Dirichlet distribution with parametric vector $(\alpha, \beta)'$. Furthermore, if a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a Dirichlet distribution with parametric vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, then the marginal distribution of one component of $\mathbf{X}$, say $X_j$, is a beta distribution with parameters $\alpha_j$ and $\Sigma_{i=1}^k \alpha_i - \alpha_j$.

A more general result is the following: Suppose that $\mathbf{X} = (X_1, \ldots, X_k)'$ has a Dirichlet distribution as above and that the indices $1, \ldots, k$ are partitioned into $m$ groups $J_1, \ldots, J_m$ $(2 \le m \le k)$ which are nonempty, mutually exclusive, and exhaustive. Also, let $Y_i = \Sigma_{j \in J_i} X_j$ $(i = 1, \ldots, m)$. Then the $m$-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ has a Dirichlet distribution with parametric vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$, where $\beta_i = \Sigma_{j \in J_i} \alpha_j$ $(i = 1, \ldots, m)$.

If a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a Dirichlet distribution with parametric vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, then any moment of the form $E(X_1^{r_1} \cdots X_k^{r_k})$, where $r_1, \ldots, r_k$ are nonnegative integers, can be found as follows: Let the set $S \subset R^{k-1}$ be defined by the equation



$$S = \Big\{(x_1, \ldots, x_{k-1})' : x_i > 0 \ (i = 1, \ldots, k-1) \text{ and } \sum_{i=1}^{k-1} x_i < 1\Big\}. \tag{3}$$

The integral over the set $S$ of the p.d.f. given in Eq. (1) must be unity. Therefore, if $\alpha_1, \ldots, \alpha_k$ are any positive numbers and if $x_k = 1 - \Sigma_{i=1}^{k-1} x_i$,

$$\int \cdots \int_S x_1^{\alpha_1 - 1} \cdots x_k^{\alpha_k - 1}\, dx_1 \cdots dx_{k-1} = \frac{\Pi_{i=1}^k \Gamma(\alpha_i)}{\Gamma(\Sigma_{i=1}^k \alpha_i)}. \tag{4}$$

It follows that

$$E(X_1^{r_1} \cdots X_k^{r_k}) = \frac{\Gamma(\Sigma_{i=1}^k \alpha_i)}{\Pi_{i=1}^k \Gamma(\alpha_i)} \cdot \frac{\Pi_{i=1}^k \Gamma(\alpha_i + r_i)}{\Gamma[\Sigma_{i=1}^k (\alpha_i + r_i)]}. \tag{5}$$

Let $\alpha_0 = \Sigma_{i=1}^k \alpha_i$. Then it follows from Eq. (5) that for $i = 1, \ldots, k$,

$$E(X_i) = \frac{\alpha_i}{\alpha_0} \tag{6}$$

and

$$\operatorname{Var}(X_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}. \tag{7}$$

These results also follow from the known facts about the beta distribution. Furthermore, when $i \ne j$,

$$\operatorname{Cov}(X_i, X_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}. \tag{8}$$

Further discussion of these and other interesting properties of the Dirichlet distribution is given by Wilks (1962).

## 5.4 THE MULTIVARIATE NORMAL DISTRIBUTION

A $k$-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a *nonsingular multivariate normal distribution with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$ if $\mathbf{X}$ has an absolutely continuous distribution whose p.d.f. $f(\cdot\,|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is specified at any point $\mathbf{x} \in R^k$ by the equation

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-k/2} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]. \tag{1}$$

In Eq. (1), the mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ is a $k$-dimensional vector whose components can be arbitrary real numbers; the $k \times k$ covariance matrix $\boldsymbol{\Sigma}$ must be symmetric and positive definite, but otherwise its components can also be arbitrary real numbers.

For reasons similar to those given in Sec. 4.7 for the univariate normal distribution, the multivariate normal distribution is very important in many phases of current work in probability and statistics. The entire body of theory known as multivariate statistical analysis deals almost exclusively with the analysis of random samples from multivariate normal distributions. An excellent book in this area is Anderson (1958).

We shall now derive the characteristic function $\zeta$ of a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ which has a multivariate normal distribution whose p.d.f. is defined by Eq. (1). For any point $\mathbf{t} = (t_1, \ldots, t_k)' \in R^k$,

$$\begin{aligned}\zeta(\mathbf{t}) &= E(e^{i\mathbf{t}'\mathbf{X}}) \\ &= \int_{R^k} \cdots \int (2\pi)^{-k/2}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp [i\mathbf{t}'\mathbf{x} \\ &\qquad - \tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})] \prod_{j=1}^{k} dx_j. \qquad (2)\end{aligned}$$

It is well known that there exists a nonsingular $k \times k$ matrix $\mathbf{B}$ such that the symmetric and positive definite matrix $\mathbf{\Sigma}$ can be expressed in the form $\mathbf{\Sigma} = \mathbf{BB}'$. The variables in the integral in Eq. (2) can now be changed by means of the transformation

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{By}. \qquad (3)$$

The Jacobian $J$ of this transformation is $J = |\mathbf{B}| = |\mathbf{\Sigma}|^{\frac{1}{2}}$, and so Eq. (2) becomes

$$\begin{aligned}\zeta(\mathbf{t}) &= \int_{R^k} \cdots \int (2\pi)^{-k/2} \exp [i\mathbf{t}'(\boldsymbol{\mu} + \mathbf{By}) - \tfrac{1}{2}\mathbf{y}'\mathbf{y}] \prod_{j=1}^{k} dy_j \\ &= e^{i\mathbf{t}'\boldsymbol{\mu}} \int_{R^k} \cdots \int (2\pi)^{-k/2} \exp [i(\mathbf{B}'\mathbf{t})'\mathbf{y} - \tfrac{1}{2}\mathbf{y}'\mathbf{y}] \prod_{j=1}^{k} dy_j. \qquad (4)\end{aligned}$$

Next, let $\mathbf{B}'\mathbf{t} = \mathbf{s} = (s_1, \ldots, s_k)'$. Then Eq. (4) can be written in the following form:

$$\zeta(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu}} \prod_{j=1}^{k} \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp (is_j y_j - \tfrac{1}{2}y_j^2)\, dy_j. \qquad (5)$$

Since each integral in Eq. (5) is simply the c.f. of a univariate standard normal distribution (see Exercise 15 of Chap. 4), it follows that

$$\begin{aligned}\zeta(\mathbf{t}) &= e^{i\mathbf{t}'\boldsymbol{\mu}} \prod_{j=1}^{k} \exp\left(\frac{-s_j^2}{2}\right) = \exp (i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{s}'\mathbf{s}) \\ &= \exp (i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\mathbf{BB}'\mathbf{t}), \qquad (6)\end{aligned}$$

or

$$\zeta(\mathbf{t}) = \exp (i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}). \qquad (7)$$



An obvious analogy can be seen between the c.f. given by Eq. (7) and the c.f. of a random variable having a univariate normal distribution, as given in Exercise 15 of Chap. 4. It should be noted that Eq. (7) provides a verification of the fact that the integral over $R^k$ of the p.d.f. given by Eq. (1) actually has the value 1, since at the point $\mathbf{t} = \mathbf{0}$ it is seen that $\zeta(\mathbf{0}) = 1$. Furthermore, it can be verified that $\boldsymbol{\mu}$ is actually the mean vector and $\mathbf{\Sigma}$ is the covariance matrix of this distribution (see Exercise 7).

