Chapter 1

Multivariate Normal and Related Distributions

In multivariate analysis we treat ordered sets (random vectors) of several random variables as a whole, rather than individual varaibles separately. First we introduce some basic concepts that describe the behavior of random vectors together with their properties: moments, cumulants, disrtibution functions, density functions, and characteristic functions.

The uniqueness theorem of characteristic function and the continuity theorem are given without proofs, and then the multivariate normal, spherical, and elliptical distributions are defined. They are important in studying asymptotic expansions of various statistics in non-normal models.

1.1 Random Vectors

1.1.1 Mean Vector and Covariance Matrix

Multivariate analysis deals with issues related to the observation of correlated random variables on units of a selected random sample. We denote a set of p random variables $X_1, \ldots, X_p$ by a vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$, which is called a *random vector*. The mean or expectation of $\boldsymbol{X}$ is defined to be the vector of expectations:

$$\mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} \mathrm{E}(X_1) \\ \vdots \\ \mathrm{E}(X_p) \end{pmatrix}.$$

A typical set of multivariate random samples, $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, arises from taking measurements on a $p \times 1$ random vector $\boldsymbol{X}$ for each of n objects or people. It is convenient to express these observation vectors in matrix form,

$$\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)' = \begin{pmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_n' \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix},$$

where the transpose of a vector $\boldsymbol{a}$ and a matrix A is denoted by $\boldsymbol{a}'$ and A', respectively.

Let $\mathbf{X}$ be a matrix of random variables, which we call a *random matrix*. Here the rows of $\mathbf{X}$ may or may not be random observations of $\boldsymbol{X}$. More generally, the expectation of a random matrix $\mathbf{X} = (X_{ij})$ is defined by the matrix whose (i, j)th element is $\mathrm{E}(X_{ij})$: namely, $\mathrm{E}(\mathbf{X}) = (\mathrm{E}(X_{ij}))$.

Theorem 1.1.1 *Let* $\mathbf{X} = (X_{ij})$ *and* $\mathbf{Y} = (Y_{ij})$ *be* $n \times p$ *random matrices. If* $\mathrm{A}, \mathrm{B},$ *and* C *are* $m \times n, p \times q,$ *and* $m \times q$ *matrices of constants, then:*

(1) $\mathrm{E}(\mathbf{X} + \mathbf{Y}) = \mathrm{E}(\mathbf{X}) + \mathrm{E}(\mathbf{Y})$.

(2) $\mathrm{E}(\mathrm{A}\mathbf{X}\mathrm{B} + \mathrm{C}) = \mathrm{A}\mathrm{E}(\mathbf{X})\mathrm{B} + \mathrm{C}$.

Proof. The (i, j)th element of the left-hand side of (1) can be expressed as

$$\mathrm{E}(X_{ij} + Y_{ij}) = \mathrm{E}(X_{ij}) + \mathrm{E}(Y_{ij}),$$

which is equal to the (i, j)th element of the right-hand side of (1). Similarly, the (i, j)th element of the left-hand side of (2) can be expressed as

$$\mathrm{E}\left(\sum_{k=1}^{n}\sum_{\ell=1}^{p} a_{ik} X_{k\ell} b_{\ell j} + c_{ij}\right) = \sum_{k=1}^{n}\sum_{\ell=1}^{p} a_{ik} \mathrm{E}(X_{k\ell}) b_{\ell j} + c_{ij},$$

which is equal to the (i, j)th element of the right-hand side of (2). □

Definition 1.1.1 If a $p \times 1$ random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ has mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$, the covariance matrix of $\boldsymbol{X}$ is defined by

$$\Sigma \equiv \mathrm{Var}(\boldsymbol{X}) = \mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'].$$

Furthermore, if a $q \times 1$ random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_q)'$ has mean $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$, the covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined by

$$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\eta})'].$$

In particular, $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}) = \mathrm{Var}(\boldsymbol{X})$.

The (i,j)th element of Σ is

$$\begin{aligned}\sigma_{ij} &= \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathrm{Cov}(X_i, X_j),\end{aligned}$$

the covariance between X_i and X_j, and the (i,i)th element is

$$\begin{aligned}\sigma_{ii} &= \mathrm{E}[(X_i - \mu_i)^2] \\ &= \mathrm{Var}(X_i),\end{aligned}$$

the variance of X_i. The variance is denoted by $\sigma_i^2 = \sigma_{ii}$ with $\sigma_i \geq 0$. The σ_i is called the *standardized deviation.* The covariance matrix of $\boldsymbol{X}$ is symmetric (i.e., $\Sigma' = \Sigma$) and positive semidefinite, as in Theorem 1.1.2.

Theorem 1.1.2 *Let Σ be the covariance matrix of a $p \times 1$ random vector.*

(1) *Σ is positive semidefinite (nonnegative definite), that is, for any $p \times 1$ fixed vector $\boldsymbol{a} = (a_1, \ldots, a_p)'$,*

$$\boldsymbol{a}'\Sigma\boldsymbol{a} = \sum_{i=1}^{p}\sum_{j=1}^{p} \sigma_{ij} a_i a_j \geq 0.$$

(2) *Let B be a $q \times p$ constant matrix and $\boldsymbol{b}$ be a $q \times 1$ constant vector. Then the covariance matrix of $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$ is*

$$\mathrm{Var}(\boldsymbol{Y}) = B\Sigma B'.$$

Proof. (1) For any constant vector $\boldsymbol{a}$, we have

$$\begin{aligned}\mathrm{Var}(\boldsymbol{a}'\boldsymbol{X}) &= \mathrm{E}[(\boldsymbol{a}'\boldsymbol{X} - \boldsymbol{a}'\boldsymbol{\mu})^2] \\ &= \mathrm{E}[\{\boldsymbol{a}'(\boldsymbol{X} - \boldsymbol{\mu})\}^2] \\ &= \mathrm{E}[\boldsymbol{a}'(\boldsymbol{X} - \boldsymbol{\mu}) \cdot (\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{a}] = \boldsymbol{a}'\Sigma\boldsymbol{a}.\end{aligned}$$

The result required follows from $\mathrm{Var}(\boldsymbol{a}'\boldsymbol{X}) \geq 0$.

(2) Using $\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}) = B(\boldsymbol{X} - \boldsymbol{\mu})$, we have

$$\mathrm{Var}(\boldsymbol{Y}) = \mathrm{E}[B(\boldsymbol{X} - \boldsymbol{\mu})\{B(\boldsymbol{X} - \boldsymbol{\mu})\}'] = B\Sigma B'. \qquad \square$$

The covariance matrix is said to be *positive definite* if

$$\boldsymbol{a}'\Sigma\boldsymbol{a} > 0 \quad \text{for any constant vector } \boldsymbol{a} \neq \boldsymbol{0}.$$

In general, the covariance matrix Σ is positive semidefinite. Therefore, it follows from [A.1.8] that the characteristic roots of Σ are nonnegative, and these are denoted by

$$\lambda_1 \geq \cdots \geq \lambda_p \geq 0.$$

For $i = 1, \ldots, p$, let γ_i be the characteristic vector corresponding to λ_i. We may assume without loss of generality that they are orthonormal: that is,

$$\gamma_i'\gamma_j = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j. \end{cases}$$

The characteristic roots and vectors satisfy

$$\Sigma\gamma_i = \lambda_i\gamma_i, \quad i = 1, \ldots, p,$$

which is expressed as

$$\Sigma\Gamma = \Gamma\Lambda, \tag{1.1.1}$$

where

$$\Lambda = \text{diag}\left(\lambda_1, \ldots, \lambda_p\right), \quad \Gamma = (\gamma_1, \ldots, \gamma_p).$$

Here the matrix Γ is normalized as $\Gamma'\Gamma = \text{I}_p$. Equation (1.1.1) implies that

$$\begin{aligned}\Sigma &= \Gamma\Lambda\Gamma' \\ &= \Gamma\Lambda^{1/2}\Lambda^{1/2}\Gamma' \\ &= \Gamma\Lambda^{1/2}\Gamma'\Gamma\Lambda^{1/2}\Gamma',\end{aligned}$$

where $\Lambda^{1/2} = \text{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p}\right)$. Defining $\Sigma^{1/2} = \Gamma\Lambda^{1/2}\Gamma'$ and putting $C = \Gamma\Lambda^{1/2}$, we can write

$$\Sigma = (\Sigma^{1/2})^2 = CC'.$$

Note that such expressions are used for any positive semidefinite matrix.

The covariance matrix Σ contains the variance of the p variables and the covariances between all pairs of variables. It is desirable to have a single measure for the overall multivariate scatter. One such is the generalized variance, defined as the *determinant* of the covariance matrix. It is expressed as

$$|\Sigma| = |\Gamma\Lambda\Gamma'| = |\Lambda| = \lambda_1 \cdots \lambda_p.$$

The other overall measure is

$$\begin{aligned}\text{tr}\ \Sigma &= \sigma_{11} + \cdots + \sigma_{pp} \\ &= \text{tr}\,\Gamma\Lambda\Gamma' = \text{tr}\,\Lambda\Gamma'\Gamma \\ &= \text{tr}\,\Lambda = \lambda_1 + \cdots + \lambda_p,\end{aligned}$$

which is called the *total variance.*

1.1.2 Characteristic Function and Distribution

Let $\boldsymbol{X}$ be a p-dimensional random vector. Suppose that the probability of the random point falling in any (measurable) set E in the p-dimensional Euclidean space R^p is expressed as

$$\Pr\{\boldsymbol{X} \in E\} = \int_E f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

where $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1 \cdots \mathrm{d}x_p$. Then the function $f(\boldsymbol{x})$ is called the *probability density function*, or simply, the *density* of $\boldsymbol{X}$. The characteristic function of $\boldsymbol{X}$ is defined by

$$C(\boldsymbol{t}) = \mathrm{E}[\mathrm{e}^{i\boldsymbol{t}'\boldsymbol{X}}],$$

where $i = \sqrt{-1}$, $\boldsymbol{t} = (t_1, \ldots, t_p)'$, and $-\infty < t_j < \infty, j = 1, \ldots, p$. It is well known that there exists, one-to-one correspondence between the distribution of $\boldsymbol{X}$ and its characteristic function. In particular, we have the following theorem.

Theorem 1.1.3 *If the $p \times 1$ random vector $\boldsymbol{X}$ has the density function $f(\boldsymbol{x})$ and the characteristic function $C(\boldsymbol{t})$, then*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-i\boldsymbol{t}'\boldsymbol{x}} C(\boldsymbol{t})\,\mathrm{d}t_1 \cdots \mathrm{d}t_p.$$

This shows that the characteristic function determines the density function uniquely. For proofs of this theorem and Theorems 1.1.4 and 1.1.5, see, for example, Lukacs (1970).

The characteristic function is also used to obtain various moments of $\boldsymbol{X}$. In fact, interchanging the integral formally with the differential, we have

$$\frac{\partial^m}{\partial t_1^{m_1} \cdots \partial t_p^{m_p}} C(\boldsymbol{t}) = \mathrm{E}\left[(iX_1)^{m_1} \cdots (iX_p)^{m_p} \mathrm{e}^{i\boldsymbol{t}'\boldsymbol{X}}\right],$$

where $m = m_1 + \cdots + m_p$, and putting $\boldsymbol{t} = \boldsymbol{0}$, we can expect to get the moment $\mathrm{E}[X_1^{m_1} \cdots X_p^{m_p}]$. The result is stated in the following theorem.

Theorem 1.1.4 *Suppose that $\mathrm{E}\left[|X_1|^{m_1} \cdots |X_p|^{m_p}\right]$ is finite; then the characteristic function is $(m_1, \ldots, m_p)$th times continuously differentiable, and*

$$\left.\frac{\partial^m}{\partial t_1^{m_1} \cdots \partial t_p^{m_p}} C(\boldsymbol{t})\right|_{\boldsymbol{t}=\boldsymbol{0}} = i^m \mathrm{E}\left[X_1^{m_1} \cdots X_p^{m_p}\right].$$

The characteristic function is useful for obtaining asymptotic distribution of a statistic when a parameter, typically the sample size, is large. The following result is known as the *continuity theorem*.

Theorem 1.1.5 *Let $\{F_n(\boldsymbol{x})\}$ be a sequence of distribution functions, and let $\{C_n(\boldsymbol{t})\}$ be the sequence of corresponding characteristic functions. A necessary and sufficient condition for $F_n(\boldsymbol{x})$ to converge to a distribution function $F(\boldsymbol{x})$ is that for every $\boldsymbol{t}$, $C_n(\boldsymbol{t})$ converges to a limit $C(\boldsymbol{t})$ that is continuous at $\boldsymbol{t} = \boldsymbol{0}$. When this condition is satisfied, the limit $C(\boldsymbol{t})$ is identical with the characteristic function of the limiting distribution function $F(\boldsymbol{x})$.*

In the distribution theory of multivariate statistics, the variable transformation method is often used, in addition to the characteristic function method. The method is based on the following well-known theorem.

Theorem 1.1.6 *Let $\boldsymbol{X}$ be a $p \times 1$ random vector having a density function $f(\boldsymbol{x})$ that is positive on a set $\mathcal{S} \subset R^p$. Suppose that the transformation $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), \ldots, y_p(\boldsymbol{x}))'$ is one-to-one of $\mathcal{S}$ onto $\mathcal{T}$, where $\mathcal{T}$ denotes the image of $\mathcal{S}$ under $\boldsymbol{y}$. Let $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{y})$ be the inverse transformation. Assume that partial derivatives $\partial x_i/\partial y_j$ $(i, j = 1, \ldots, p)$ exist and are continuous on $\mathcal{T}$. Then the density function of the random vector $\boldsymbol{Y} = \boldsymbol{y}(\boldsymbol{X})$ is given by*

$$g(\boldsymbol{y}) = f(\boldsymbol{x}(\boldsymbol{y}))|J(\boldsymbol{y} \to \boldsymbol{x})|,$$

where $J(\boldsymbol{y} \to \boldsymbol{x})$ is a determinant of the Jacobian matrix of the transformation from $\boldsymbol{y}$ to $\boldsymbol{x}$:

$$J(\boldsymbol{y} \to \boldsymbol{x}) = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_1}{\partial y_p} \\ \vdots & & \vdots \\ \dfrac{\partial x_p}{\partial y_1} & \cdots & \dfrac{\partial x_p}{\partial y_p} \end{vmatrix}.$$

1.2 Multivariate Normal Distribution

1.2.1 Bivariate Normal Distribution

Let X be distributed as a normal distribution with mean μ and variance σ^2 [i.e., $\mathrm{N}(\mu, \sigma^2)$]. Then, the probability of the event $a < X \leq b$ is expressed as

$$\Pr\{a < X \leq b\} = \int_a^b f(x)\,\mathrm{d}x,$$

where

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

Here $-\infty < \mu < \infty$ and $\sigma > 0$. The function $f(x)$ is called the *probability density function (pdf)*, or simply the *density* of X, and is often denoted in full by $f(x\colon \mu, \sigma^2)$.

The distribution of two-dimensional variables (X_1, X_2) is defined by the density function $f(x_1, x_2)$, satisfying

$$\Pr\{a < X_1 \le b,\ c < X_2 \le d\} = \int_a^b \int_c^d f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

for any $a < b$ and $c < d$. For the bivariate normal distribution, its density function is defined by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)}Q\right\}, \tag{1.2.1}$$

where

$$Q(x_1, x_2) = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2.$$

If (X_1, X_2) follows the bivariate normal distribution with density function (1.2.1), the mean and variance of X_j are μ_j and σ_j^2, respectively. The covariance and correlation between X_1 and X_2 are $\rho\sigma_1\sigma_2$ and ρ, respectively. The covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Furthermore, each of the distributions of X_1 and X_2 is normal. The conditional distribution of X_2 given $X_1 = x_1$ is also normal. The conditional means and conditional variance are

$$\mu_2 + \rho(\sigma_2/\sigma_1)\cdot(x_1 - \mu_1) \quad \text{and} \quad \sigma_2^2(1-\rho^2),$$

respectively.

These properties are summarized as follows:

(1) $\mathrm{E}(X_1) = \mu_1,\ \mathrm{Var}(X_1) = \sigma_1^2,$
$\mathrm{E}(X_2) = \mu_1,\ \mathrm{Var}(X_2) = \sigma_2^2.$

(2) $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2), \quad X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2).$

(3) $\mathrm{Cov}(X_1, X_2) = \sigma_1\sigma_2\rho,\ \mathrm{Corr}(X_1, X_2) = \rho.$

(4) $\mathrm{E}(X_2|X_1 = x_1) = \mu_2 + \rho(\sigma_2/\sigma_1)(x_1 - \mu_1),$
$\mathrm{E}(X_1|X_2 = x_2) = \mu_1 + \rho(\sigma_1/\sigma_2)(x_2 - \mu_2).$

(5) $\mathrm{Var}(X_2|X_1 = x_1) = \sigma_2^2(1-\rho^2),\ \ \mathrm{Var}(X_1|X_2 = x_2) = \sigma_1^2(1-\rho^2).$

(6) $X_2|X_1 = x_1 \sim \mathrm{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x_1 - \mu_1), \sigma_2^2(1-\rho^2)),$
$X_1|X_2 = x_2 \sim \mathrm{N}(\mu_1 + \rho(\sigma_1/\sigma_2)(x_2 - \mu_2), \sigma_1^2(1-\rho^2)).$

For the derivation of these properties, it is essential that the exponent Q in the density function of the *bivariate normal distribution* can be put in the following quadratic form:

$$Q = (1-\rho^2)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left\{\left(\frac{x_2-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right\}^2,$$

and hence the probability density function is expressed as

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left\{-\frac{1}{2\sigma_1^2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2\sigma_2^2\sqrt{1-\rho^2}}\left\{x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)\right\}^2\right]. \end{aligned}$$

Properties (1) through (6) can all be shown by using the decomposition above.

1.2.2 Definitions

The density function of a random variable Z with the standard normal distribution $\mathrm{N}(0,1)$ is

$$\varphi(z) \equiv \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}, \quad -\infty < z < \infty. \tag{1.2.2}$$

A random variable X with a general normal distribution with mean μ and variance σ^2 is obtained by the linear transformation

$$X = \sigma Z + \mu.$$

In fact, starting from (1.2.2) we can show that the density function of X is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

The approach is generalized as follows. The probability density function of $\boldsymbol{Z} = (Z_1,\ldots,Z_p)'$, where $Z_1,\ldots,Z_p$ are independent and identically distributed $\mathrm{N}(0,1)$, is given by

$$\prod_{i=1}^{p}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^p \mathrm{e}^{-\boldsymbol{z}'\boldsymbol{z}/2}. \tag{1.2.3}$$

Consider the transformation

$$\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}.$$

Then it is easy to see that

$$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu} \quad \text{and} \quad \mathrm{Cov}(\boldsymbol{X}) = \Sigma.$$

Furthermore, the Jacobian of transformation from $\boldsymbol{X}$ to $\boldsymbol{Z}$ is (see [A.3.1])

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = |\Sigma^{-1/2}| = |\Sigma|^{-1/2},$$

and the density function of $\boldsymbol{X}$ is given by

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \tag{1.2.4}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$, $-\infty < \mu_j < \infty$ and $\Sigma > 0$.

Definition 1.2.1 A p-dimensional random $\boldsymbol{X}$ is said to have a *p-variate normal distribution* with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ if its density function is given by (1.2.4).

Theorem 1.2.1 *Let $\boldsymbol{\mu}$ be a p-dimensional fixed vector and Σ be a $p \times p$ positive definite matrix. Then the following two statements are equivalent.*

(1) $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$.

(2) $\boldsymbol{Z} \equiv \Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$.

Proof. If $f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)$ is the probability density function of $\boldsymbol{X}$, the transformation $\boldsymbol{X} \to \boldsymbol{Z}$ gives

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = |\Sigma|^{-1/2} f(\boldsymbol{z}; \boldsymbol{0}, \mathrm{I}_p).$$

with the Jacobian equal to $|\Sigma|^{-1/2}$, which means (1) $\Longrightarrow$ (2). A similar argument can apply to proving (2) $\Longrightarrow$ (1). □

Suppose that $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. If Σ is singular, its density function is not defined. On the other hand, its characteristic function (see Theorem 1.2.2)

$$C_{\boldsymbol{X}}(\boldsymbol{t}) = i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t} \tag{1.2.5}$$

is still valid. To have a natural definition of a singular normal distribution, consider the following decomposition of Σ with rank r by [A.1.11]:

$$\Sigma = \Gamma \begin{pmatrix} \mathrm{D}_\lambda & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{pmatrix} \Gamma', \quad \Gamma = (\Gamma_1, \Gamma_2),$$

where Γ is a $p \times p$ orthogonal matrix, Γ_1 is an $a \times r$ matrix, and $\mathrm{D}_\lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_r)$, $\lambda_j > 0$, $j = 1, \ldots, r$. Putting $\mathrm{B} = \Gamma_1 \mathrm{D}_\lambda^{1/2}$, we have $\Sigma = \mathrm{BB}'$.

Note that B is not unique, since B can be replaced by BL, where L is any $r \times r$ orthogonal matrix. Using one such matrix, B, and $\boldsymbol{Z} \sim \mathrm{N}_r(\mathbf{0}, \mathrm{I}_r)$, consider

$$\tilde{\boldsymbol{X}} = \mathrm{B}\boldsymbol{Z} + \boldsymbol{\mu}. \tag{1.2.6}$$

Then the characteristic function of $\tilde{\boldsymbol{X}}$ is

$$\begin{aligned} C_{\tilde{\boldsymbol{X}}}(\boldsymbol{t}) &= \mathrm{E}[\exp\{i\boldsymbol{t}'(\mathrm{B}\boldsymbol{Z} + \boldsymbol{\mu})\}] \\ &= \exp\left\{i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'(\mathrm{BB}')\boldsymbol{t}\right\} = C_{\boldsymbol{X}}(\boldsymbol{t}). \end{aligned}$$

From this result a general p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ with rank $r \leq p$ is defined as the distribution of $\mathrm{B}\boldsymbol{Z} + \boldsymbol{\mu}$, where $\Sigma = \mathrm{BB}'$ and $\boldsymbol{Z} \sim \mathrm{N}_r(\mathbf{0}, \mathrm{I}_r)$. In this book it is assumed that the covariance matrix of a multivariate normal variable is nonsingular. However, we note that most properties are also valid for *singular normal distributions.*

1.2.3 Some Properties

Theorem 1.2.2 *Let $\boldsymbol{X}$ be distributed as $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Then:*

(1) $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}, \mathrm{Var}(\boldsymbol{X}) = \Sigma.$

(2) *The characteristic function of $\boldsymbol{X}$ is*

$$C_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\left(i\boldsymbol{\mu}'\boldsymbol{t} - \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right), \tag{1.2.7}$$

where $\boldsymbol{t} = (t_1, \ldots, t_p)'$ and t_j is a real number.

(3) *Let* B *be a $q \times p$ constant matrix with* $\mathrm{rank}(\mathrm{B}) = q$, *and $\boldsymbol{b}$ a q constant vector. Then $\boldsymbol{Y} = \mathrm{B}\boldsymbol{X} + \boldsymbol{b} \sim \mathrm{N}_q(\mathrm{B}\boldsymbol{\mu} + \boldsymbol{b}, \mathrm{B}\Sigma\mathrm{B}')$.*

Proof. From Theorem 1.2.1 we can write

$$\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu},$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$ and $Z_1, \ldots, Z_p \sim$ *i.i.d.* $\mathrm{N}(0, 1)$. Using this fact, it is easy to show (1). Writing $\boldsymbol{d} = (d_1, \ldots, d_p)' = \Sigma^{1/2}\boldsymbol{t}$, the characteristic function of $\boldsymbol{X}$ can be put in the form

$$\begin{aligned} C_{\boldsymbol{X}}(\boldsymbol{t}) &= \mathrm{E}[\mathrm{e}^{i\boldsymbol{t}'(\Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu})}] = \mathrm{e}^{i\boldsymbol{t}'\boldsymbol{\mu}}\mathrm{E}[\mathrm{e}^{i\boldsymbol{t}'\Sigma^{1/2}\boldsymbol{Z}}] = \mathrm{e}^{i\boldsymbol{t}'\boldsymbol{\mu}}\prod_{j=1}^{p}\mathrm{E}[\mathrm{e}^{id_jZ_j}] \\ &= \exp\left(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\sum_{j=1}^{p}d_j^2\right), \end{aligned}$$

which coincides with (1.2.7). For a proof of (3), consider the characteristic function of $\boldsymbol{Y} = \mathrm{B}\boldsymbol{X} + \boldsymbol{b}$ given by

$$\begin{aligned} \mathrm{E}\big[\exp\{i\boldsymbol{t}'(\mathrm{B}\boldsymbol{X} + \boldsymbol{b})\}\big] &= \exp(i\boldsymbol{t}'\boldsymbol{b})\exp\Big\{i(\mathrm{B}'\boldsymbol{t})'\boldsymbol{\mu} - \frac{1}{2}(\mathrm{B}'\boldsymbol{t})'\Sigma(\mathrm{B}'\boldsymbol{t})\Big\} \\ &= \exp\Big\{i\boldsymbol{t}'(\mathrm{B}\boldsymbol{\mu} + \boldsymbol{b}) - \frac{1}{2}\boldsymbol{t}'\mathrm{B}\Sigma\mathrm{B}'\boldsymbol{t}\Big\}. \end{aligned}$$

Result (3) follows from (2) and the fact that the distribution is determined (see Theorem 1.1.3) by its characteristic function. □

Theorem 1.2.3 *For $i = 1, \ldots, n$, let $\boldsymbol{X}_i$ be a sequence of mutually independent random vectors following $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$. Then, for any constant vector $\boldsymbol{a} = (a_1, \ldots, a_n)'$,*

$$\boldsymbol{Y} = a_1\boldsymbol{X}_1 + \cdots + a_n\boldsymbol{X}_n \sim \mathrm{N}_p(\boldsymbol{\eta}, \boldsymbol{a}'\boldsymbol{a}\Sigma),$$

where $\boldsymbol{\eta} = a_1\boldsymbol{\mu}_1 + \cdots + a_n\boldsymbol{\mu}_n$. In particular,

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma).$$

Proof. Let $C_{\boldsymbol{X}_j}(\boldsymbol{t})$ be the characteristic function of $\boldsymbol{X}_j$. Then, using Theorem 1.2.2(2), the characteristic function of $\boldsymbol{Y}$ can be expressed as

$$\prod_{j=1}^{n} C_{\boldsymbol{X}_j}(a_j\boldsymbol{t}) = \exp\Big(i\,\boldsymbol{\eta}'\boldsymbol{t} - \frac{1}{2}\boldsymbol{a}'\boldsymbol{a}\boldsymbol{t}'\Sigma\boldsymbol{t}\Big),$$

which proves the result required. □

The central moments of $\boldsymbol{X}$ can be obtained by using Theorem 1.1.4 and the fact that the characteristic function of $\boldsymbol{X} - \boldsymbol{\mu}$ is $\exp(-\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t})$. Some moments are given in the following theorem.

Theorem 1.2.4 *When $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$:*

(1) $\mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ij}$.

(2) $\mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_\ell - \mu_\ell)] = \sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}$.

(3) $\mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_\ell - \mu_\ell)(X_s - \mu_s)(X_t - \mu_t)]$

$$\begin{aligned} &= \sigma_{ij}\sigma_{k\ell}\sigma_{st} + \sigma_{ij}\sigma_{ks}\sigma_{\ell t} + \sigma_{ij}\sigma_{kt}\sigma_{\ell s} \\ &\quad + \sigma_{ik}\sigma_{j\ell}\sigma_{st} + \sigma_{ik}\sigma_{js}\sigma_{\ell t} + \sigma_{ik}\sigma_{jt}\sigma_{\ell s} \\ &\quad + \sigma_{i\ell}\sigma_{jk}\sigma_{st} + \sigma_{i\ell}\sigma_{js}\sigma_{kt} + \sigma_{i\ell}\sigma_{jt}\sigma_{kt} \\ &\quad + \sigma_{is}\sigma_{jk}\sigma_{\ell t} + \sigma_{is}\sigma_{j\ell}\sigma_{kt} + \sigma_{is}\sigma_{jt}\sigma_{k\ell} \\ &\quad + \sigma_{it}\sigma_{jk}\sigma_{\ell s} + \sigma_{it}\sigma_{j\ell}\sigma_{ks} + \sigma_{it}\sigma_{js}\sigma_{k\ell}. \end{aligned}$$

(4) *Every central moment of odd order is* 0.

Theorem 1.2.5 *Let $\boldsymbol{X}$ be $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ and decompose $\boldsymbol{X}$, $\boldsymbol{\mu}$, and Σ as*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\boldsymbol{X}_i : p_i \times 1$, $\boldsymbol{\mu}_i : p_i \times 1$, and $\Sigma_{ij} : p_i \times p_j$. Then the conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ is normal with mean $\mathrm{E}(\boldsymbol{X}_2|\boldsymbol{X}_1)$ and covariance matrix $\mathrm{Var}(\boldsymbol{X}_2|\boldsymbol{X}_1)$, where

$$\begin{aligned} \mathrm{E}(\boldsymbol{X}_2|\boldsymbol{X}_1) &= \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1), \\ \mathrm{Var}(\boldsymbol{X}_2|\boldsymbol{X}_1) &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \equiv \Sigma_{22\cdot 1}. \end{aligned}$$

Proof. From Theorem 1.2.2 (3)

$$\boldsymbol{X}_1 \sim \mathrm{N}(\boldsymbol{\mu}_1, \Sigma_{11}).$$

Put

$$\mathrm{M} = \begin{pmatrix} \mathrm{I}_{p_1} & \mathrm{O} \\ -\Sigma_{21}\Sigma_{11}^{-1} & \mathrm{I}_{p_2} \end{pmatrix}.$$

Then

$$\mathrm{M}\Sigma\mathrm{M}' = \begin{pmatrix} \Sigma_{11} & \mathrm{O} \\ \mathrm{O} & \Sigma_{22\cdot 1}. \end{pmatrix}.$$

and hence $|\Sigma| = |\Sigma_{11}| \cdot |\Sigma_{22\cdot 1}|$. Furthermore, from [A.1.2],

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{pmatrix} + \begin{pmatrix} -\Sigma_{11}^{-1}\Sigma_{12} \\ \mathrm{I}_{p_2} \end{pmatrix} \Sigma_{22\cdot 1}^{-1} \Big(-\Sigma_{21}\Sigma_{11}^{-1} \ \mathrm{I}_{p_2} \Big).$$

These imply that the density function of $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ can be decomposed as

$$\begin{aligned} f_p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) &= f_{p_1}(\boldsymbol{x}_1; \boldsymbol{\mu}_1, \Sigma_{11}) \\ &= \times f_{p_2}(\boldsymbol{x}_2; \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22\cdot 1}). \end{aligned}$$

Using the decomposition, we can get the result required. □

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent p-dimensional normal vectors with means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$ and the same covariance matrix Σ. Put $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$, and consider the transformation

$$\mathbf{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)' = \mathrm{H}\mathbf{X}, \tag{1.2.8}$$

where H is an $n \times n$ orthogonal matrix. Then $\mathbf{Y}$ has the same properties as $\mathbf{X}$ except that the mean of $\mathbf{Y}$ is changed to $\mathrm{E}(\mathbf{Y}) = \mathrm{HE}(\mathbf{X})$. More precisely, the result is stated in the following theorem.

Theorem 1.2.6 *Let* $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$ *be an* $n \times p$ *random matrix whose rows are independent normal variates with mean* $\mathrm{E}(\mathbf{X}) = \mathrm{M}$ *and the same covariance matrix* Σ. *Let* $\mathbf{Y}$ *be the random matrix in* (1.2.8) *obtained by an* $n \times n$ *orthogonal transformation* H. *Then the rows of* $\mathbf{Y}$ *are independent normal variates with mean* $\mathrm{E}(\mathbf{Y}) = \mathrm{HM}$ *and the same covariance matrix* Σ.

Proof. The density function of $\mathbf{X}$ is expressed as

$$\prod_{i=1}^{n}(2\pi)^{-p/2}|\Sigma|^{-1/2}\,\mathrm{etr}\Big\{-\frac{1}{2}\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_i)(\boldsymbol{x}_i - \boldsymbol{\mu}_i)'\Big\}$$
$$= (2\pi)^{-pn/2}|\Sigma|^{-n/2}\,\mathrm{etr}\Big\{-\frac{1}{2}\Sigma^{-1}(\mathrm{X}-\mathrm{M})'(\mathrm{X}-\mathrm{M})\Big\},$$

where $\mathrm{etr}\,\bullet$ represents $\exp(\mathrm{tr}\,\bullet)$ and $\mathrm{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ is an $n \times p$ matrix. From the transformation (1.2.8) we have $\mathrm{X} = \mathrm{H}'\mathrm{Y}$, or equivalently

$$\begin{pmatrix} \boldsymbol{x}_{(1)} \\ \vdots \\ \boldsymbol{x}_{(p)} \end{pmatrix} = \mathrm{Diag}\,(\mathrm{H}', \ldots, \mathrm{H}') \begin{pmatrix} \boldsymbol{y}_{(1)} \\ \vdots \\ \boldsymbol{y}_{(p)} \end{pmatrix}, \tag{1.2.9}$$

where, for $j = 1, \ldots, p$, $\boldsymbol{x}_{(j)}$ and $\boldsymbol{y}_{(j)}$ denote the jth column vector of X and Y, respectively, that is, $\boldsymbol{x}_{(j)} = (x_{1j}, \ldots, x_{nj})'$ and $\boldsymbol{y}_{(j)} = (y_{1j}, \ldots, y_{nj})'$. Note, in particular, that the both sides of (1.2.9) are np-vectors. Then the Jacobian of the transformation is

$$\frac{\partial \mathrm{X}}{\partial \mathrm{Y}} = |\mathrm{H}'|^n = \pm 1.$$

Therefore the density function of $\mathbf{Y}$ is given by

$$(2\pi)^{-pn/2}|\Sigma|^{-n/2}\,\mathrm{etr}\{-\frac{1}{2}\Sigma^{-1}(\mathrm{H}'\mathrm{Y}-\mathrm{M})'(\mathrm{H}'\mathrm{Y}-\mathrm{M})\}$$
$$= (2\pi)^{-pn/2}|\Sigma|^{-n/2}\,\mathrm{etr}\{-\frac{1}{2}\Sigma^{-1}(\mathrm{Y}-\mathrm{HM})'(\mathrm{Y}-\mathrm{HM})\},$$

since

$$\begin{aligned}(\mathrm{H}'\mathrm{Y}-\mathrm{M})'(\mathrm{H}'\mathrm{Y}-\mathrm{M}) &= \{\mathrm{H}'(\mathrm{Y}-\mathrm{HM})\}'\{\mathrm{H}'(\mathrm{Y}-\mathrm{HM})\}' \\ &= (\mathrm{Y}-\mathrm{HM})'\mathrm{HH}'(\mathrm{Y}-\mathrm{HM}) \\ &= (\mathrm{Y}-\mathrm{HM})'(\mathrm{Y}-\mathrm{HM}).\end{aligned}$$

This gives the result required. □

Definition 1.2.2 The distribution of an $n \times p$ random matrix $\mathbf{X} = (\boldsymbol{X}_{(1)}, \ldots, \boldsymbol{X}_{(p)})$ is said to be normal if the random vector defined by

$$\mathrm{vec}(\mathbf{X}) \equiv \begin{pmatrix} \boldsymbol{X}_{(1)} \\ \vdots \\ \boldsymbol{X}_{(p)} \end{pmatrix}$$

follows an np-variate normal distribution. If this is the case we simply write $\mathbf{X} \sim \mathrm{N}_{n\times p}(\Xi, \Psi)$ with $\Xi = \mathrm{E}(\mathbf{X})$ and $\Psi = \mathrm{Var}(\mathrm{vec}(\mathbf{X}))$.