Equation (7) represents the c.f. of a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ even if the matrix $\mathbf{\Sigma}$ is nonnegative definite but not positive definite. The random vector $\mathbf{X}$ is said to have a *singular multivariate normal distribution* if the c.f. of $\mathbf{X}$ is given by Eq. (7), where $\mathbf{\Sigma}$ is a symmetric, nonnegative definite, singular matrix. It is still true that $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{X}) = \mathbf{\Sigma}$. However, since $\mathbf{\Sigma}$ is singular, there must be certain linear relations among the components $X_1, \ldots, X_k$, and therefore these $k$ random variables cannot have a joint $k$-dimensional p.d.f. If some of the components are deleted until there are no linear relations among the ones that remain, then these remaining components will have a nonsingular multivariate normal distribution.

Suppose that $\mathbf{X} = (X_1, \ldots, X_k)'$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, where $\mathbf{\Sigma}$ might be either singular or nonsingular. Let $\mathbf{A}$ be a given $m \times k$ matrix, and let the $m$-dimensional random vector $\mathbf{Y}$ be defined as $\mathbf{Y} = \mathbf{AX}$. At any point $\mathbf{t} \in R^m$, the c.f. $\zeta_Y$ of $\mathbf{Y}$ has the following value:

$$\zeta_Y(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{Y}}) = E(e^{i\mathbf{t}'\mathbf{AX}}) = E(e^{i(\mathbf{A}'\mathbf{t})'\mathbf{X}}) = \zeta(\mathbf{A}'\mathbf{t}).$$

Hence, from Eq. (7),

$$\zeta_Y(\mathbf{t}) = \exp (i\mathbf{t}'\mathbf{A}\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\mathbf{A\Sigma A}'\mathbf{t}). \qquad (8)$$

A direct comparison of Eqs. (7) and (8) reveals that $\mathbf{Y}$ itself has a multivariate normal distribution for which the mean vector is $\mathbf{A}\boldsymbol{\mu}$ and the covariance matrix is $\mathbf{A\Sigma A}'$.

One consequence of this derivation is the following basic result, which can be demonstrated by choosing the matrix $\mathbf{A}$ appropriately: The marginal joint distribution of any subset of the random variables $X_1, \ldots, X_k$ will again be normal, and the corresponding subvector of $\boldsymbol{\mu}$ and the corresponding submatrix of $\mathbf{\Sigma}$ will be the mean vector and the covariance matrix of that distribution (see Exercise 8).

We shall now derive the conditional distribution of some of the components of $\mathbf{X}$ when the values of the other components are given. It is assumed that $\mathbf{X} = (X_1, \ldots, X_k)'$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and with nonsingular covariance matrix $\mathbf{\Sigma}$.

Suppose that the $k$-dimensional random vector $\mathbf{X}$ is partitioned as in the form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \tag{9}$$

where $\mathbf{X}_1$ comprises the first $k_1$ components of $\mathbf{X}$, $\mathbf{X}_2$ comprises the last $k_2$ components of $\mathbf{X}$, and $k_1 + k_2 = k$. Suppose also that the mean vector $\boldsymbol{\mu}$, the covariance matrix $\boldsymbol{\Sigma}$, and its inverse $\boldsymbol{\Sigma}^{-1}$ are partitioned as follows:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}. \tag{10}$$

Here, $\boldsymbol{\mu}_1$ is a $k_1$-dimensional vector and $\boldsymbol{\mu}_2$ is a $k_2$-dimensional vector; both $\boldsymbol{\Sigma}_{11}$ and $\mathbf{T}_{11}$ are $k_1 \times k_1$ matrices; both $\boldsymbol{\Sigma}_{22}$ and $\mathbf{T}_{22}$ are $k_2 \times k_2$ matrices; both $\boldsymbol{\Sigma}_{12}$ and $\mathbf{T}_{12}$ are $k_1 \times k_2$ matrices; and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$ and $\mathbf{T}_{12} = \mathbf{T}_{21}'$. Note that $E(\mathbf{X}_i) = \boldsymbol{\mu}_i$ and $\text{Cov}(\mathbf{X}_i) = \boldsymbol{\Sigma}_{ii}$ for $i = 1, 2$ and that the elements of $\boldsymbol{\Sigma}_{12}$ are the covariances of the various components of $\mathbf{X}_1$ with those of $\mathbf{X}_2$. The conditional distribution of $\mathbf{X}_1$ when the value of $\mathbf{X}_2$ is given can now be developed in terms of the elements which are displayed in Eq. (10).

The joint p.d.f. of $\mathbf{X}_1$ and $\mathbf{X}_2$ is simply the p.d.f. of $\mathbf{X}$, as given by Eq. (1). Therefore, for any points $\mathbf{x}_1 \in R^{k_1}$ and $\mathbf{x}_2 \in R^{k_2}$, the value $f(\mathbf{x}_1, \mathbf{x}_2)$ of this joint p.d.f. is specified by Eq. (1) with

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \tag{11}$$

Furthermore, the marginal distribution of $\mathbf{X}_2$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_{22}$. Therefore, at any point $\mathbf{x}_2 \in R^{k_2}$, the value $g(\mathbf{x}_2)$ of the marginal p.d.f. of $\mathbf{X}_2$ is specified by the equation

$$g(\mathbf{x}_2) = (2\pi)^{-k_2/2}|\boldsymbol{\Sigma}_{22}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{x}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right]. \tag{12}$$

It follows from Eqs. (10) and (11) that

$$\begin{aligned}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) &= (\mathbf{x}_1' - \boldsymbol{\mu}_1', \mathbf{x}_2' - \boldsymbol{\mu}_2') \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix} \\ &= (\mathbf{x}_1 - \boldsymbol{\mu}_1)'\mathbf{T}_{11}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ &\quad + (\mathbf{x}_1 - \boldsymbol{\mu}_1)'\mathbf{T}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &\quad + (\mathbf{x}_2 - \boldsymbol{\mu}_2)'\mathbf{T}_{21}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ &\quad + (\mathbf{x}_2 - \boldsymbol{\mu}_2)'\mathbf{T}_{22}(\mathbf{x}_2 - \boldsymbol{\mu}_2). \end{aligned} \tag{13}$$

It can be shown (see Exercise 9) that

$$\boldsymbol{\Sigma}_{22}^{-1} = \mathbf{T}_{22} - \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{T}_{12}. \tag{14}$$



This identity permits the value specified in Eq. (13) to be rewritten as

$$\begin{aligned}[(\mathbf{x}_1 - \boldsymbol{\mu}_1) + \mathbf{T}_{11}^{-1}\mathbf{T}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)]'\mathbf{T}_{11}[(\mathbf{x}_1 - \boldsymbol{\mu}_1) + \mathbf{T}_{11}^{-1}\mathbf{T}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)] \\ + (\mathbf{x}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2). \end{aligned} \tag{15}$$

It is also shown in Exercise 9 that

$$\mathbf{T}_{11}^{-1}\mathbf{T}_{12} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \tag{16}$$

and

$$\mathbf{T}_{11}^{-1} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{17}$$

Therefore, if we let

$$Q_1 = (\mathbf{x}_1 - \boldsymbol{\nu}_1)'(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}(\mathbf{x}_1 - \boldsymbol{\nu}_1), \tag{18}$$

where

$$\boldsymbol{\nu}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \tag{19}$$

and also let

$$Q_2 = (\mathbf{x}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \tag{20}$$

then it follows that expression (15) is equal to $Q_1 + Q_2$.

Hence, for any points $\mathbf{x}_1 \in R^{k_1}$ and $\mathbf{x}_2 \in R^{k_2}$,

$$f(\mathbf{x}_1, \mathbf{x}_2) = (2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{Q_1 + Q_2}{2}\right). \tag{21}$$

Furthermore, by Eqs. (12) and (20),

$$g(\mathbf{x}_2) = (2\pi)^{-k_2/2}|\boldsymbol{\Sigma}_{22}|^{-\frac{1}{2}} \exp\left(-\frac{Q_2}{2}\right). \tag{22}$$

It therefore follows that the conditional p.d.f. $h(\cdot\,|\mathbf{x}_2)$ of $\mathbf{X}_1$, when $\mathbf{X}_2 = \mathbf{x}_2$, is specified at the point $\mathbf{x}_1$ by the equation

$$h(\mathbf{x}_1|\mathbf{x}_2) = (2\pi)^{-k_1/2}\left(\frac{|\boldsymbol{\Sigma}_{22}|}{|\boldsymbol{\Sigma}|}\right)^{\frac{1}{2}} \exp\left(-\frac{Q_1}{2}\right). \tag{23}$$

By examination of Eq. (18) and a direct comparison of the conditional p.d.f. given by Eq. (23) and the multivariate normal p.d.f. given by Eq. (1), it can be seen that the conditional distribution of $\mathbf{X}_1$, when $\mathbf{X}_2 = \mathbf{x}_2$, is a $k_1$-dimensional multivariate normal distribution whose mean vector is $\boldsymbol{\nu}_1$, as given by Eq. (19), and whose covariance matrix is $\mathbf{T}_{11}^{-1}$, as given by Eq. (17).

Furthermore, by comparing the constant that appears in Eq. (23) with the constant that usually appears in the p.d.f. of a multivariate normal distribution whose covariance matrix is given by Eq. (17), we obtain the following relation:

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{22}| \cdot |\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|. \tag{24}$$

The *precision matrix* $\mathbf{T}$ of a nonsingular multivariate normal distribution is defined to be the inverse of the covariance matrix; that is,

$$\mathbf{T} = \mathbf{\Sigma}^{-1}. \tag{25}$$

As stated in Sec. 4.7 in regard to a univariate normal distribution, it will be more convenient in much of our later work to specify a nonsingular multivariate normal distribution by its mean vector and its precision matrix rather than by its mean vector and its covariance matrix. Thus, if a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and precision matrix $\mathbf{T}$, its p.d.f. $g(\cdot|\boldsymbol{\mu}, \mathbf{T})$ is specified at any point $\mathbf{x}$ ($\mathbf{x} \in R^k$) by the equation

$$g(\mathbf{x}|\boldsymbol{\mu}, \mathbf{T}) = (2\pi)^{-k/2}|\mathbf{T}|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{T}(\mathbf{x} - \boldsymbol{\mu})\right]. \tag{26}$$

Note that if it is stated that a random vector has a multivariate normal distribution with a specified precision matrix, then this distribution must be nonsingular and it must have a p.d.f. of the form given in Eq. (26).

## 5.5 THE WISHART DISTRIBUTION

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample of $k$-dimensional random vectors from a multivariate normal distribution for which the mean vector is $\mathbf{0}$ and the $k \times k$ covariance matrix is $\mathbf{\Sigma}$. Also, let $\mathbf{V}$ denote the random symmetric $k \times k$ matrix which is defined by the equation

$$\mathbf{V} = \sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'. \tag{1}$$

This random matrix $\mathbf{V}$ has a *Wishart distribution with n degrees of freedom and parametric matrix* $\mathbf{\Sigma}$. It will be assumed that $n > k - 1$ and that the matrix $\mathbf{\Sigma}$ is nonsingular. Under these conditions, the Wishart distribution is called *nonsingular* and the distribution of $\mathbf{V}$ can be represented by a p.d.f. in the following way:

Since $\mathbf{V}$ is a symmetric matrix, it can be expressed in the form

$$\mathbf{V} = \begin{bmatrix} V_{11} & \cdots & V_{1k} \\ \cdots & \cdots & \cdots \\ V_{k1} & \cdots & V_{kk} \end{bmatrix}, \tag{2}$$

where $V_{ij} = V_{ji}$ ($i, j = 1, \ldots, k$). Hence, the random matrix $\mathbf{V}$ is completely specified by the $k(k+1)/2$ distinct random variables $V_{ij}$ which lie on or above its main diagonal. When $n > k - 1$ and $\mathbf{\Sigma}$ is nonsingular, it can be shown that the joint distribution of the random variables $V_{ij}$ ($i, j = 1, \ldots, k$; $i \leq j$) is absolutely continuous and can therefore be represented by their joint p.d.f. in $R^{k(k+1)/2}$.