By using the notation of *matrix normal distribution*, Theorem 1.2.6 can be stated as follows:

$$\mathbf{X} \sim \mathrm{N}_{n\times p}(\Xi, \Sigma \otimes \mathrm{I}_n) \Longrightarrow \mathbf{Y} = \mathrm{H}\mathbf{X} \sim \mathrm{N}_{n\times p}(\mathrm{H}\Xi, \Sigma \otimes \mathrm{I}_n),$$

where H is an orthogonal matrix. Here $\otimes$ denotes the Kronecker or direct product; that is, for matrices $\mathrm{A} = (a_{ij})$ and B, $\mathrm{A} \otimes \mathrm{B} = (a_{ij}\mathrm{B})$.

Theorem 1.2.7 *Let* $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$ *be a n-vector consisting of mutually independent* $\boldsymbol{X}_i$*'s such that* $\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\xi}_i, \Sigma)$. *Let* B *be a* $q \times n$ *matrix with rank* q. *Then the distribution of* $\mathbf{Y} = \mathrm{B}\mathbf{X}$ *is normal with* $\mathrm{E}(\mathbf{Y}) = \mathrm{B}\Xi$ *and*

$$\mathrm{Cov}(\boldsymbol{Y}_{(i)}, \boldsymbol{Y}_{(j)}) = \sigma_{ij}\mathrm{BB}',$$

where $\Xi = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)'$, $\mathbf{Y} = (\boldsymbol{Y}_{(1)}, \ldots, \boldsymbol{Y}_{(p)})$, *and* $\mathrm{V} = \mathrm{BB}' = (v_{ij})$.

Proof. The density function of $\mathbf{X}$ is expressed as

$$f(\mathrm{X}) = 2^{pn/2}|\Sigma|^{-n/2}\,\mathrm{etr}\Big\{-\frac{1}{2}\Sigma^{-1}(\mathrm{X} - \Xi)'(\mathrm{X} - \Xi)\Big\}.$$

Let $\mathrm{T} = (t_{ij})$ be a $q \times p$ real matrix. Then the characteristic function of $\mathbf{Y}$ is

$$C_{\mathbf{Y}}(\mathrm{T}) = \int_{\mathrm{X}} \mathrm{etr}\big(i\mathrm{T}'\mathrm{BX}\big) f(\mathrm{X})\,\mathrm{dX}.$$

Note that

$$\begin{aligned}\mathrm{tr}\,\Sigma^{-1}(\mathrm{X} &- \Xi)'(\mathrm{X} - \Xi) - 2i\,\mathrm{tr}\,\mathrm{T}'\mathrm{BX} \\ &= \mathrm{tr}\,\Sigma^{-1}\{\mathrm{X} - (\Xi + i\mathrm{B}'\mathrm{T}\Sigma)\}'\{\mathrm{X} - (\Xi + i\mathrm{B}'\mathrm{T}\Sigma)\} \\ &\qquad -2i\ \mathrm{tr}\,\mathrm{T}'\mathrm{B}\Xi + \mathrm{tr}\,\Sigma\mathrm{T}'\mathrm{BB}'\mathrm{T}.\end{aligned}$$

Therefore, we have

$$C_{\mathbf{Y}}(\mathrm{T}) = \exp\Big\{i\ \mathrm{tr}\,\mathrm{T}'\mathrm{B}\Xi - \frac{1}{2}\,\mathrm{tr}\,\Sigma\mathrm{T}'\mathrm{BB}'\mathrm{T}\Big\}.$$

The required result is obtained by noting that

$$\mathrm{tr}\,\Sigma\mathrm{T}'\mathrm{BB}'\mathrm{T} = \mathrm{vec}(\mathrm{T})'[\Sigma \otimes \mathrm{BB}']\,\mathrm{vec}(\mathrm{T}). \qquad \square$$

1.3 Spherical and Elliptical Distributions

It is easily seen that the density function of a p-variate normal distribution $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ is constant on the ellipsoids

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = k$$

for every positive value of k in a p-dimensional Euclidean space. A general class of distribution with this property is the class of elliptically countered distributions with density function of the form

$$\Lambda^{-1/2} g[(\boldsymbol{x}-\boldsymbol{\nu})'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\nu})], \tag{1.3.1}$$

where Λ is a positive definite matrix, $g(\cdot) \geq 0$, and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{z}'\boldsymbol{z}) \mathrm{d}z_1 \cdots \mathrm{d}z_p = 1. \tag{1.3.2}$$

Further, if $\boldsymbol{X}$ has the density function (1.3.1), then considering the transformed variate

$$\boldsymbol{Z} = \Lambda^{-1/2}(\boldsymbol{X} - \boldsymbol{\nu}), \tag{1.3.3}$$

we can see that the density function of $\boldsymbol{Z}$ is $g(\boldsymbol{z}'\boldsymbol{z})$, and the contours of constant density of $\boldsymbol{Z}$ are spheres centered at the origin. In other words, the distribution of $\boldsymbol{Z}$ belongs to the class of spherically contoured distributions. From (1.3.3) $\boldsymbol{X}$ is expressed in terms of $\boldsymbol{Z}$ as

$$\boldsymbol{X} = \boldsymbol{\nu} + \Lambda^{1/2}\boldsymbol{Z}. \tag{1.3.4}$$

Therefore, the class of elliptically contoured distributions is, in fact, derived from the smaller class of spherically contoured distributions through relation (1.3.4).

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$ be distributed as $\mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$. Then $\boldsymbol{U} = \boldsymbol{Z}/\|\boldsymbol{Z}\|$ is a distribution on the sphere $\boldsymbol{u}'\boldsymbol{u} = 1$ satisfying
"The distributions of $\boldsymbol{U}$ and $\mathrm{H}\boldsymbol{U}$ are identical for all $\mathrm{H} \in \mathrm{O}(p)$",
where $\mathrm{O}(p)$ is the set of orthogonal matrices of order p. Such a distribution is called a *spherical distribution.* It is known [see Muirhead, 1982] that the spherical distribution on the unit sphere is determined uniquely.

Definition 1.3.1 A $p \times 1$ random vector $\boldsymbol{Z}$ is said to have a spherical distribution if $\boldsymbol{Z}$ and $\mathrm{H}\boldsymbol{Z}$ have the same distribution for all $p \times p$ orthogonal matrices H.

In the following we treat a spherical distribution with a density function.

Theorem 1.3.1 *The following statements are equivalent.*

(1) *$\boldsymbol{Z}$ has a spherical distribution.*

(2) *The density function of $\boldsymbol{Z}$ is expressed as $g(\boldsymbol{z}'\boldsymbol{z})$ for some scalar function $g(\cdot)$.*

(3) *The characteristic function of $\boldsymbol{Z}$ is of the form $\phi(\boldsymbol{t}'\boldsymbol{t})$ for some scalar function $\phi(\cdot)$.*

Proof. Let $f(\boldsymbol{z})$ be the density function of $\boldsymbol{Z}$. For any orthogonal matrix H, the density function of $\boldsymbol{X} = \mathrm{H}\boldsymbol{Z}$ is given by $f(\mathrm{H}'\boldsymbol{x})$. If (1) is true, $f(\boldsymbol{z}) = f(\mathrm{H}'\boldsymbol{z})$. Taking H whose first column is $(\boldsymbol{z}'\boldsymbol{z})^{-1/2}\boldsymbol{z}$, we can see that (2) is true, since $\mathrm{H}'\boldsymbol{z} = ((\boldsymbol{z}'\boldsymbol{z})^{1/2}, 0, \ldots, 0)'$. The converse is shown similarly. If (1) is true, the characteristic functions of $\boldsymbol{Z}$ and $\mathrm{H}\boldsymbol{Z}$ are the same, i.e. $C_{\boldsymbol{Z}}(\boldsymbol{t}) = C_{\mathrm{H}\boldsymbol{Z}}(\boldsymbol{t}) = C_{\boldsymbol{Z}}(\mathrm{H}'\boldsymbol{t})$. Therefore, we have (3). If (3) is true, it is easy to see that $C_{\boldsymbol{Z}}(\boldsymbol{t}) = C_{\mathrm{H}\boldsymbol{Z}}(\boldsymbol{t})$, and hence (1). □

Theorem 1.3.2 *Suppose that $\boldsymbol{X}$ has a p-variate spherical distribution with* $\Pr\{\boldsymbol{X} = \boldsymbol{0}\} = 0$. *Then*

$$R \equiv (\boldsymbol{X}'\boldsymbol{X})^{1/2} \quad and \quad \boldsymbol{U}(\boldsymbol{X}) = R^{-1}\boldsymbol{X}$$

are mutually independent and the latter follows a spherical distribution on $\mathcal{S}_p = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} = 1\}$.

Proof. The second assertion follows from the identities

$$\boldsymbol{U}(\mathrm{H}\boldsymbol{X}) = (\boldsymbol{X}'\mathrm{H}'\mathrm{H}\boldsymbol{X})^{-1/2}\mathrm{H}\boldsymbol{X} = \mathrm{H}\boldsymbol{U}(\boldsymbol{X}),$$

which hold for any orthogonal matrix H. For the independence, see Muirhead (1982). □

Definition 1.3.2 Let $\boldsymbol{\mu}$ be a $p \times 1$ constant vector, and A be a $p \times p$ constant matrix. Random vector $\boldsymbol{X}$ is said to have an *elliptical distribution* with parameter $\boldsymbol{\mu}$ and $\Lambda = \mathrm{A}'\mathrm{A}$, if it can be put in the form $\boldsymbol{X} = \boldsymbol{\mu} + \mathrm{A}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a random vector following a spherical distribution. If this is the case, we write $\boldsymbol{X} \sim \mathrm{E}_p(\boldsymbol{\mu}, \Lambda)$.

Theorem 1.3.3 *Statements* (1) *through* (3) *are equivalent.*

(1) $\boldsymbol{X} \sim \mathrm{E}_p(\boldsymbol{\mu}, \Lambda)$.

(2) *The probability density function of $\boldsymbol{X}$ is of the form* $|\Lambda|^{-1/2}g((\boldsymbol{x} - \boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x} - \boldsymbol{\mu}))$.

(3) *The characteristic function of $\boldsymbol{X}$ is of the form* $\exp\left(i\boldsymbol{t}'\boldsymbol{\mu}\right)\phi(\boldsymbol{t}'\Lambda\boldsymbol{t})$.

Proof. By expression $\boldsymbol{X} = \boldsymbol{\mu} + R\Lambda\boldsymbol{U}$, the (1), (2), and (3) can be stated in terms of $\boldsymbol{Z} = R\boldsymbol{U}$. The results follow from Theorem 1.3.1. □

The coordinate transformation in the case of p-variate is defined by

$$\begin{aligned} x_1 &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-2}\sin\theta_{p-1}, \\ x_2 &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-2}\cos\theta_{p-1}, \\ x_3 &= r\sin\theta_1\sin\theta_2\cdots\cos\theta_{p-2}, \\ &\vdots \\ x_{p-1} &= r\sin\theta_1\cos\theta_2, \\ x_p &= r\cos\theta_1, \end{aligned} \tag{1.3.5}$$

where $r > 0$, $0 < \theta_i < \pi$ $(i = 1, \ldots, p-2)$, $0 < \theta_{p-1} < 2\pi$.

Theorem 1.3.4 *If $\boldsymbol{X}$ is $\mathrm{E}_p(\mathbf{0}, \mathrm{I}_p)$ with density function $c_p h(\boldsymbol{x}'\boldsymbol{x})$, the random variables R, $\Theta_1, \ldots, \Theta_{p-1}$ defined by* (1.3.5) *are independent. Further, the marginal density of Θ_k is*

$$\frac{\Gamma[\frac{1}{2}(p-k)]\cos^{p-k-1}\theta}{\{\Gamma(\frac{1}{2})\Gamma[\frac{1}{2}(p-k-1)]\}}, \quad k = 1, 2, \ldots, p-2,$$

and Θ_{p-1} is $1/(2\pi)$. The marginal density of R^2 is

$$f(y) = \frac{c_p\pi^{p/2}}{\Gamma(\frac{1}{2}p)}y^{p/2-1}h(y), \quad y > 0. \tag{1.3.6}$$

Proof. The Jacobian of the transformation from $x_1, \ldots, x_p$ to $r, \theta_1, \ldots,$ θ_{p-1} given by (1.3.5) is $r^{p-1}\sin^{p-2}\theta_1\sin^{p-3}\theta_2\cdots\sin\theta_{p-2}$ (see Problem 1.7). It follows that the joint density function of R^2, $\Theta_1, \ldots, \Theta_{p-1}$ is

$$\frac{1}{2}c_p(r^2)^{p/2-1}\sin^{p-2}\theta_1\sin^{p-3}\theta_2\cdots\sin\theta_{p-2}h(r^2). \tag{1.3.7}$$

This shows that R^2, $\Theta_1, \ldots, \Theta_{p-1}$ are mutually independent and Θ_k has density function proportional to $\sin^{p-1-k}\theta_k$. Integration (1.3.7) with respect to $\theta_1, \ldots, \theta_{p-1}$ yields the factor $2\pi^{p/2}/\Gamma(\frac{1}{2}p)$, which is the surface area of a sphere of unit radius in R^p. Therefore, R^2 has the density function given by (1.3.6). The marginal density function of the Θ_i's is found by noting that

$$\int_{-\pi/2}^{\pi/2}\cos^{h-1}\theta\,\mathrm{d}\theta = \frac{\Gamma(\frac{1}{2}h)\Gamma(\frac{1}{2})}{\Gamma[\frac{1}{2}(h+1)]}.$$ □

When $\boldsymbol{X} \sim \mathrm{E}_p(\boldsymbol{\mu}, \Lambda)$, we can write

$$\boldsymbol{X} = \boldsymbol{\mu} + R\Lambda\boldsymbol{U}, \tag{1.3.8}$$

where $R \geq 0$, $\boldsymbol{U}$ has a spherical distribution, and R and $\boldsymbol{U}$ are independent. Further from the uniqueness of a spherical distribution we may let

$$\boldsymbol{U} = \boldsymbol{Z}/\|\boldsymbol{Z}\|, \quad \boldsymbol{Z} \sim \mathrm{N}(\mathbf{0}, \mathrm{I}_p).$$

Using (1.3.8), we have the following theorem.

Theorem 1.3.5 *If* $\boldsymbol{X} \sim \mathrm{E}_p(\boldsymbol{\mu}, \Lambda)$, *then*

(1) $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$.

(2) $\mathrm{Var}(\boldsymbol{X}) = (1/p)\mathrm{E}(R^2)\Lambda = \Sigma$.

Further, using that the characteristic function in Theorem 1.3.3(3), we have

$$\mathrm{E}(R^2) = -2p\phi'(0).$$

Theorem 1.3.6 *Let* $\boldsymbol{X} \sim \mathrm{E}_p(\boldsymbol{\mu}, \Lambda)$, *and* B *be a* $q \times p$ *constant matrix with rank* q. *Then*

$$\boldsymbol{Y} = \mathrm{B}\boldsymbol{X} \sim \mathrm{E}_p(\mathrm{B}\boldsymbol{\mu}, \mathrm{B}\Lambda\mathrm{B}').$$

Proof. The characteristic function of $\boldsymbol{Y}$ is expressed as

$$C_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{E}[\mathrm{e}^{i\boldsymbol{t}'\mathrm{B}\boldsymbol{X}}] = \mathrm{e}^{i\boldsymbol{t}'\mathrm{B}\boldsymbol{\mu}}\phi(\boldsymbol{t}'\mathrm{B}\Lambda\mathrm{B}\boldsymbol{t}).$$

The desired result follows from the uniqueness of the characteristic function. □

Theorem 1.3.7 *Let* $\boldsymbol{X} \sim \mathrm{E}_p(\boldsymbol{\mu}, \Lambda)$ *and decompose* $\boldsymbol{X}$, $\boldsymbol{\mu}$, Σ *as*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\boldsymbol{X}_i : p_i \times 1$, $\boldsymbol{\mu}_i : p_i \times 1$, $\Sigma_{ij} : p_i \times p_j$. *Then the conditional distribution of* $\boldsymbol{X}_1$ *given* $\boldsymbol{X}_2$ *is an elliptical distribution. The conditional mean and the conditional variance are given by*

$$\begin{aligned} \mathrm{E}(\boldsymbol{X}_1|\boldsymbol{X}_2) &= \boldsymbol{\mu}_1 + \Lambda_{12}\Lambda_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \\ \mathrm{Var}(\boldsymbol{X}_1|\boldsymbol{X}_2) &= h(\boldsymbol{X}_2)\Lambda_{11\cdot 2}. \end{aligned}$$

where $\Lambda_{11\cdot 2} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$.

Proof. From Theorem 1.3.6 it follows that $\boldsymbol{X}_i \sim \mathrm{E}_{p_i}(\boldsymbol{\mu}_i, \Lambda_{ii})$, $i = 1, 2$. Note that

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = a(\boldsymbol{x}_1|\boldsymbol{x}_2) + (\boldsymbol{x}_2-\boldsymbol{\mu}_2)'\Lambda_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2),$$

where

$$\begin{aligned} a(\boldsymbol{x}_1|\boldsymbol{x}_2) &= [\boldsymbol{x}_1-\boldsymbol{\mu}_1-\Lambda_{12}\Lambda_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)]' \\ &\quad \times\Lambda_{22\cdot1}^{-1}[\boldsymbol{x}_1-\boldsymbol{\mu}_1-\Lambda_{12}\Lambda_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)]. \end{aligned}$$

Therefore, the conditional density function of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is written as

$$\frac{g((\boldsymbol{x}-\boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}{g((\boldsymbol{X}_2-\boldsymbol{\mu}_2)'\Lambda_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}_2))} = g_{1|2}(a(\boldsymbol{x}_1|\boldsymbol{X}_2)).$$

This proves the result required. □

Example 1.3.1 (Elliptical Distributions)

(1) Multivariate t-distribution. Let

$$\boldsymbol{X} = \boldsymbol{\mu} + \mathrm{A}\boldsymbol{Z}, \quad \boldsymbol{Z} = \frac{1}{S}\boldsymbol{Y},$$

where $mS^2 \sim \chi^2(m)$, $\boldsymbol{Y} \sim \mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$, and S and $\boldsymbol{Y}$ are independent. Then the pdf of $\boldsymbol{X}$ is given by

$$\frac{\Gamma(\frac{1}{2}(m+p))}{\Gamma(\frac{1}{2}m)m^{p/2}\pi^{p/2}} \mid \Lambda \mid^{-1/2} \left[1 + \frac{1}{m}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-(m+p)/2}.$$

(2) The ϵ-contaminated p-variate normal. This is a mixture distribution of $\mathrm{N}_p(\boldsymbol{\mu}, \Lambda)$ and $\mathrm{N}_p(\boldsymbol{\mu}, c\Lambda)$ where the pdf is given by

$$\begin{aligned} &(1-\epsilon)(2\pi)^{-p/2} \mid \Lambda \mid^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \\ &\qquad +\epsilon(2\pi)^{-p/2} \mid c\Lambda \mid^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(c\Lambda)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \end{aligned}$$

1.4 Multivariate Cumulants

In general, the cumulants of a statistic play an important role in the derivation of asymptotic expansions for its distribution. In this section we consider expressing the cumulants of a random vector as a population distribution in terms of its moments.

First we consider the case of a univariate variable X. The characteristic function of X is defined as

$$C(t) = \mathrm{E}\left[\mathrm{e}^{itX}\right]$$

for all real numbers. In the following it is assumed for simplicity that all the moments exist. In general, if X has an rth moment, the results up to the r moments or cumulants are assured. Using

$$\mathrm{e}^z = 1 + z + \frac{1}{2!}(z)^2 + \frac{1}{3!}(z)^3 + \cdots,$$

the characteristic function can be expanded as

$$C(t) = \mathrm{E}\left[\sum_{r=0}^{\infty} \frac{1}{r!}(itX)^r\right] = \sum_{r=0}^{\infty} \mu_r \frac{(it)^r}{r!},$$

where μ_r is the rth moment of x, $\mu_r = \mathrm{E}[X^r]$. The *cumulants* of X are the coefficients κ_r in

$$\log C(t) = \sum_{r=1}^{\infty} \kappa_r \frac{(it_j)^r}{r!}.$$

The first four cumulants in terms of the moments are

$$\begin{aligned}
\kappa_1 &= \mu_1 = \mu, \\
\kappa_2 &= \mu_2 - \mu_1^2 = \sigma^2, \\
\kappa_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3, \\
\kappa_4 &= \mu_3 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4.
\end{aligned}$$

The skewness $\kappa_{[3]}$ and kurtosis $\kappa_{[4]}$ of X are

$$\kappa_{[3]} = \kappa_3/(\kappa_2)^{3/2} \quad \text{and} \quad \kappa_{[4]} = \kappa_4/(\kappa_2)^2.$$

The r-th *moments* and *cumulants* of X_j in $\boldsymbol{X} = (X_1, \ldots, X_p)'$ are expressed as $\mu_r^j = \mathrm{E}[X_j^r]$ and κ_r^j. The *skewness* and *kurtosis* of X_j are $\kappa_{[3]}^j$ and $\kappa_{[4]}^j$, respectively.

Next we consider the cumulants of the joint distribution of $\boldsymbol{X} = (X_1, \ldots, X_p)'$. The characteristic function of $\boldsymbol{X}$ is defined as

$$C(\boldsymbol{t}) = \mathrm{E}\left[\mathrm{e}^{i(t_1X_1 + \cdots + t_pX_p)}\right] = \mathrm{E}\left[\mathrm{e}^{i\boldsymbol{t}'\boldsymbol{X}}\right],$$

where t_j, $j = 1, \ldots, p$ are real numbers and $\boldsymbol{t} = (t_1, \ldots, t_p)'$. The characteristic function can be expanded as

$$\begin{aligned}
C(\boldsymbol{t}) &= \mathrm{E}\left[\sum_{r=0}^{\infty} \frac{1}{r!}(i\boldsymbol{t}'\boldsymbol{X})^r\right] \\
&= \sum_{r=0}^{\infty} \sum_{r_1+\cdots+r_p=r} \mu_{r_1\cdots r_p}^{1\cdots p} \times \frac{(it_1)^{r_1}\cdots(it_p)^{r_p}}{r_1!\cdots r_p!},
\end{aligned}$$

where $\mu_{r_1\cdots r_p}^{1\cdots p}$ is the moment of $\boldsymbol{X}$ defined by

$$\mu_{r_1\cdots r_p}^{1\cdots p} = \mathrm{E}\Big[X_1^{r_1}\cdots X_p^{r_p}\Big].$$

The *cumulants* of $\boldsymbol{X}$ are defined as the coefficients $\kappa_{r_1\cdots r_p}^{1\cdots p}$ in the expansion

$$\begin{aligned}\log C(\boldsymbol{t}) &= \log\left\{1+\sum_{r=1}^{\infty}\sum_{r_1+\cdots+r_p=r}\mu_{r_1\cdots r_p}^{1\cdots p}\frac{(it_1)^{r_1}\cdots(it_p)^{r_p}}{r_1!\cdots r_p!}\right\}\\ &= \sum_{r=1}^{\infty}\sum_{r_1+\cdots+r_p=r}\kappa_{r_1\cdots r_p}^{1\cdots p}\frac{(it_1)^{r_1}\cdots(it_p)^{r_p}}{r_1!\cdots r_p!}.\end{aligned}$$

In these notations, we write simply

$$\begin{aligned}&\mu_{r_1 0\cdots 0}^{12\cdots p} = \mu_{r_1}^{1}, \quad \mu_{r_1 r_2\cdots 0}^{12\cdots p} = \mu_{r_1 r_2}^{12},\\ &\kappa_{r_1 0\cdots 0}^{12\cdots p} = \kappa_{r_1}^{1}, \quad \kappa_{r_1 r_2\cdots 0}^{12\cdots p} = \kappa_{r_1 r_2}^{12}, \quad \text{etc.}\end{aligned}$$

If $\boldsymbol{X}$ is distributed as $\mathrm{E}_p(\boldsymbol{\mu},\Lambda)$, then from Theorem 1.3.3(3) we have

$$\log C(\boldsymbol{t}) = i\boldsymbol{t}'\boldsymbol{\mu} + \log\phi(\boldsymbol{t}'\Sigma\boldsymbol{t}).$$

Therefore,

$$\begin{aligned}&\kappa_1^j = \mu_1^j,\\ &\kappa_{11}^{ij} = \sigma_{ij} = -2\phi'(0)\lambda_{ij}, \quad \Sigma = -2\phi'(0)\Lambda,\\ &\kappa_{1\cdots p}^{r_1\cdots r_p} = 0, \quad \text{if } r_1+\cdots+r_p (\geq 3) \text{ is odd},\\ &\kappa_{1111}^{ijk\ell} = \kappa(\sigma_{ij}\sigma_{k\ell}+\sigma_{ik}\sigma_{j\ell}+\sigma_{i\ell}\sigma_{jk}).\end{aligned}$$

The skewness and kurtosis of X_j are given by

$$\begin{aligned}\kappa_{[3]}^j &= \frac{\kappa_3^j}{(\kappa_2^j)^{3/2}} = 0,\\ \kappa_{[4]}^j &= \frac{\kappa_4^j}{(\kappa_2^j)^2} = 3\cdot\frac{[\phi''(0)-\phi'(0)^2]}{\phi'(0)^2}\\ &= 3\kappa.\end{aligned}$$

If, in particular, $\boldsymbol{X}$ follows $\mathrm{N}_p(\boldsymbol{\mu},\Sigma)$, the logarithm of its characteristc funci-tion takes most simple form:

$$\log C(\boldsymbol{t}) = i\boldsymbol{t\mu} + \frac{1}{2}i^2\boldsymbol{t}'\Sigma\boldsymbol{t},$$

which immediately implies

$$\begin{aligned}&\kappa_1^j = \mu_1^j,\\ &\kappa_{11}^{ij} = \sigma_{ij},\\ &\kappa_{r_1\cdots r_p}^{1\cdots p} = 0 \quad (r_1+\cdots+r_p \geq 3).\end{aligned}$$

The moments and cumulants are also denoted as follows:

$$\mu^{1\cdots p}_{r_1\cdots r_p} = \mu_{\underbrace{1\cdots 1}_{r_1}\cdots\underbrace{p\cdots p}_{r_p}} \quad \text{and} \quad \kappa^{1\cdots p}_{r_1\cdots r_p} = \kappa_{\underbrace{1\cdots 1}_{r_1}\cdots\underbrace{p\cdots p}_{r_p}}.$$

Then the characteristic function of $\boldsymbol{X}$ is expanded as

$$C(\boldsymbol{t}) = 1 + \sum_{r=1}^{\infty} \frac{i^r}{r!}\mu_{j_1\cdots j_r} t_{j_1}\cdots t_{j_r}$$
$$\log C(\boldsymbol{t}) = \sum_{r=1}^{\infty} \frac{i^r}{r!}\kappa_{j_1\cdots j_r} t_{j_1}\cdots t_{j_r}.$$

We have a formal relationship between moments and cumulants. As a rigorous statement, we have the following result.

Theorem 1.4.1 *If the sth moment of a random vector $\boldsymbol{X}$ exists, the characteristic function and the cumulant function of $\boldsymbol{X}$ can be expanded as*

$$C(\boldsymbol{t}) = 1 + \sum_{r=1}^{s} \frac{i^r}{r!}\mu_{j_1\cdots j_r} t_{j_1}\cdots t_{j_r} + \mathrm{o}(\|\, \boldsymbol{t}\, \|),$$
$$\log C(\boldsymbol{t}) = \sum_{r=1}^{s} \frac{i^r}{r!}\kappa_{j_1\cdots j_r} t_{j_1}\cdots t_{j_r} + \mathrm{o}(\|\, \boldsymbol{t}\, \|),$$

where $\|\, \boldsymbol{t}\, \| = \sqrt{t_1^2 + \cdots + t_p^2}$.

For *skewness* $\kappa_{[3]}$ and *kurtosis* $\kappa_{[4]}$ of a random vector $\boldsymbol{X}$, there are some definitions. Mardia (1970) defined

$$\kappa_3^{(1)} = \mathrm{E}\big[\{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^3\big],$$
$$\kappa_4^{(1)} = \mathrm{E}\big[\{(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\}^2\big].$$

Let a random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ have mean $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and covariance matrix $\mathrm{Cov}(\boldsymbol{X}) = \Sigma = (\sigma_{ij})$. Let the cumulants of $\boldsymbol{X} - \boldsymbol{\mu}$ be denoted by $\gamma_{j_1\cdots j_r}$ instead of $\kappa_{j_1\cdots j_r}$.

$$\begin{aligned}
\gamma_a &= 0,\\
\gamma_{ab} &= \sigma_{ab},\\
\gamma_{abc} &= E[(X_a-\mu_a)(X_b-\mu_b)(X_c-\mu_c)],\\
\gamma_{abcd} &= E[(X_a-\mu_a)(X_b-\mu_b)(X_c-\mu_c)(X_d-\mu_d)]\\
&\quad - (\sigma_{ab}\sigma_{cd} + \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}).
\end{aligned}$$

The quantities $\kappa_{[3]}^{(1)}$ and $\kappa_{[4]}^{(1)}$ can be expressed in terms of the cumulants of $\boldsymbol{X} - \boldsymbol{\mu}$ as

$$\begin{aligned}
\kappa_{[3]}^{(1)} &= \sum_{a,b,c,a',b',c'} \gamma_{abc}\gamma_{a'b'c'}\sigma^{aa'}\sigma^{bb'}\sigma^{cc'}, \\
\kappa_{[4]}^{(1)} &= \sum_{a,b,c,d} \gamma_{abcd}\sigma^{ab}\sigma^{cd},
\end{aligned}$$

where $\Sigma^{-1} = (\sigma^{ij})$. An alternative definition for skewness is

$$\kappa_{[3]}^{(2)} = \sum_{a,b,c,a',b',c'} \gamma_{abc}\gamma_{a'b'c'}\sigma^{ab}\sigma^{ca'}\sigma^{b'c'}.$$

For the standardized variate $\boldsymbol{Y} = (Y_1, \ldots, Y_p)' = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$, we put

$$\mu_{i_1 \cdots i_k} = \mathrm{E}(Y_{i_1} \cdots Y_{i_k})$$

and denote the corresponding cumulants by $\kappa_{i_1 \cdots i_k}$. Then

$$\begin{aligned}
\mu_a &= \kappa_a \ (= 0), \\
\mu_{ab} &= \kappa_{ab} \ (= \delta_{ab}), \\
\mu_{abc} &= \kappa_{abc}, \\
\mu_{abcd} &= \kappa_{abcd} + \sum_{[3]} \delta_{ab}\delta_{cd}, \\
\mu_{abcde} &= \kappa_{abcde} + \sum_{[10]} \kappa_{abc}\delta_{de}, \\
\mu_{abcdef} &= \kappa_{abcdef} + \sum_{[15]} \kappa_{abcd}\delta_{ef} + \sum_{[10]} \kappa_{abc}\kappa_{def} + \sum_{[15]} \delta_{ab}\delta_{cd}\delta_{ef},
\end{aligned}$$

where $\sum_{[3]} \delta_{ab}\delta_{cd} = \delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}$. The multivariate skewness and kurtosis are expressed as

$$\begin{aligned}
\kappa_{[3]}^{(1)} &= \sum_{a,b,c} (\kappa_{abc})^2, \\
\kappa_{[3]}^{(2)} &= \sum_{a} \left(\sum_{b} \kappa_{abb} \right)^2 = \sum_{a,b,c} \kappa_{aab}\kappa_{bcc}, \\
\kappa_{[4]}^{(1)} &= \sum_{a,b} \kappa_{aabb}.
\end{aligned}$$

Problems

1.1 Suppose that $\boldsymbol{X} = (X_1, X_2)'$ is distributed as a bivariate distribution with mean and variance given by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

respectively. The density function is a constant on ellipsoids

$$(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = c$$

for every positive value of c in two-dimensional Euclidian space. After transforming the coordinates by $(x_i - \mu_i)/\sigma_i = y_i$, $i = 1, 2$, show that the loci of constant density are ellipses defined by

$$\frac{1}{1-\rho^2}(y_1^2 - 2\rho y_1 y_2 + y_2^2) = k$$

for $k > 0$. Show that

(1) for $\rho > 0$, the principal axes are along the lines $y_1 = y_2, y_1 = -y_2$ with lengths $2\sqrt{k(1+\rho)}$ and $2\sqrt{k(1-\rho)}$, respectively, and that

(2) for $\rho < 0$, the principal axes are along the lines $y_1 = -y_2, y_1 = y_2 < 0$ with lengths $2\sqrt{k(1-\rho)}$ and $2\sqrt{k(1+\rho)}$, respectively.

Hint: Consider an orthogonal transformation

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

By choosing an appropriate θ, we can write the ellipse as

$$\frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} = k',$$

where λ_1 and λ_2 are solutions of $(1-\lambda)^2 - \rho^2 = 0$: that is $\lambda_i = 1 \pm |\rho|$.

1.2 Let $\boldsymbol{Y}$ be a random vector defined by

$$\boldsymbol{Y} = \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{B} + \boldsymbol{\varepsilon},$$

where $\mathrm{X}: p \times q$ and $\mathrm{Z}: p \times r$ are fixed matrices, $\boldsymbol{\beta}: q \times 1$ is a parameter vector, $\boldsymbol{B}$ and $\boldsymbol{\varepsilon}$ are random vectors having $\mathrm{N}_r(\mathbf{0}, \Delta)$ and $\mathrm{N}_p(\mathbf{0}, \sigma^2\mathrm{I}_p)$, respectively, and $\boldsymbol{B}$ and $\boldsymbol{\varepsilon}$ are independent. Then show the following.

(a) $\boldsymbol{y} \sim \mathrm{N}_p(\mathrm{X}\boldsymbol{\beta}, \mathrm{W}), \quad \mathrm{W} = \mathrm{Z}\Delta\mathrm{Z}' + \sigma^2\mathrm{I}_p.$

(b) $\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{B} \end{pmatrix} \sim \mathrm{N}_{p+r}\left(\begin{pmatrix} \mathrm{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathrm{W} & \mathrm{Z}\Delta \\ \Delta\mathrm{Z}' & \Delta \end{pmatrix}\right).$

(c) $\mathrm{E}(\boldsymbol{B}\,|\,\boldsymbol{Y}) = \Delta\mathrm{Z}\mathrm{W}^{-1}(\boldsymbol{Y} - \mathrm{X}\boldsymbol{\beta})$.