Each specification of the values $v_{ij}$ of the random variables $V_{ij}$ ($i, j = 1, \ldots, k$; $i \leq j$) can be regarded either as characterizing a vector $(v_{11}, v_{12}, v_{22}, \ldots, v_{kk})' \in R^{k(k+1)/2}$ or, equivalently, as characterizing a symmetric $k \times k$ matrix

$$\mathbf{v} = \begin{bmatrix} v_{11} & \cdots & v_{1k} \\ \cdots & \cdots & \cdots \\ v_{1k} & \cdots & v_{kk} \end{bmatrix}. \tag{3}$$

Accordingly, the value $f(\mathbf{v}|n, \mathbf{\Sigma})$ of the joint p.d.f. of the $k(k+1)/2$ distinct random variables $V_{ij}$ can be specified for each symmetric matrix $\mathbf{v}$ of the form given in Eq. (3). Furthermore, it is convenient to refer to the joint p.d.f. $f(\cdot|n, \mathbf{\Sigma})$ of these random variables simply as the p.d.f. of the random symmetric matrix $\mathbf{V}$. It is important to keep in mind that although the function $f(\cdot|n, \mathbf{\Sigma})$ will be called the p.d.f. of the random matrix $\mathbf{V}$ and will be treated as a function of symmetric matrices $\mathbf{v}$, it is, in fact, the joint p.d.f. of only $k(k+1)/2$ distinct random variables and is a function of vectors in $R^{k(k+1)/2}$.

It can be shown that when $n > k - 1$ and $\mathbf{\Sigma}$ is nonsingular, the probability is 1 that the random matrix $\mathbf{V}$ defined by Eq. (2) will be positive definite. Therefore, for any matrix $\mathbf{v}$ which is not symmetric and positive definite, $f(\mathbf{v}|n, \mathbf{\Sigma}) = 0$. Furthermore, for any $k \times k$ matrix $\mathbf{v}$ which is symmetric and positive definite,

$$f(\mathbf{v}|n, \mathbf{\Sigma}) = c|\mathbf{\Sigma}|^{-n/2}|\mathbf{v}|^{(n-k-1)/2} \exp\left[-\tfrac{1}{2}\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{v})\right]. \tag{4}$$

In this equation, tr $(\mathbf{\Sigma}^{-1}\mathbf{v})$ denotes the trace of the matrix $\mathbf{\Sigma}^{-1}\mathbf{v}$ and the constant $c$ has the following value:

$$c = \left[2^{nk/2}\pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma\left(\frac{n+1-j}{2}\right)\right]^{-1}. \tag{5}$$

Equations (4) and (5) together describe the p.d.f. of a nonsingular Wishart distribution of dimension $k \times k$, with $n$ degrees of freedom and parametric matrix $\mathbf{\Sigma}$.

Let $S$ denote the set of symmetric, positive definite $k \times k$ matrices. Since each matrix in $S$ is specified by $k(k+1)/2$ distinct elements, the set $S$ can be regarded as a subset of $R^{k(k+1)/2}$. Since the integral of the p.d.f. $f(\cdot|n, \mathbf{\Sigma})$ over the space $R^{k(k+1)/2}$ must have the value 1, Eq. (5) is equivalent to the following equation:

$$c^{-1}|\mathbf{\Sigma}|^{n/2} = \int_S |\mathbf{v}|^{(n-k-1)/2} \exp\left[-\tfrac{1}{2}\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{v})\right] dv_{11}\, dv_{12} \cdots dv_{kk}. \tag{6}$$

For a derivation of Eqs. (4) to (6), as well as of the other properties of the Wishart distribution that are mentioned in this section, the reader is referred to Anderson (1958), Wilks (1962), or Rao (1965).

Much of the importance of the Wishart distribution stems from the following famous result. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample of $k$-dimensional random vectors from a multivariate normal distribution for which the mean vector is $\boldsymbol{\mu}$ and the covariance matrix is $\boldsymbol{\Sigma}$. Also, let the vector $\bar{\mathbf{X}}$ and the $k \times k$ matrix $\mathbf{S}$ be defined as follows:

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \qquad \text{and} \qquad \mathbf{S} = \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'.$$

Then the random vector $\bar{\mathbf{X}}$ and the random matrix $\mathbf{S}$ are independent; $\bar{\mathbf{X}}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}/n$, and $\mathbf{S}$ has a Wishart distribution with $n - 1$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$. It is clear from this result as well as from the representation given in Eq. (1) that the Wishart distribution is essentially a multivariate generalization of the $\chi^2$ distribution.

Suppose that a $k \times k$ random matrix $\mathbf{V}$ has a Wishart distribution with $n$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$. The following three properties are easily obtained from the representation in Eq. (1) (see Exercise 15): (1) The expectation of $\mathbf{V}$ is given by the relation $E(\mathbf{V}) = n\boldsymbol{\Sigma}$. (2) If $\mathbf{A}$ is an $m \times k$ matrix, then the $m \times m$ random matrix $\mathbf{AVA}'$ has a Wishart distribution with $n$ degrees of freedom and parametric matrix $\mathbf{A\Sigma A}'$. (3) Suppose that the matrices $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are partitioned as follows:

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \tag{7}$$

where $\mathbf{V}_{11}$ and $\boldsymbol{\Sigma}_{11}$ are square matrices with the same dimension. Then the random matrix $\mathbf{V}_{11}$ has a Wishart distribution with $n$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}_{11}$.

Furthermore (see Exercise 16), suppose that $\mathbf{V}_1, \ldots, \mathbf{V}_r$ are independent $k \times k$ random matrices and that $\mathbf{V}_i$ has a Wishart distribution with $n_i$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$ $(i = 1, \ldots, r)$. Then the sum $\mathbf{V}_1 + \cdots + \mathbf{V}_r$ also has a Wishart distribution with $n_1 + \cdots + n_r$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$.

If a $k \times k$ random matrix $\mathbf{V}$ has a Wishart distribution whose p.d.f. is given by Eq. (4), then the characteristic function $\zeta$ of $\mathbf{V}$ can be given in a simple form. The correct interpretation of the function $\zeta$ involves essentially the same considerations as those involved in the interpretation of the p.d.f. given in Eq. (4). Thus, $\zeta$ is really the joint characteristic function of the $k(k+1)/2$ distinct random variables $V_{ij}$ which define $\mathbf{V}$, and $\zeta$ is therefore a function of $k(k+1)/2$ distinct real variables $t_{ij}$ $(i, j = 1, \ldots, k;\ i \le j)$. However, each specification of the values of the variables $t_{ij}$ can be regarded as characterizing a symmetric matrix $\mathbf{t}$ defined as follows:



$$\mathbf{t} = \begin{bmatrix} 2t_{11} & t_{12} & \cdots & t_{1k} \\ t_{12} & 2t_{22} & \cdots & t_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ t_{1k} & t_{2k} & \cdots & 2t_{kk} \end{bmatrix}. \tag{8}$$