(d) $\boldsymbol{Y}|\boldsymbol{B} \sim \mathrm{N}_p(\mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\boldsymbol{B}, \sigma^2\mathrm{I}_p)$.

1.3 Show that Theorem 1.2.2 holds for a singular multivariate normal variable.

Hint: Suppose that $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ and Σ is singular with rank r. Then use the fact that $\boldsymbol{X}$ is expressed as $\boldsymbol{X} = \mathrm{B}\boldsymbol{Z}$, where $\mathrm{BB}' = \Sigma$ and $\boldsymbol{Z} \sim \mathrm{N}_r(\mathbf{0}, \mathrm{I}_r)$.

1.4 If $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, the characteristic function of $\boldsymbol{X}'\mathrm{A}\boldsymbol{X}$ is given by

$$\begin{aligned}
\mathrm{E}\big[e^{it\boldsymbol{X}'\mathrm{A}\boldsymbol{X}}\big] &= |\mathrm{I}_p - 2it\mathrm{A}|^{-1/2}\exp\big(it\boldsymbol{\mu}'\mathrm{A}(\mathrm{I}_p - 2it\mathrm{A})\boldsymbol{\mu}\big) \\
&\qquad -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + it\boldsymbol{x}'\mathrm{A}\boldsymbol{x} \\
&= -\frac{1}{2}\Big\{\boldsymbol{x} - (\Sigma^{-1} - 2it\mathrm{A})^{-1}\Sigma^{-1}\boldsymbol{\mu}\Big\}(\Sigma^{-1} - 2it\mathrm{A}) \\
&\qquad \times\Big\{\boldsymbol{x} - (\Sigma^{-1} - 2it\mathrm{A})^{-1}\Sigma^{-1}\boldsymbol{\mu}\Big\} - \frac{1}{2}\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} \\
&\qquad + \frac{1}{2}\boldsymbol{\mu}'\Sigma^{-1}(\Sigma^{-1} - 2it\mathrm{A})^{-1}\Sigma^{-1}\boldsymbol{\mu}.
\end{aligned}$$

1.5 Let $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$ be a random matrix such that $\boldsymbol{X}_i, i = 1, \ldots, n$ are independent and $\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$. Show that the density function of $\mathbf{X}$ is written as

$$f(\mathrm{X}) = |(2\pi)\Sigma|^{-n/2}\,\mathrm{etr}\Big\{-\frac{1}{2}\Sigma^{-1}(\mathrm{X} - \mathrm{M})'(\mathrm{X} - \mathrm{M})\Big\},$$

where $\mathrm{X} = (\boldsymbol{x}_1, \ldots \boldsymbol{x}_n)'$ and $\mathrm{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)'$ are $p \times n$ matrices, and the characteristic function of $\mathbf{X}$ is given by

$$C_{\mathbf{X}}(\mathrm{T}) = \mathrm{etr}\Big\{i\mathrm{T}'\mathrm{M} - \frac{1}{2}\mathrm{T}'\Sigma\mathrm{T}\Big\},$$

where $\mathrm{T} : p \times n$ is any real matrix.

Hint: The density function of $\mathbf{X}$ can be obtained by computing

$$f(\mathrm{X}) = \prod_{j=1}^{n} |(2\pi)\Sigma|^{-1/2}\,\mathrm{etr}\Big\{-\frac{1}{2}\Sigma^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}_j)(\boldsymbol{x}_j - \boldsymbol{\mu}_j)'\Big\}.$$

Note that for any $p \times n$ matrix $\mathrm{T} = (t_{ij})$, we have

$$\sum_{j=1}^{p}\sum_{k=1}^{n} t_{jk}x_{jk} = \mathrm{tr}\,\mathrm{TX}$$

and

$$
\begin{aligned}
-\frac{1}{2}\,\mathrm{tr}\,\Sigma^{-1}(\mathrm{X}\ -\ &\mathrm{M})'(\mathrm{X}-\mathrm{M}) + i\ \mathrm{tr}\,\mathrm{TX} \\
&= -\frac{1}{2}\,\mathrm{tr}\,\Sigma^{-1}\left\{\mathrm{X}-(\mathrm{M}+i\mathrm{T}\Sigma)\right\}'\left\{\mathrm{X}-(\mathrm{M}+i\mathrm{T}\Sigma)\right\} \\
&\qquad + i\ \mathrm{tr}\,\mathrm{T}'\mathrm{M} - \frac{1}{2}\,\mathrm{tr}\,\mathrm{T}'\Sigma\mathrm{T}.
\end{aligned}
$$

1.6 Suppose that $\boldsymbol{X}$ is distributed as $\mathrm{N}_p(\mathbf{0}, \mathrm{I}_p)$. Then, show that the mean of $||\boldsymbol{X}||^{-(p-1)}$ exists.

Hint: Use the coordinate transformation in Problem 1.9.

1.7 Let the density function of $\boldsymbol{X} = (X_1, \ldots, X_p)'$ be defined by

$$
f(\boldsymbol{x}) = \frac{\Gamma(\frac{1}{2}p+1)}{|\Sigma|^{1/2}\{(p+2)\pi\}^{p/2}}, \quad \mathrm{if}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \le p+2,
$$

and $f(\boldsymbol{x}) = 0$, otherwise. Then, show that $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$, and $\mathrm{Var}(\boldsymbol{X}) = \Sigma$.

1.8 (a) The volume of p-dimensional unit ball is given by

$$
\int_{\sum_{i=1}^{p} z_i^2 \le 1} \cdots \int \mathrm{d}z_1 \cdots \mathrm{d}z_p = \frac{\pi^{p/2}}{\Gamma((p+2)/2)}
$$

(b) the volume of the ellipsoid $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c^2$ $(c > 0)$ is

$$
\frac{1}{\Gamma((p+1)/2)}|\Sigma|^{1/2}c^p\pi^{p/2}.
$$

1.9 (a) Let (x_1, x_2, x_3) be written in polar coordinates as

$$
x_1 = r\sin\theta_1\sin\theta_2, \quad x_2 = r\sin\theta_1\cos\theta_2, \quad x_3 = r\cos\theta_1,
$$

where $r > 0, 0 < \theta_1 < \pi$, and $0 < \theta_2 < 2\pi$. Show that the Jacobian of the transformation is given by

$$
\frac{\partial(x_1, x_2, x_3)}{\partial(r, \theta_1, \theta_2)} = r^2\sin\theta_1.
$$

(b) The coordinate transformation in the case of p-variate is defined by

$$
\begin{aligned}
x_1 &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-2}\sin\theta_{p-1}, \\
x_2 &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-2}\cos\theta_{p-1}, \\
x_3 &= r\sin\theta_1\sin\theta_2\cdots\cos\theta_{p-2}, \\
&\ \vdots \\
x_{p-1} &= r\sin\theta_1\cos\theta_2, \\
x_p &= r\cos\theta_1,
\end{aligned}
$$

where $r > 0, 0 < \theta_i < \pi$ $(i = 1, \ldots, p-2), 0 < \theta_{p-1} < 2\pi$. Show that the Jacobian of the transformation is given by

$$J = r^{p-1} \sin^{p-2}\theta_1 \sin^{p-3}\theta_2 \cdots \sin\theta_{p-2}.$$

1.10 Show that the volume of the ball B_a with radius a in R^p is

$$V = \frac{(a\sqrt{\pi})^p}{\Gamma((p+2)/2)} = \begin{cases} a^p \frac{\pi^{\frac{p}{2}} 2^{\frac{p}{2}}}{2 \cdot 4 \cdots p}, & (p : \text{even}) \\ a^p \frac{\pi^{\frac{p-1}{2}} 2^{\frac{p+1}{2}}}{1 \cdot 3 \cdots p}, & (p : \text{odd}) \end{cases}$$

Hint: The result follows from

$$\int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} J \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{p-2} \mathrm{d}\theta_{p-1} = \frac{2\pi^{p/2}}{\Gamma(p/2)} \cdot \frac{a^p}{p}.$$

1.11 Suppose that a p-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ has the uniform distribution on the unit ball $S = \{\boldsymbol{x} = (x_1, \ldots, x_p)' : \|\boldsymbol{x}\| \leq 1\}$ in R^p. Show that

$$\Pr\{\|\boldsymbol{X}\| \leq a\} = \frac{\int_{B_a} \mathrm{d}\boldsymbol{x}}{\int_S \mathrm{d}\boldsymbol{x}} = a^p, \quad |a| < 1.$$

1.12 For the multivariate t-distribution (see Example 1.3.1) with the density, show that $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\mathrm{Var}(\boldsymbol{X}) = [m/(m-2)]\Lambda$.

1.13 Show that the first four cumulants are given, in terms of the moments, by

$$\begin{aligned} \kappa_1 &= \mu_1 = \mu, \\ \kappa_2 &= \mu_2 - \mu_1^2 = \sigma^2, \\ \kappa_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3, \\ \kappa_4 &= \mu_3 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4. \end{aligned}$$

1.14 Let $\boldsymbol{X}$ be distributed as $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Then, for a $p \times 1$ constant vector $\boldsymbol{a}$, show that

$$\mathrm{E}[\{\boldsymbol{a}'(\boldsymbol{X} - \boldsymbol{\mu})\}^j] = \begin{cases} \dfrac{(2m)!}{2^m m!}(\boldsymbol{a}'\Sigma\boldsymbol{a})^m & \text{if } j = 2m \ \ (\text{even}), \\ 0 & \text{if } j = 2m-1 \ \ (\text{odd}). \end{cases}$$

1.15 For $i = i = 1, \ldots, n$,

(1) let X_i be mutually independent random variables such that $X_i \sim \mathrm{N}(\mu_i, \sigma^2)$. What is the distribution of $\boldsymbol{X} = (X_1, \ldots, X_n)$? Find the mean vector and covariance matrix of $\boldsymbol{X}$. Similarly,

(2) let $\boldsymbol{X}_i$ be mutually independent random vectors such that $\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$. What is the distribution of $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$? Find the mean vector and covariance matrix of $\mathrm{vec}(\mathbf{X})$.

1.16 Suppose that $\boldsymbol{X}$ is distributed as a singular multivariate normal distribution $\mathrm{N}(\mathbf{0}, \Sigma)$ with $\mathrm{rank}(\Sigma) = k$.

(1) Using [A.1.12], show that we can express $\Sigma = \Gamma_1 \Lambda_1 \Gamma_1'$, where $\Lambda_1 = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_k), \lambda_1 \geq \cdots \geq \lambda_k > 0$ and Γ_1 satisfies $\Gamma_1'\Gamma_1 = \mathrm{I}_k$.

(2) Let $\Sigma^+ = \Gamma_1 \Lambda_1^{-1} \Gamma_1$. Then, show that Σ^+ is the Moore–Penrose inverse matrix, that is, Σ^+ satisfies the following four conditions: (i) $\Sigma\Sigma^+\Sigma = \Sigma$, (ii) $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$, (iii) $(\Sigma^+\Sigma)' = \Sigma^+\Sigma$, (iv) $(\Sigma\Sigma^+)' = \Sigma\Sigma^+$.

(3) Show that $\boldsymbol{X}'\boldsymbol{X} \sim \chi^2(k)$.

Chapter 2

Wishart Distribution

The Wishart distribution is a multivariate generalization of the chi-square distribution. The distributions of sample covariance matrix and various sums of squares and products matrices are of Wishart, provided the underlying distribution is normal.

In this book we use the generalized definition of Wishart distribution that includes both nonsingular and singular cases, with the probability density function for nonsingular case.

We provide some basic properties of the Wishart distribution; sampling distributions; Wishartness and independence of quadratic forms; asymptotic distributions.

2.1 Definition

Let $\boldsymbol{X}$ be a p-dimensional random vector that is distributed as $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample of $\boldsymbol{X}$ with size n. Then the sample mean vector and covariance matrices are defined by

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

respectively. From Theorem 1.2.3 we know that $\bar{\boldsymbol{X}} \sim \mathrm{N}(\boldsymbol{\mu}, (1/n)\Sigma)$. Now we consider an orthogonal transformation

$$\begin{aligned} \mathbf{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)' &= \mathrm{H}'(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)' \\ &= \mathrm{H}'\mathbf{X}, \end{aligned}$$

where H is an $n \times n$ orthogonal matrix. Then from Theorem 1.2.6, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are mutually independent and each is distributed as $\mathrm{N}_p(\cdot, \Sigma)$. The mean of $\mathbf{Z}$

is $\mathrm{E}(\mathbf{Z}) = \mathrm{H}'E(\mathbf{X})$. If we choose the first column of H as $(1/\sqrt{n})\mathbf{1}_n$, then

$$\boldsymbol{Z}_1 \sim \mathrm{N}_p(\sqrt{n}\,\boldsymbol{\mu}, \Sigma), \quad \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n \sim \text{i.i.d. } \mathrm{N}_p(\mathbf{0}, \Sigma).$$

Furthermore,

$$\bar{\boldsymbol{X}} = \frac{1}{\sqrt{n}}\boldsymbol{Z}_1, \quad \mathbf{W} = (n-1)\mathbf{S} = \sum_{i=2}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i'.$$

The expression of $\mathbf{W}$ shows that $\mathbf{W}$ is distributed as a Wishart distribution $\mathrm{W}_p(n-1, \Sigma)$. In general, a Wishart distribution is defined as follows.

Definition 2.1.1 If a $p \times p$ random matrix $\mathbf{W}$ is expressed as

$$\mathbf{W} = \sum_{j=1}^{n} \boldsymbol{Z}_j \boldsymbol{Z}_j',$$

where $\boldsymbol{Z}_j \sim \mathrm{N}_p(\boldsymbol{\mu}_j, \Sigma)$ and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent, $\mathbf{W}$ is said to have a *noncentral Wishart distribution* with n degrees of freedom, covariance matrix Σ, and noncentrality matrix $\Delta = \boldsymbol{\mu}_1\boldsymbol{\mu}_1' + \cdots + \boldsymbol{\mu}_n\boldsymbol{\mu}_n'$. We write that $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma; \Delta)$. In the special case $\Delta = \mathrm{O}$, $\mathbf{W}$ is said to have a *Wishart distribution*, denoted by $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$.

The noncentrality matrix is also defined as $\Delta = \Sigma^{-1}\mathrm{M}'\mathrm{M}$, not $\mathrm{M}'\mathrm{M}$, where $\mathrm{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)'$. Using the $\boldsymbol{Z}_j$ in the definition of a Wishart distribution $\mathrm{W}_p(n, \Sigma)$, let $\mathbf{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)'$. Then $\mathbf{W} = \mathbf{Z}'\mathbf{Z}$. When $p = 1$, $W/\sigma^2 \sim \chi^2(n)$, where $\Sigma = \sigma^2$.

The Wishart distribution $\mathrm{W}_p(n, \Sigma)$ is defined as the distribution of $\mathbf{W} = \mathbf{Z}'\mathbf{Z}$, where the rows of $Z : n \times p$ are independent identically distributed as $\mathrm{N}_p(\mathbf{0}, \Sigma)$. When $n < p$, $\mathbf{W}$ is called *singular* and the $\mathrm{W}_p(n, \Sigma)$ distribution does not have a density function. When $n \geq p$, we derive the density function of $\mathbf{W}$ by using the following theorem due to Hsu (1940).

Theorem 2.1.1 *Let* $\mathbf{X}$ *be an* $n \times p$ *random matrix where* $n \geq p$. *If the density function of* $\mathbf{X}$ *has the form* $f(\mathrm{X}) = g(\mathrm{X}'\mathrm{X})$, *the density function of* $\mathbf{V} = \mathbf{X}'\mathbf{X}$ *is*

$$h(\mathrm{V}) = \frac{\pi^{pn/2}}{\Gamma_p(n/2)}|\mathrm{V}|^{(n-p-1)/2}g(\mathrm{V}),$$

where $\Gamma_p(a) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[a-(i-1)/2]$.

Proof. First we make the transformation $\mathbf{X} = \mathbf{ZT}$, where $\mathbf{T}$ is a $p \times p$ upper triangular matrix with positive diagonal elements $T_{ii} > 0$ and $\mathbf{Z}$ is an $n \times p$ column-orthogonal matrix (i.e., $\mathbf{Z}'\mathbf{Z} = \mathrm{I}_p$). Using [A.1.15], the transformation is one-to-one. Its Jacobian is $\mathrm{dX} = [2^p\pi^{pn/2}/\Gamma(n/2)]\prod_{i=1}^{p} t_{ii}^{n-i}\mathrm{dT} \cdot \mathrm{d(H)}$, where $\mathrm{dX} = \prod_{i=1}^{n}\prod_{j=1}^{p} \mathrm{d}x_{ij}$, $\mathrm{dT} = \prod_{i<j} \mathrm{d}t_{ij}$, and $\mathrm{d(H)}$ is the invariant measure defined over the totality of p orthogonal frames in n-space normalized so that $\int_{\mathrm{H'H=I}_p} \mathrm{d(H)} = 1$. This Jacobian was given by James (1954). Hence, the joint density function of $\mathbf{T}$ and $\mathbf{Z}$ is

$$\frac{2^p \pi^{pn/2}}{\Gamma_p(n/2)} \prod_{i=1}^{p} t_{ii}^{n-i} g(\mathrm{T'T}),$$

which is also the marginal density function of $\mathbf{T}$, since the joint density function does not depend on $\mathbf{Z}$. Now we make the transformation from $\mathbf{T}$ to $\mathbf{V} = \mathbf{T'T}$ whose Jacobian is, by [A.3.6], $J(\mathrm{T} \to \mathrm{V}) = 2^p \prod_{i=1}^{p} t_{ii}^{p-i+1}$ and hence $d\mathrm{T} = 2^{-p}\prod_{i=1}^{p} t_{ii}^{-p+i-1} d\mathrm{V}$. The final result is obtained by noting that $\prod_{i=1}^{p} t_{ii} = |\mathrm{V}|^{1/2}$. □

We consider the characteristic function of the noncentral Wishart distribution. Let $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma; \Omega)$. Then without loss of generality we may write $\mathbf{W}$ as

$$\mathbf{W} = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i',$$

where the random vectors $\boldsymbol{X}_i$ are mutually independent, each following the normal distribution $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$ and $\Omega = \sum_{i=1}^{n} \boldsymbol{\mu}_i \boldsymbol{\mu}_i'$. The characteristic function of $\mathbf{W}$ is the expectation of

$$\begin{aligned}
&\exp\{i(t_{11}W_{11} + \cdots + t_{pp}W_{pp} + t_{12}W_{12} + \cdots + t_{p-1,p}W_{p-1,p})\}\\
&\quad = \exp(i \operatorname{tr} \mathrm{T}\mathbf{W})\\
&\quad = \exp(i\boldsymbol{X}_1'\mathrm{T}\boldsymbol{X}_1)\cdots\exp(i\boldsymbol{X}_n'\mathrm{T}\boldsymbol{X}_n),
\end{aligned}$$

where $\mathrm{T} = ((1+\delta_{ij})t_{ij}/2)$ is a symmetric matrix and δ_{ij} is the Kroneker delta [i.e., $\delta_{ij} = 1$, $\delta_{ij} = 0 (i \neq j)$]. Note that if $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$,

$$\mathrm{E}[\exp(i\boldsymbol{X}'\mathrm{T}\boldsymbol{X})] = |\mathrm{I} - 2i\Sigma\mathrm{T}|^{-1/2} \exp\{i\boldsymbol{\mu}'\mathrm{T}(\mathrm{I} - 2i\Sigma\mathrm{T})^{-1}\boldsymbol{\mu}\}.$$

The result is obtained by using

$$\begin{aligned}
&(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) - 2i\boldsymbol{X}'\mathrm{T}\boldsymbol{X}\\
&\quad = (\boldsymbol{X} - \mathrm{A}\Sigma^{-1}\boldsymbol{\mu})'\mathrm{A}^{-1}(\boldsymbol{X} - \mathrm{A}\Sigma^{-1}\boldsymbol{\mu}) + \boldsymbol{\mu}'(\Sigma^{-1} - \Sigma^{-1}\mathrm{A}\Sigma^{-1})\boldsymbol{\mu},
\end{aligned}$$

where $\mathrm{A} = (\Sigma^{-1} - 2i\mathrm{T})^{-1}$. Therefore, the characteristic function of $\mathbf{W}$ is expressed as

$$C_{\mathbf{W}}(\mathrm{T}) = |\mathrm{I}_p - 2i\Sigma\mathrm{T}|^{-n/2} \exp\left\{i\sum_{j=1}^{n} \boldsymbol{\mu}_j'\mathrm{T}(\mathrm{I}_p - 2i\Sigma\mathrm{T})^{-1}\boldsymbol{\mu}_j\right\},$$

which gives the following result.

Theorem 2.1.2 *Let* $\mathbf{W}$ *be a noncentral Wishart matrix having* $\mathrm{W}_p(n, \Sigma; \Omega)$. *Then the characteristic function of* $\mathbf{W}$ *is given by*

$$C_{\mathbf{W}}(\mathrm{T}) = |\mathrm{I}_p - 2i\Sigma\mathrm{T}|^{-n/2}\,\mathrm{etr}\{i\Omega\mathrm{T}(\mathrm{I}_p - 2i\Sigma\mathrm{T})^{-1}\},$$

where $\mathrm{etr}(\mathrm{A}) = \exp(\mathrm{tr}\,\mathrm{A})$. *In particular, if* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$,

$$C_{\mathbf{W}}(\mathrm{T}) = |\mathrm{I}_p - 2i\mathrm{T}|^{-n/2}.$$

Theorem 2.1.3 *If* $\mathbf{W}$ *is* $\mathrm{W}_p(n, \Sigma)$ $(n \geq p)$, *the density function of* $\mathbf{W}$ *is*

$$\frac{1}{2^{pn/2}\Gamma_p(\frac{1}{2}n)|\Sigma|^{n/2}}\,\mathrm{etr}\Big(-\frac{1}{2}\Sigma^{-1}\mathrm{W}\Big)|\mathrm{W}|^{(n-p-1)/2}, \quad \mathrm{W} > \mathrm{O}, \qquad (2.1.1)$$

where $\mathrm{etr}(\mathrm{A}) = \exp(\mathrm{tr}\,\mathrm{A})$.

Proof. Starting from the definition of $\mathbf{W}$, in Theorem 2.1.2 we have seen that the characteristic function of $\mathbf{W}$ is $|\mathrm{I}_p - 2i\mathrm{T}|^{-n/2}$. On the other hand, using the density function (2.1.1) we can obtain the characteristic function of $\mathbf{W}$ as $|\mathrm{I}_p - 2i\mathrm{T}|^{-n/2}$. The result follows from Theorem 1.1.3 on one-to-one correspondence between the characteristic function and its density function. □

2.2 Some Basic Properties

Theorem 2.2.1

(1) *If* $\mathbf{W}_1 \sim \mathrm{W}_p(n_1, \Sigma; \Delta_1)$, $\mathbf{W}_2 \sim \mathrm{W}_p(n_2, \Sigma; \Delta_2)$ *and they are independent, then* $\mathbf{W}_1 + \mathbf{W}_2 \sim \mathrm{W}_p(n_1 + n_2, \Sigma; \Delta_1 + \Delta_2)$.

(2) *If* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma; \Delta)$ *and* B *is a* $q \times p$ *matrix with rank* q, *then* $\mathrm{B}\mathbf{W}\mathrm{B}' \sim \mathrm{W}_q(n, \mathrm{B}\Sigma\mathrm{B}'; \mathrm{B}\Delta\mathrm{B}')$.

Proof. Result (1) is obtained easily from the definition of Wishart distribution. We can write $\mathbf{W}$ as

$$\mathbf{W} = \sum_{i=1}^{n} \boldsymbol{Z}_i\boldsymbol{Z}_i', \quad \boldsymbol{Z}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$$

and hence

$$\mathrm{B}\mathbf{W}\mathrm{B}' = \sum_{i=1}^{n}(\mathrm{B}\boldsymbol{Z}_i)(\mathrm{B}\boldsymbol{Z}_i)'.$$

Result (2) follows from the fact that $\mathrm{B}\boldsymbol{Z}_i \sim \mathrm{N}_q(\mathrm{B}\boldsymbol{\mu}_i, \mathrm{B}\Sigma\mathrm{B}')$.

Theorem 2.2.2 *Let* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$ *and* $\mathbf{W}$ *and* Σ *be partitioned as*

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

respectively, where $\mathbf{W}_{ij} : p_i \times p_j$ *and* $\Sigma_{ij} : p_i \times p_j$. *Then:*

(1) $\mathbf{W}_{ii} \sim \mathrm{W}_{p_i}(n, \Sigma_{ii})$, $i = 1, 2$.

(2) $\mathbf{W}_{11\cdot 2} = \mathbf{W}_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21} \sim \mathrm{W}_{p_1}(n - p_2, \Sigma_{11\cdot 2})$.

(3) $\mathbf{W}_{11\cdot 2}$ *and* $\{\mathbf{W}_{21}, \mathbf{W}_{22}\}$ *are independent.*

(4) *The conditional distribution of* $\mathbf{W}_{12} = (\boldsymbol{B}_1, \ldots . \boldsymbol{B}_{p_2})$ *is normal with mean* $\mathrm{E}(\mathbf{W}_{12}|\mathbf{W}_{22}) = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{W}_{22}$ *and*

$$\mathrm{Cov}(\boldsymbol{B}_i, \boldsymbol{B}_j|\mathbf{W}_{22}) = (\mathbf{W}_{22})_{ij}\Sigma_{11\cdot 2},$$

where $(\mathrm{A})_{ij}$ *denotes the* (i, j)*th element of* A.

Proof. We may write $\mathbf{W}$ as

$$\mathbf{W} = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' = \sum_{i=1}^{n} \begin{pmatrix} \boldsymbol{X}_{i1} \\ \boldsymbol{X}_{i2} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_{i1} \\ \boldsymbol{X}_{i2} \end{pmatrix}',$$

where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim$ i.i.d. $\mathrm{N}_p(\mathbf{0}, \Sigma)$, and $\boldsymbol{X}_{i1} : p_i \times 1$. Then

$$\mathbf{W}_{11} = \sum_{i=1}^{n} \boldsymbol{X}_{i1}\boldsymbol{X}_{i1}', \quad \boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in} \sim \text{i.i.d. } \mathrm{N}_p(\mathbf{0}, \Sigma_{11}).$$

This shows (1). Let

$$\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)' = (\mathbf{X}_1 \ \mathbf{X}_2), \quad \mathbf{X}_1 : n \times p_1;$$

that is, $\mathbf{X}_1 = (\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n})'$ and $\mathbf{X}_2 = (\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n})'$. Consider the distribution of

$$\mathbf{W}_{11\cdot 2} = \mathbf{X}_1'\{\mathrm{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2\}\mathbf{X}_1$$

given $\mathbf{X}_2$. Then $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{n1}$ are independent and

$$\boldsymbol{X}_{i1}|\boldsymbol{X}_{i2} \sim \mathrm{N}_{p_1}(\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{X}_{i2}, \Sigma_{11\cdot 2}).$$

Let $\mathbf{H} = (\mathbf{H}_1 \ \mathbf{H}_2)$ be the orthogonal matrix such that $\mathbf{H}_1 = \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1/2}$, and consider

$$\mathbf{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)' = \mathbf{H}'\mathbf{X}_1.$$

From Theorem 1.2.6 it is seen that $\boldsymbol{Z}_{p_2+1}, \ldots, \boldsymbol{Z}_n \sim$ i.i.d. $\mathrm{N}_{p_2}(\mathbf{0}, \Sigma_{11\cdot 2})$. Further,

$$\mathbf{W}_{11\cdot 2} = \mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1\mathbf{H}_1\mathbf{H}_1'\mathbf{X}_1' = \sum_{i=p_2+1}^{n} \boldsymbol{Z}_i\boldsymbol{Z}_i'.$$

This implies that $\mathbf{W}_{11\cdot 2} \sim \mathrm{W}_{p_1}(n-p_2, \Sigma_{11\cdot 2})$, and that $\mathbf{W}_{11\cdot 2}$ does not depend on $\mathbf{X}_2$. The latter shows that $\mathbf{W}_{11\cdot 2}$ and $\mathbf{W}_{22}$ are independent. Further, we can express

$$\mathbf{W}_{11\cdot 2} = (\mathbf{X}_1'\mathbf{H}_2)\cdot(\mathbf{X}_1'\mathbf{H}_2)', \quad \mathbf{W}_{12} = \mathbf{X}_1\mathbf{H}_1 \cdot (\mathbf{X}_2'\mathbf{X}_2)^{1/2},$$

and hence the conditional distributions of $\mathbf{W}_{11\cdot 2}$ and $\mathbf{W}_{12}$ given $\mathbf{W}_{22}$ are independent. These imply (3). For a proof of (4), since $\mathbf{W}_{22} = \mathbf{X}_1'\mathbf{X}_2$, it is easily seen that the conditional distribution of $\mathbf{W}_{12}$ given $\mathbf{X}_2$ is normal with

$$\mathrm{E}(\mathbf{W}_{12}|\mathbf{X}_2) = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2'\mathbf{X}_2 = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{W}_{22}.$$

To compute its covariance matrix, put

$$\mathbf{X}_2 = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_{p_2}), \quad \boldsymbol{U}_i = (U_{1i}, \ldots, U_{ni})'.$$

Then using

$$\boldsymbol{B}_i = U_{1i}X_{11} + U_{2i}X_{21} + \cdots + U_{ni}X_{n1}, \quad i = 1, \ldots, p_2,$$

we obtain the desired result. □

From Theorem 2.2.2 we have:

(4)′ the conditional distribution of $\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2} = (\tilde{\boldsymbol{B}}_1, \ldots, \tilde{\boldsymbol{B}}_{p_2})$ given $\mathbf{W}_{22}$ is normal with mean $\mathrm{E}(\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2}|\mathbf{W}_{22}) = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{W}_{22}^{1/2}$ and covariance matrix

$$\mathrm{Var}(\tilde{\boldsymbol{B}}_i|\mathbf{W}_{22}) = \Sigma_{11\cdot 2}, \quad \mathrm{Cov}(\tilde{\boldsymbol{B}}_i, \tilde{\boldsymbol{B}}_j|\mathbf{W}_{22}) = \mathrm{O}, \quad i \neq j.$$

Theorem 2.2.3 *If* $\Sigma_{12} = \mathrm{O}$ *in Theorem 2.2.2, then:*

(1) $\mathbf{W}_{11\cdot 2}$, $\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2}$, *and* $\mathbf{W}_{22}$ *are mutually independent.*

(2) $\mathbf{W}_{11\cdot 2} \sim \mathrm{W}_{p_1}(n - p_2, \Sigma_{11})$, $\mathbf{W}_{22} \sim \mathrm{W}_{p_2}(n, \Sigma_{22})$, *and the columns of* $\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2}$ *are independent and identically distributed as* $\mathrm{N}_{p_1}(\mathbf{0}, \Sigma_{11})$, *and hence* $\mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21} \sim \mathrm{W}_{p_1}(n, \Sigma_{11})$.

Proof. For a proof of (1), from Theorem 2.2.2(3) it is sufficient to show that $\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2}$ is independent of $\mathbf{W}_{22}$. This follows from the fact that the conditional distribution of $\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2}$ given $\mathbf{W}_{22}$ does not depend on $\mathbf{W}_{22}$. The second result follows the results of Theorem 2.2.2 and (4)′ in the case of $\Sigma_{12} = \mathrm{O}$. □

For an extension of Theorem 2.2.3 to the case when $\mathbf{W}$ has a noncentral Wishart distribution, see Kabe (1964) and Problem 2.8.

Theorem 2.2.4 *If* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$, *then*

$$\mathrm{E}[|\mathbf{W}|^h] = 2^{ph}|\Sigma|^h \prod_{i=1}^{p} \frac{\Gamma[\frac{1}{2}(n-i+1)+h]}{\Gamma[\frac{1}{2}(n-i+1)]}, \quad \Re(h) > -\frac{1}{2}n + \frac{1}{2}(p-1).$$

Proof. Let $C(p, n; \Sigma) = [2^{pn/2}\Gamma_p(\frac{1}{2}n)|\Sigma|^{n/2}]^{-1}$. Then, using Theorem 2.1.3

$$\begin{aligned}\mathrm{E}[|\mathbf{W}|^h] &= C(p, n; \Sigma) \int_{\mathrm{W}} |\mathrm{W}|^{(n+2h-p-1)/2} \times \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}\mathrm{V}\right) d\mathrm{V} \\ &= \frac{C(p, n; \Sigma)}{C(p, n+2h; \Sigma)}\end{aligned}$$

which is equal to the required expression. Note that the above integral is extended to a complex number such that $\Re(n + 2h) > p - 1$. $\square$

The first two *moments of Wishart* and *inverse Wishart matrices* are given in the following Theorems 2.2.5 through 2.2.8. For higher moments, see, for example Gupta and Nagar (2000) and Mathai et al. (1995).

Theorem 2.2.5 *Let* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$ *and let* A *be a* $p \times p$ *constant matrix. Then*

(1) $\mathrm{E}(\mathbf{W}) = n\Sigma$.

(2) $\mathrm{E}(\mathbf{WAW}) = n\Sigma\mathrm{A}'\Sigma + n\,\mathrm{tr}\,(\Sigma\mathrm{A})\Sigma + n^2\Sigma\mathrm{A}\Sigma$.

Theorem 2.2.6 *Let* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$ *and let* A *and* B *be* $p \times p$ *constant matrices. Then*

(1) $\mathrm{E}(\,\mathrm{tr}\,\mathrm{A}\mathbf{W}) = n\,\mathrm{tr}\,\mathrm{A}\Sigma$.

(2) $\mathrm{E}[\,\mathrm{tr}\,(\mathrm{A}\mathbf{W})\,\mathrm{tr}\,(\mathrm{B}\mathbf{W})] = n\,\mathrm{tr}\,(\mathrm{A}\Sigma\mathrm{B}\Sigma) + n\,\mathrm{tr}\,(\mathrm{A}'\Sigma\mathrm{B}\Sigma) + n^2\,\mathrm{tr}\,(\mathrm{A}\Sigma)\,\mathrm{tr}\,(\mathrm{B}\Sigma)$.

(3) $\mathrm{E}[\,\mathrm{tr}\,(\mathrm{A}\mathbf{W}\mathrm{B}\mathbf{W})] = n\,\mathrm{tr}\,(\mathrm{A}\Sigma\mathrm{B}'\Sigma) + n\,\mathrm{tr}\,(\mathrm{A}\Sigma)\,\mathrm{tr}\,(\mathrm{B}\Sigma) + n^2\,\mathrm{tr}\,(\mathrm{A}\Sigma\mathrm{B}\Sigma)$.

Theorem 2.2.7 *Let* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$, *and let* A *be a* $p \times p$ *constant matrix. Then*

(1) $\mathrm{E}(\mathbf{W}^{-1}) = \dfrac{1}{n-p-1}\Sigma, \quad n - p - 1 > 0.$

(2) $\mathrm{E}(\mathbf{W}^{-1}\mathrm{A}\mathbf{W}^{-1}) = c_1\Sigma^{-1}\mathrm{A}\Sigma^{-1} + c_2[\Sigma^{-1}\mathrm{A}'\Sigma^{-1} + \,\mathrm{tr}\,(\mathrm{A}\Sigma^{-1})\Sigma^{-1}]$,

where $c_1 = (n-p-2)c_2$ *and* $c_2 = \{(n-p)(n-p-1)(n-p-3)\}^{-1}$.

Theorem 2.2.8 *Let* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$*, and let* A *and* B *be* $p \times p$ *constant matrices. Let* c_1 *and* c_2 *be the constants in Theorem 2.2.7. Then:*

$$(1)\quad \mathrm{E}[\mathrm{tr}\,(\mathrm{A}\mathbf{W}^{-1})] = \frac{1}{n-p-1}\,\mathrm{tr}\,(\mathrm{A}\Sigma), \quad n-p-1>0.$$

$$(2)\quad \mathrm{E}[(\mathrm{tr}\,\mathrm{A}\mathbf{W}^{-1})(\mathrm{tr}\,\mathrm{B}\mathbf{W}^{-1})] = c_1\,\mathrm{tr}\,(\mathrm{A}\Sigma^{-1})\,\mathrm{tr}\,(\mathrm{B}\Sigma^{-1}) + c_2[\mathrm{tr}\,(\mathrm{B}\Sigma^{-1}\mathrm{A}\Sigma^{-1} + \mathrm{tr}\,(\mathrm{B}\Sigma^{-1}\mathrm{A}'\Sigma^{-1})].$$

$$(3)\quad \mathrm{E}[\mathrm{tr}\,(\mathrm{A}\mathbf{W}^{-1}\mathrm{B}\mathbf{W}^{-1})] = c_1\,\mathrm{tr}\,(\mathrm{A}\Sigma^{-1}\mathrm{B}\Sigma^{-1}) + c_2[\mathrm{tr}\,(\mathrm{A}\Sigma^{-1}\mathrm{B}'\Sigma^{-1}) + \mathrm{tr}\,(\mathrm{A}\Sigma^{-1})\,\mathrm{tr}\,(\mathrm{B}\Sigma^{-1})].$$

Theorem 2.2.9 *Let* $\mathbf{W}$ *be a Wishart matrix having* $\mathrm{W}_p(n, \Sigma)$ *and* B *be a* $p \times q$ *constant matrix with rank* q*. Put*

$$\mathbf{Q} = (\mathrm{B}'\mathbf{W}^{-1}\mathrm{B})^{-1}\mathrm{B}'\mathbf{W}^{-1}\Sigma\mathbf{W}^{-1}\mathrm{B}(\mathrm{B}'\mathbf{W}^{-1}\mathrm{B})^{-1}.$$

Then

$$\mathrm{E}(\mathbf{Q}) = \frac{n-1}{n-(p-q)-1}(\mathrm{B}'\Sigma^{-1}\mathrm{B})^{-1}.$$

Proof. We may obtain the expectation of

$$\mathbf{Q} = (\mathrm{C}'\mathbf{W}^{-1}\mathrm{C})^{-1}\mathrm{C}'\mathbf{W}^{-2}\mathrm{C}(\mathrm{C}'\mathbf{W}^{-1}\mathrm{C})^{-1},$$

when $\mathbf{W} \sim \mathrm{W}_p(n, \mathrm{I}_p)$, where $\mathrm{C} = \Sigma^{-1/2}\mathrm{B}$. Let $\mathrm{H} = (\mathrm{H}_1\ \mathrm{H}_2)$ be an orthogonal matrix such that $\mathrm{H}_1 = \mathrm{C}(\mathrm{C}'\mathrm{C})^{-1/2}$. Then $\mathbf{V} = \mathrm{H}'\mathbf{W}\mathrm{H} \sim \mathrm{W}_p(m, \mathrm{I}_p)$. We decompose $\mathbf{V}$ as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \quad \mathbf{V}_{12} : q \times p,$$

and $\mathbf{Q}$ is expressed as

$$\mathbf{Q} = (\mathrm{B}'\Sigma^{-1}\mathrm{B})^{-1/2}\{\mathrm{I}_q + \mathbf{V}_{12}\mathbf{V}_{22}^{-2}\mathbf{V}_{21}\}(\mathrm{B}'\Sigma^{-1}\mathrm{B})^{-1/2}.$$

The required result is obtained by using Theorem 2.2.3. □

2.3 Functions of Wishart Matrices

Some properties of a Wishart matrix can be obtained by decomposing it to a triangular matrix. Let $\mathbf{W}$ be a $p \times p$ positive definite random matrix. Then, using [A.1.13] there exists a lower triangular matrix $\mathbf{T}$ such that

$$\mathbf{W} = \mathbf{T}\mathbf{T}', \quad \mathbf{T} = \begin{pmatrix} T_{11} & & & \\ T_{21} & T_{22} & & \text{\Large 0} \\ \vdots & \vdots & \ddots & \\ T_{p1} & T_{p2} & \cdots & T_{pp} \end{pmatrix}.$$

Theorem 2.3.1 *Let* $\mathbf{W}$ *be* $\mathrm{W}_p(n, \Sigma)$, *where* $n \geq p$, *and put* $\mathbf{W} = \mathbf{TT}'$, *where* $\mathbf{T}$ *is a lower triangular matrix with positive diagonal elements. Then the pdf of* $\mathbf{T}$ *is given by*

$$f(\mathrm{T}) = \frac{2^p}{K} \prod_{i=1}^{p} t_{ii}^{n-i} \operatorname{etr}\Big(-\frac{1}{2}\Sigma^{-1}\mathrm{TT}'\Big),$$

where $K = 2^{np/2}|\Sigma|^{n/2}\Gamma_p(n/2)$.

Proof. The density of $\mathbf{W}$ is given (Theorem 2.1.3) by

$$\frac{1}{K} \operatorname{etr}\Big(-\frac{1}{2}\Sigma^{-1}\mathrm{W}\Big) |\mathrm{W}|^{(n-p-1)/2}.$$

Put $\mathrm{W} = \mathrm{TT}'$, where T is a lower triangular matrix.

$$\operatorname{tr}\Sigma^{-1}\mathrm{W} = \operatorname{tr}\Sigma^{-1}\mathrm{TT}',$$
$$|\mathrm{W}| = |\mathrm{TT}'| = |\mathrm{T}|^2 = \prod_{i=1}^{k} t_{ii}^2.$$

The required result is proved by seeing that the Jacobian is [see, e.g. Muirhead, 1982; Siotani et al., 1985]

$$\left| \frac{\partial w_{ij}}{\partial t_{k\ell}} \right| = 2^p \prod_{i=1}^{p} t_{ii}^{p-i+1}. \qquad \square$$

Corollary 2.3.1 *If* $\Sigma = \mathrm{I}_p$, *the elements* $T_{ij} (p \geq i \geq j \geq 1)$ *are all independent,* $T_{ii}^2 \sim \chi^2(n-i+1)(i = 1, \ldots, p)$, *and* $T_{ij} \sim N(0,1)(p \geq i > j \geq 1)$.

Proof. Since $\Sigma = \mathrm{I}_p$,

$$\operatorname{tr}\Sigma^{-1}\mathrm{TT}' = \operatorname{tr}\mathrm{TT}' = \sum_{i \geq j}^{p} t_{ij}^2.$$

Therefore, the joint density of $T_{ij} (i \geq j)$ can be written in the form

$$\prod_{i=1}^{p} \left\{ \frac{1}{2^{n-i+1}\Gamma[\frac{1}{2}(n-i+1)]} \exp\left(-\frac{1}{2} t_{ii}^2 \right) (t_{ij}^2)^{(n-i-1)/2} \right\}$$
$$\times \prod_{i>j}^{p} \left\{ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t_{ij}^2 \right) \right\},$$

which is the product of the marginal density functions for the elements of $\mathbf{T}$ stated in the corollary. $\square$

The result in Corollary 2.3.1, due to Bartlett (1933), is called *Bartlett's decomposition.*

Let $\mathbf{S}$ be the sample covariance matrix based on a sample of size n from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. The sample generalized variance is defined by $|\mathbf{S}|$. Then since $(n-1)\mathbf{S} \sim \mathrm{W}_p(n-1, \Sigma)$, the distribution of $|\mathbf{S}|$ will be obtained based on the following theorem.

Theorem 2.3.2 *If* $\mathbf{W} \sim \mathrm{W}(n, \Sigma)$, *where* $n \geq p$, *then* $|\mathbf{W}|/|\Sigma|$ *has the same distribution as that of* $\prod_{i=1}^p \chi_{n-i+1}^2$, *where for* $i = 1, \ldots, p$, χ_{n-i+1}^2 *denote mutually independent random variables, each following the chi-square distribution with n-i+1 degrees of free.*

Proof. Note that

$$\frac{|\mathbf{W}|}{|\Sigma|} = |\Sigma^{-1/2}\mathbf{W}\Sigma^{-1/2}|$$

and $\Sigma^{-1/2}\mathbf{W}\Sigma^{-1/2} \sim \mathrm{W}_p(n, \mathrm{I}_p)$. Let T be a lower triangular matrix such that $\Sigma^{-1/2}\mathbf{W}\Sigma^{-1/2} = \mathbf{TT}'$. Then the result follows from Corollary 2.3.1 by taking $\mathbf{TT}' = \prod_{i=1}^p T_{ii}^2 = \prod_{i=1}^p \chi_{n-i+1}^2$. □

Theorem 2.3.3 *If* $\mathbf{W}$ *is distributed as* $\mathrm{W}_p(n, \Sigma)$ *and* M *is a* $q \times p$ *matrix of rank* q, *then* $(\mathrm{M}\mathbf{W}^{-1}\mathrm{M}')^{-1}$ *is distributed as* $\mathrm{W}_q(n-p+q, (\mathrm{M}\Sigma^{-1}\mathrm{M}')^{-1})$.

Proof. Put $\tilde{\mathbf{W}} = \Sigma^{-1/2}\mathbf{W}\Sigma^{-1/2}$, which is distributed [Theorem 2.2.1 (2)] as $\mathrm{W}_p(n, \mathrm{I}_p)$. Then

$$(\mathrm{M}\mathbf{W}^{-1}\mathrm{M}')^{-1} = (\mathrm{R}\tilde{\mathbf{W}}^{-1}\mathrm{R}')^{-1},$$

where $\mathrm{R} = \mathrm{M}\Sigma^{-1/2}$. Since $(\mathrm{RR}')^{-1} = (\mathrm{M}\Sigma^{-1}\mathrm{M}')^{-1}$, it is sufficient to show that $(\mathrm{R}\tilde{W}^{-1}\mathrm{R}')^{-1} \sim \mathrm{W}_q(n-p+q, (\mathrm{RR}')^{-1})$. Consider an orthogonal matrix $\mathrm{H} = (\mathrm{H}_1\ \mathrm{H}_2)$, whose first q columns H_1 are $\mathrm{H}_1 = \mathrm{R}'(\mathrm{RR}')^{-1/2}$. Then $\mathbf{V} = \mathrm{H}'\tilde{\mathbf{W}}\mathrm{H} \sim \mathrm{W}_p(n, \mathrm{I}_p)$ and let $\mathbf{V}$ partition as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \quad \mathbf{V}_{12} : q \times (p-q).$$

Noting that the first $q \times q$ submatrix of $\mathbf{V}^{-1}$ is expressed (see [A.1.2]) as $\mathbf{V}_{11\cdot 2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$,

$$\begin{aligned}(\mathrm{R}\tilde{\mathbf{W}}^{-1}\mathrm{R}')^{-1} &= \{(\mathrm{RH})\mathbf{V}^{-1}(\mathrm{RH})'\}^{-1} \\ &= (\mathrm{RR}')^{-1/2}\{(\mathrm{I}\ \mathrm{O})\mathbf{V}^{-1}(\mathrm{I}\ \mathrm{O})'\}^{-1}(\mathrm{RR}')^{-1/2} \\ &= (\mathrm{RR}')^{-1/2}\mathbf{V}_{11\cdot 2}(\mathrm{RR}')^{-1/2}.\end{aligned}$$

Therefore, the required result is obtained by using Theorems 2.2.2(2) and 2.2.1(2). □

2.4 Cochran's Theorem

Cochran's theorem [Cochran, 1934] is useful in proving that certain vector quadratic forms $\boldsymbol{X}'\mathrm{A}_1\boldsymbol{X}, \ldots, \boldsymbol{X}'\mathrm{A}_k\boldsymbol{X}$ are independently distributed as central or noncentral chi-squares distributions, where $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \sigma^2\mathrm{I}_n)$ and A_i's are $n \times n$ symmetric matrices. There are various versions of *Cochran's theorem*, which include the following:

Theorem 2.4.1 *Let* $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$ *be an n-vector such that* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ *are independent and* $\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, \ldots, n$. *Suppose that the random variable* $\mathbf{Q} = \mathrm{X}'\mathrm{A}\mathbf{X}$ *is decomposed into a sum of k quadratic forms* $Q_i(\mathbf{X}) = \mathbf{X}'\mathrm{A}_i\mathbf{X}$, $i = 1, \ldots, k$:

$$Q(\mathbf{X}) = Q_1(\mathbf{X}) + \cdots + Q_k(\mathbf{X}),$$

where A *and* A_i $(i = 1, \ldots, k)$ *are* $n \times n$ *symmetric matrices of ranks* r *and* r_i. *If* A *is an idempotent matrix, then* $Q_i(\mathbf{X})$, $i = 1, \ldots, k$ *are independent and each has the noncentral Wishart* $\mathrm{W}_p(r_i, \Sigma; \Omega_i)$ *if and only if* $r = r_1 + \cdots + r_k$, *where* $\Omega_i = \mathrm{M}'\mathrm{A}_i\mathrm{M}$ *and* $\mathrm{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)'$.

As is well known, Theorem 2.4.1 is a statistical statement for a matrix decomposition

$$\mathrm{A} = \mathrm{A}_1 + \cdots + \mathrm{A}_k.$$

For its details and a proof of Theorem 2.4.1, see, for example, Siotani et al. (1985). Here, we state a simple multivariate version of Cochran's theorem, which will be used often.

Theorem 2.4.2 *Let* $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$, *where* $\boldsymbol{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, \ldots, n$ *and* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ *are independent. Let* A, A_1, *and* A_2 *be* $n \times n$ *symmetric matrices and* B *be a* $q \times n$ *matrix with rank* q. *Then:*

(1) $\mathbf{X}'\mathrm{A}\mathbf{X} \sim \mathrm{W}_p(k, \Sigma; \Delta) \Longleftrightarrow \mathrm{A}^2 = \mathrm{A}, \operatorname{tr}\mathrm{A} = k, \Delta = \mathrm{E}(\mathbf{X})'\mathrm{A}\mathrm{E}(\mathbf{X})$.

(2) $\mathbf{X}'\mathrm{A}_1\mathbf{X}$ *and* $\mathbf{X}'\mathrm{A}_2\mathbf{X}$ *are independent* $\Longleftrightarrow \mathrm{A}_1\mathrm{A}_2 = \mathrm{O}$.

(3) $\mathbf{X}'\mathrm{A}\mathbf{X}$ *and* $\mathrm{B}\mathbf{X}$ *are independent* $\Longleftrightarrow \mathrm{B}\mathrm{A} = \mathrm{O}$.

Proof. For a $p \times 1$ constant vector $\boldsymbol{a}$, $\boldsymbol{Y} = \mathbf{X}\boldsymbol{a} \sim N_n(\boldsymbol{\eta}, \sigma^2\mathrm{I}_p)$, where $\boldsymbol{\eta} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)'\boldsymbol{a}$ and $\sigma^2 = \boldsymbol{a}'\Sigma\boldsymbol{a}$. The necessity ($\Longrightarrow$) of each of (1), (2), and (3) follows from the results in the univariate case. Now we prove the sufficiency of (1). For an $n \times n$ symmetric matrix A, there exist an orthogonal matrix H and a diagonal matrix $\mathrm{L} = \operatorname{diag}(\ell_1, \ldots, \ell_p)$, $\ell_1 \geq \cdots \geq \ell_n$ such that $\mathrm{A} = \mathrm{H}\mathrm{L}\mathrm{H}'$. Since $\mathrm{A}^2 = \mathrm{A}$ and $\operatorname{tr} A = k$, $\ell_1 = \cdots = \ell_k = 1$ and $\ell_{k+1} = \cdots = \ell_p = 0$. Put

$$\mathbf{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)' = \mathrm{H}'\mathbf{X}.$$

Then, $\boldsymbol{Z}_i$'s are independent and normal with the same covariance matrix Σ. Further

$$\mathbf{X}'\mathrm{A}\mathbf{X} = \mathbf{Z}'\mathrm{L}\mathbf{Z} = \sum_{i=1}^{k} \boldsymbol{Z}_i\boldsymbol{Z}_i' \sim \mathrm{W}_p(k, \Sigma; \Omega),$$

where $\Omega = \sum_{i=1}^{k} \mathrm{E}(\boldsymbol{Z}_i)\mathrm{L}\mathrm{E}(\boldsymbol{Z}_i)'$. The proof is completed by seeing that $\Omega = \mathrm{E}(\mathbf{Z})'\mathrm{L}\mathrm{E}(\mathbf{Z}) = \mathrm{E}(\mathbf{X})'\mathrm{A}\mathrm{E}(\mathbf{X})$. Next we prove the sufficiency ($\Leftarrow$) of (2). We can write $\mathrm{A}_i = \mathrm{H}_i\mathrm{L}_i\mathrm{H}_i$, $i = 1, 2$, where $\mathrm{H}_i'\mathrm{H}_i = \mathrm{I}_{n_i}$ and $\mathrm{L}_i = \mathrm{diag}\,(\ell_{i1}, \ldots, \ell_{im_i})$ with $\ell_{ij} \neq 0$. Since $\mathrm{A}_1'\mathrm{A}_2 = \mathrm{O}$, we can construct an orthogonal matrix $\mathrm{H} = (\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3)$. Put $\mathrm{H}_i'\mathbf{X} = \mathbf{Z}_i$, $i = 1, 2, 3$. Then $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ are independent. The result is obtained by noting that $\mathbf{X}'\mathrm{A}_1\mathbf{X} = \mathbf{Z}_1'\mathrm{L}_1\mathbf{Z}_1$ and $\mathbf{X}'\mathrm{A}_2\mathbf{X} = \mathbf{Z}_2'\mathrm{L}_2\mathbf{Z}_2$. Similarly, the sufficiency ($\Leftarrow$) of (3) is proved. $\square$

As an example of the use of Theorem 2.4.2, let $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ be the sample mean vector and sample covariance matrix based on a random sample $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$ of size n from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Then

$$\bar{\boldsymbol{X}} = \mathbf{X}' \cdot \frac{1}{n}\mathbf{1}_n, \quad (n-1)\mathbf{S} = \mathbf{X}'(\mathrm{I}_n - \mathrm{P}_0)\mathbf{X},$$

where $\mathbf{1}_n = (1, \ldots, 1)'$ and $\mathrm{P}_0 = (1/n)\mathbf{1}_n\mathbf{1}_n'$. It is easy to see that $\mathrm{I}_n - \mathrm{P}_0$ is an idempotent matrix and $(\mathrm{I}_n - \mathrm{P}_0)\mathbf{1}_n = \mathbf{0}$. Further, $\mathrm{tr}\,(\mathrm{I}_n - \mathrm{P}_0) = n - 1$ and $\Omega = \mathrm{M}'(\mathrm{I}_n - \mathrm{P}_0)\mathrm{M} = \mathrm{O}$ when $\mathrm{M} = (\boldsymbol{\mu}, \ldots, \boldsymbol{\mu})' = \mathbf{1}_n\boldsymbol{\mu}'$. Therefore, $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ are independent and $(n-1)\mathbf{S} \sim \mathrm{W}_p(n-1, \Sigma)$.

2.5 Asymptotic Distributions

Let $n\mathbf{S} \sim \mathrm{W}_p(n, \Sigma)$ and consider *asymptotic distribution* of

$$\mathbf{V} = \sqrt{n}(\mathbf{S} - \Sigma). \tag{2.5.1}$$

The characteristic function of $\mathbf{V}$ is expressed as

$$\begin{aligned} C_{\mathbf{V}}(\mathrm{T}) &= \mathrm{E}[e^{i\,\mathrm{tr}\,\mathrm{T}\mathbf{V}}] \\ &= \exp(-\sqrt{n}i\,\mathrm{tr}\,\mathrm{T}\Sigma) \times \mathrm{E}[e^{i\,\mathrm{tr}\,(1/\sqrt{n})\mathrm{T}\cdot n\mathbf{S}}], \end{aligned}$$

where $\mathrm{T} = (1 + \delta_{ij})t_{ij}/2$. Using Theorem 2.1.2, we have

$$C_{\mathbf{V}}(\mathrm{T}) = \exp(-\sqrt{n}i\,\mathrm{tr}\,\Sigma\mathrm{T}) \times \left|\mathrm{I}_p - \frac{2i}{\sqrt{n}}\Sigma\mathrm{T}\right|^{-n/2}.$$

For any symmetric matrix A,

$$-\log|\mathrm{I}_p - \varepsilon\mathrm{A}| = \varepsilon\,\mathrm{tr}\,\mathrm{A} + \frac{1}{2}\varepsilon^2\,\mathrm{tr}\,\mathrm{A}^2 + \frac{1}{3}\varepsilon^3\,\mathrm{tr}\,\mathrm{A}^3 + \cdots. \tag{2.5.2}$$

Therefore

$$\begin{aligned}\log C_{\mathbf{V}}(\mathrm{T}) &= -\sqrt{n}\, i \operatorname{tr} \mathrm{T}\Sigma - \frac{n}{2} \log \left| \mathrm{I}_p - \frac{2i}{\sqrt{n}} \Sigma \mathrm{T} \right| \\ &= -\operatorname{tr}(\Sigma\mathrm{T})^2 + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

and hence

$$C_{\mathbf{V}}(\mathrm{T}) = \exp\{-\operatorname{tr}(\Sigma\mathrm{T})^2\} + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right).$$

The asymptotic covariance between v_{ij} and $v_{k\ell}$ is:

$$\text{The coefficient of } t_{ij}t_{k\ell} \text{ in } 2\operatorname{tr}(\Sigma\mathrm{T})^2 = \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}.$$

Theorem 2.5.1 *Let $n\mathbf{S} \sim \mathrm{W}_p(n, \Sigma)$ and put $\mathbf{S} = (S_{ij})$. Then the limiting distribution of $\mathbf{V} = \sqrt{n}(\mathbf{S} - \Sigma)$ [i.e., the limiting joint distribution of $\sqrt{n}(S_{ij} - \sigma_{ij})$] is the $\{p(p+1)/2\}$-variate normal distribution with mean zero and covariance matrix with elements*

$$\mathrm{Cov}(\sqrt{n}(S_{ij} - \sigma_{ij}), \sqrt{n}(S_{k\ell} - \sigma_{k\ell})) = \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}. \tag{2.5.3}$$

For asymptotic distribution of a function of a Wishart matrix, we can use the following theorem [see Serfling, 1980, Anderson, 2003, and Theorem 4.2.3].

Theorem 2.5.2 *Let $\{\boldsymbol{U}(n)\}$ be a sequence of m-component random vectors and $\boldsymbol{b}$ be a fixed vector such that $\sqrt{n}(\boldsymbol{U}(n) - \boldsymbol{b})$ has the limiting distribution $\mathrm{N}_m(\mathbf{0}, \Psi)$ as $n \to \infty$. Let $\boldsymbol{f}(\boldsymbol{u}) = (f_1(\boldsymbol{u}), \ldots, f_k(\boldsymbol{u}))$, $k \le m$ be a vector-valued function of $\boldsymbol{u}$ such that each component $f_i(\boldsymbol{u})$ is continuously differentiable in a neighborhood of $\boldsymbol{u} = \boldsymbol{b}$. Then the limiting distribution of $\sqrt{n}\{\boldsymbol{f}(\boldsymbol{U}(n)) - \boldsymbol{f}(\boldsymbol{b})\}$ is $\mathrm{N}_k(\mathbf{0}, \mathrm{A}\Psi\mathrm{A}')$, where $\mathrm{A} = (a_{ij})$ is a $k \times m$ matrix with $a_{ij} = \partial f_i(\boldsymbol{u})/\partial u_j|_{\boldsymbol{u}=\boldsymbol{b}}$.*

Proof. The proof is based on the result that for every $\varepsilon > 0$ there exists a neighborhood $N_\varepsilon(\boldsymbol{b})$ such that

$$\left| f_j(\boldsymbol{u}) - f_j(\boldsymbol{b}) - \sum_{i=1}^{m} \partial f_j(\boldsymbol{u})/\partial u_i(\boldsymbol{u}_j - \boldsymbol{b}_j) \right| \le \varepsilon \|\boldsymbol{u} - \boldsymbol{b}\|$$

for all $\boldsymbol{u} \in N_\varepsilon(\boldsymbol{b})$. □

Now we consider asymptotic variance (A-Var) of the random variable $f(\mathbf{S})$, where $n\mathbf{S} \sim \mathrm{W}_p(n, \Sigma)$. Suppose that $f(\mathrm{S})$ is continuously differentiable in a neighborhood of $\mathrm{S} = \Sigma$. Then the asymptotic distribution of $f(\mathbf{S})$ is normal with mean $f(\Sigma)$ and the following variance:

$$\text{A-Var}\{f(\mathbf{S})\} = \frac{2}{n} \operatorname{tr}(\Gamma\Sigma)^2, \tag{2.5.4}$$

where $\Gamma = (\gamma_{ij})$,

$$\gamma_{ij} = \partial_{ij} f(\Sigma), \quad \partial_{ij} \equiv \frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial \sigma_{ij}}.$$

The expression (2.5.4) is as follows. From Theorems 2.5.1 and 2.5.2,

$$\begin{aligned}
&\text{A-Var}\{f(\mathbf{S})\} \\
&\quad = \sum_{i,j,k,\ell} \frac{1}{2}(1+\delta_{ij})\frac{\partial f(\Sigma)}{\partial \sigma_{ij}} \cdot \frac{1}{2}(1+\delta_{k\ell})\frac{\partial f(\Sigma)}{\partial \sigma_{k\ell}} \times \frac{1}{n}(\sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}) \\
&\quad = \frac{1}{n}\sum_{i,j,k,\ell} \gamma_{ij}\gamma_{k\ell}(\sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}) \\
&\quad = \frac{2}{n}\,\mathrm{tr}\,(\Gamma\Sigma)^2.
\end{aligned}$$

Let R be a sample correlation coefficient based on a random sample of size $N = n+1$ from a normal population $\mathrm{N}_2(\boldsymbol{\mu}, \Sigma)$. Then we may write

$$R = S_{12}/\sqrt{S_{11}S_{22}} = V_{12}/\sqrt{V_{11}V_{22}},$$

where $\mathbf{S} = (S_{ij})$, $\mathbf{V} = (V_{ij})$, and $n\mathbf{S} = \mathbf{V} \sim \mathrm{W}_2(n, \Sigma)$. Then by Theorem 2.5.2, R is asymptotically normal with mean $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ and variance calculated by Theorem 2.5.2 on (2.5.4). In the following we obtain the asymptotic variance in two ways.

Since R is invariant under the transformation $\tilde{S}_{ij} = S_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, without loss of generality we may assume that

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \equiv \Sigma_0.$$

For simplicity, let

$$\boldsymbol{U}(n) = \frac{1}{n}\begin{pmatrix} V_{11} \\ V_{22} \\ V_{12} \end{pmatrix} = \frac{1}{n}\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

and hence $R = U_3 U_1^{-1/2} U_2^{-1/2}$. Then, by Theorem 2.5.1 the random vector $\sqrt{n}\big(\boldsymbol{U}(n) - \boldsymbol{b}\big)$ has a limiting normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\Phi = \begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix}.$$

The function $r = u_3 u_1^{-1/2} u_2^{-1/2}$ satisfies the conditions of the theorem and

$$\begin{aligned}
\boldsymbol{a}' &= \left(\frac{\partial r}{\partial u_1}, \frac{\partial r}{\partial u_2}, \frac{\partial r}{\partial u_3}, \right)\bigg|_{\boldsymbol{u}=\boldsymbol{b}} \\
&= \left(-\frac{1}{2}\rho, -\frac{1}{2}\rho, 1\right).
\end{aligned}$$

Therefore, the variance of the limiting distribution of $\sqrt{n}(R-\rho)$ is

$$\boldsymbol{a}'\Phi\boldsymbol{a} = (1-\rho^2)^2,$$

and the asymptotic variance of R is $(1-\rho^2)^2/n$.

Next we consider the formula (2.5.4). The elements of Γ are

$$\begin{aligned}
\gamma_{11} &= \frac{\partial}{\partial\sigma_{11}}\sigma_{12}\sigma_{11}^{-1/2}\sigma_{22}^{-1/2}\Big|_{\Sigma=\Sigma_0} = -\frac{1}{2}\rho,\\
\gamma_{22} &= \frac{\partial}{\partial\sigma_{22}}\sigma_{12}\sigma_{11}^{-1/2}\sigma_{22}^{-1/2}\Big|_{\Sigma=\Sigma_0} = -\frac{1}{2}\rho,\\
\gamma_{12} = \gamma_{21} &= \frac{1}{2}\frac{\partial}{\partial\sigma_{12}}\sigma_{12}\sigma_{11}^{-1/2}\sigma_{22}^{-1/2}\Big|_{\Sigma=\Sigma_0} = -\frac{1}{2}.
\end{aligned}$$

These yield the asymptotic variance

$$\frac{2}{n}\,\mathrm{tr}\,(\Gamma\Sigma_0)^2 = \frac{1}{n}(1-\rho^2)^2.$$

Problems

2.1 (Noncentral chi-square distribution) Assuming that $Z_k\ (k=1,\dots,n)$ are independent and distributed as $\mathrm{N}(\mu_k,1)$, show that the density function of $X=\sum_{k=1}^n Z_k^2$ is given by

$$f(x) = \mathrm{e}^{-\delta^2/2}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{\delta^2}{2}\right)^k \frac{x^{(n+2k)/2-1}\mathrm{e}^{-x/2}}{\Gamma((n+2k)/2)2^{(n+2k)/2}},\quad x>0.$$

This distribution, $\chi^2(n;\delta^2)$, is called the *noncentral chi-square distribution* with degrees of freedom n and *noncentrality parameter* δ^2.

Hint: If $Z\sim\mathrm{N}(\mu,1)$, the characteristic function is

$$\begin{aligned}
C_{Z^2}(t) &= \mathrm{E}[\exp(itZ^2)]\\
&= (1-2it)^{-1/2}\exp\left\{-\frac{1}{2}\mu^2+\frac{1}{2}(1-2it)^{-1}\mu^2\right\}.
\end{aligned}$$

Therefore, the characteristic function of $X=\sum_{k=1}^n Z_k^2$ can be put in the following form:

$$\begin{aligned} C_X(t) &= \prod_{k=1}^{n} \mathrm{E}[\exp(itZ_k^2)] \\ &= \prod_{k=1}^{n} (1-2it)^{-1/2} \exp\left\{ -\frac{1}{2}\mu_k^2 + \frac{1}{2}(1-2it)^{-1}\mu_k^2 \right\} \\ &= (1-2it)^{-1/2} \exp\left\{ -\frac{1}{2}\delta^2 + \frac{1}{2}(1-2it)^{-1}\delta^2 \right\} \\ &= \mathrm{e}^{-\delta^2/2} \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{1}{2}\delta^2\right)^k (1-2it)^{-(n+2k)/2}. \end{aligned}$$

Inverting this characteristic function, we obtain the probability density function.

2.2 Let K be a random variable with Poisson distribution $\mathrm{Po}(\frac{1}{2}\delta^2)$. Then check that the density function and characteristic function of X with $\chi^2(n;\delta^2)$ are written as

$$\begin{aligned} f_X(x) &= \mathrm{E}\left[\frac{x^{(n+2K)/2-1}\mathrm{e}^{-x/2}}{\Gamma((n+2K)/2)2^{(n+2k)/2}}\right], \\ C_X(t) &= \mathrm{E}\left[(1-2it)^{-(n+2K)/2}\right], \end{aligned}$$

respectively. The density function is also written in terms of a hypergeometric function as

$$f_X(x) = \frac{x^{n/2-1}\mathrm{e}^{-x/2}}{\Gamma(n/2)2^{n/2}}\mathrm{e}^{-\delta/2}{}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\delta x\right),$$

where

$${}_0F_1(b;z) = \sum_{k=0}^{\infty} \frac{1}{(b)_k}\frac{z^k}{k!}, \quad (b)_k = b(b+1)\cdots(b+k-1).$$

2.3 Let $\mathbf{W} \sim \mathrm{W}_p(n, \mathrm{I}_p)$. Show that $\mathrm{tr}\,\mathbf{W}$ has the chi-aquare distribution with np degrees of freedom.

2.4 Let $\mathbf{W} \sim \mathrm{W}_p(n,\Sigma)$. Show that $\boldsymbol{a}'\mathbf{W}\boldsymbol{a}$ and $\boldsymbol{b}'\mathbf{W}\boldsymbol{b}$ are independent if and only if $\boldsymbol{a}'\Sigma\boldsymbol{b} = 0$.

2.5 Let $\mathbf{W}$ be distributed as $\mathrm{W}_p(n,\Sigma)$. Then, show that the pdf of $\mathbf{U} = \mathbf{W}^{-1}$ is given by

$$\frac{1}{2^{pn/2}\Gamma_p(\frac{1}{2}n)|\Sigma|^{n/2}}|\mathrm{U}|^{-(n+p+1)/2}\,\mathrm{etr}\left(-\frac{1}{2}\Sigma^{-1}\mathrm{U}^{-1}\right).$$

Hint: Use [A.3.5].

2.6 Suppose that $\mathbf{W}$ is distribution as a Wishart distribution $\mathrm{W}_p(n, \Sigma)$, and let $\mathbf{W}$ and Σ partition as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Put

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix},$$

where $\mathbf{U}_{11} = \mathbf{W}_{11\cdot 2} = \mathbf{W}_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}$, $\mathbf{U}_{12} = \mathbf{U}_{21}' = \mathbf{W}_{11}\mathbf{W}_{22}^{-1/2}$ and $\mathbf{U}_{22} = \mathbf{W}_{22}$.

(1) Show that the transformation from $\mathbf{W}$ and $\mathbf{U}$ is one-to-one.

(2) When $\Sigma = \mathrm{I}_p$, obtain the joint pdf of $\mathbf{U}$.

Hint: Use Theorem 2.2.3, or obtain the Jacobian of the inverse transformation.