Therefore, the characteristic function $\zeta$ can conveniently be regarded as a function of symmetric matrices $\mathbf{t}$ which have the form given in Eq. (8). With this convention, it can be shown that

$$\zeta(\mathbf{t}) = E\left[\exp\left(i \sum_{\beta=1}^{k} \sum_{\alpha=1}^{\beta} t_{\alpha\beta} V_{\alpha\beta}\right)\right] = \left(\frac{|\boldsymbol{\Sigma}^{-1}|}{|\boldsymbol{\Sigma}^{-1} - i\mathbf{t}|}\right)^{n/2}. \tag{9}$$

The *precision matrix* $\mathbf{T}$ of a nonsingular Wishart distribution with parametric matrix $\boldsymbol{\Sigma}$ is defined by the relation

$$\mathbf{T} = \boldsymbol{\Sigma}^{-1}. \tag{10}$$

As stated in connection with the normal distribution, it will be more convenient in much of our later work to specify a Wishart distribution by its degrees of freedom $n$ and its precision matrix $\mathbf{T}$, rather than by $n$ and $\boldsymbol{\Sigma}$. Thus, if a random $k \times k$ matrix $\mathbf{V}$ has a Wishart distribution with $n$ degrees of freedom $(n > k - 1)$ and precision matrix $\mathbf{T}$, its p.d.f. $g(\cdot|n, \mathbf{T})$ is specified for any symmetric, positive definite $k \times k$ matrix $\mathbf{v}$ by the equation

$$g(\mathbf{v}|n, \mathbf{T}) = c|\mathbf{T}|^{n/2}|\mathbf{v}|^{(n-k-1)/2} \exp\left[-\tfrac{1}{2}\operatorname{tr}(\mathbf{Tv})\right]. \tag{11}$$

The constant $c$ is given by Eq. (5). Furthermore, for any symmetric matrix $\mathbf{v}$ which is not positive definite, $g(\mathbf{v}|n, \mathbf{T}) = 0$.

## 5.6 THE MULTIVARIATE $t$ DISTRIBUTION

Suppose that the $k$-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_k)'$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and precision matrix $\mathbf{T}$, that the random variable $Z$ has a $\chi^2$ distribution with $n$ degrees of freedom, and that $\mathbf{Y}$ and $Z$ are independent. Suppose also that $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ is any given vector in $R^k$. Let a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ be defined by the equation

$$X_i = Y_i\left(\frac{Z}{n}\right)^{-\frac{1}{2}} + \mu_i \qquad i = 1, \ldots, k. \tag{1}$$

Then the distribution of $\mathbf{X}$ is called a *multivariate $t$ distribution with $n$ degrees of freedom, location vector $\boldsymbol{\mu}$, and precision matrix* $\mathbf{T}$. We shall now derive the p.d.f. $f(\cdot|n, \boldsymbol{\mu}, \mathbf{T})$ of the vector $\mathbf{X}$.

At any point $\mathbf{y} \in R^k$, the value $g_Y(\mathbf{y})$ of the p.d.f. of the random vector $\mathbf{Y}$ is

$$g_Y(\mathbf{y}) = (2\pi)^{-k/2}|\mathbf{T}|^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\mathbf{y}'\mathbf{T}\mathbf{y}\right). \tag{2}$$

Furthermore, for any number $z$ ($z > 0$), the value $g_Z(z)$ of the p.d.f. of the random variable $Z$ is

$$g_Z(z) = \left[2^{n/2}\Gamma\left(\frac{n}{2}\right)\right]^{-1} z^{(n/2)-1}e^{-z/2}. \tag{3}$$

Since $\mathbf{Y}$ and $Z$ are independent, the value $g_{Y,Z}(\mathbf{y}, z)$ of their joint p.d.f. at any point $(\mathbf{y}, z)$ such that $\mathbf{y} \in R^k$ and $z > 0$ is the product of the values given in Eqs. (2) and (3). Moreover, $g_{Y,Z}(\mathbf{y}, z) = 0$ at any other point $(\mathbf{y}, z)$.

The joint p.d.f. $g_{X,Z}$ of the random vector $\mathbf{X}$ and the random variable $Z$ can now be computed. From Eq. (1),

$$Y_i = \left(\frac{Z}{n}\right)^{\frac{1}{2}}(X_i - \mu_i) \qquad i = 1, \ldots, k. \tag{4}$$

Therefore, the Jacobian $J$ of the transformation from the $k + 1$ random variables $\{X_1, \ldots, X_k, Z\}$ to the $k + 1$ random variables $\{Y_1, \ldots, Y_k, Z\}$ is the determinant of a triangular matrix, and its value is

$$J = \left(\frac{z}{n}\right)^{k/2}. \tag{5}$$

By substituting the values of $Y_1, \ldots, Y_k$ given by Eq. (4) in the joint p.d.f. $g_{Y,Z}$, multiplying by $J$, and combining terms, we obtain for any point $\mathbf{x} \in R^k$ and any number $z$ ($z > 0$),

$$g_{X,Z}(\mathbf{x}, z) = c'z^{(n+k-2)/2} \exp\left\{-\tfrac{1}{2}\left[1 + \frac{1}{n}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{T}(\mathbf{x} - \boldsymbol{\mu})\right]z\right\}. \tag{6}$$

Here

$$c' = |\mathbf{T}|^{\frac{1}{2}}\left[2^{(n+k)/2}(n\pi)^{k/2}\Gamma\left(\frac{n}{2}\right)\right]^{-1}. \tag{7}$$

The desired p.d.f. $f(\cdot\,|n, \boldsymbol{\mu}, \mathbf{T})$ of $\mathbf{X}$ can now be obtained as a marginal p.d.f. by integrating the joint p.d.f. given in Eq. (6) over all positive values of $z$. From the definition of the gamma function given in Eq. (2) of Sec. 4.8, for any number $Q$ ($Q > 0$),

$$\int_0^\infty z^{(n+k-2)/2} \exp(-Qz)\, dz = \Gamma\left(\frac{n+k}{2}\right)Q^{-(n+k)/2}. \tag{8}$$

Hence, for any point $\mathbf{x} \in R^k$,

$$f(\mathbf{x}|n, \boldsymbol{\mu}, \mathbf{T}) = c\left[1 + \frac{1}{n}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{T}(\mathbf{x} - \boldsymbol{\mu})\right]^{-(n+k)/2}, \tag{9}$$



where

$$c = \frac{\Gamma[(n+k)/2]\,|\mathbf{T}|^{\frac{1}{2}}}{\Gamma(n/2)(n\pi)^{k/2}}. \tag{10}$$

Equations (9) and (10) define the p.d.f. of the $k$-dimensional multivariate $t$ distribution with $n$ degrees of freedom, location vector $\boldsymbol{\mu}$, and precision matrix T.

This distribution is a $k$-dimensional generalization of the univariate $t$ distribution discussed in Sec. 4.12; when $k = 1$, the p.d.f. given by Eqs. (9) and (10) is simply the univariate p.d.f. given by Eq. (5) of Sec. 4.12. Of course, this result could have been anticipated from the definition of the multivariate $t$ distribution given at the beginning of this section and the corresponding property of the univariate $t$ distribution, as given in Exercise 32 of Chap. 4.