2.7 Let $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i} \sim$ i.i.d.$\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, 2$, and let

$$\bar{\boldsymbol{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{X}_{ij}, \quad i = 1, 2, \qquad \bar{\boldsymbol{X}} = \frac{1}{n}\{n_1\bar{\boldsymbol{X}}_1 + n_2\bar{\boldsymbol{X}}_2\},$$

where $n = n_1 + n_2$. Examine that

$$\mathbf{B} = \sum_{i=1}^{2} n_i(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})' \sim \mathrm{W}_p(1, \Sigma; \Omega),$$

by the following two methods, where $\Omega = (n_1n_2/n)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'$.
Method 1: Use $\mathbf{B} = (n_1n_2/n)(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'$.
Method 2: Express $\mathbf{B} = \mathbf{X}'\mathrm{P}_b\mathbf{X}$, and use Cochran theorem, where $\mathbf{X} = (\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}, \boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2})$,

$$\mathrm{P}_b = \begin{pmatrix} \frac{1}{n_1}\mathrm{G}_{n_1} & \mathrm{O} \\ \mathrm{O} & \frac{1}{n_2}\mathrm{G}_{n_2} \end{pmatrix} - \frac{1}{n}\mathrm{G}_n,$$

and G_m is the $m \times m$ matrix whose elements are all one.

2.8 Let $\mathbf{W} \sim \mathrm{W}_p(n, \mathrm{I}_p; \Omega)$ and $\mathbf{W}$ and Ω be partitioned as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{pmatrix},$$

respectively, where $\mathbf{W}_{ij} : p_i \times p_j$ and $\Omega_{11} : p_1 \times p_1$. Then show that:

(1) $\mathbf{W}_{22\cdot 1}$, $\mathbf{W}_{21}\mathbf{W}_{11}^{-1/2}$, and $\mathbf{W}_{11}$ are mutually independent.

(2) $\mathbf{W}_{22\cdot 1} \sim \mathrm{W}_{p_2}(n - p_1, \mathrm{I}_{p_2})$, $\mathbf{W}_{11} \sim \mathrm{W}_{p_1}(n, \Omega_{11})$, and the columns of $\mathbf{W}_{12}\mathbf{W}_{22}^{-1/2}$ are independent and identically distributed as $\mathrm{N}_{p_2}(\mathbf{0}, \mathrm{I}_{p_2})$, and hence $\mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{12} \sim \mathrm{W}_{p_2}(n, \mathrm{I}_{p_1})$.

2.9 Suppose that $n\mathbf{S}$ is distributed as a Wishart distribution $\mathrm{W}_p(n, \Sigma)$. Obtain the asymptotic variances of $\sqrt{n}(\operatorname{tr}\mathbf{S} - \operatorname{tr}\Sigma)$ and $\sqrt{n}(\operatorname{tr}\mathbf{S}^2 - \operatorname{tr}\Sigma^2)$.

2.10 (Continuation) Show that the asymptotic distribution of $|\mathbf{S}|$ is normal $\mathrm{N}(|\Sigma|, (2p/n)|\Sigma|^2)$ or, what is the same thing, that the asymptotic distribution of $\sqrt{n}(|\mathbf{S}|/|\Sigma| - 1)$ is $\mathrm{N}(0, 2p)$.

2.11 Let $\mathbf{W}$ have a Wishart distribution $\mathrm{W}_p(n, \Sigma)$, and let $\mathbf{W}_i$ and Σ_i be the matrices consisting of the first i rows and columns of $\mathbf{W}$ and Σ, respectively with $\mathbf{W}_0 = 1$ and $\Sigma_0 = 1$. Show that

$$V_i = \frac{|\mathbf{W}_i|}{|\mathbf{W}_{i-1}|}\left(\frac{|\Sigma_i|}{|\Sigma_{i-1}|}\right)^{-1} \sim \chi^2(n - i + 1),$$

and that $V_1, \ldots, V_p$ are independent.

2.12 Suppose that $n\mathbf{S}$ is distributed as a Wishart distribution $\mathrm{W}_p(n, \Sigma)$. Then show that

$$\begin{pmatrix} \sqrt{n}\left(\operatorname{tr}\mathbf{S} - \operatorname{tr}\Sigma\right) \\ \sqrt{n}\left(\operatorname{tr}\mathbf{S}^2 - tr\Sigma^2\right) \end{pmatrix} \xrightarrow{d} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\operatorname{tr}\Sigma^2 & 4\operatorname{tr}\Sigma^3 \\ 4\operatorname{tr}\Sigma^3 & 8\operatorname{tr}\Sigma^4 \end{pmatrix}\right).$$

Chapter 3

Hotelling's T^2 and Lambda Statistics

In this chapter we treat the *Hotelling* T^2 and Lambda statistics. They are statistics measuring the differences between two mean vectors and among several mean vectors, respectively. The exact distributions are given. Their decompositions are studied. We also study the likelihood ratio test for additional information when several mean vectors are compared.

3.1 T^2 Statistic

3.1.1 Distribution of the T^2 Statistic

Let $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ be the sample mean vector and the sample covariance matrix based on a sample of size n from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Then it is important to know the distribution of

$$T_{(1)}^2 = n\bar{\boldsymbol{X}}'S^{-1}\bar{\boldsymbol{X}},$$

which is a test statistic for $\boldsymbol{\mu} = \mathbf{0}$. It is also important to know the distribution of the statistic obtained from $T_{(1)}^2$ by replacing $\bar{\boldsymbol{X}}$ by $\bar{\boldsymbol{X}} - \boldsymbol{\mu}$. Note that the latter distribution is obtained as the distribution of $T_{(1)}^2$ with $\boldsymbol{\mu} = \mathbf{0}$ (i.e., the *null distribution* of $T_{(1)}^2$).

For comparison of two mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, let $\bar{\boldsymbol{X}}_i$ and $\mathbf{S}_i$ $(i = 1, 2)$ be the sample mean vectors and the sample covariance matrices based on samples of sizes n_i from $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$, $i = 1, 2$. Let $\mathbf{S}$ be a pooled sample covariance matrix defined by

$$\mathbf{S} = \frac{1}{n-2}\{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2\},$$

where $n = n_1 + n_2$. Then it is also important to know the distribution of

$$T_{(2)}^2 = \sqrt{\frac{n_1 n_2}{n}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\mathbf{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2).$$

The statistics $T_{(1)}^2$ and $T_{(2)}^2$ are called T^2 statistic, which was introduced by Hotelling (1936). 43 The distributions of these statistics are treated as

$$T^2 = \boldsymbol{Z}'\left(\frac{1}{m}\mathbf{W}\right)^{-1}\boldsymbol{Z}, \tag{3.1.1}$$

where $\boldsymbol{Z} \sim \mathrm{N}_p(\boldsymbol{\nu}, \Sigma)$, $m\mathbf{W} \sim \mathrm{W}_p(m, \Sigma)$, and $\boldsymbol{Z}$ and $\mathbf{W}$ are independent. The distribution of T^2 is called the T^2 *distribution.* For $T_{(1)}^2$, $m = n - 1$, and for $T_{(2)}^2$, $m = n_1 + n_2 - 2$. We denote the distribution of T^2 by $T(p, m; \delta^2)$, where $\delta^2 = \boldsymbol{\nu}'\Sigma^{-1}\boldsymbol{\nu}$, and denote $T(p, m; 0)$ simply by $T(p, m)$.

Theorem 3.1.1 *Let T^2 be the random variable defined by* (3.1.1)*. Then*

$$\frac{1}{m}T^2 \cdot \frac{m-p+1}{p} \sim F(p, m-p+1; \delta^2),$$

where $F(a, b; \delta^2)$ denotes a noncentral F-distribution with degrees of freedom a, b and noncentrality parameter δ^2.

Proof. When we consider the distribution of T^2, without loss of generality we may assume that $\Sigma = \mathrm{I}_p$ and $\boldsymbol{Z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathrm{I}_p)$, $\boldsymbol{\mu} = \Sigma^{-1}\boldsymbol{\nu}$. Consider the conditional distribution of T^2 given $\boldsymbol{Z}$. The conditional distribution of $\mathbf{W}$ given $\boldsymbol{Z}$ is $\mathrm{W}_p(m, \mathrm{I}_p)$ since $\mathbf{W}$ is independent of $\boldsymbol{Z}$. Let H be an orthogonal matrix whose first column is $\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1/2}$. Then the conditional distribution of $\mathbf{V} = \mathrm{H}'\mathbf{W}\mathrm{H}$ is $W_p(m, \mathrm{I}_p)$, which does not depend on $\boldsymbol{Z}$. Therefore, $\mathbf{V} \sim \mathrm{W}_p(m, \mathrm{I}_p)$ and is independent of $\boldsymbol{Z}$. Decompose $\mathbf{V}$ as

$$V = \begin{pmatrix} V_{11} & \boldsymbol{V}_{21}' \\ \boldsymbol{V}_{21} & \mathbf{V}_{22} \end{pmatrix}.$$

Note that the $(1, 1)$ element of $\mathbf{V}^{-1}$ is $V_{11\cdot 2}^{-1}$, where $V_{11\cdot 2} = V_{11} - \boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}$. Then

$$\begin{aligned}\frac{1}{m}\mathrm{T}^2 &= (\mathrm{H}\boldsymbol{Z})'(\mathrm{H}'\mathbf{W}\mathrm{H})^{-1}(\mathrm{H}'\boldsymbol{Z}) \\ &= ((\boldsymbol{Z}'\boldsymbol{Z})^{1/2}, 0, \ldots, 0)\mathbf{V}^{-1}((\boldsymbol{Z}'\boldsymbol{Z})^{1/2}, 0, \ldots, 0)' \\ &= \boldsymbol{Z}'\boldsymbol{Z} \cdot V_{11\cdot 2}^{-1}.\end{aligned}$$

The required result follows from the fact that $\boldsymbol{Z}'\boldsymbol{Z} \sim \chi^2(p; \delta^2)$ and $V_{11\cdot 2} \sim \chi^2(n - (p-1))$ (see Theorem 2.2.2). □

The T^2 *statistic* $T^2_{(2)}$ in the *two-sample case* is closely related to the *Mahalanobis distance* between two population, defined by

$$D^2 = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\mathbf{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2). \tag{3.1.2}$$

In fact, $T^2_{(2)}$ is just a constant times D^2; that is,

$$T^2_{(2)} = cD^2, \quad c = n_1 n_2/(n_1 + n_2).$$

The population Mahalanobis distance is defined by

$$\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Theorem 3.1.2 *Suppose that D^2 is the sample Mahalanobis distance between two normal populations* $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, $i = 1, 2$. *Then:*

(1) $\mathrm{E}(D^2) = \dfrac{m}{m-p-1}\left(\Delta^2 + \dfrac{p}{c}\right).$

(2) $\mathrm{Var}(D^2) = \dfrac{2m^2}{(m-p-1)^2(m-p-3)}\left[\dfrac{p(m-1)}{c^2} + \dfrac{2(m-1)}{c}\Delta^2 + \Delta^4\right],$

where $m = n_1 + n_2 - 2$.

Proof. The results are obtained by direct computation based on

$$\begin{aligned} D^2 = \frac{1}{c}\mathrm{T}^2 &= \frac{mp}{c(m-p-1)}F(p, m-p+1; c\Delta^2) \\ &= \frac{p}{c} \cdot \frac{\chi^2(p; c\Delta^2)}{\chi^2(m-p+1)}, \end{aligned}$$

which follows from Theorem 3.1.1. □

3.1.2 Decomposition of $\mathbf{T}^2$ and $\mathbf{D}^2$

Let $\boldsymbol{Z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, $m\mathbf{S} \sim \mathrm{W}_p(m, \Sigma)$, and $\boldsymbol{Z}$ and $\mathbf{S}$ be independent. Let $\boldsymbol{\mu}$, Σ, $\boldsymbol{Z}$, and $\mathbf{S}$ be

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix},$$

where $\boldsymbol{\mu}_i$: $p_i \times 1$, Σ_{ij}: $p_i \times p_j$, $\boldsymbol{Z}_i$: $p_i \times 1$, and $\mathbf{S}_{ij}$: $p_i \times p_j$. We also use the notations:

$$\begin{aligned} &\mathcal{B} = \Sigma_{21}\Sigma_{11}^{-1}, \quad \mathbf{B} = \mathbf{S}_{21}\mathbf{S}_{11}^{-1}, \\ &\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{22} - \mathcal{B}\Sigma_{11}\mathcal{B}', \\ &\mathbf{S}_{22\cdot1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} = \mathbf{S}_{22} - \mathbf{B}\mathbf{S}_{11}\mathbf{B}'. \end{aligned}$$

Now we consider the following decompositions of

$$\Delta^2 = \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} \quad \text{and} \quad \mathrm{T}^2 = \boldsymbol{Z}'\mathbf{S}^{-1}\boldsymbol{Z}$$

Theorem 3.1.3 *Let* $\Delta^2 = \boldsymbol{\mu}'\Sigma\boldsymbol{\mu}$, $\Delta_1 = \boldsymbol{\mu}_1'\Sigma_{11}^{-1}\boldsymbol{\mu}_1$, $T^2 = \boldsymbol{Z}'\mathbf{S}^{-1}\boldsymbol{Z}$ *and* $T_1^2 = \boldsymbol{Z}_1'\mathbf{S}_{11}^{-1}\boldsymbol{Z}_1$. *Then*

$$\Delta^2 = \Delta_1^2 + \Delta_{2\cdot 1}^2, \quad T^2 = T_1^2 + T_{2\cdot 1}^2,$$

where

$$\Delta_{2\cdot 1}^2 = \boldsymbol{\mu}_{2\cdot 1}'\Sigma_{22\cdot 1}^{-1}\boldsymbol{\mu}_{2\cdot 1}, \quad \boldsymbol{\mu}_{2\cdot 1} = \boldsymbol{\mu}_2 - \mathcal{B}\boldsymbol{\mu}_1,$$
$$T_{2\cdot 1}^2 = \boldsymbol{Z}_{2\cdot 1}'\mathbf{S}_{22\cdot 1}^{-1}\boldsymbol{Z}_{2\cdot 1}, \quad \boldsymbol{Z}_{2\cdot 1} = \boldsymbol{Z}_2 - \mathbf{B}\boldsymbol{Z}_1.$$

Proof. The results are obtained by applying [A.1.2] to Σ^{-1} and $\mathbf{S}^{-1}$. □

The quantity

$$U = \frac{T^2 - \mathrm{T}_1^2}{n + T_1^2} = \frac{T_{2\cdot 1}^2}{n + T_1^2} \tag{3.1.3}$$

is called *Rao's U-statistic.* We now consider the sampling distribution of $U = T_{2\cdot 1}^2/(n + T_1^2)$, where $T_{2\cdot 1}^2$ and T_1^2 may be defined in terms of $\mathbf{V} = n\mathbf{S} \sim \mathrm{W}_p(n, \Sigma)$ as

$$T_{2\cdot 1}^2 = n\boldsymbol{Z}_{2\cdot 1}'\mathbf{V}_{22\cdot 1}^{-1}\boldsymbol{Z}_{2\cdot 1}, \quad T_1^2 = n\boldsymbol{Z}_1'\mathbf{V}_{11}\boldsymbol{Z}_1,$$

where $\boldsymbol{Z}_{2\cdot 1} = \boldsymbol{Z}_2 - \mathbf{B}\boldsymbol{Z}_1$ and $\mathbf{V}$ is partitioned similarly to $\mathbf{S}$. First, we note that $\mathbf{V}_{22\cdot 1} \sim \mathrm{W}_{p_2}(n - p_1, \Sigma_{22\cdot 1})$; $\mathbf{V}_{22\cdot 1}$ and $(\mathbf{B}, \mathbf{V}_{11})$ are independent. Let $\mathbf{V}$ denote, $\mathbf{V} = \mathbf{X}'\mathbf{X} = (\mathbf{X}_1\ \mathbf{X}_2)'(\mathbf{X}_1\ \mathbf{X}_2)$, where $\mathbf{X}_1 : n \times p_1$ and the rows of $\mathbf{X}$ are independently distributed as $\mathrm{N}_p(\mathbf{0}, \Sigma)$. The conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1$ is normal such that the rows are independent with the same covariance matrix $\Sigma_{22\cdot 1}$ and $\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1) = \mathbf{X}_1\mathcal{B}$. Note that the conditional distribution of $\mathbf{B}(\mathbf{X}_1'\mathbf{X}_1)^{-1/2} = \mathbf{X}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1/2}$ is normal such that its rows are independent with the same covariance matrix $\Sigma_{22\cdot 1}$ and $\mathrm{E}(\mathbf{B}(\mathbf{X}_1'\mathbf{X}_1)^{-1/2}|\mathbf{X}_1) = \mathcal{B}(\mathbf{X}_1'\mathbf{X}_1)^{1/2}$. Therefore, we check that the conditional distribution $\boldsymbol{Z}_{2\cdot 1} = \boldsymbol{Z}_2 - \mathbf{B}(\mathbf{X}_1'\mathbf{X}_1)^{1/2}(\mathbf{X}_1'\mathbf{X}_1)^{-1/2}\boldsymbol{Z}_1$ given that $(\boldsymbol{Z}_1, \mathbf{X}_1)$ is normal with mean

$$\mathrm{E}(\boldsymbol{Z}_{2\cdot 1}|\boldsymbol{Z}_1, \mathbf{X}_1) = \boldsymbol{\mu}_2 + \mathcal{B}(\boldsymbol{Z}_1 - \boldsymbol{\mu}_1) - \mathcal{B}\boldsymbol{Z}_1 = \boldsymbol{\mu}_{2\cdot 1}$$

and covariance matrix

$$\begin{aligned}\mathrm{Var}(\boldsymbol{Z}_{2\cdot 1}|\boldsymbol{Z}_1, \mathbf{X}_1) &= \Sigma_{22\cdot 1} + \boldsymbol{Z}_1'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\boldsymbol{Z}_1 \cdot \Sigma_{22\cdot 1} \\ &= \left(1 + \frac{1}{n}T_1^2\right)\Sigma_{22\cdot 1}.\end{aligned}$$

These imply the following result.

Theorem 3.1.4 *Let* $\boldsymbol{Z} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$, *and* $\boldsymbol{Z}$ *and* $\mathbf{W}$ *be independently distributed. Using the same notation as in Theorem 3.1.2, the conditional distribution of*

$$\frac{n-p+1}{p_2}U = \frac{n-p+1}{p_2} \cdot \frac{T_{2\cdot 1}^2}{n+T_1^2}$$

given that T_1^2 *is the noncentral F-distribution with* $(p_2, n-p+1)$ *degrees of freedom and noncentrality parameter*

$$\lambda^2 = \frac{n}{n+T_1^2}\Delta_{2\cdot 1}^2.$$

In particular, if $\Delta_{2\cdot 1}^2 = 0$ *(i.e.,* $\Delta^2 = \Delta_1^2$*), then* $\{(n-p+1)/p_2\}U$ *is distributed independent of* T_1^2 *as the central F-distribution with* $(p_2, n-p+1)$ *degrees of freedom.*

Theorem 3.1.4 can be applied to Hotelling's T^2-statistic $T_{(2)}^2$ in a two-sample case. Now we write $\boldsymbol{\mu}^{(i)}$ and $\bar{\boldsymbol{X}}^{(i)}$ instead of $\boldsymbol{\mu}_i$ and $\bar{\boldsymbol{X}}_i$. Then the Hotelling statistic is

$$T^2 = c(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)})'\mathbf{S}^{-1}(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)}) = cD^2.$$

Here $c = \sqrt{n_1 n_2/n}$, $m\mathbf{S} \sim \mathrm{W}_p(m, \Sigma)$, and $m = n-2$. In the test of $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$, we are interesting in deciding whether $\boldsymbol{X}_2$ has additional information in the presence of $\boldsymbol{X}_1$. This is can be examined by testing the hypothesis

$$\boldsymbol{\mu}_{2\cdot 1}^{(1)} = \boldsymbol{\mu}_{2\cdot 1}^{(2)},$$

where $\boldsymbol{\mu}_{2\cdot 1}^{(i)} = \boldsymbol{\mu}_2^{(i)} - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1^{(i)}$, $i = 1, 2$. Then a reasonable test statistic is

$$U = c\frac{D^2 - D_1^2}{n + cD_1^2}, \tag{3.1.4}$$

which is shown to be a likelihood ratio test. The null distribution of $\{(m-p+1)/p_2\}U$ is an F-distribution with $(p_2, m-p+1)$ degrees of freedom.

Example 3.1.1 We consider the data used by Fisher (1936), which are cited in Mardia et al. (1979). Measurements were made on 50 specimens of flowers from each of two species of iris, *setosa* and *versicolors*, found growing together in the same colony. We denote these two populations by Π_1 and Π_2. Four flower measurements were taken: X_1, sepal length; X_2, sepal width; X_3, petal length; and X_4, petal width. Let the samples from Π_1 and Π_2 be denoted by $\boldsymbol{x}_j^{(1)}$, $j = 1, \ldots, n_1$ and $\boldsymbol{x}_j^{(2)}$, $j = 1, \ldots, n_2$, where $n_1 = n_2 = 50$. The sample mean vectors and covariance matrices are defined by

$$\bar{\boldsymbol{x}}^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{x}_j^{(i)}, \quad \mathrm{S}^{(i)} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\boldsymbol{x}_j^{(i)} - \bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_j^{(i)} - \bar{\boldsymbol{x}}^{(i)})',$$

and the pooled sample covariance matrix is defined by

$$\mathrm{S} = \frac{1}{n-2}\{(n_1-1)\mathrm{S}^{(1)} + (n_2-1)\mathrm{S}^{(2)}\}, \quad n = n_1 + n_2.$$

The sample mean vectors $\bar{\boldsymbol{x}}^{(i)}$ observed are

	X_1	X_2	X_3	X_4
Π_1	5.006	3.428	1.462	0.246
Π_2	5.936	2.770	4.260	1.326

The sample covariance matrices (upper triangular elements are correlation coefficients) are found as

Π_1	X_1	X_2	X_3	X_4
X_1	0.124	0.743	0.267	0.278
X_2	0.099	0.144	0.178	0.233
X_3	0.016	0.012	0.030	0.332
X_4	0.010	0.009	0.006	0.011

and

Π_2	X_1	X_2	X_3	X_4
X_1	0.266	0.526	0.754	0.546
X_2	0.085	0.098	0.561	0.664
X_3	0.183	0.083	0.221	0.787
X_4	0.056	0.041	0.073	0.039

The pooled sample covariance matrix is given by

Π	X_1	X_2	X_3	X_4
X_1	0.266	0.526	0.754	0.546
X_2	0.085	0.098	0.561	0.664
X_3	0.183	0.083	0.221	0.787
X_4	0.056	0.041	0.073	0.039

We assume that the populations Π_1 and Π_2 are $\mathrm{N}_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathrm{N}_p(\boldsymbol{\mu}_2, \Sigma)$, respectively. For testing $H : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, we have

$$\begin{aligned}\mathrm{T}^2 &= \frac{n_1 n_2}{n}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})'\mathrm{S}^{-1}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}) \\ &= \frac{n_1 n_2}{n} D^2 = 2580.84\end{aligned}$$

and

$$F = \frac{1}{m}T^2 \cdot \frac{m-p+1}{p}$$
$$= \frac{1}{98} \times 2580.84 \times \frac{98-4+1}{4} = 625.46.$$

The p-value is 0.000, and we reject the hypothesis. Next we consider whether X_1 has additional information in the presence of $\omega = \{X_2, X_3, X_4\}$. Let the Mahalanobis distances be based on $\Omega = \{X_1, \dots, X_4\}$ and ω by D_Ω and D_ω, respectively. Then

$$D_\Omega = 10.16 \quad \text{and} \quad D_\omega = 10.12.$$

The test statistic is given by

$$T_{(\Omega-\omega)\cdot\omega} = \frac{m-p+1}{p-k_\omega}\frac{c(D_\Omega^2 - D_\omega^2)}{n-2+cD_\omega^2}$$

which is distributed as an F-distribution with $p - k_\omega$ and $m - p + 1$, where $c = \sqrt{n_1 n_2/n}$ and k_ω is the number of variables in ω. The value of $T_{(\Omega-\omega)\cdot\omega}$ is 0.72, the p-value is 0.328, and hence X_1 has no additional information in the presence of $\{X_2, X_3, X_4\}$.

3.2 Lambda Statistic

3.2.1 Motivation of the Lambda Statistic

Let n_i samples

$$\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i}, \quad i = 1, \dots, k \tag{3.2.1}$$

be available from k populations with mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ, and denote the sample mean vectors and the sample covariance matrices by

$$\bar{\boldsymbol{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{X}_{ij} \quad \text{and} \quad \mathbf{S}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)'.$$

Further, let $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ be the overall mean vector and the pooled sample covariance matrix, defined by

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{j=1}^{k} n_i\bar{\boldsymbol{X}}_i \quad (n = n_1 + \cdots + n_k) \quad \text{and} \quad \mathbf{S} = \frac{1}{n-k}\sum_{j=1}^{n_i}(n_i - 1)\mathbf{S}_i,$$

respectively. For testing

$$H_0\colon \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k, \tag{3.2.2}$$

we have two basic matrices given by

$$\mathbf{S}_b = \sum_{i=1}^{k} n_i(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})', \quad \mathbf{S}_w = \sum_{i=1}^{k}(n_i - 1)\mathbf{S}_i. \tag{3.2.3}$$

The matrices $\mathbf{S}_b$ and $\mathbf{S}_w$ are called *between-group and within-group matrices of sums of squares and products*, respectively. Note that

$$\begin{aligned}\mathbf{S}_t &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}})(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}})' \\ &= \mathbf{S}_b + \mathbf{S}_w.\end{aligned}$$

Then a typical test is based on a monotone function of $|\mathbf{S}_w|/|\mathbf{S}_w + \mathbf{S}_b|$ which is called the *lambda statistic.* If, in particular, the populations are all normal, $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$, it reduces to the likelihood ratio test.

Theorem 3.2.1 *Let the samples* (3.2.1) *be from a p-variate normal population* $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma), i = 1, \ldots, k$. *Then the likelihood ratio criterion* λ *for testing the hypothesis* (3.2.2) *can be expressed as*

$$-2\log\lambda = -n\log\frac{|\mathbf{S}_w|}{|\mathbf{S}_w + \mathbf{S}_b|}.$$

Proof. Let $g(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \Sigma)$ be -2 log likelihood for samples (3.2.1). Then

$$\begin{aligned}g(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \Sigma) &= n\log|\Sigma| + np\log(2\pi) \\ &\quad + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathrm{tr}\,\Sigma^{-1}(\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i)(\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i)'.\end{aligned}$$

Note that

$$\begin{aligned}&\mathrm{tr}\,\Sigma^{-1}(\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i)(\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i)' \\ &\quad = \mathrm{tr}\,\Sigma^{-1}\mathrm{S}_w + \sum_{i=1}^{k} n_i\ \mathrm{tr}\,\Sigma^{-1}(\bar{\boldsymbol{x}}_i - \boldsymbol{\mu}_i)(\bar{\boldsymbol{x}}_i - \boldsymbol{\mu}_i)',\end{aligned}$$

which has the minimum at $\boldsymbol{\mu}_i = \hat{\boldsymbol{\mu}}_i = \bar{\boldsymbol{x}}_i, i = 1, \ldots, k$. Therefore,

$$\begin{aligned}\min_{\boldsymbol{\mu}_i, \Sigma} g(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \Sigma) &= \min_{\Sigma} g(\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_k, \Sigma) \\ &= \min_{\Sigma}\{n\log|\Sigma| + \mathrm{tr}\,\Sigma^{-1}\mathrm{S}_w\} + np\log(2\pi).\end{aligned}$$

The minimum for Σ of the last expression is attained at $\Sigma = \hat{\Sigma} = (1/n)\mathrm{S}_w$ by using [A.2.10], and hence

$$\min_{\boldsymbol{\mu}_i, \Sigma} g(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \Sigma) = n\log\left|\frac{1}{n}\mathrm{S}_w\right| + np(\log(2\pi) + 1).$$

The $-2\log$ likelihood under H_0 is expressed as $g_0(\boldsymbol{\mu}, \Sigma) = g(\boldsymbol{\mu}, \ldots, \boldsymbol{\mu}, \Sigma)$. Similarly, its minimum is attained at

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}, \quad \Sigma = \hat{\Sigma} = \frac{1}{n} S_t = \frac{1}{n}(\mathrm{S}_w + \mathrm{S}_b)$$

and

$$\min_{\boldsymbol{\mu}_i, \Sigma} g_0(\boldsymbol{\mu}, \Sigma) = n \log \left| \frac{1}{n}(\mathrm{S}_w + \mathrm{S}_b) \right| + np(\log(2\pi) + 1).$$

Therefore,

$$\begin{aligned} -2\log\lambda &= \min g_0(\boldsymbol{\mu}, \Sigma) - \min g(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \Sigma) \\ &= -n \log \frac{|\mathrm{S}_w|}{|\mathrm{S}_w + \mathrm{S}_b|}. \end{aligned}$$

□

3.2.2 Distribution of the Lambda Statistic

Theorem 3.2.2 *Let* $\mathbf{S}_b$ *and* $\mathbf{S}_w$ *be the between group and within group matrices of sums of squares and products given by* (3.2.3), *based on* k p*-variate normal populations* $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$. *Then* $\mathbf{S}_b$ *and* $\mathbf{S}_w$ *are independently distributed as* $\mathrm{W}_p(k-1, \Sigma; \Omega)$ *and* $\mathrm{W}_p(n-k, \Sigma)$, *respectively, where* $\Omega = \sum_{i=1}^k n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'$ *and* $\bar{\boldsymbol{\mu}} = (1/n)\sum_{i=1}^k n_i \boldsymbol{\mu}_i$.

Proof. We have seen that $(n_i - 1)\mathbf{S}_i \sim \mathrm{W}_p(n_i - 1, \Sigma)$. Further, the $\mathbf{S}_i$ are independent, and hence $\mathbf{S}_w = \sum_{i=1}^k (n_i - 1)\mathbf{S}_i \sim \mathrm{W}_p(n-k, \Sigma)$. Put $\boldsymbol{X}_i = \sqrt{n_i}\,\bar{\boldsymbol{Y}}_i$, and then

$$\begin{aligned} \mathbf{S}_b &= \sum_{i=1}^k n_i \bar{\boldsymbol{Y}}_i \bar{\boldsymbol{Y}}_i' - n\bar{\boldsymbol{Y}}\bar{\boldsymbol{Y}}' \\ &= \sum_{i=1}^k \boldsymbol{X}_i \boldsymbol{X}_i' - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}', \end{aligned}$$

where $\tilde{\boldsymbol{X}} = \sum_{i=1}^k \sqrt{n_i/n}\,\boldsymbol{X}_i$. Let H be a $k \times k$ orthogonal matrix whose first column is $(\sqrt{n_1/n}, \ldots, \sqrt{n_k/n})'$, and consider the transformation

$$\mathbf{Z} = \begin{pmatrix} \boldsymbol{Z}_1' \\ \vdots \\ \boldsymbol{Z}_k' \end{pmatrix} = \mathrm{H}' \begin{pmatrix} \boldsymbol{X}_1' \\ \vdots \\ \boldsymbol{X}_k' \end{pmatrix} = \mathrm{H}'\mathbf{X}.$$

Then $\sum_{i=1}^k \boldsymbol{Z}_i \boldsymbol{Z}_i' = \sum_{i=1}^k \boldsymbol{X}_i \boldsymbol{X}_i'$ and $\boldsymbol{Z}_1 = \tilde{\boldsymbol{X}}$, and

$$\mathbf{S}_b = \sum_{i=1}^k \boldsymbol{Z}_i \boldsymbol{Z}_i' - \boldsymbol{Z}_1 \boldsymbol{Z}_1' = \sum_{i=2}^k \boldsymbol{Z}_i \boldsymbol{Z}_i'.$$

By Theorem 1.2.6, the $\boldsymbol{Z}_i$'s are independently and normally distributed with the same covariance matrix Σ. Therefore, $\mathbf{S}_b \sim \mathrm{W}_p(k-1, \Sigma; \Psi)$. The non-centrality matrix Ψ is

$$\begin{aligned}\Psi &= \sum_{i=2}^{k} \mathrm{E}(\boldsymbol{Z}_i)\mathrm{E}(\boldsymbol{Z}_i)' \\ &= \mathrm{E}(\mathbf{Z})'\mathrm{E}(\mathbf{Z}) - \mathrm{E}(\boldsymbol{Z}_1)\mathrm{E}(\boldsymbol{Z}_1)' \\ &= \mathrm{E}(\mathbf{X})'\mathrm{E}(\mathbf{X}) - n\bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}' \\ &= \sum_{i=1}^{k} n_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i' - n\bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}' = \Omega. \qquad \square\end{aligned}$$

The null distribution of $|\mathbf{S}_w|/|\mathbf{S}_w + \mathbf{S}_b|$ is called lambda distribution, which is formally defined as follows:

Definition 3.2.1 Let $\mathbf{B}$ and $\mathbf{W}$ be independent random matrices following the Wishart distribution $\mathrm{W}_p(q, \Sigma)$ and $\mathrm{W}_p(n, \Sigma)$, respectively, with $n \geq p$. Then the distribution of

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}$$

is said to be the p-dimensional *Lambda distribution* with (q, n)-degrees of freedom, and is denoted by $\Lambda_p(q, n)$. In what follows, we write $\Lambda_{p;q,n}$ to denote a random variable with this distribution:

$$\Lambda_{p;q,n} \sim \Lambda_p(q, n)$$

When $p = 1$,

$$\Lambda_{1;q,n} = \frac{\chi_n^2}{\chi_n^2 + \chi_q^2}.$$

This implies that $\Lambda_1(q, n)$ is distributed as a beta distribution $\mathrm{Be}(n, q)$ and

$$\frac{1 - \Lambda_{1;q,n}}{\Lambda_{1;q,n}} \cdot \frac{n}{q} = \frac{\chi_q^2/q}{\chi_n^2/n} \sim F(q, n).$$

The distribution of $\Lambda_p(1, n)$ is also simplified. In this case we can write

$$\Lambda_{p;1,n} = \frac{|\mathbf{W}|}{|\mathbf{W} + \boldsymbol{X}\boldsymbol{X}'|},$$

where $\mathbf{W}$ and $\boldsymbol{X}$ are independent and distributed as $\mathrm{W}_p(n, \mathrm{I}_p)$ and $\mathrm{N}_p(\mathbf{0}, \mathrm{I}_p)$, respectively. Note

$$|\mathbf{W} + \boldsymbol{X}\boldsymbol{X}'| = |\mathbf{W}|(1 + \boldsymbol{X}'\mathbf{W}^{-1}\boldsymbol{X}),$$

which is obtained by computing

$$\begin{vmatrix} 1 & \boldsymbol{X}' \\ -\boldsymbol{X} & \mathbf{W} \end{vmatrix}$$

in two ways (see [A.1.1] (2)). Using the formula above,

$$\Lambda_{p;1,n} = (1 + \boldsymbol{X}'\mathbf{W}^{-1}\boldsymbol{X})^{-1}$$

and hence

$$\frac{1 - \Lambda_{p;1,n}}{\Lambda_{p:1,n}} = (1 + \boldsymbol{X}'\mathbf{W}^{-1}\boldsymbol{X})^{-1}.$$

The distribution of $\boldsymbol{X}'\mathbf{W}^{-1}\boldsymbol{X}$ is given in Theorem 3.1.1. Summarizing these results, we have the following theorem.

Theorem 3.2.3 *For the Lambda distribution, it holds that:*

(1) $\dfrac{1 - \Lambda_{1;q,n}}{\Lambda_{1;q,n}} \cdot \dfrac{n}{q} \sim F(q, n).$

(2) $\dfrac{1 - \Lambda_{p;1,n}}{\Lambda_{p;1,n}} \cdot \dfrac{n + 1 - p}{p} \sim F(p, n + 1 - p).$

The distribution of $\Lambda_p(q, n) = |\mathbf{W}|/|\mathbf{W}+\mathbf{B}|$ is defined by two independent $p \times p$ matrices $\mathbf{B}$ and $\mathbf{W}$, where $\mathbf{B} \sim \mathrm{W}_p(q, \mathrm{I}_p)$ and $\mathbf{W} \sim \mathrm{W}_p(n, \mathrm{I}_p)$. Next we shall see that it may be defined based on two $q \times q$ Wishart matrices. The result was given by Wakaki et al. (2003).

Theorem 3.2.4 *Let* $\mathbf{B} = \mathbf{X}'\mathbf{X}$ *and* $\mathbf{W}$ *be independently distributed as* $\mathrm{W}_p(q, \mathrm{I}_p; \mathrm{M}'\mathrm{M})$ *and* $\mathrm{W}_p(n, \mathrm{I}_p)$*, respectively. Here* X *is a* $q \times p$ *random matrix whose elements are independent normal variates with* $\mathrm{E}(\mathbf{X}) = \mathrm{M}$ *and the common variance* 1*. Put*

$$\mathbf{U} = \mathbf{X}\mathbf{X}' \quad \textit{and} \quad \mathbf{V} = \mathbf{U}^{1/2}(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}')^{-1}\mathbf{U}^{1/2}.$$

Then:

(1) $\mathbf{U}$ *and* $\mathbf{V}$ *are independently distributed as* $\mathrm{W}_q(p, \mathrm{I}_q; \mathrm{M}'\mathrm{M})$ *and* $\mathrm{W}_q(n - p + q, \mathrm{I}_q)$*, respectively.*

(2) $\dfrac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \dfrac{|\mathbf{V}|}{|\mathbf{V} + \mathbf{U}|}.$

Proof. By the definition of Wishart distribution, it follows that $\mathbf{U} \sim \mathrm{W}_q(p, \mathrm{I}_q; \mathrm{M}'\mathrm{M})$. On the other hand, Theorem 2.3.3 asserts that the conditional distribution of $(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}$ given $\mathbf{X}$ is $\mathrm{W}_q(n - p + q, (\mathbf{X}'\mathbf{X})^{-1})$.

Therefore, $\mathbf{V} \sim \mathrm{W}_q(n-p+q, \mathrm{I}_q)$ and hence $\mathbf{V}$ is independent of $\mathbf{U}$. For the second result, note that

$$\frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|} = \{(1+\ell_1)\cdots(1+\ell_p)\}^{-1},$$

where $\ell_1 \geq \cdots \geq \ell_1 \geq 0$. The nonzero characteristic roots of $\mathbf{UV}^{-1}$ are the same as those of $\mathbf{XW}^{-1}\mathbf{X}'$ and hence $\mathbf{BW}^{-1}$. This completes the proof. □

Theorem 3.2.5 *For the Lambda distribution* $\Lambda_p(q, n)$,

$$\Lambda_p(q, n) = \Lambda_q(p, n-p+q)$$

holds.

Proof. This is a direct consequence of Theorem 3.2.4. □

3.3 Test for Additional Information

In the test of $H_0\colon \boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(k)}$ in (3.2.2), we are interested in whether a subset of variables $X_1, \ldots, X_p$ is sufficient, or is the set of remainder variables has no additional information or redundant. The distributions of tests based on $\mathbf{W}(\mathbf{W}+\mathbf{B})^{-1}$ depend on the parameters $\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(k)}$ and Σ through $\Sigma^{-1}\Omega$, where Ω is the noncentrality matrix of $\mathbf{B}$, defined by

$$\Omega = \sum_{j=1}^{k} n_j(\boldsymbol{\mu}^{(j)} - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}^{(j)} - \bar{\boldsymbol{\mu}})' = \Xi\Xi',$$

where $\bar{\boldsymbol{\mu}} = (1/n)\sum_{j=1}^{k} n_j \boldsymbol{\mu}^{(j)}$ and $\Xi = (n_1(\boldsymbol{\mu}^{(1)} - \bar{\boldsymbol{\mu}}), \ldots, n_k(\boldsymbol{\mu}^{(k)} - \bar{\boldsymbol{\mu}}))$. Note that Ω denotes departures from the null hypothesis H_0. A single measure for the overall departures is

$$\delta^2 = \operatorname{tr}\Sigma^{-1}\Omega, \tag{3.3.1}$$

which may be considered as an information quantity on departures from the null hypothesis H_0. Without loss of generality we consider the sufficiency of a subvector $\boldsymbol{X}_1 = (X_1, \ldots, X_q)'$ of $\boldsymbol{X}$, or redundancy of the remainder vector $\boldsymbol{X}_2 = (X_{q+1}, \ldots, X_p)'$. Let $\boldsymbol{\mu}^{(j)}$, $\bar{\boldsymbol{\mu}}$, and Ω partition in the same manner as the partition of $\boldsymbol{X}$:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu}^{(j)} = \begin{pmatrix} \boldsymbol{\mu}_1^{(j)} \\ \boldsymbol{\mu}_2^{(j)} \end{pmatrix}, \quad \bar{\boldsymbol{\mu}} = \begin{pmatrix} \bar{\boldsymbol{\mu}}_1 \\ \bar{\boldsymbol{\mu}}_2 \end{pmatrix},$$
$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

Then the information quantity for $\boldsymbol{X}_1$ on departures from the null hypothesis is expressed as

$$\delta_1^2 = \mathrm{tr}\,\Sigma_{11}^{-1}\Omega_{11}.$$

Then we say that $\boldsymbol{X}_1$ is sufficient or $\boldsymbol{X}_2$ is redundant if $\delta^2 = \delta_1^2$; it is called a sufficiency of $\boldsymbol{X}_1$. Using [A.1.2], we can write

$$\delta^2 = \delta_1^2 + \mathrm{tr}\,\Sigma_{22\cdot 1}^{-1}\sum_{j=1}^{k} n_j \tilde{\boldsymbol{\mu}}_{2\cdot 1}^{(j)}\tilde{\boldsymbol{\mu}}_{2\cdot 1}^{(j)\prime},$$

where $\tilde{\boldsymbol{\mu}}_{2\cdot 1}^{(j)} = \boldsymbol{\mu}_2^{(j)} - \bar{\boldsymbol{\mu}}_2 - \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{\mu}_1^{(j)} - \bar{\boldsymbol{\mu}}_1)$. Therefore, the sufficiency condition of $\boldsymbol{X}_1$ is expressed as $\tilde{\boldsymbol{\mu}}_{2\cdot 1}^{(j)} = \mathbf{0}$, $j = 1, \ldots, q$, which is equivalent to

$$H_{2\cdot 1} : \boldsymbol{\mu}_{2\cdot 1}^{(1)} = \cdots = \boldsymbol{\mu}_{2\cdot 1}^{(q)}, \tag{3.3.2}$$

where $\boldsymbol{\mu}_{2\cdot 1}^{(j)} = \boldsymbol{\mu}_2^{(j)} - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1^{(j)}$, $j = 1, \ldots, q$. The hypothesis $H_{2\cdot 1}$ was introduced by Rao (1948, 1970). To obtain a likelihood ratio for $H_{2\cdot 1}$, we partition the observation matrix as

$$\begin{aligned}\mathbf{X} &= (\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}, \ldots, \boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{kn_k})' \\ &= \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}, \quad \mathbf{X}_1 : n \times k.\end{aligned}$$

Then the conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1$ is normal such that the rows of $\mathbf{X}_2$ are independently distributed with covariance matrix $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, and the conditional mean is given by

$$\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1) = \mathrm{A}\Theta + \mathbf{X}_1\Sigma_{11}^{-1}\Sigma_{12}, \tag{3.3.3}$$

where

$$\mathrm{A} = \begin{pmatrix} \mathbf{1}_{n_1} & & \mathrm{O} \\ & \ddots & \\ \mathrm{O} & & \mathbf{1}_{n_k} \end{pmatrix}, \quad \Theta' = (\boldsymbol{\mu}_{2\cdot 1}^{(1)}, \ldots, \boldsymbol{\mu}_{2\cdot 1}^{(q)}).$$

We use the notation $\mathbf{W}$ and $\mathbf{T}$ for $\mathbf{S}_w$ and $\mathbf{S}_t$, respectively, and partition $\mathbf{W}$ and $\mathbf{T}$ as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix}, \tag{3.3.4}$$

where $\mathbf{W}_{12}$: $q \times (p-q)$ and $\mathbf{T}_{12}$: $q \times (p-q)$. Then, using [A.2.11], we have that the maximum likelihood estimations of $\Sigma_{22\cdot 1}$ and Σ_{11} under $H_{2\cdot 1}$ are

$$\begin{aligned}n\hat{\Sigma}_{22\cdot 1} &= \mathbf{X}_2'\left\{\mathrm{I}_n - \mathrm{P}_0 - \mathrm{P}_{(\mathrm{I}_n - \mathrm{P}_0)\mathbf{X}_1}\right\}\mathbf{X}_2 \\ &= \mathbf{T}_{22} - \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{T}_{12} = \mathbf{T}_{22\cdot 1}, \\ n\hat{\Sigma}_{11} &= \mathbf{X}_1'(\mathrm{I}_n - \mathrm{P}_{\mathrm{A}})\mathbf{X}_1 \\ &= \mathbf{W}_{11},\end{aligned}$$

where $\mathrm{P}_0 = (1/n)\mathbf{1}_n\mathbf{1}_n'$ and for an $n \times m$ matrix B with $\mathrm{rank}(B) = m$, $\mathrm{P_B} = \mathrm{B}(\mathrm{B'B})^{-1}\mathrm{B'}$. Therefore,

$$-2\log\max_{H_{2\cdot 1}} f(\mathrm{X};\boldsymbol{\mu}^{(1)},\ldots,\boldsymbol{\mu}^{(k)},\Sigma) = n\log\left|\frac{1}{n}\mathrm{T}_{22\cdot 1}\right| \\ +n\log\left|\frac{1}{n}\mathrm{W}_{11}\right| + np\{\log 2\pi + 1\}.$$

On the other hand, under no restriction

$$-2\log\max f(\mathrm{X};\boldsymbol{\mu}^{(1)},\ldots,\boldsymbol{\mu}^{(k)},\Sigma) = n\log\left|\frac{1}{n}\mathrm{W}\right| + np\{\log 2\pi + 1\}.$$

These imply that the likelihood ratio test criterion for $H_{2\cdot 1}$ is given by

$$\lambda = \left(\frac{|\mathbf{W}|}{|\mathbf{W}_{11}|\cdot|\mathbf{T}_{22\cdot 1}|}\right)^{n/2} = \left(\frac{|\mathbf{W}_{22\cdot 1}|}{|\mathbf{T}_{22\cdot 1}|}\right)^{n/2}.$$

Theorem 3.3.1 *Suppose that* (3.2.1) *is a set of samples from* $\mathrm{N}_p(\boldsymbol{\mu}^{(j)},\Sigma)$, $j = 1,\ldots,k$. *Then the likelihood ratio criterion* λ *for the hypothesis* $H_{2\cdot 1}$ *in* (3.3.2) *is given by*

$$\lambda = \left(\frac{|\mathbf{W}_{22\cdot 1}|}{|\mathbf{T}_{22\cdot 1}|}\right)^{n/2},$$

where $\mathbf{W} = \mathbf{S}_w$ *and* $\mathbf{T} = \mathbf{S}_t$ *are given by* (3.3.4). *Further, under* $H_{2\cdot 1}$,

$$\frac{|\mathbf{W}_{22\cdot 1}|}{|\mathbf{T}_{22\cdot 1}|} \sim \Lambda_{p-k}(q, n-q-k).$$

Proof. We consider the conditional distribution of $\mathbf{W}_{22\cdot 1}$ and $\mathbf{T}_{22\cdot 1}$ given $\mathbf{X}_1$. The matrix $\mathbf{W}$ is expressed as $\mathbf{X}'(\mathrm{I}_n - \mathrm{P_A})\mathbf{X}$. Note that, in general, for a matrix A, $\mathrm{P_A} = \mathrm{A}(\mathrm{A'A})^{-1}\mathrm{A'}$ denotes the projection matrix on the space spanned by the column vectors of A. Hence

$$\mathbf{W}_{22\cdot 1} = \mathbf{X}_2'\mathbf{Q}_1\mathbf{X}_2, \quad \mathbf{Q}_1 = \mathrm{I}_n - \mathrm{P_A} - \mathrm{P}_{(\mathrm{I}_n - \mathrm{P_A})\mathbf{X}_1}.$$

It is easy to see that $\mathbf{Q}_1^2 = \mathbf{Q}_1$, $\mathrm{rank}(\mathbf{Q}_1) = \mathrm{tr}\,\mathbf{Q}_1 = n - q - k$, $\mathbf{Q}_1\mathrm{A} = \mathrm{O}$, $\mathbf{Q}_1\mathbf{X}_1 = \mathrm{O}$, and

$$\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1)'\mathbf{Q}_1\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1) = \mathrm{O}.$$

This implies that $\mathbf{W}_{22\cdot 1}|\mathbf{X}_1 \sim \mathrm{W}_{p-k}(n-q-k, \Sigma_{22\cdot 1})$ and hence $\mathbf{W}_{22\cdot 1} \sim \mathrm{W}_{p-k}(n-q-k, \Sigma_{22\cdot 1})$. For $\mathbf{T}_{22\cdot 1}$, we have

$$\mathbf{T}_{22\cdot 1} = \mathbf{X}_2'\mathbf{Q}_2\mathbf{X}_2, \quad \mathbf{Q}_2 = \mathrm{I}_n - \mathrm{P}_0 - \mathrm{P}_{(\mathrm{I}_n - \mathrm{P}_0)\mathbf{X}_1},$$

and hence

$$\mathbf{S}_h = \mathbf{T}_{22\cdot 1} - \mathbf{W}_{22\cdot 1} = \mathbf{X}_2'(\mathbf{Q}_2 - \mathbf{Q}_1)\mathbf{X}_2.$$

Similarly $\mathbf{Q}_2$ is idempotent. Using $\mathrm{P_0P_A} = \mathrm{P_AP_0} = \mathrm{P_0}$, we have $\mathbf{Q}_1\mathbf{Q}_2 = \mathbf{Q}_2\mathbf{Q}_1 = \mathbf{Q}_1$, and hence

$$(\mathbf{Q}_2 - \mathbf{Q}_1)^2 = \mathbf{Q}_2 - \mathbf{Q}_1, \qquad \mathbf{Q}_1 \cdot (\mathbf{Q}_2 - \mathbf{Q}_1) = \mathrm{O}.$$

Further, under $H_{2\cdot 1}$,

$$\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1)'(\mathbf{Q}_2 - \mathbf{Q}_1)\mathrm{E}(\mathbf{X}_2|\mathbf{X}_1) = \mathrm{O}.$$

The required result is obtained by using the Cochran Theorem in Theorem 2.2.3. □

3.3.1 Decomposition of the Lambda Statistic

Theorem 3.3.2 *Let* $\mathbf{B}$ *and* $\mathbf{W}$ *be independent,* $\mathbf{B} \sim \mathrm{W}_p(q, \Sigma)$, *and* $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$ *with* $n \geq p$. *Put* $\mathbf{T} = \mathbf{W} + \mathbf{B}$, *and partition* $\mathbf{W}$ *and* $\mathbf{T}$ *as*

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix},$$

where $\mathbf{W}_{ij}$: $p_i \times p_j$ *and* $\mathbf{T}_{ij}$: $p_i \times p_j$. *Then* $\Lambda = |\mathbf{W}|/|\mathbf{T}|$ *can be decomposed as*

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}_{11}|}{|\mathbf{T}_{11}|} \cdot \frac{|\mathbf{W}_{22\cdot 1}|}{|\mathbf{T}_{22\cdot 1}|} = \Lambda^{(1)} \cdot \Lambda^{(2)}.$$

Further, it holds that:

(1) $\Lambda^{(1)} \sim \Lambda_{p_1}(q, n), \quad \Lambda^{(2)} \sim \Lambda_{p_2}(q, n - p_1)$.

(2) $\Lambda^{(1)}$ *and* $\Lambda^{(2)}$ *are independent.*

Proof. From Theorem 2.2.2 it follows that

$$\begin{aligned} &\mathbf{W}_{11} \sim \mathrm{W}_{p_1}(n, \Sigma_{11}), \quad \mathbf{W}_{22\cdot 1} \sim \mathrm{W}_{p_2}(n - p_1, \Sigma_{22\cdot 1}), \\ &\mathbf{T}_{11} \sim \mathrm{W}_{p_2}(n, \Sigma_{11}), \quad \mathbf{T}_{22\cdot 1} \sim \mathrm{W}_{p_2}(n + q - p_1, \Sigma_{22\cdot 1}). \end{aligned}$$

$\mathbf{W}_{22\cdot 1}$ is independent of $\mathbf{W}_{11}$, and $\mathbf{T}_{22\cdot 1}$ is independent of $\mathbf{T}_{11}$. Let $\mathbf{B}$ be partitioned in the same way as in $\mathbf{W}$ or $\mathbf{T}$. Then

$$\Lambda^{(1)} = \frac{|\mathbf{W}_{11}|}{|\mathbf{T}_{11}|} = \frac{|\mathbf{W}_{11}|}{|\mathbf{W}_{11} + \mathbf{B}_{11}|} \sim \Lambda_{p_1}(q, n),$$

and $\Lambda^{(2)}$ is independent of $\Lambda^{(1)}$. In the following we show that $\mathbf{S}_h = \mathbf{T}_{22\cdot 1} - \mathbf{W}_{22\cdot 1} \sim \mathrm{W}_{p_2}(q, \Sigma_{22\cdot 1})$ and $\mathbf{S}_h$ is independent of $\mathbf{W}_{22\cdot 1}$. We can write $\mathbf{B}$ and $\mathbf{W}$ as $\mathbf{B} = \mathbf{X}'\mathbf{X}$ and $\mathbf{W} = \mathbf{U}'\mathbf{U}$, respectively, where the rows of $\mathbf{X}$: $q \times p$ and

$\mathbf{U}$: $n \times p$ are independently distributed as $\mathrm{N}_p(\mathbf{0}, I_p)$. Let $\mathbf{X}$ and $\mathbf{U}$ partition as $\mathbf{X} = (\mathbf{X}_1\ \mathbf{X}_2)$, $\mathbf{X}_i$: $q \times p_i$ and $\mathbf{U} = (\mathbf{U}_1\ \mathbf{U}_2)$, $\mathbf{U}_i$: $n \times p_i$. Then

$$\mathbf{B} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{U}_1'\mathbf{U}_1 & \mathbf{U}_1'\mathbf{U}_2 \\ \mathbf{U}_2'\mathbf{U}_1 & \mathbf{U}_2'\mathbf{U}_2 \end{pmatrix},$$

and $\mathbf{W}_{22\cdot 1} = \mathbf{U}_2'(\mathrm{I}_n - \mathbf{U}_1(\mathbf{U}_1'\mathbf{U}_1)^{-1}\mathbf{U}_1')\mathbf{U}_2$. Noting that $(\mathrm{I}_q + \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{X}_1')^{-1} = \mathrm{I}_q - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1 + \mathbf{W}_{11})^{-1}\mathbf{X}_1'$ (use [A.1.4]),

$$\begin{aligned} \mathbf{S}_h = {} & (\mathbf{X}_2 - \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{W}_{12})' \\ & \times (\mathrm{I}_q + \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{X}_1')^{-1}(\mathbf{X}_2 - \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{W}_{12}). \end{aligned}$$

Let

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{U} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \\ \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix} = (\mathbf{Z}_1\ \mathbf{Z}_2).$$

We can write $\mathbf{V}_{22\cdot 1} = \mathbf{Z}_2'\mathbf{F}\mathbf{Z}_2$, $\mathbf{S}_h = \mathbf{Z}_2'\mathbf{G}\mathbf{Z}_2$, where

$$\begin{aligned} \mathbf{F} &= \begin{pmatrix} \mathrm{O} & \mathrm{O} \\ \mathrm{O} & \mathrm{I}_n - \mathbf{U}_1(\mathbf{U}_1'\mathbf{U}_1)^{-1}\mathbf{U}_1' \end{pmatrix}, \\ \mathbf{G} &= (\mathrm{I}_q - \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{U}_1')'(\mathrm{I}_q + \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{X}_1')(\mathrm{I}_q - \mathbf{X}_1\mathbf{W}_{11}^{-1}\mathbf{U}_1'). \end{aligned}$$

Given $\mathbf{Z}_1$, the rows of the $\mathbf{Z}_2$ are independently distributed as a p_2-variate normal distribution with the common covariance matrix $\Sigma_{22\cdot 1}$ and means $\mathrm{E}(\mathbf{Z}_2|\mathbf{Z}_1) = \mathbf{Z}_1\Sigma_{11}^{-1}\Sigma_{12}$. Note that $\mathbf{F}^2 = \mathbf{F}$, $\mathbf{G}^2 = \mathbf{G}$ and $\mathbf{FG} = \mathrm{O}$. Therefore, from Theorem 2.2.3, $\mathbf{S}_h \sim \mathrm{W}_{p_2}(q, \Sigma_{11\cdot 2})$ and $\mathbf{S}_h$ is independent of $W_{22\cdot 1}$. This completes the proof. □

In Theorem 3.3.2, let $p_2 = p - 1$, and write

$$\begin{aligned} &\mathbf{W}_{11} \to \mathbf{W}_{(1\cdots p-1)(1\cdots p-1)}, \quad W_{22\cdot 1} \to W_{pp\cdot 1\cdots p-1}, \\ &\mathbf{T}_{11} \to \mathbf{T}_{(1\cdots p-1)(1\cdots p-1)}, \quad T_{22\cdot 1} \to T_{pp\cdot 1\cdots p-1}. \end{aligned}$$

Then

$$\begin{aligned} \Lambda &= \frac{|\mathbf{W}_{(1\cdots p-1)(1\cdots p-1)}|}{|\mathbf{T}_{(1\cdots p-1)(1\cdots p-1)}|} \cdot \frac{W_{pp\cdot 1\cdots p-1}}{T_{pp\cdot 1\cdots p-1}}, \\ \frac{W_{pp\cdot 1\cdots p-1}}{T_{pp\cdot 1\cdots p-1}} &\sim \frac{\chi^2(n-(p-1))}{\chi^2(n-(p-1)) + \chi^2(q)} = \mathrm{Be}\left(\frac{1}{2}\{n-(p-1)\}, \frac{1}{2}q\right). \end{aligned}$$

Applying a similar decomposition to $|\mathbf{W}_{(1\cdots p-1)(1\cdots p-1)}|/|\mathbf{T}_{(1\cdots p-1)(1\cdots p-1)}|$, we obtain the decomposition theorem of Λ:

Theorem 3.3.3 *Suppose that $\Lambda = |\mathbf{W}|/|\mathbf{T}|$ is distributed as a lambda distribution $\Lambda_p(q,n)$, $\mathbf{W} \sim \mathrm{W}_p(n, \Sigma)$, $\mathbf{T} \sim \mathrm{W}_p(n+q, \Sigma)$, and $\mathbf{W}$ and $\mathbf{T}$ are independent. Then the distribution of Λ is the distribution of the product*

$\prod_{i=1}^p V_i$, *where* $V_1, \ldots, V_p$ *are independent and* $V_i \sim \mathrm{Be}(\frac{1}{2}(n-i+1), \frac{1}{2}q)$. *The variables* V_i *may be defined by*

$$V_1 = \frac{W_{11}}{T_{11}}, \quad V_i = \frac{W_{ii\cdot 1\cdots i-1}}{T_{ii\cdot 1\cdots i-1}}, \quad i = 2, \ldots, p,$$

where W_{11} *and* T_{11} *are the* $(1,1)$ *elements of* $\mathbf{W}$ *and* $\mathbf{T}$, *respectively. Let the first* $i \times i$ *matrices of* $\mathbf{W}$ *and* $\mathbf{T}$ *be denoted by*

$$\begin{pmatrix} \mathbf{W}_{11}^{(i)} & \boldsymbol{W}_{12}^{(i)} \\ \boldsymbol{W}_{12}^{(i)\prime} & W_{ii} \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} \mathbf{T}_{11}^{(i)} & \boldsymbol{T}_{12}^{(i)} \\ \boldsymbol{T}_{12}^{(i)\prime} & T_{ii} \end{pmatrix},$$

respectively. Then,

$$\begin{aligned} W_{ii\cdot 1\cdots i-1} &= W_{ii} - \boldsymbol{W}_{12}^{(i)\prime}\{\mathbf{W}_{11}^{(i)}\}^{-1}\boldsymbol{W}_{12}^{(i)}, \\ T_{ii\cdot 1\cdots i-1} &= T_{ii} - \boldsymbol{T}_{12}^{(i)\prime}\{\mathbf{T}_{11}^{(i)}\}^{-1}\boldsymbol{T}_{12}^{(i)}. \end{aligned}$$

Theorem 3.3.4 *When* $p = 2$*:*

(1) *The distribution of* $\Lambda_{2;q,n}$ *is the same as that of* Y^2, *where* $Y \sim \mathrm{Be}(n, q)$.

(2) $\dfrac{1-\sqrt{\Lambda_{2;q,n}}}{\sqrt{\Lambda_{2;q,n}}} \cdot \dfrac{2(n-1)}{2q} \sim F(2q, 2(n-1))$.

(3) $\dfrac{1-\sqrt{\Lambda_{2;q,n}}}{\sqrt{\Lambda_{2;q,n}}} \cdot \dfrac{2(n+1-p)}{2p} \sim F(2p, 2(n+1-p))$.

Proof. From Theorem 3.3.3 we may write

$$\Lambda_{2;q,n} = V_1 V_2, \quad V_1 \sim \mathrm{Be}\left(\frac{1}{2}n, \frac{q}{2}\right), \quad V_2 \sim \mathrm{Be}\left(\frac{1}{2}(n-1), \frac{q}{2}\right),$$

and V_1 and V_2 are independent. The hth moment of $\Lambda_{2;q,n}$ is

$$\mathrm{E}[\Lambda_{2;q,n}^h] = \frac{\Gamma[\frac{1}{2}n+h]\Gamma[\frac{1}{2}(n+q)]}{\Gamma[\frac{1}{2}n]\Gamma[\frac{1}{2}(n+q)+h]} \times \frac{\Gamma[\frac{1}{2}(n-1)+h]\Gamma[\frac{1}{2}(n+q-1)]}{\Gamma[\frac{1}{2}(n-1)]\Gamma[\frac{1}{2}(n+q-1)+h]}.$$

Using the duplication formula

$$\begin{aligned} \Gamma\left(\alpha+\frac{1}{2}\right)\Gamma(\alpha+1) &= \frac{\sqrt{\pi}\,\Gamma(2\alpha+1)}{2^{2\alpha}}, \\ \mathrm{E}[\Lambda_{2;q,n}^h] &= \frac{\Gamma[q+n-1]\Gamma[n-1+h]}{\Gamma[q+n-3+2h]\Gamma[n+1]}, \end{aligned}$$

which is equal to $\mathrm{E}[Y^{2h}]$. This proves (1). From (1) we can write

$$\sqrt{\Lambda_{2;q,n}} = \frac{\chi_{2n}^2}{\chi_{2n}^2 + \chi_{2q}^2},$$

which gives (2). Result (3) follows from (2) and Theorem 3.2.5. □

Problems

3.1 (t-Distribution) Let X and Y be independent random variables following $\mathrm{N}(0,1)$ and $\chi^2(n)$, respectively. Then show that the density function is of $T = X/\sqrt{Y/n}$ is given by

$$f_T(t) = \frac{\Gamma\big((n+1)/2\big)}{\sqrt{n\pi}\,\Gamma\big(n/2\big)}\left(1+\frac{1}{n}t^2\right)^{-(n+1)/2}.$$

This distribution is called a *t-distribution* with n degrees of freedom.

Hint: The joint density function of (X,Y) is

$$f(x,y) = \frac{1}{\sqrt{n}}\exp\left(-\frac{x^2}{2}\right)\cdot\frac{1}{2^{n/2}}y^{n/2-1}\exp\Big(-\frac{y}{2}\Big).$$

Consider the transformation $(X,Y)\to(T,S); T = X/\sqrt{Y/n},\ S = Y$, and let

$$t = \frac{x}{\sqrt{y/n}}, \qquad s = y.$$

The Jacobian of the transformation is $\sqrt{s/n}$, and we have the joint density function $g(t,s)$ of (T,S). The result is obtained by integrating $g(t,s)$ with respect to s.

3.2 (Noncentral t-distribution) In problem 2.1, let the variable Z follow the normal distribution $\mathrm{N}(\lambda,1)$. Then, show that the density function $T = Z/\sqrt{Y/n}$ is given by

$$f(t) = \exp\Big(-\frac{1}{2}\lambda^2\Big)\frac{n^{n/2}}{\sqrt{\pi}\,\Gamma(n/2)}\sum_{k=0}^{\infty}\frac{\Gamma((n+k+1)/2)}{k!}\frac{(\sqrt{2}\,t\lambda)^k}{(n+t^2)^{(n+k+1)/2}}.$$

This distribution is called a *noncentral t-distribution* with n degrees of freedom and noncentrality parameter λ^2, which is denoted as $t(n;\lambda)$.

Hint: The conditional distribution of T given the condition $Y=y$ is the normal distribution $\mathrm{N}(\lambda(y/n)^{-1/2},(y/n)^{-1})$. Therefore, the density function of T is expressed as

$$\begin{aligned}
f(t) &= \mathrm{E}_Y\left[\frac{\sqrt{Y/n}}{\sqrt{2\pi}}\exp\left\{-\frac{Y}{2n}\Big(t-\frac{\lambda}{\sqrt{Y/n}}\Big)^2\right\}\right]\\
&= \int_0^\infty \frac{\sqrt{y/n}}{\sqrt{2\pi}}\exp\left\{-\frac{y}{2n}\Big(t-\frac{\lambda}{\sqrt{y/n}}\Big)^2\right\}\cdot\frac{y^{n/2-1}\mathrm{e}^{-y/2}}{\Gamma(n/2)2^{n/2}}\,\mathrm{d}y\\
&= \mathrm{e}^{-\lambda^2/2}\int_0^\infty \frac{\sqrt{y/n}}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(1+t^2/n)y\right\}\exp\left(\sqrt{y/n}\,\lambda t\right)\\
&\qquad\times\frac{y^{n/2-1}\mathrm{e}^{-y/2}}{\Gamma(n/2)2^{n/2}}\,\mathrm{d}y.
\end{aligned}$$

The result is obtained by expanding $\exp\left(\sqrt{y/n}\,\lambda t\right)$ in a power series and integrating term by term.

3.3 (F-distribution) Let X and Y be independent random variables following $\chi^2(m)$ and $\chi^2(n)$, respectively. Then show that the density function of $V = X/m/Y/n$ is

$$f_V(v) = \frac{\Gamma\big((m+n)/2\big)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} v^{m/2-1}\left(1+\frac{m}{n}v\right)^{-(m+n)/2}, \qquad v > 0.$$

The distribution is called an F-distribution with (m, n) degrees of freedom and is denoted by $\mathrm{F}(m, n)$.

Hint: The density function of (X, Y) is

$$f(x, y) = \frac{x^{m/2-1}\exp\big(-x/2\big)}{2^{m/2}\Gamma\big(m/2\big)} \cdot \frac{y^{n/2-1}\exp\big(-y/2\big)}{2^{n/2}\Gamma\big(n/2\big)}$$

Consider the transformation from (X, Y) to (V, U) : $V = (X/m)/(Y/n)$, $U = Y$, and set

$$v = \frac{x/m}{y/n}, \quad u = y.$$

The transformation is one-to-one and its Jacobian is $(m/n)u$. Therefore, the density function of (V, U) is

$$\begin{aligned} g(v, u) &= f\left(\frac{m}{n}vu, u\right)\left|\frac{m}{n}\right| \\ &= \frac{m^{m/2}n^{-n/2}v^{m/2-1}u^{(m+n)/2-1}\exp\big\{-\big((m/n)v+1\big)u/2\big\}}{2^{m/2}2^{n/2}\Gamma\big(m/2\big)\Gamma\big(n/2\big)}. \end{aligned}$$

The density function of V is obtained by integrating out $g(v, u)$ with respect to u: $g(v) = \int_0^\infty g(v, u)\,\mathrm{d}u$.

3.4 (Noncentral F-distribution) In Problem 2.5, let X be distributed as the noncental chi-squared distribution $\chi^2(m; \delta^2)$. Then show that the density function of V is given by

$$f(x) = \mathrm{e}^{-\delta^2/2}\sum_{k=0}^{\infty}\frac{(\delta^2/2)^k}{k!} \cdot \frac{(m/n)^{k+m/2}}{\mathrm{B}(k+m/2, n/2)} \cdot \frac{x^{m/2+k-1}}{\big(1+(m/n)x\big)^{k+(m+n)/2}},$$

where $\mathrm{B}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. This distribution is called a *noncentral* F-*distribution* with (m, n) degrees of freedom and noncentrality parameter δ^2, which is denoted by $\mathrm{F}(m, n; \delta^2)$. Also, the density function is expressed as $f_{m,n}(x; \delta^2) = \mathrm{E}_K[f_{m+2K,n}(x)]$, where $K \sim \mathrm{Po}(\frac{1}{2}\delta^2)$.

Hint: The result is obtained by the same method as that used in deriving the F-distribution.

3.5 Consider the null distribution of $T^2_{2\cdot 1}$ appearing in Rao's U-statistic in (3.1.3):

$$U = \frac{T^2 - \mathrm{T}_1^2}{n + T_1^2} = \frac{T^2_{2\cdot 1}}{n + T_1^2},$$

which gives the expression

$$T^2_{2\cdot 1} = U \cdot (n + \mathrm{T}_1^2).$$

Show that the distribution of $T^2_{2\cdot 1}$ can be regarded as a scale mixture of the F-distribution by chcking that U and T_1^2 are mutually independent.

Note: Approximations to the scale mixture of the F-distribution will be discussed in Chapter 14].

3.6 Let $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ be the sample mean vector and sample covariance matrix based on a random sample of size $N = n + 1$ from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Prove the following three propositions:

(1) $$\max_{\boldsymbol{a}} \frac{N\{\boldsymbol{a}'(\bar{\boldsymbol{X}} - \boldsymbol{\mu})\}}{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}} = T^2 \equiv N(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}).$$

(2) $$\frac{1}{n}T^2 \cdot \frac{n-p+1}{p} \sim \mathrm{F}(p, n-p+1).$$

(3) $$\mathrm{P}\{\boldsymbol{a}'\bar{\boldsymbol{X}} - c\sqrt{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}} \le \boldsymbol{a}'\boldsymbol{\mu} \le \boldsymbol{a}'\bar{\boldsymbol{X}} + c\sqrt{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}, \ \text{for any} \ \ \boldsymbol{a}\} \ge 1 - \alpha,$$

where $c^2 = \dfrac{np(n-p+1)}{n+1}\mathrm{F}(p, n-p+1; \alpha)$.