If $\mathbf{X}$ has a multivariate $t$ distribution whose p.d.f. is given by Eqs. (9) and (10), it can be shown that for $n > 2$, the mean vector $E(\mathbf{X})$ and the covariance matrix $\mathrm{Cov}(\mathbf{X})$ of $\mathbf{X}$ exist and their values are (see Exercise 18)

$$E(\mathbf{X}) = \boldsymbol{\mu} \qquad \text{and} \qquad \mathrm{Cov}(\mathbf{X}) = \frac{n}{n-2}\mathbf{T}^{-1}. \tag{11}$$

Furthermore, the marginal distribution of any subset of the components of $\mathbf{X}$ can be obtained from the representation in Eq. (1) and the corresponding properties of the multivariate normal distribution, as discussed in Sec. 5.4. The result is as follows (see Exercise 19): Suppose that the random vector $X$ is partitioned as in the form

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix}. \tag{12}$$

Here the dimension of $\mathbf{X}_i$ is $k_i$ ($i = 1, 2$) and $k_1 + k_2 = k$. Also, suppose that the location vector $\boldsymbol{\mu}$ and the precision matrix $\mathbf{T}$ are partitioned as in the forms

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \qquad \mathbf{T} = \begin{bmatrix}\mathbf{T}_{11} & \mathbf{T}_{12}\\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{bmatrix}. \tag{13}$$

Here the dimension of $\boldsymbol{\mu}_i$ is $k_i$ ($i = 1, 2$) and the dimension of the submatrix $\mathbf{T}_{ij}$ is $k_i \times k_j$ ($i, j = 1, 2$). Then the marginal distribution of $\mathbf{X}_1$ is a $k_1$-dimensional multivariate $t$ distribution with $n$ degrees of freedom, its location vector is $\boldsymbol{\mu}_1$, and its precision matrix is $\mathbf{T}_{11} - \mathbf{T}_{12}\mathbf{T}_{22}^{-1}\mathbf{T}_{21}$.

The conditional distribution of $\mathbf{X}_1$ when $\mathbf{X}_2 = \mathbf{x}_2$ ($\mathbf{x}_2 \in R^{k_2}$) is also a $k_1$-dimensional multivariate $t$ distribution, but it is rather complicated. The degrees of freedom change, and both the location vector of the conditional distribution and the precision matrix depend on the given point $\mathbf{x}_2$. Specifically (see Exercise 20), the conditional distribution has

$n + k_2$ degrees of freedom, the location vector is

$$\boldsymbol{\mu}_1 - \mathbf{T}_{11}^{-1}\mathbf{T}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \tag{14}$$

and the precision matrix is

$$\frac{n + k_2}{n + (\mathbf{x}_2 - \boldsymbol{\mu}_2)'(\mathbf{T}_{22} - \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{T}_{12})(\mathbf{x}_2 - \boldsymbol{\mu}_2)} \mathbf{T}_{11}. \tag{15}$$

The following fact will also be used later in this book. If a random vector **X** has a $k$-dimensional multivariate $t$ distribution whose p.d.f. is given by Eqs. (9) and (10), then the random variable

$$\frac{1}{k}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{T}(\mathbf{X} - \boldsymbol{\mu}) \tag{16}$$

has an $F$ distribution with $k$ and $n$ degrees of freedom, as defined in Sec. 4.13 (see Exercise 21).

A bibliography on multivariate normal distributions and multivariate $t$ distributions has been compiled by Gupta (1963).

## 5.7 THE BILATERAL BIVARIATE PARETO DISTRIBUTION

We shall conclude this chapter with the definition of a bivariate distribution such that each of its two individual univariate marginal distributions is essentially a Pareto distribution, although one of these is on an interval of the form $(-\infty, r_1)$ while the other is on an interval of the form $(r_2, \infty)$. Suppose that $r_1$, $r_2$, and $\alpha$ are three numbers such that $r_1 < r_2$ and $\alpha > 0$. We shall say that the joint distribution of the random variables $X_1$ and $X_2$ is a *bilateral bivariate Pareto distribution with parameters* $r_1$, $r_2$, *and* $\alpha$ if the distribution is absolutely continuous and the joint p.d.f. $f(\cdot|r_1, r_2, \alpha)$ is as follows:

At any point $(x_1, x_2) \in R^2$ such that $x_1 < r_1$ and $x_2 > r_2$,

$$f(x_1, x_2|r_1, r_2, \alpha) = \frac{\alpha(\alpha + 1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}}. \tag{1}$$

Furthermore, at any other point $(x_1, x_2) \in R^2$, $f(x_1, x_2|r_1, r_2, \alpha) = 0$.

If the joint p.d.f. of $X_1$ and $X_2$ is given by Eq. (1), it can be easily verified that both the marginal distribution of $r_2 - X_1$ and the marginal distribution of $X_2 - r_1$ are univariate Pareto distributions with parameters $r_2 - r_1$ and $\alpha$, as defined in Sec. 4.11 (see Exercise 23).

Thus, the following means and variances of $X_1$ and $X_2$ can be computed from Eqs. (2) and (3) of Sec. 4.11:

$$E(X_1) = \frac{\alpha r_1 - r_2}{\alpha - 1}, \qquad E(X_2) = \frac{\alpha r_2 - r_1}{\alpha - 1}, \tag{2}$$



and

$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \frac{\alpha(r_2 - r_1)^2}{(\alpha - 1)^2(\alpha - 2)}. \tag{3}$$

Furthermore, it can be shown that the correlation between $X_1$ and $X_2$ is $-1/\alpha$ (see Exercise 24).

## EXERCISES

1. Suppose that $X_1, \ldots, X_k$ are independent random variables and that $X_i$ has a Poisson distribution with mean $\lambda_i$ $(i = 1, \ldots, k)$. Let $n$ be any fixed positive integer. Show that the conditional distribution of the random vector $\mathbf{X} = (X_1, \ldots, X_k)'$, given that $\Sigma_{i=1}^k X_i = n$, is multinomial with parameters $n$ and $\mathbf{p} = (p_1, \ldots, p_k)'$, where $p_i = \lambda_i/(\Sigma_{j=1}^k \lambda_j)$ for $i = 1, \ldots, k$.