3.7 Let

$$T^2 = \boldsymbol{Z}'\left(\frac{1}{m}\mathbf{W}\right)^{-1}\boldsymbol{Z} \quad \text{and} \quad L = \frac{|\mathbf{W}|}{|\mathbf{W} + \boldsymbol{Z}\boldsymbol{Z}'|},$$

where $\boldsymbol{Z} \sim \mathrm{N}_p(\boldsymbol{\nu}, \Sigma)$, $m\mathbf{W} \sim \mathrm{W}_p(m, \Sigma)$, and $\boldsymbol{Z}$ and $\mathbf{W}$ are independent. Then, show that the distribution of $(1 - L)/L$ is the same as one of $(1/m)T^2$.

Hint: Use $|\mathbf{W} + \boldsymbol{Z}\boldsymbol{Z}'| = |\mathbf{W}| \cdot (1 + \boldsymbol{Z}'\mathbf{W}^{-1}\boldsymbol{Z})$.

3.8 Consider testing hypothesis $\mathrm{H}_0 : \mathrm{D}'\boldsymbol{\mu} = \mathbf{0}$ on the basis of a random sample of size $N = n+1$ from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, where Σ is unknown, and D is a given matrix of size $p \times r$ with rank r. Let $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ be the sample mean vector and sample covariance matrix, respectively.

(1) Show that the hypothesis H_0 is equivalent to '$\boldsymbol{\mu} = \mathrm{Q}\boldsymbol{\theta}$', where Q is a $p \times q\,(= p - r)$ matrix with rank q satisfying $\mathrm{D}'\mathrm{Q} = \mathrm{O}$, and $\boldsymbol{\theta}$ is an unknown vector. Further, show (see Sections 12.2.1 and 12.4.3) that the LR test is based on

$$T^2 = N(\mathrm{D}'\bar{\mathrm{X}})'(\mathrm{D}'\mathrm{S}\mathrm{D})^{-1}\mathrm{D}'\bar{\mathrm{X}}.$$

(2) Show that under the hypothesis, $(1/n)T^2 \cdot [(n-r+1)/r]$ is ditributed as $\mathrm{F}(r, n-r+1)$.

3.9 Let $\Lambda = |\mathbf{W}|/|\mathbf{T}|$ be distributed as a lambda distribution $\Lambda_p(q, n)$, and cosider the decomposition $\Lambda = V_1 V_2 \cdots V_p$ given in Theorem 3.3.3. Let $\Lambda(X_1, \ldots, X_i)$ be the Λ corresponding to the first i variables $X_1, \ldots, X_i$, and put

$$\Lambda(X_i|X_1, \ldots, X_{i-1}) = \Lambda(X_1, \ldots, X_i)/\Lambda(X_1, \ldots, X_{i-1}), \ i = 2, \ldots, p.$$

Then, show that $V_1 = \Lambda(X_1), \ V_i = \Lambda(X_i|X_1, \ldots, X_{i-1}), i = 2, \ldots, p.$

This page intentionally left blank

Chapter 4

Correlation Coefficients

In this chapter we define several correlation coefficients between variables or random vectors in connection with linear regression models. Their sampling counterparts are obtained by the least squares and maximum likelihood methods. The sampling distributions and statistical inferences of these correlations are discussed. A covariance selection model that is related to partial correlations is discussed as well. The canonical correlation, which is the correlation between random vectors, will be discussed later in Chapter 11.

4.1 Ordinary Correlation Coefficients

4.1.1 Population Correlation

The correlation between two variables X_1 and X_2 is defined by

$$\rho = \rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}. \tag{4.1.1}$$

Let the mean vector, covariance matrix, and correlation matrix of $\boldsymbol{X} = (X_1, X_2)'$ be denoted by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \quad \mathrm{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and let $\sigma_{11} = \sigma_1^2$ $(\sigma_1 \geq 0)$ and $\sigma_{22} = \sigma_2^2$ $(\sigma_2 \geq 0)$. Then the correlation between X_1 and X_2 is denoted by

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}.$$

The advantage of the correlation is that it is independent of the scale. More precisely, consider the transformation

$$Z_1 = aX_1 + b, \quad Z_2 = cX_2 + d,$$

where a, b, c, and d are constants and $a > 0$, $c > 0$. Then

$$\begin{aligned}&\mathrm{Var}(Z_1) = a^2\mathrm{Var}(X_1), \quad \mathrm{Var}(Z_2) = c^2\mathrm{Var}(X_2),\\ &\mathrm{Cov}(Z_1, Z_2) = ac\mathrm{Cov}(X_1, X_2).\end{aligned}$$

Therefore,

$$\rho(aX_1 + b, cX_2 + d) = \rho(X_1, X_2),$$

which shows that ρ is invariant under location and scale transformation.

It is well know that

$$|\rho| \leq 1, \tag{4.1.2}$$

which is also shown from a consequence of Theorem 4.1.1.

Consider predicting X_1 by a linear function of X_2 (i.e., $\alpha + \beta X_2$). Then we have the following result.

Theorem 4.1.1

$$\min_{\alpha,\beta} \mathrm{E}[(X_1 - \alpha - \beta X_2)^2] = \sigma_1^2(1 - \rho^2).$$

The equality holds when $\beta = \rho\sigma_2/\sigma_1 \equiv b$ *and* $\alpha = \mu_1 - b\mu_2$.

Proof. Consider a reduction,

$$X_1 - \alpha - \beta X_2 = \{X_1 - \mu_1 - \beta(X_2 - \mu_2)\} + (\mu_1 - \alpha - \beta\mu_2).$$

Then

$$\begin{aligned}&\mathrm{E}[(X_1 - \alpha - \beta X_2)^2]\\ &\quad = \mathrm{E}[\{X_1 - \mu_1 - \beta(X_2 - \mu_2)\}^2] + (\mu_1 - \alpha - \beta\mu_2)^2\\ &\quad = \sigma_{11} - 2\beta\sigma_{21} + \beta\sigma_{22} + (\mu_1 - \alpha - \beta\mu_2)^2\\ &\quad = \sigma_{22}(\beta - \sigma_{21}/\sigma_{22})^2 + \sigma_{11} - \sigma_{21}^2/\sigma_{22}\\ &\qquad + (\mu_1 - \alpha - \beta\mu_2)^2\\ &\quad \geq \sigma_{11}(1 - \rho^2).\end{aligned}$$

The equality condition is easily obtained. □

From Theorem 4.1.1 we have an optimum predictor

$$\hat{X}_1 = \mu_1 + b(X_2 - \mu_2), \quad b = \rho\sigma_2/\sigma_1, \tag{4.1.3}$$

which is called the *best linear predictor*. It satisfies

$$\mathrm{E}[(X_1 - \hat{X}_1)^2] = \sigma_1^2(1 - \rho^2). \tag{4.1.4}$$

This shows that $|\rho| \leq 1$, since the left-hand side is nonnegative.

4.1.2 Sample Correlation

Let $\boldsymbol{X}_1 = (X_{11}, X_{12})', \ldots, \boldsymbol{X}_N = (X_{N1}, X_{N2})$ be a random sample drawn from a population with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Let the sample mean vector and the sample covariance matrix be denoted by

$$\bar{\boldsymbol{X}} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

Then the sample correlation between X_1 and X_2 is defined by

$$R = R(X_1, X_2) = \frac{S_{12}}{\sqrt{S_{11}S_{22}}} = \frac{S_{12}}{S_1 S_2},$$

where $S_{ii} = S_i^2$, $S_i \geq 0$, $i = 1, 2$.

It is easy to see that:

(1) $|R| \leq 1$.

(2) $R(X_1, X_2) = R(aX_1 + b, cX_2 + d)$, where $a > 0$, $c > 0$.

The best linear predictor in the sample version is defined by

$$\hat{X}_1 = \bar{X}_1 + B(X_2 - \bar{X}_2), \quad B = RS_2/S_1.$$

The predictor can also be obtained by considering

$$\min_{\alpha,\beta} \sum_{i=1}^{N} (X_{i1} - \alpha - \beta X_{i2})^2.$$

Theorem 4.1.2 *Let R be the sample correlation coefficient of a sample of size $N = n + 1$ drawn from a bivariate normal distribution with correlation ρ. If $\rho = 0$, then*

$$\sqrt{n-1}\frac{R}{\sqrt{1-R^2}}$$

has the t-distribution with $n - 1$ degrees of freedom.

Proof. Since $n\mathbf{S} \sim \mathrm{W}_2(n, \Sigma)$, without loss of generality we may write

$$R = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2 \sum_{i=1}^n Y_i^2}} = \frac{\boldsymbol{X}'\boldsymbol{Y}}{\sqrt{\|\boldsymbol{X}\|^2 \cdot \|\boldsymbol{Y}\|^2}} = \frac{\boldsymbol{A}'\boldsymbol{Y}}{\sqrt{\|\boldsymbol{Y}\|^2}},$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)'$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{A} = (1/\|\boldsymbol{X}\|)\boldsymbol{X}$, and

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim \text{i.i.d. } \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Under the assumption $\rho = 0$, $\boldsymbol{Y}$ is independent of $\boldsymbol{X}$. The conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}$ is $\mathrm{N}_n(\mathbf{0}, \mathrm{I}_n)$. For any orthogonal random matrix $\mathbf{H}$ depending on $\boldsymbol{X}$ only, the distributions of $\boldsymbol{Y}$ and $\mathbf{H}'\boldsymbol{Y}$ are the same. We choose $\mathbf{H}$ as its first column $\boldsymbol{A}$. Then, by writing $\mathbf{H}'\boldsymbol{Y}$ as $\boldsymbol{Y}$, we have

$$\frac{R}{\sqrt{1-R^2}} = \frac{Y_1}{\sqrt{\sum_{i=1}^2 Y_i^2}}$$

which gives our conclusion. □

It is known [see Kariya and Eaton, 1977] that the normality assumption is not important as long as one set of these samples has a spherical distribution.

From Theorem 4.1.2 we can write

$$\frac{\sqrt{n-1}\,R}{(1-R^2)^{1/2}} = \frac{Z_1}{\sqrt{\chi^2(n-1)/(n-1)}},$$

where $Z_1 \sim \mathrm{N}(0,1)$, and Z_1 and $\chi^2(n-1)$ are independent. This shows that $R^2 = \chi_1^2/(\chi_1^2 + \chi_{n-1}^2)$, $\chi_1^2 = Z^2$, and hence R^2 is distributed as a beta distribution with parameters $1/2$ and $(n-1)/2$. Therefore, the density function of R is given by

$$\frac{\Gamma(n/2)}{\pi^{1/2}\Gamma[(n-1)/2]}(1-r^2)^{(n-3)/2}, \quad -1 < r < 1.$$

The sample correlation is the MLE of ρ when the population is normal. This is shown by checking that:

(1) There exists the MLE of Σ.

(2) There exists a one-to-one correspondence between Σ and $(\sigma_{11}, \sigma_{22}, \rho)$.

The second condition can be relaxed as follows:

(2)′ $\rho = \rho(\Sigma)$ is a mapping from $\mathcal{S} = \{\Sigma \colon \Sigma > O\}$ to $(-1, 1)$.

Such a general result was given by Zehna (1966) as follows.

Lemma 4.1.1 *Let $L(\boldsymbol{\theta})$ be the likelihood function of a sample. Suppose that the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ exists so that $L(\hat{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Omega$. Let $g(\boldsymbol{\theta})$ be an arbitrary mapping of Ω to Ω^*, where Ω^* is the range of $g(\boldsymbol{\theta})$. Then $g(\hat{\boldsymbol{\theta}})$ is a maximum (induced) likelihood estimator of $g(\boldsymbol{\theta})$.*

Proof. See Zehna (1966) or Siotani et al. (1985). □

To obtain the distribution of R, we start with $\mathbf{V} = n\mathbf{S} \sim \mathrm{W}_2(n, \Sigma)$. Let

$$\mathbf{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}.$$

Then the density function of $\mathbf{V}$ (i.e., the joint density function of V_{11}, V_{12}, and V_{22}) is

$$\frac{\mathrm{etr}(-\frac{1}{2}\Sigma^{-1}\mathrm{V})|\mathrm{V}|^{(n-3)/2}}{2^n|\Sigma|^{n/2}\Gamma_2(\frac{1}{2}n)}.$$

Since V is invariant under the transformation

$$\mathrm{V} \to \begin{pmatrix} \sigma_1^{-1} & 0 \\ 0 & \sigma_2^{-1} \end{pmatrix} \mathrm{V} \begin{pmatrix} \sigma_1^{-1} & 0 \\ 0 & \sigma_2^{-1} \end{pmatrix},$$

we can assume without loss of generality that

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Then

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix},$$

so that

$$\delta = \frac{1}{2}\,\mathrm{tr}\,\Sigma^{-1}\mathrm{V} = \frac{1}{2(1-\rho^2)}(v_{11} + v_{22} - 2\rho v_{12}).$$

Therefore, the joint density function of V_{11}, V_{12}, and V_{22} is

$$\frac{\mathrm{e}^{-\delta}(v_{11}v_{22} - v_{12}^2)^{(n-3)/2}}{2^n(1-\rho^2)^{n/2}\Gamma(\frac{1}{2}n)}.$$

Consider the transformation $w_1 = \sqrt{v_{11}}$ and $w_2 = \sqrt{v_{22}}$. Noting that the Jacobian of its inverse transformation is $2^2 w_1 w_2$, the distribution of R is expressed as

$$\begin{aligned} f(r) = 4 \cdot \frac{1}{2^n\Gamma(\frac{1}{2}n)}(1-\rho^2)^{-n/2}(1-r^2)^{(n-3)/2} \\ \times \int_0^\infty \int_0^\infty (w_1w_2)^{n-1} \exp\left\{-\frac{w_1^2 - 2\rho r w_1 w_2 + w_2^2}{2(1-\rho^2)}\right\} \mathrm{d}w_1\,\mathrm{d}w_2. \end{aligned}$$

By expanding $\exp\{\rho w_1 w_2/(1-\rho^2)\}$ in a power series and integrating term by term, we obtain

$$f(r) = \frac{1}{\Gamma(\frac{1}{2}n)}(1-\rho^2)^{n/2}(1-r^2)^{(n-3)/2}\sum_{j=0}^\infty \frac{(2\rho r)^j}{j!}\left\{\Gamma\left(\frac{n+j}{2}\right)\right\}^2.$$

Since, by the duplicate formula $\Gamma(a)\Gamma(a+1) = \sqrt{\pi}\,\Gamma(2a)/2^{2a-1}$,

$$\Gamma_2\Big(\frac{1}{2}n\Big) = \sqrt{\pi}\,\Gamma\Big(\frac{1}{2}n\Big)\Gamma\Big(\frac{1}{2}(n-1)\Big) = 2^{-n+2}\pi\Gamma(n-1),$$

we finally obtain the density function of R as

$$f(r) = \frac{2^{n-2}}{\pi\Gamma(n-1)}(1-\rho^2)^{n/2}(1-r^2)^{(n-3)/2}\sum_{j=0}^{\infty}\frac{(2\rho r)^j}{j!}\Big\{\Gamma\Big(\frac{n+j}{2}\Big)\Big\}^2.$$

Theorem 4.1.3 *Let R be a sample correlation coefficient based on a random sample of size $N = n+1$ drawn from a bivariate normal population with correlation ρ. Then R is distributed asymptotically as* $\mathrm{N}(\rho, (1-\rho^2)^2/n)$, *or the limiting distribution of $\sqrt{n}(R-\rho)$ is* $\mathrm{N}(\rho, (1-\rho^2)^2)$.

Proof. See Section 4.2.5. □

Based on Theorem 4.1.3, it holds asymptotically that with probability $1-\alpha$,

$$r - \frac{1}{\sqrt{n}}z_\alpha(1-\rho^2) \le \rho \le r + \frac{1}{\sqrt{n}}z_\alpha(1-\rho^2),$$

where z_α is the upper α-point of $\mathrm{N}(0,1)$. A conventional confidence interval is constructed by substituting r for the ρ in the endpoints. However, such a confidence interval will not be accurate. It is natural to look for a transformation of R such that the transformed variable converges in law to the normality faster than that of R itself and with a variance independent of ρ. Let $g(r)$ be an analytic function of r at $r = \rho$. Then by Theorem 2.5.2, $\sqrt{n}\{g(R) - g(\rho)\}$ is asymptotically normal with 0 and variance

$$\left\{\left.\frac{\mathrm{d}g(r)}{\mathrm{d}r}\right|_{r=\rho}\right\}^2 (1-\rho^2)^2.$$

We wish to determine the function $g(\cdot)$ so that this variance is a constant, say, 1. Then we have a differential equation for $g(r)$ to satisfy

$$g'(\rho) = \frac{1}{1-\rho^2} = \frac{1}{2}\left(\frac{1}{1+\rho} + \frac{1}{1-\rho}\right).$$

Solving the equation, the transformation is given as

$$g(\rho) = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) = \tanh^{-1}\rho \equiv \xi. \tag{4.1.5}$$

Let

$$Z = \frac{1}{2}\log\left(\frac{1+R}{1-R}\right) = \tanh^{-1}R, \tag{4.1.6}$$

which was introduced by Fisher (1921) and is called *Fisher's z-transformation.*

Theorem 4.1.4 *Let Z and ξ be defined by* (4.1.6) *and* (4.1.5), *respectively. Then the limiting distribution of $\sqrt{n}(Z-\rho)$ is* $\mathrm{N}(0,1)$.

Konishi (1978) gave an interpretation of Fisher's z-transformation with an asymptotic expansion of the distribution function of R.

4.2 Multiple Correlation Coefficient

4.2.1 Population Multiple Correlation

In this section we consider a measure for the relationship between X_1 and $\boldsymbol{X}_2 = (X_2, \ldots, X_p)'$. Put $\boldsymbol{X} = (X_1, \boldsymbol{X}_2')'$, and let the mean vector and the covariance matrix of $\boldsymbol{X}$ decompose as

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{21}' \\ \boldsymbol{\sigma}_{21} & \Sigma_{22} \end{pmatrix}.$$

The multiple correlation coefficient can be characterized in various ways. Consider predicting X_1 by a linear predictor $\alpha + \boldsymbol{\beta}'\boldsymbol{X}_2$ of $\boldsymbol{X}_2$. Then the multiple correlation is the maximum correlation between X_1 and any linear function $\alpha + \boldsymbol{\beta}'\boldsymbol{X}_2$ of $\boldsymbol{X}_2$. The quantity is given as in the following.

Definition 4.2.1 The multiple correlation coefficient between X_1 and $\boldsymbol{X}_2 = (X_2, \ldots, X_p)'$, denoted by $\rho_{1(2\cdots p)}$, is defined by

$$\rho_{1(2\cdots p)} = \sqrt{\frac{\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}}{\sigma_{11}}}.$$

Theorem 4.2.1 *For linear predictor $\alpha + \boldsymbol{\beta}'\boldsymbol{X}_2$, it holds that:*

(1) $\min_{\alpha, \boldsymbol{\beta}} \mathrm{E}[(X_1 - \alpha - \boldsymbol{\beta}'\boldsymbol{X}_2)^2] = \sigma_1^2(1 - \rho_{1(2\cdots p)}^2)$.

(2) $\max_{\alpha, \boldsymbol{\beta}} \rho(X_1, \alpha + \boldsymbol{\beta}'\boldsymbol{X}_2) = \max_{\boldsymbol{\beta}} \rho(X_1, \boldsymbol{\beta}'\boldsymbol{X}_2) = \rho_{1(2\cdots p)}$.

The equality in (1) holds when

$$\boldsymbol{\beta} = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21} \equiv \boldsymbol{b}, \quad \alpha = \mu_1 - \boldsymbol{b}'\boldsymbol{\mu}_2 \equiv a.$$

The equality in (2) holds when $\boldsymbol{\beta} = c\boldsymbol{b}$, c is a proportional constant, and α is any constant.

Proof. The first result is proved by extending the proof of Theorem 4.1.1. In fact,

$$\begin{aligned}
&\mathrm{E}[(X_1 - \alpha - \boldsymbol{\beta}'\boldsymbol{X}_2)^2] \\
&\quad = \mathrm{E}[\{X_1 - \mu_1 - \boldsymbol{\beta}'(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)\}]^2 + (\mu_1 - \alpha - \boldsymbol{\beta}'\boldsymbol{\mu}_2)^2 \\
&\quad \geq \sigma_{11} - 2\boldsymbol{\sigma}_{21}'\boldsymbol{\beta} + \boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta}.
\end{aligned}$$

The equality in the last reduction folds when $\alpha = \mu_1 - \alpha - \boldsymbol{\beta}'\boldsymbol{\mu}_2$. Further, we have

$$\begin{aligned}
&\sigma_{11} - 2\boldsymbol{\sigma}_{21}'\boldsymbol{\beta} + \boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta} \\
&\quad = (\boldsymbol{\beta} - \Sigma_{22}\boldsymbol{\sigma}_{21})'\Sigma_{22}(\boldsymbol{\beta} - \Sigma_{22}\boldsymbol{\sigma}_{21}) + \sigma_{11} - \boldsymbol{\sigma}_{21}'\Sigma_{22}\boldsymbol{\sigma}_{21} \\
&\quad \geq \sigma_{11}(1 - \rho_{1(2\cdots p)}^2).
\end{aligned}$$

The equality in the last reduction holds when $\boldsymbol{\beta} = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}$. For the second result, it is easy to see that

$$\begin{aligned}
\rho(X_1, \alpha + \boldsymbol{\beta}'\boldsymbol{X}_2) &= \rho(X_1, \boldsymbol{\beta}'\boldsymbol{X}_2) \\
&= \boldsymbol{\sigma}_{21}'\boldsymbol{\beta}/[\sigma_{11}(\boldsymbol{\beta}'\Sigma_{22}\boldsymbol{\beta})^{1/2}] = \phi(\boldsymbol{\beta}).
\end{aligned}$$

By [A.2.2] we have

$$\max_{\boldsymbol{\beta}} \phi^2(\boldsymbol{\beta}) = \frac{1}{\sigma_{11}}\boldsymbol{\sigma}_{21}'\Sigma^{-1}\boldsymbol{\sigma}_{21} = \rho_{1(2\cdots p)}^2,$$

and the equality holds when $\boldsymbol{\beta} = c\Sigma^{-1}\boldsymbol{\sigma}_{21}$ (c is a proportional constant). This completes the proof. □

From Theorem 4.2.1 we have an optimum predictor

$$\hat{X}_1 = \ell(\boldsymbol{X}_2) = \boldsymbol{\mu}_2 + \boldsymbol{b}'(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \quad \boldsymbol{b} = \Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}, \tag{4.2.1}$$

which is called the *best linear predictor*, and

$$\begin{aligned}
&\mathrm{E}[(X_1 - \hat{X}_1)^2] = \sigma_{11}(1 - \rho_{1(2\cdots p)}^2), \\
&\rho(X_1, \hat{X}_1) = \rho_{1(2\cdots p)}.
\end{aligned} \tag{4.2.2}$$

In the linear regression, $\ell(\boldsymbol{X}_2)$ given by (4.2.2) may be interpreted as the best predictor of X_1 in terms of $\boldsymbol{X}_2$ in the sense that it minimizes the mean square error $\mathrm{E}[(X_1 - \ell(\boldsymbol{X}_2))^2]$. The power of prediction may be measured by $\rho_{1(2\cdots p)}$.

4.2.2 Sample Multiple Correlation

Let

$$\boldsymbol{X}_1 = \begin{pmatrix} X_{11} \\ \boldsymbol{X}_{12} \end{pmatrix}, \ldots, \boldsymbol{X}_N = \begin{pmatrix} X_{N1} \\ \boldsymbol{X}_{N2} \end{pmatrix}$$

be a random sample drawn from a population with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Let the sample mean vector, the sample covariance matrix, and the sample correlation matrix be denoted by

$$\bar{\boldsymbol{X}} = \begin{pmatrix} \bar{X}_1 \\ \bar{\boldsymbol{X}}_2 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} S_{11} & \boldsymbol{S}'_{21} \\ \boldsymbol{S}_{21} & \mathbf{S}_{22} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & \boldsymbol{R}'_{21} \\ \boldsymbol{R}_{21} & \mathbf{R}_{22} \end{pmatrix}.$$

The sample multiple correlation between X_1 and $\boldsymbol{X}_2$ is defined by

$$R_{1(2\cdots p)} = \sqrt{\frac{\boldsymbol{S}'_{21}\mathbf{S}_{22}^{-1}\boldsymbol{S}_{21}}{S_{11}}} = \sqrt{\boldsymbol{R}'_{21}\mathbf{R}_{22}^{-1}\boldsymbol{R}_{21}}, \tag{4.2.3}$$

which is the one obtained from Definition 4.2.1 by substituting $\mathbf{S}$ to Σ. The sample multiple correlation is the maximum sample correlation between X_1 and the linear function $\alpha + \boldsymbol{\beta}'\boldsymbol{X}_2$ of $\boldsymbol{X}_2$, that is,

$$\max_{\alpha,\boldsymbol{\beta}} r(X_1, \alpha + \boldsymbol{\beta}'\boldsymbol{X}_2) = \max_{\boldsymbol{\beta}} r(X_1, \boldsymbol{\beta}'\boldsymbol{X}_2) = R_{1(2\cdots p)}.$$

The maximum is attained at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} = \mathbf{S}_{22}^{-1}\boldsymbol{S}_{21}$. Similarly, consider predicting X_1 by the linear function $\alpha + \boldsymbol{\beta}'\boldsymbol{X}_2$ of $\boldsymbol{X}_2$. Then we have

$$\min_{\alpha,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{N}(X_{i1} - \alpha - \boldsymbol{\beta}'\boldsymbol{X}_{i2})^2 = S_{11}(1 - R_{1(2\cdots p)}^2).$$

The equality holds at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\alpha = \hat{\alpha}$, where

$$\hat{\boldsymbol{\beta}} = \mathbf{S}_{22}^{-1}\boldsymbol{S}_{21} \equiv \boldsymbol{B}, \quad \hat{\alpha} = \bar{X}_1 - \boldsymbol{B}'\bar{\boldsymbol{X}}_2.$$

The best predictor is

$$\ell(\boldsymbol{X}_2) = \bar{X}_1 + \boldsymbol{S}'_{21}\mathbf{S}_{22}^{-1}(\boldsymbol{X}_2 - \bar{\boldsymbol{X}}_2).$$

When the population is a p-variate normal distribution $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, the sample multiple correlation is the MLE of $\rho_{1(2\cdots p)}$. This is obtained from [A.2.2].

For simplicity, put $\bar{\rho} \equiv \rho_{1(2\cdots p)}$ and $\bar{R} = R_{1(2\cdots p)}$. We first consider the distribution of $\bar{R}$ under $\bar{\rho} = 0$. Put $\mathbf{V} = n\mathbf{S}$; then $\mathbf{V} \sim \mathrm{W}_p(n, \Sigma)$. If we partition $\mathbf{V}$ similarly to Σ or $\mathbf{S}$ as

$$\mathbf{V} = \begin{pmatrix} V_{11} & \boldsymbol{V}'_{21} \\ \boldsymbol{V}_{21} & \mathbf{V}_{22} \end{pmatrix},$$

and the sample multiple correlation is given by

$$\bar{R}^2 = \frac{\boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}}{V_{11}} = \frac{\boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}}{V_{11\cdot 2} + \boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}},$$

where $V_{11\cdot 2} = V_{11} - \boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}$. Note that

$$\bar{\rho} = 0 \quad \Leftrightarrow \quad \boldsymbol{\sigma}_{21} = \mathbf{0}.$$

From Theorem 2.2.3 we know that

$$V_{11\cdot 2}/\sigma_{11} \sim \chi^2(n-(p-1)) \quad \text{and} \quad \boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}/\sigma_{11} \sim \chi^2(p-1),$$

and that they are independent, which implies that

$$\bar{R}^2 = \frac{\chi_{p-1}^2}{\chi_{p-1}^2 + \chi_{n-(p-1)}^2},$$

which, in turn, gives the following theorem.

Theorem 4.2.2 *Let $\bar{R}$ be the sample multiple correlation between X_1 and $\boldsymbol{X}_2 = (X_2, \ldots, X_p)'$ based on a sample of size $N = n+1$ of $\boldsymbol{X} = (X_1, \boldsymbol{X}_2')'$ whose distribution is $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. If the population multiple correlation $\bar{\rho}$ is zero, then*

$$\frac{n-(p-1)}{p-1} \cdot \frac{\bar{R}^2}{1-\bar{R}^2}$$

has an F-distribution with $(p-1, n-(p-1))$ degrees of freedom.

Next we consider the distribution of $\bar{R}^2$ when $\bar{\rho}$ is nonzero. We know that

$$\frac{\bar{R}^2}{1-\bar{R}^2} = \frac{\boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}}{V_{11\cdot 2}}.$$

From Theorem: 2.2.2 we have

(1) $V_{11\cdot 2}$ and $\boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}$ are independent,
(2) $V_{11\cdot 2}/\sigma_{11\cdot 2} \sim \chi^2(n-p+1)$, and
(3) the conditional distribution of $\boldsymbol{V}_{21}'\mathbf{V}_{22}^{-1}\boldsymbol{V}_{21}/\sigma_{11\cdot 2}$ given $\mathbf{V}_{22}$ is $\chi^2(p-1;\xi^2)$, where $\xi^2 = \boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\mathbf{V}_{22}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}/\sigma_{11\cdot 2}$.

Therefore,

$$Z \equiv \frac{n-(p-1)}{p-1} \cdot \frac{\bar{R}^2}{1-\bar{R}^2} = \frac{\chi_{p-1}^2(\xi^2)/(p-1)}{\chi_{n-p+1}^2/(n-p+1)},$$

which is the noncentral F-distribution with $(p-1, n-p+1)$ degrees of freedom and noncentrality parameter ξ^2. The conditional density of Z given ξ^2 is given (see Problem 4.) by

$$\mathrm{e}^{-\xi^2/2} {}_1F_1\left(\frac{1}{2}n; \frac{1}{2}(p-1); \frac{(1/2)m\xi^2 z}{1+mz}\right) m^{(p-1)/2} \\ \times \frac{\Gamma(\frac{1}{2}n) z^{(p-3)/2}}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(n-p+1)](1+mz)^{n/2}},$$

where $m = (p-1)/(n-p+1)$ and ${}_1F_1$ denotes a hypergeometric function. The general hypergeometric function (or series) is

$${}_aF_b(a_1, \ldots, a_p; b_1, \ldots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \cdot \frac{z^k}{k!}, \tag{4.2.4}$$

where $(a)_k = a(a+1)\cdots(a+k-1)$. By changing variables from Z to $\bar{R}^2$, the conditional distribution of $\bar{R}^2$ given $\mathbf{W}_{22}$ is

$$f(\bar{r}^2|\xi^2) = \mathrm{e}^{-\xi^2/2} {}_1F_1\left(\frac{1}{2}n; \frac{1}{2}(p-1); \frac{1}{2}\xi^2\bar{r}^2\right) \cdot \frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(n-p+1)]} \\ \times (\bar{r}^2)^{(n-3)/2}(1-\bar{r}^2)^{(n-p-1)/2}, \qquad 0 < \bar{r}^2 < 1.$$

The (unconditional) density function is obtained by taking the expectation of $f(\bar{r}^2|\xi^2)$ with respect to ξ^2. Now $\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\mathbf{V}_{22}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21} \sim \mathrm{W}_1(n, \boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21})$, and hence

$$U = \frac{\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\mathbf{V}_{22}\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}}{\boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}} \sim \chi^2(n).$$

We can write $\xi^2 = \delta^2 U$, where $\delta^2 = \boldsymbol{\sigma}_{21}'\Sigma_{22}^{-1}\boldsymbol{\sigma}_{21}/\sigma_{11\cdot 2}$. Replacing ξ^2 in $f(\bar{r}^2|\xi^2)$ by $\xi^2 = \delta^2 U$, and taking expectation with respect to U, we have the following theorem.

Theorem 4.2.3 *Under Theorem 4.2.2, if $\bar{\rho} \neq 0$, the density function of $\bar{R}^2$ is*

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(n-p+1)]}(\bar{r}^2)^{(n-3)/2}(1-\bar{r}^2)^{(n-p-1)/2} \\ \times (1-\bar{\rho}^2)^{n/2} {}_1F_1\left(\frac{1}{2}n; \frac{1}{2}n; \bar{\rho}^2\bar{r}^2\right), \qquad 0 < \bar{r}^2 < 1.$$

Example 4.2.1 Consider the marks from an examination on five subjects for 88 students given in Mardia et al. (1979). The five subjects are: X_1, mechanics; X_2, vector; X_3, algebra; X_4, analysis; and X_5, statistics. The correlation coefficients between these subjects are given in Table 4.2.1. Based on the

Table 4.2.1 Correlation matrix of five subjects

	X_1	X_2	X_3	R_4	R_5
X_1	1.000	–			
X_2	0.533	1.000	–		
X_3	0.547	0.610	1.000	–	
X_4	0.409	0.485	0.711	1.000	–
X_5	0.389	0.437	0.665	0.607	1.000

X_1	X_2	X_3	X_4	X_5
0.602	0.658	0.821	0.735	0.692

correlation matrix, we obtain multiple correlations between any subject and the other subjects as follows (see Table 4.2.1). The maximum value is 0.821, which is the multiple correlation between X_3 and the reminder variables.