2. If the random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a multinomial distribution with parameters $n$ and $\mathbf{p} = (p_1, \ldots, p_k)'$, show that its characteristic function $\zeta$ is specified at any point $\mathbf{t} = (t_1, \ldots, t_k)' \in R^k$ by the equation

$$\zeta(\mathbf{t}) = \left(\sum_{j=1}^k p_j e^{it_j}\right)^n.$$

Also, show that $E(\mathbf{X}) = n\mathbf{p}$, that $\mathrm{Var}(X_i) = np_i(1 - p_i)$ for $i = 1, \ldots, k$, and that $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$. Is the covariance matrix of **X** singular or nonsingular?

3. If $\mathbf{X}_1, \ldots, \mathbf{X}_r$ are independent random vectors and $\mathbf{X}_i$ has a multinomial distribution with parameters $n_i$ and $\mathbf{p}$ $(i = 1, \ldots, r)$, show that the sum $\mathbf{X}_1 + \cdots + \mathbf{X}_r$ also has a multinomial distribution with parameters $n_1 + \cdots + n_r$ and the same vector $\mathbf{p}$.

4. Let $S$ be the subset of $R^{k-1}$ which is defined by Eq. (3) of Sec. 5.3. For any point in $S$, define $x_k = 1 - \Sigma_{i=1}^{k-1} x_i$. Prove that the p.d.f. of the Dirichlet distribution, as defined by Eq. (1) of Sec. 5.3, is properly normed by showing that

$$\int \cdots \int_S \left(\prod_{i=1}^k x_i^{\alpha_i - 1}\right) dx_1 \cdots dx_{k-1} = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \cdots + \alpha_k)}.$$

*Hint:* Change variables by means of the following equations:

$$y_1 = x_1,$$
$$y_2 = \frac{x_2}{1 - x_1},$$
$$y_3 = \frac{x_3}{1 - x_1 - x_2},$$
$$\cdots\cdots\cdots\cdots$$
$$y_{k-1} = \frac{x_{k-1}}{1 - x_1 - x_2 - \cdots - x_{k-2}}.$$

5. (*a*) Suppose that $X_1, \ldots, X_n$ are independent random variables and that $X_i$ has a gamma distribution with parameters $\alpha_i$ and $\beta$ $(i = 1, \ldots, n)$. Also, let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ be a random vector defined as follows:

$$Y_i = \frac{X_i}{X_1 + \cdots + X_n} \qquad i = 1, \ldots, n.$$

Show that $\mathbf{Y}$ and the random variable $X_1 + \cdots + X_n$ are independent and that $\mathbf{Y}$ has a Dirichlet distribution with parametric vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$.

(*b*) Suppose that the random vector $(X_1, \ldots, X_k)'$ has a Dirichlet distribution with parametric vector $\alpha = (\alpha_1, \ldots, \alpha_k)'$. For any given integer $r$ such that $2 \leq r \leq k$, let $\mathbf{Y} = (Y_1, \ldots, Y_r)'$ be a random vector defined as follows:

$$Y_i = \frac{X_i}{X_1 + \cdots + X_r} \qquad i = 1, \ldots, r.$$

Show that $\mathbf{Y}$ and the random variable $X_1 + \cdots + X_r$ are independent and that $\mathbf{Y}$ has a Dirichlet distribution with parametric vector $\alpha_r = (\alpha_1, \ldots, \alpha_r)'$.

6. Suppose that $X_1, \ldots, X_n$ is a random sample from the uniform distribution on the interval (0, 1), and let $Y_1 \leq \cdots \leq Y_n$ be the order statistics of this sample. Also, let $\mathbf{Z} = (Z_1, \ldots, Z_{n+1})'$ be the random vector defined as follows:

$$\begin{aligned} Z_1 &= Y_1, \\ Z_2 &= Y_2 - Y_1, \\ &\cdots\cdots\cdots \\ Z_n &= Y_n - Y_{n-1}, \\ Z_{n+1} &= 1 - Y_n. \end{aligned}$$

Show that $\mathbf{Z}$ has a Dirichlet distribution.

7. Suppose that $\mathbf{X} = (X_1, \ldots, X_k)'$ is a random vector whose p.d.f. is given by Eq. (1) of Sec. 5.4. Prove that $E(\mathbf{X}) = \boldsymbol{\mu}$ and that $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$.

8. Suppose that the random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $r$ be any integer such that $1 \leq r \leq k$. Also, let $\boldsymbol{\mu}_r$ be the subvector of $\boldsymbol{\mu}$ comprising the first $r$ components of $\boldsymbol{\mu}$, and let $\boldsymbol{\Sigma}_r$ be the submatrix of $\boldsymbol{\Sigma}$ comprising the elements in the first $r$ rows and first $r$ columns of $\boldsymbol{\Sigma}$. Show that the random vector $\mathbf{X}_r = (X_1, \ldots, X_r)'$ also has a normal distribution with mean vector $\boldsymbol{\mu}_r$ and covariance matrix $\boldsymbol{\Sigma}_r$.

9. Suppose that a nonsingular matrix $\boldsymbol{\Sigma}$ and its inverse $\boldsymbol{\Sigma}^{-1}$ have been partitioned as in Eq. (10) of Sec. 5.4. Show that

$$\mathbf{T}_{11}^{-1}\mathbf{T}_{12} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$$

and that

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

10. Assuming that the joint distribution of the random variables $X_1, \ldots, X_k$ is multivariate normal and that the correlation between any two of them is 0, show that $X_1, \ldots, X_k$ are independent.

11. Suppose that $\mathbf{X} = (X_1, \ldots, X_k)'$, where $X_1, \ldots, X_k$ are independent random variables each of which has a normal distribution with variance $\sigma^2$. Let $\mathbf{A}$ be an orthogonal $k \times k$ matrix (i.e., a matrix $\mathbf{A}$ such that $\mathbf{A}' = \mathbf{A}^{-1}$), and let the random vector $\mathbf{Y} = (Y_1, \ldots, Y_k)'$ be defined by the transformation $\mathbf{Y} = \mathbf{AX}$. Show that $Y_1, \ldots, Y_k$ are also independent random variables each of which has a normal distribution with the same variance $\sigma^2$.

12. Suppose that a $k$-dimensional random vector $\mathbf{X}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and nonsingular covariance matrix $\boldsymbol{\Sigma}$. Show that the random variable $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has a $\chi^2$ distribution with $k$ degrees of freedom.

13. Suppose that in a certain population of married couples, the heights of husbands and wives have a bivariate normal distribution. Express, in terms of the univariate standard normal distribution, the probability that if a couple is chosen at random, the husband's height will be greater than his wife's height. Assuming that the wife's height is $y$, find a similar expression for the conditional probability of this event. For what value of $y$ are the two probabilities equal?



14. Suppose that a random matrix $\mathbf{V}$ has a Wishart distribution whose p.d.f. is given by Eqs. (4) and (5) of Sec. 5.5. Show that for any positive integer $r$,

$$E(|\mathbf{V}|^r) = 2^{rk} \left\{ \prod_{j=1}^{k} \frac{\Gamma[(n+1-j+2r)/2]}{\Gamma[(n+1-j)/2]} \right\} |\boldsymbol{\Sigma}|^r.$$

15. Suppose that a $k \times k$ random matrix $\mathbf{V}$ has a Wishart distribution with $n$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$. Prove the following three properties:

(*a*) $E(\mathbf{V}) = n\boldsymbol{\Sigma}$.

(*b*) If $\mathbf{A}$ is any $m \times k$ matrix of constants, then the $m \times m$ random matrix $\mathbf{AVA}'$ has a Wishart distribution with $n$ degrees of freedom and parametric matrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$.

(*c*) If the matrices $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are partitioned as in Eq. (7) of Sec. 5.5, then the random matrix $\mathbf{V}_{11}$ has a Wishart distribution with $n$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}_{11}$.

16. Suppose that the $k \times k$ random matrices $\mathbf{V}_1, \ldots, \mathbf{V}_r$ are independent and that $\mathbf{V}_i$ has a Wishart distribution with $n_i$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$ $(i = 1, \ldots, r)$. Show that the sum $\mathbf{V}_1 + \cdots + \mathbf{V}_r$ also has a Wishart distribution with $n_1 + \cdots + n_r$ degrees of freedom and parametric matrix $\boldsymbol{\Sigma}$.

17. Let $g(\cdot | n, \mathbf{T})$ denote the p.d.f. of the Wishart distribution with $n$ degrees of freedom and precision matrix $\mathbf{T}$, as given by Eq. (11) of Sec. 5.5. Let $\boldsymbol{\Sigma} = \mathbf{T}^{-1}$, and assume that Eq. (6) of Sec. 5.5 holds even when the matrix $\mathbf{T}$ is replaced by any matrix of complex numbers whose real parts form a positive definite matrix. Under this assumption, derive the characteristic function $\zeta$ of the Wishart distribution, as given by Eq. (9) of Sec. 5.5.

18. Suppose that a random vector $\mathbf{X}$ has a multivariate $t$ distribution with $n$ degrees of freedom, location vector $\boldsymbol{\mu}$, and precision matrix $\mathbf{T}$. Prove that for $n > 2$, the mean vector $E(\mathbf{X})$ and the covariance matrix $\text{Cov}(\mathbf{X})$ exist and that their values are given by Eq. (11) of Sec. 5.6. *Hint:* To compute $E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']$, use the representation in Eq. (1) of Sec. 5.6 and the independence of $\mathbf{Y}$ and $Z$.

19. Suppose that a $k$-dimensional random vector $\mathbf{X}$ has a multivariate $t$ distribution with $n$ degrees of freedom, location vector $\boldsymbol{\mu}$, and precision matrix $\mathbf{T}$. If $\mathbf{X}$, $\boldsymbol{\mu}$, and $\mathbf{T}$ are partitioned as in Eqs. (12) and (13) of Sec. 5.6, show that the marginal distribution of $\mathbf{X}_1$ is a multivariate $t$ distribution with $n$ degrees of freedom, location vector $\boldsymbol{\mu}_1$, and precision matrix $\mathbf{T}_{11} - \mathbf{T}_{12}\mathbf{T}_{22}^{-1}\mathbf{T}_{21}$. *Hint:* From the discussion at the beginning of Sec. 5.6 and the discussion in Sec. 5.4, it follows that the conditional distribution of $\mathbf{X}_1$ when $Z$ is given is a multivariate normal distribution having the appropriate mean vector and the appropriate precision matrix.

20. Suppose that the $k$-dimensional random vector $\mathbf{X}$ has a multivariate $t$ distribution as specified in Exercise 19, and suppose that $\mathbf{X}$, $\boldsymbol{\mu}$, and $\mathbf{T}$ are partitioned as also specified in that exercise. Show that the conditional distribution of $\mathbf{X}_1$, when $\mathbf{X}_2 = \mathbf{x}_2$ is given $(\mathbf{x}_2 \in R^{k_2})$, is a multivariate $t$ distribution with $n + k_2$ degrees of freedom and that the location vector and the precision matrix are given by the expressions (14) and (15), respectively, of Sec. 5.6. *Hint:* Rewrite the quadratic form that appears in the joint p.d.f. of $\mathbf{X}_1$ and $\mathbf{X}_2$ as was done in Eq. (15) of Sec. 5.4.

21. If a random vector $\mathbf{X}$ has a multivariate $t$ distribution whose p.d.f. is given by Eqs. (9) and (10) of Sec. 5.6, show that the random variable which is displayed in the expression (16) of Sec. 5.6 has an $F$ distribution with $k$ and $n$ degrees of freedom. *Hint:* Use the representation in Eq. (1) of Sec. 5.6.

22. Suppose that the random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ has a $k$-dimensional multivariate $t$ distribution with $n$ degrees of freedom, location vector $\boldsymbol{\mu}$, and precision matrix $\mathbf{T}$. Also, let $\mathbf{A}$ be an $m \times k$ matrix such that $\mathbf{AT}^{-1}\mathbf{A}'$ is nonsingular, and let the random vector $\mathbf{U} = (U_1, \ldots, U_m)'$ be defined as $\mathbf{U} = \mathbf{AX}$. Show that $\mathbf{U}$ has an $m$-dimensional multivariate $t$ distribution with $n$ degrees of freedom, location vector $\mathbf{A}\boldsymbol{\mu}$, and precision matrix $(\mathbf{AT}^{-1}\mathbf{A}')^{-1}$.

23. Suppose that the joint distribution of two random variables $X_1$ and $X_2$ is a bilateral bivariate Pareto distribution whose p.d.f. is given by Eq. (1) of Sec. 5.7. Show that both the marginal distribution of $r_2 - X_1$ and the marginal distribution of $X_2 - r_1$ are Pareto distributions with parameters $r_2 - r_1$ and $\alpha$.

24. (*a*) Suppose that the random variables $X_1$ and $X_2$ have a bilateral bivariate Pareto distribution with parameters $r_1$, $r_2$, and $\alpha$. Show that if $\alpha > 2$,

$$E[(X_2 - X_1)^2] = \frac{\alpha(\alpha + 1)(r_2 - r_1)^2}{(\alpha - 1)(\alpha - 2)}.$$

(*b*) Show that if $\alpha > 2$, the correlation between $X_1$ and $X_2$ is $-1/\alpha$.

25. Suppose that $X_1$ and $X_2$ have a bilateral bivariate Pareto distribution with parameters $r_1$, $r_2$, and $\alpha$. For fixed values of $r_1$ and $r_2$, describe the limiting behavior of the joint distribution of $X_1$ and $X_2$ as $\alpha \to \infty$.

PART TWO

# subjective probability and utility