4.3 Partial Correlation

4.3.1 Population Partial Correlation

Consider the correlation ρ_{12} between two random variables X_1 and X_2. If X_1 and X_2 are to be considered in conjunction with $p-2$ other variables $X_3, \ldots, X_p$, we may regard ρ_{12} as a mixture of a direct correlation between X_1 and X_2 and an indirect portion due to the presence of other variables correlating with X_1 and X_2. The partial correlation measuring the direct portion of the total correlation is defined as a correlation between X_1 and X_2 after the removal by linear regression of effects due to other variables. Let $\boldsymbol{X} = (X_1, X_2, \boldsymbol{X}'_3)'$, and let the corresponding partitions of the mean vector $\boldsymbol{\mu}$ and the covariance matrix of $\boldsymbol{X}$ be

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \boldsymbol{\mu}_3 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \boldsymbol{\sigma}'_{31} \\ \sigma_{21} & \sigma_{22} & \boldsymbol{\sigma}'_{32} \\ \boldsymbol{\sigma}_{31} & \boldsymbol{\sigma}_{32} & \Sigma_{33} \end{pmatrix}. \tag{4.3.1}$$

The best linear predictor of $X_i (i = 1, 2)$ by a linear function of $\boldsymbol{X}_3$ is

$$\ell_i(\boldsymbol{X}_3) = \mu_i + \boldsymbol{\sigma}'_{i3}\Sigma_{33}^{-1}(\boldsymbol{X}_3 - \boldsymbol{\mu}_3), \quad i = 1, 2.$$

The residuals

$$X_i^* = X_i - \ell_i(\boldsymbol{X}_3), \quad i = 1, 2, \tag{4.3.2}$$

are regarded as the portions of X_i remaining after removing the effects due to the presence of $\boldsymbol{X}_3$. Then

$$\boldsymbol{X}^* = \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} = \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\sigma}_{31}' \\ \boldsymbol{\sigma}_{32}' \end{pmatrix} \Sigma_{33}^{-1}(\boldsymbol{X}_3 - \boldsymbol{\mu}_3)$$
$$= \begin{pmatrix} 1 & 0 & -\boldsymbol{\sigma}_{31}'\Sigma_{33}^{-1} \\ 0 & 1 & -\boldsymbol{\sigma}_{32}'\Sigma_{33}^{-1} \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \boldsymbol{X}_3 - \boldsymbol{\mu}_3 \end{pmatrix}$$

and hence

$$\mathrm{Var}(\boldsymbol{X}^*) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\sigma}_{31}' \\ \boldsymbol{\sigma}_{32}' \end{pmatrix} \Sigma_{33}^{-1}(\boldsymbol{\sigma}_{31}\ \boldsymbol{\sigma}_{32})$$
$$\equiv \Sigma_{(12)(12)\cdot 3\cdots p}.$$

We denote

$$\Sigma_{(12)(12)\cdot 3\cdots p} = \begin{pmatrix} \sigma_{11\cdot 3\cdots p} & \sigma_{12\cdot 3\cdots p} \\ \sigma_{21\cdot 3\cdots p} & \sigma_{22\cdot 3\cdots p} \end{pmatrix}. \tag{4.3.3}$$

Definition 4.3.1 The partial correlation between X_1 and X_2 after the removal of effects due to $\boldsymbol{X}_3$ is

$$\rho_{12\cdot 3\cdots p} = \frac{\sigma_{12\cdot 3\cdots p}}{\sqrt{\sigma_{11\cdot 3\cdots p}\sigma_{22\cdot 3\cdots p}}},$$

which is also called the *partial correlation* (*coefficient*) between X_1 and X_2 given $\boldsymbol{X}_3$.

The partial correlation may be defined in terms of the correlation matrix

$$\mathrm{R} = (\rho_{ij}) = \begin{pmatrix} \rho_{11} & \rho_{12} & \boldsymbol{\rho}_{31}' \\ \rho_{21} & \rho_{22} & \boldsymbol{\rho}_{32}' \\ \boldsymbol{\rho}_{31} & \boldsymbol{\rho}_{32} & \mathrm{R}_{33} \end{pmatrix} \tag{4.3.4}$$

by substituting R for Σ. That is,

$$\rho_{12\cdot 3\cdots p} = \frac{\rho_{12}^*}{\sqrt{\rho_{11}^*\rho_{22}^*}},$$

where

$$\mathrm{R}_{(12)(12)\cdot 3\cdots p} = \begin{pmatrix} \rho_{11}^* & \rho_{12}^* \\ \rho_{21}^* & \rho_{22}^* \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\rho}_{31}' \\ \boldsymbol{\rho}_{32}' \end{pmatrix} \mathrm{R}_{33}^{-1}(\boldsymbol{\rho}_{31}\ \boldsymbol{\rho}_{32}).$$

In particular, when $p = 3$,

$$\rho_{12\cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}.$$

4.3.2 Sample Partial Correlation

Let $\boldsymbol{X}$, $\mathbf{S}$, and $\mathbf{R}$ be versions of $\boldsymbol{\mu}$, Σ, and R based on a sample of size $N = n+1$ from a p-variate population with mean $\boldsymbol{\mu}$ and covariance matrix Σ. We partition $\boldsymbol{X}$, $\mathbf{S}$, and $\mathbf{R}$ in the same manner as in the partition of $\boldsymbol{\mu}$, Σ, and R. Similarly, let

$$
\begin{aligned}
\mathbf{S}_{(12)(12)\cdot 3\cdots p} &= \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{S}'_{31} \\ \boldsymbol{S}'_{32} \end{pmatrix} \mathbf{S}_{33}^{-1} (\boldsymbol{S}_{31}\ \boldsymbol{S}_{32}) \\
&= \begin{pmatrix} S^*_{11} & S^*_{12} \\ S^*_{21} & S^*_{22} \end{pmatrix}, \\
\mathbf{R}_{(12)(12)\cdot 3\cdots p} &= \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{R}'_{31} \\ \boldsymbol{R}'_{32} \end{pmatrix} \mathbf{R}_{33}^{-1} (\boldsymbol{R}_{31}\ \boldsymbol{R}_{32}) \\
&= \begin{pmatrix} R^*_{11} & R^*_{12} \\ R^*_{21} & R^*_{22} \end{pmatrix}.
\end{aligned}
$$

Then the sample partial correlation $R_{12\cdot 3\cdots p}$ between X_1 and X_2 given $\boldsymbol{X}_3$ is defined from Definition 4.3.1 by substituting $\mathbf{S}$ or $\mathbf{R}$ for Σ. That is,

$$
R_{12\cdot 3\cdots p} = \frac{S^*_{12}}{\sqrt{S^*_{11} S^*_{22}}} = \frac{R^*_{12}}{\sqrt{R^*_{11} R^*_{22}}}. \tag{4.3.5}
$$

For the definition we may consider a regression approach by the best linear predictors of X_1 and X_2 based on $\boldsymbol{X}_3$, which are given by

$$
\begin{aligned}
\hat{X}_1 &= \bar{X}_1 + \boldsymbol{S}'_{31} \mathbf{S}_{33}^{-1} (\boldsymbol{X}_3 - \bar{\boldsymbol{X}}_3), \\
\hat{X}_2 &= \bar{X}_2 + \boldsymbol{S}'_{32} \mathbf{S}_{33}^{-1} (\boldsymbol{X}_3 - \bar{\boldsymbol{X}}_3).
\end{aligned}
$$

The residuals are

$$
X_1^* = X_1 - \hat{X}_1, \quad X_2^* = X_1 - \hat{X}_2.
$$

Then we can see that

$$
R_{12\cdot 3\cdots p} = R(X_1^*, X_2^*).
$$

We note that $R_{12\cdot 3\cdots p}$ is the MLE of $\rho_{12\cdot 3\cdots p}$ when the population is normal. For this, we have that there is a one-to-one correspondence between Σ and $(\Sigma_{(12)(12)\cdot 3\cdots p}, (\boldsymbol{\sigma}_{31}, \boldsymbol{\sigma}_{32}), \Sigma_{33})$, and the MLE of $\Sigma_{(12)(12)\cdot 3\cdots p}$ is $(n/N)\mathbf{S}_{(12)(12)\cdot 3\cdots p}$. Further, the partial correlation $\rho_{12\cdot 3\cdots p}$ is defined through $\Sigma_{(12)(12)\cdot 3\cdots p}$ [i.e., $\rho_{12\cdot 3\cdots p} = g(\Sigma_{(12)(12)\cdot 3\cdots p})$]. Therefore, from Lemma 4.1.1 the MLE of $\rho_{12\cdot 3\cdots p}$ is obtained by $g((n/N)\mathbf{S}_{(12)(12)\cdot 3\cdots p})$, which is

$$
g(\mathbf{S}_{(12)(12)\cdot 3\cdots p}) = R_{12\cdot 3\cdots p}.
$$

As seen in (4.3.5), the sample partial correlation $R_{12\cdot 3\cdots p}$ is regarded as a simple correlation defined from $\mathbf{S}_{(12)(12)\cdot 3\cdots p}$ instead of $\mathbf{S}_{(12)(12)}$. The distributions of $\mathbf{S}_{(12)(12)}$ and $\mathbf{S}_{(12)(12)\cdot 3\cdots p}$ are $\mathrm{W}_2(n, \Sigma_{(12)(12)})$ and $\mathrm{W}_2(n-(p-2), \Sigma_{(12)(12)\cdot 3\cdots p})$, respectively. Therefore, there is nothing new to be added to the results obtained for the simple correlation coefficient except for the replacement of n with $n-q$. In particular, the significance of $\rho_{12\cdot 3\cdots p}$ can be tested, based on

$$\sqrt{n-(p-2)-1}\frac{\bar{R}_{12\cdot 3\cdots p}}{\sqrt{1-\bar{R}^2_{12\cdot 3\cdots p}}} \sim \mathrm{t}(n-(p-2)-1)$$

under $\rho_{12\cdot 3\cdots p} = 0$.

4.3.3 Covariance Selection Model

For the relationship among p variables $X_1, \ldots, X_p$, it is interesting to examine whether two variables are independent, or conditionally independent, given the remainder variables. We denote the independence of X_i and X_j by $X_i \perp X_j$, and the conditional independence of X_i and X_j given $\boldsymbol{X}_r$ by $X_i \perp X_j | \boldsymbol{X}_r$, where $\boldsymbol{X}_r$ is a subvector of $\boldsymbol{X}$ obtained by deleting X_i and X_j.

In this section we assume that $\boldsymbol{X} = (X_1, \ldots, X_p)'$ follow $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Then we know that:

(a) $X_1 \perp X_2 \Leftrightarrow \sigma_{12} = 0 \quad (\rho_{12} = 0)$.

(b) $X_1 \perp X_2 \,|\, \boldsymbol{X}_3 \Leftrightarrow \sigma_{12\cdot 3\cdots p} = 0 \quad (\rho_{12\cdot 3\cdots p} = 0)$.

More generally:

(1) $X_i \perp X_j \Leftrightarrow \sigma_{ij} = 0 \quad (\rho_{ij} = 0)$.

(2) $X_i \perp X_j \,|\, \boldsymbol{X}_r \Leftrightarrow \sigma_{ij\cdot \mathrm{rest}} = 0 \quad (\rho_{12\cdot \mathrm{rest}} = 0)$.

Here $\sigma_{ij\cdot\mathrm{rest}}(\rho_{12\cdot\mathrm{rest}})$ is the residual covariance (correlation) between X_i and X_j after removing the effects due to the remaining $(p-2)$ variables, or the conditional covariance (correlation) given the remaining p variables.

It is shown (see the following theorem) that statement (2) is equivalent to $\sigma^{ij} = 0$ ($\rho^{ij} = 0$), where $\Sigma^{-1} = (\sigma^{ij})$, and $\mathrm{R}^{-1} = (\rho^{ij})$. A covariance matrix with $\sigma^{ij} = 0$ for a set of (i, j) called a *covariance selection model* was formulated by Dempster (1972).

Theorem 4.3.1 *Let* $\mathrm{R} = (\rho_{ij})$ *be the correlation matrix of* $\boldsymbol{X} = (X_1, X_2, \boldsymbol{X}_3')'$.

Then the partial correlation between X_1 *and* X_2 *given* $\boldsymbol{X}_3$ *is expressed as*

$$\rho_{12\cdot 3\cdots p} = -\frac{\rho^{12}}{\sqrt{\rho^{11}\rho^{22}}},$$

where $\mathrm{R}^{-1} = (\rho^{ij})$.

Proof. Let Q be the 2×2 matrix constructed by the first two rows and columns of R^{-1}. Then, from [A.1.2] we have

$$\begin{aligned} \mathrm{Q} &= \mathrm{R}^{-1}_{(12)(12)\cdot 3\cdots p} \\ &= \begin{pmatrix} \rho_{11\cdot 3\cdots p} & \rho_{12\cdot 3\cdots p} \\ \rho_{21\cdot 3\cdots p} & \rho_{22\cdot 3\cdots p} \end{pmatrix}^{-1}, \end{aligned}$$

and hence

$$\rho^{11} = \frac{\rho_{22\cdot 3\cdots p}}{|\mathrm{R}_{(12)(12)\cdot 3\cdots p}|}, \ \rho^{22} = \frac{\rho_{11\cdot 3\cdots p}}{|\mathrm{R}_{(12)(12)\cdot 3\cdots p}|}, \ \rho^{12} = \rho^{21} = \frac{-\rho_{12\cdot 3\cdots p}}{|\mathrm{R}_{(12)(12)\cdot 3\cdots p}|}.$$

These complete the proof. □

From Theorem 4.3.1 we have

$$\rho_{12\cdot 3\cdots p} = 0 \iff \rho^{12} = 0. \tag{4.3.6}$$

The result (4.3.6) is extended for the partial correlation between X_i and X_j given the remainder variables; that is,

$$\rho_{ij\cdot \mathbf{rest}} = 0 \iff \rho^{ij} = 0. \tag{4.3.7}$$

As a covariance selection model, we consider a model M such that all the partial correlations between every component of $\boldsymbol{X}_2$ and of $\boldsymbol{X}_3$ are zero, where $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \boldsymbol{X}_3')'$ and $\boldsymbol{X}_i$: $p_i \times 1$. In accordance with the partition of $\boldsymbol{X}$, the covariance matrix Σ and its inverse $\Lambda = \Sigma^{-1}$ can be put in the form

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix} \text{ and } \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix}, \tag{4.3.8}$$

respectively, where Σ_{ij}: $p_i \times p_j$ and Λ_{ij}: $p_i \times p_j$. Our model is M: $\Lambda_{23} = \mathrm{O}$, which is equivalent to

$$M\colon \Sigma_{23\cdot 1} \equiv \Sigma_{23} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13} = \mathrm{O}. \tag{4.3.9}$$

This equivalence follows from

$$\Lambda_{(23)(23)} \equiv \begin{pmatrix} \Lambda_{22} & \Lambda_{23} \\ \Lambda_{32} & \Lambda_{33} \end{pmatrix} = \Sigma^{-1}_{(23)(23)\cdot 1} \equiv \begin{pmatrix} \Sigma_{22\cdot 1} & \Sigma_{23\cdot 1} \\ \Sigma_{32\cdot 1} & \Sigma_{33\cdot 1} \end{pmatrix}^{-1},$$

which is a consequence of [A.1.2] and Problem 4.7. Here $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, and so on. Similar notation will be used in the following whenever there is fear of confusion. The model is equivalent (see Chapter 11) to the redundancy condition of $\boldsymbol{X}_2$ in canonical correlation analysis between $(\boldsymbol{X}_1', \boldsymbol{X}_2')'$ and $\boldsymbol{X}_3$.

The likelihood ratio criterion is derived by using the fact that M is a model for the conditional independence between $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$ given $\boldsymbol{X}_1$. It is a monotone function (see, Problem 4.8) of

$$L = \frac{|\mathbf{S}|}{|\mathbf{S}_{11}| \cdot |\mathbf{S}_{22\cdot1}| \cdot |\mathbf{S}_{33\cdot1}|} = \frac{|\mathbf{S}_{(23)(23)\cdot1}|}{|\mathbf{S}_{22\cdot1}| \cdot |\mathbf{S}_{33\cdot1}|}. \tag{4.3.10}$$

The statistic can be written as

$$L = \frac{|\mathbf{W}_{(23)(23)}|}{|\mathbf{W}_{22}| \cdot |\mathbf{W}_{33}|} = \frac{|\mathbf{W}_{33\cdot2}|}{|\mathbf{W}_{33}|}$$

in terms of

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{22} & \mathbf{W}_{23} \\ \mathbf{W}_{32} & \mathbf{W}_{33} \end{pmatrix} = n \begin{pmatrix} \mathbf{S}_{22\cdot1} & \mathbf{S}_{23\cdot1} \\ \mathbf{S}_{32\cdot1} & \mathbf{S}_{33\cdot1} \end{pmatrix},$$

where $\mathbf{W}_{33\cdot2} = \mathbf{W}_{33} - \mathbf{W}_{32}\mathbf{W}_{22}^{-1}\mathbf{W}_{23}$. Using Theorems 2.2.2 and 2.2.3, we know that $\mathbf{W} \sim \mathrm{W}_{p_2+p_3}(n - p_1, \Sigma_{(23)(23)\cdot1})$. We also know that, under the model M, two variables $\mathbf{W}_{33\cdot2}$ and $\mathbf{W}_{32}\mathbf{W}_{22}^{-1}\mathbf{W}_{23}$ are independently distributed as $\mathrm{W}_{p_3}(n - p_1 - p_3, \Sigma_{33\cdot12})$ and $\mathrm{W}_{p_3}(p_2, \Sigma_{33\cdot12})$, respectively. Here we use

$$\Sigma_{33\cdot12} = \Sigma_{33\cdot1} - \Sigma_{32\cdot1}\Sigma_{22\cdot1}^{-1}\Sigma_{23\cdot1},$$

which is shown in Problem 4.1. Therefore, under M,

$$L \sim \Lambda_{p_3}(p_2, n - p_1 - p_2).$$

Let $\mathbf{R}$ be the sample correlation matrix, and partition it in the same way as those of $\mathbf{S}$. Then the statistic L is expressed as

$$L = \frac{|\mathbf{R}_{(23)(23)\cdot1}|}{|\mathbf{R}_{22\cdot1}| \cdot |\mathbf{R}_{33\cdot1}|}$$

in terms of the correlation matrix $\mathbf{R}$, since L is invariant under the transformation from $\mathbf{S}$ to $\mathbf{R} = \mathbf{DSD}$, where $\mathbf{D} = \mathrm{diag}\,(1/\sqrt{S_{11}}, \ldots, 1/\sqrt{S_{pp}})$ and $\mathbf{S} = (S_{ij})$.

The above mentioned result is summarized as follows.

Theorem 4.3.2 *Suppose that the covariance matrix* Σ *of a random vector* $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \boldsymbol{X}_3')'$, $\boldsymbol{X}_i : p_i \times 1$ *and its inverse* Λ *are partitioned as in* (4.3.8). *Let* $\mathbf{S}$ *be the sample covariance matrix based on a sample of size* $N = n + 1$ *from* $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. *Then, the likelihood statistic for* $M : \Lambda_{23} = \mathrm{O}$ *is given by* L *in* (4.3.10) *and its null distribution is* $\Lambda_{p_3}(p_2, n - p_1 - p_2)$.

Example 4.3.1 We examine the relationship between the five mathematics subjects considered in Example 4.2.1. The partial correlation coefficients between x_i and x_j given the remaining three variables are given in Table 4.3.1. From the table we may suggest a covariance selection model:

$$M: \rho^{14} = \rho^{24} = \rho^{15} = \rho^{25} = 0,$$

which is equivalent to

$$X_1 \perp X_4 | (X_2, X_3, X_5), \qquad X_2 \perp X_4 | (X_1, X_3, X_5),$$

$$X_1 \perp X_5 | (X_2, X_3, X_4), \qquad X_2 \perp X_5 | (X_1, X_3, X_4).$$

Table 4.3.1 Partial correlation matrix of five subjects

X_1	–				
X_2	0.33	–			
X_3	0.23	0.28	–		
X_4	0.00	0.08	0.43	–	
X_5	0.02	0.02	0.36	0.25	–

The covariance selection model means that all the partial correlations between each of $\{X_1, X_2\}$ and of $\{X_4, X_5\}$ given the reminder variables are zero. The likelihood ratio test is

$$L = \frac{|\mathbf{R}_{(1245)(1245)\cdot 3}|}{|\mathbf{R}_{(12)(12)\cdot 3}| \cdot |\mathbf{R}_{(45)(45)\cdot 3}|},$$

whose null distribution is $\Lambda_2(2, 87-1-2) = \Lambda_2(2, 84)$. Therefore, writing $L = \Lambda_{2;2,84}$ to make its distributional property clear, that is $\Lambda_{2;2,84} \sim \Lambda_2(2, 84)$, we obtain from Theorem 3.3.4,

$$F = \frac{1 - \sqrt{\Lambda_{2;2,84}}}{\sqrt{\Lambda_{2;2,84}}} \cdot \frac{2(87-1)}{2 \cdot 2} \sim F(4, 172).$$

The matrices in L are computed as

$$\mathrm{R}_{(1245)(1245)\cdot 3} = \begin{pmatrix} 1.000 & 0.533 & 0.409 & 0.389 \\ 0.533 & 1.000 & 0.485 & 0.437 \\ 0.409 & 0.485 & 1.000 & 0.607 \\ 0.389 & 0.437 & 0.607 & 1.000 \end{pmatrix}$$

$$- \begin{pmatrix} 0.547 \\ 0.610 \\ 0.711 \\ 0.665 \end{pmatrix} \begin{pmatrix} 0.547 & 0.610 & 0.711 & 0.665 \end{pmatrix}$$

$$= \begin{pmatrix} 0.701 & 0.220 & 0.021 & 0.026 \\ 0.220 & 0.628 & 0.052 & 0.031 \\ 0.021 & 0.052 & 0.495 & 0.135 \\ 0.026 & 0.031 & 0.135 & 0.558 \end{pmatrix}.$$

Further, $|\mathrm{R}_{(1245)(1245)\cdot 3}| = 0.100$, and

$$|\mathrm{R}_{(12)(12)\cdot 3}| = \begin{vmatrix} 0.701 & 0.220 \\ 0.220 & 0.628 \end{vmatrix} = 0.392,$$

$$|\mathrm{R}_{(45)(45)\cdot 3}| = \begin{vmatrix} 0.495 & 0.135 \\ 0.135 & 0.558 \end{vmatrix} = 0.258.$$

Therefore. $L = 0.990, F = 0.219$, and its p-value is 0.927. This shows that the covariance selection model M is not rejected.

Problems

4.1 Let S be a $p \times p$ positive definite matrix, and partition it as

$$\mathrm{S} = \begin{pmatrix} \mathrm{S}_{11} & \mathrm{S}_{12} & \mathrm{S}_{13} \\ \mathrm{S}_{21} & \mathrm{S}_{22} & \mathrm{S}_{23} \\ \mathrm{S}_{31} & \mathrm{S}_{32} & \mathrm{S}_{33} \end{pmatrix}, \quad \mathrm{S}_{ij}\colon p_i \times p_j.$$

Prove the following:

$$\mathrm{S}_{3\cdot 12} \equiv \mathrm{S}_{33} - \mathrm{S}_{3(12)}\mathrm{S}_{(12)(12)}^{-1}\mathrm{S}_{(12)3} = \mathrm{S}_{33\cdot 1} - \mathrm{S}_{31\cdot 1}\mathrm{S}_{22\cdot 1}^{-1}\mathrm{S}_{23\cdot 1},$$

where $\mathrm{S}_{3(12)} = (\mathrm{S}_{31}\ \mathrm{S}_{32})$, $\mathrm{S}_{33\cdot 1} = \mathrm{S}_{33} - \mathrm{S}_{31}\mathrm{S}_{11}^{-1}\mathrm{S}_{13}$, and so on.

Hint: Let $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \boldsymbol{X}_3')'$ be a random vector having a normal distribution $\mathrm{N}_p(\mathbf{0}, \Sigma)$. Then the result is shown from the fact that the conditional covariance matrix of $\boldsymbol{X}_3$ given $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is the same as the conditional covariance matrix computed from the conditional distribution of $(\boldsymbol{X}_2', \boldsymbol{X}_3')'$ given $\boldsymbol{X}_1$. This fact is obtained from the fact that the conditional density function of $\boldsymbol{X}_3$ given $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is expressed as

$$\begin{aligned} f(\boldsymbol{x}_3|\boldsymbol{x}_1, \boldsymbol{x}_2) &\equiv f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)/f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)/f(\boldsymbol{x}_1)}{f(\boldsymbol{x}_1, \boldsymbol{x}_2)/f(\boldsymbol{x}_1)} \\ &= f(\boldsymbol{x}_2, \boldsymbol{x}_3|\boldsymbol{x}_1)/f(\boldsymbol{x}_2|\boldsymbol{x}_1). \end{aligned}$$

4.2 Let $h(\mathbf{S})$ be the sample multiple correlation between X_1 and $\boldsymbol{X}_2 = (X_2, \ldots, X_p)$, based on a sample of size $N = n + 1$, where $\mathbf{S}$ is the sample covariance matrix. Then show that $h(\mathbf{S})$ is invariant under the transformation from $\mathbf{S}$ to $\tilde{\mathbf{S}} = \mathrm{DSD}$, where $\mathrm{D} = \mathrm{diag}\,(d_1, \ldots, d_p)$ with $d_i > 0$.

4.3 In a bivariate normal sample, the regression coefficient defined by $B = V_{12}/V_{11} = R(V_{22}/V_{11})^{1/2}$ is the MLE of the population regression coefficient $\beta = \rho\sigma_2/\sigma_1$, where

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2(\boldsymbol{\mu}, \Sigma), \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2 \end{pmatrix},$$

and the joint distribution of V_{11}, V_{12}, and V_{22} is $\mathrm{W}_2(n, \Sigma)$.

(a) Show that the pdf of B is

$$f(b) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi}\,\Gamma(n/2)} |\Sigma|^{1/2n} [\sigma_1^{n-1}(\sigma_2^2 - 2\rho\sigma_1\sigma_2 b + \sigma_1^2 b^2)]^{-1/2(n+1)}.$$

(b) Prove that $\mathrm{E}(B) = \beta$ and $\mathrm{Var}(B) = (\sigma_2^2/\sigma_1^2)(1-\rho^2)/(n-2)$.

4.4 (Continuation) Show that the random variable

$$T_1 = \{\sqrt{n}\,\sigma_1/\sigma_2(1-\rho^2)^{1/2}\}(B - \beta)$$

is distributed as Student's t-distribution with n degrees of freedom (d.f.).

4.5 (Continuation) Prove that the random variable T defined by

$$T = \sqrt{n-1}\{V_{11}/V_{22}(1-R^2)\}^{1/2}(B - \beta)$$

has Student's t-distribution with $n-1$ d.f. [*Note*: This is due to Bartlett (1933), and we can obtain a test and confidence interval for β by using this distribution.]

4.6 Let

$$\boldsymbol{X}_1 = \begin{pmatrix} X_{11} \\ \boldsymbol{X}_{12} \end{pmatrix}, \ldots, \boldsymbol{X}_N = \begin{pmatrix} X_{N1} \\ \boldsymbol{X}_{N2} \end{pmatrix}$$

be a random sample of size $N = n + 1$ drawn from a population. Let the sample mean vector and the sample covariance matrix be denoted by

$$\bar{\boldsymbol{X}} = \begin{pmatrix} \bar{X}_1 \\ \bar{\boldsymbol{X}}_2 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} S_{11} & \boldsymbol{S}_{21}' \\ \boldsymbol{S}_{21} & \mathbf{S}_{22} \end{pmatrix}.$$

Then, observe that

(1) $\displaystyle r^2(X_1, \alpha + \boldsymbol{\beta}'\boldsymbol{X}_2) = \frac{(\boldsymbol{\beta}'\boldsymbol{S}_{21})^2}{S_{11}\boldsymbol{\beta}'\mathbf{S}_{22}\boldsymbol{\beta}}.$

(2) $$\frac{1}{n}\sum_{i=1}^{N}(X_{i1} - \alpha - \boldsymbol{\beta}'\boldsymbol{X}_{i2})^2 = (\bar{X}_1 - \alpha - \boldsymbol{\beta}'\bar{\boldsymbol{X}}_2)^2 + S_{11\cdot 2} + (\boldsymbol{\beta} - \mathbf{S}_{22}^{-1}\boldsymbol{S}_{21})'\mathbf{S}_{22}(\boldsymbol{\beta} - \mathbf{S}_{22}^{-1}\boldsymbol{S}_{21}).$$

Using these results, prove that

(3) $\max_{\alpha,\boldsymbol{\beta}} r(X_1, \alpha + \boldsymbol{\beta}'\boldsymbol{X}_2) = R_{1(2\cdots p)}$.

(4) $\min_{\alpha,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{N}(X_{i1} - \alpha - \boldsymbol{\beta}'\boldsymbol{X}_{i2})^2 = S_{11}(1 - R^2_{1(2\cdots p)})$.

4.7 For a symmetric and nonsigular matrix A, let

$$\mathrm{A}^{-1} \equiv \begin{pmatrix} \mathrm{A}_{11} & \mathrm{A}_{12} \\ \mathrm{A}_{21} & \mathrm{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathrm{B}_{11} & \mathrm{B}_{12} \\ \mathrm{B}_{21} & \mathrm{B}_{22} \end{pmatrix} \equiv \mathrm{B}.$$

Then, show that $\mathrm{B}_{12} = \mathrm{O} \Leftrightarrow \mathrm{A}_{12} = \mathrm{O}$.

4.8 Show that the statistic L given by (4.3.10) is an LR test statistic for testing $M : \Lambda_{23} = \mathrm{O} \Leftrightarrow \Sigma_{23\cdot 1} = \mathrm{O}$, along the following lines. Let f_1 and $f_{2|1}$ be the likelihoods of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$, respectively. Then, observe that

$$\begin{aligned}
-2\log f_1 &= N\log|\Sigma| + Np_1 \log 2\pi \\
&\quad + \mathrm{tr}\,\Sigma^{-1}\{n\mathbf{S}_{11} + N(\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1)(\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1)'\} \\
&\equiv g_1(\boldsymbol{\mu}_1, \Sigma_{11}), \\
-2\log f_{2|1} &= N\log|\Sigma_{22\cdot 1}| + N\log 2\pi \\
&\quad + \mathrm{tr}\,\Sigma_{22\cdot 1}^{-1}\left\{n\mathbf{S}_{22\cdot 1} + (\mathcal{B} - \Sigma_{21}\Sigma_{11}^{-1})\mathbf{S}_{11}(\mathcal{B} - \Sigma_{21}\Sigma_{11}^{-1})'\right. \\
&\quad \left. + N(\bar{\boldsymbol{x}}_2 - \mathcal{B}\bar{\boldsymbol{X}}_1 - \boldsymbol{\mu}_{2\cdot 1})(\bar{\boldsymbol{x}}_2 - \mathcal{B}\bar{\boldsymbol{X}}_1 - \boldsymbol{\mu}_{2\cdot 1})'\right\} \\
&= g_{2|1}(\boldsymbol{\mu}_{2\cdot 1}, \mathcal{B}, \Sigma_{22\cdot 1}),
\end{aligned}$$

where $\mathcal{B} = \Sigma_{21}\Sigma_{11}^{-1}$ and $\boldsymbol{\mu}_{2\cdot 1} = \boldsymbol{\mu}_2 - \mathcal{B}\boldsymbol{\mu}_1$. Using these results, show that

$$\begin{aligned}
\max\{-2\log f_{(12)}\} &= \min g_1(\boldsymbol{\mu}_1, \Sigma_{11}) + \min g_{2|1}(\boldsymbol{\mu}_{2\cdot 1}, \mathcal{B}, \Sigma_{22\cdot 1}) \\
&= N\log|(n/N)\mathbf{S}_{11}| + +Np_1 \log 2\pi \\
&\quad + N\log|(n/N)\mathbf{S}_{22\cdot 1}| + Np_2 \log 2\pi.
\end{aligned}$$

Let $f_{(123)}$ be the likelihood of $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \boldsymbol{X}_3')'$. Then, noting that under M,

$$f_{(123)} = f_1 \cdot f_{(23)|1} = f_1 \cdot f_{2|1} \cdot f_{3|1},$$

obtain the conclusion.

4.9 Show that

$$1 - \rho^2_{1(2,\ldots,p)} = (1 - \rho^2_{12})(1 - \rho^2_{12\cdot 3})\cdots(1 - \rho^2_{12\cdot 34\cdots p}).$$

This page intentionally left blank

Chapter 5

Asymptotic Expansions for Multivariate Basic Statistics

In statistical inference it is of fundamental importance to obtain the sampling distribution of statistics. However, we often encounter situations where the exact distribution cannot be obtained in closed form, or even if it is obtained, it might be of little use because of its complexity. One practical way of getting around the problem is to provide reasonable approximations of the distribution function, along with extra information on their possible errors. In this chapter, we summarize some methods and theories on asymptotic expansions of multivariate statistics. Both high-dimensional and large-sample approximations are discussed. These include topics on the Edgeworth, Cornish–Fisher, and bootstrap approximations, and their validity. Perturbation methods on the characteristic roots and vectors of some random matrices are also discussed.

5.1 Edgeworth Expansion and its Validity

Let $\{F_n(x)\}$ be a set of distribution functions indexed by a parameter n: typically, sample size. Suppose, for example, that $F_n(x)$ is approximated by the first k-terms of asymptotic expansion:

$$G_{k,n}(x) = G(x) + \sum_{j=1}^{k-1} n^{-j/2} p_j(x) g(x), \tag{5.1.1}$$

where $G(x)$ is the limiting distribution function of $F_n(x)$, $g(x)$ is the density function of $G(x)$, and the $p_j(x)$ are suitable polynomials. For a wide class of statistics it is known that the error $R_{k,n}(x) = F_n(x) - G_{k,n}(x)$ satisfies the order condition [see Bhattacharya and Ghosh, 1978]

$$R_{k,n}(x) = \mathrm{O}(n^{-k/2}) \quad \text{uniformly in } x. \tag{5.1.2}$$

In this chapter we are concerned with either the methods of finding the first few terms of the asymptotic expansion or the validity of the expansion in the sense that the order condition (5.1.2) holds true. In this section we focus on a class of statistics having valid asymptotic expansions. The problem of obtaining its error bound is discussed in Chapters 13 through 16.

Let $\{\boldsymbol{Z}_j\}$, $j = 1, 2, \ldots$ be a sequence of *i.i.d.* random m-vectors from a distribution with mean $\boldsymbol{\xi}$ and covariance matrix $\Psi > \mathrm{O}$. Consider the distribution of

$$\boldsymbol{V}_n = \sqrt{n}(\bar{\boldsymbol{Z}}_n - \boldsymbol{\xi}), \tag{5.1.3}$$

where $\bar{\boldsymbol{Z}}_n = (1/n)\sum_{i=1}^n \boldsymbol{Z}_i$. Then it follows that the limiting distribution of $\boldsymbol{V}_n$ is $\mathrm{N}_p(\mathbf{0}, \Psi)$. The result, called the *multivariate central limit theorem*, is shown as follows. Let the characteristic function of $\boldsymbol{Z}_1$ be denoted by

$$C_{\boldsymbol{Z}_1}(\boldsymbol{t}) = \mathrm{E}[\mathrm{e}^{i\boldsymbol{t}'\boldsymbol{Z}_1}],$$

where $\boldsymbol{t} = (t_1, \ldots, t_m)'$ and the t_j are real numbers. Then the characteristic function of $\boldsymbol{V}_n$ is written as

$$\begin{aligned} C_{\boldsymbol{V}_n}(\boldsymbol{t}) &= \mathrm{E}[\mathrm{e}^{i\boldsymbol{t}'\boldsymbol{V}_n}] \\ &= \exp(-\sqrt{n}\,i\boldsymbol{t}'\boldsymbol{\xi})\mathrm{E}[\exp\{i\sum_{j=1}^n \boldsymbol{t}'\boldsymbol{Z}_j/\sqrt{n}\}] \\ &= \exp(-\sqrt{n}\,i\boldsymbol{t}'\boldsymbol{\xi})[C_{\boldsymbol{Z}_1}((1/\sqrt{n})\boldsymbol{t})]^n, \end{aligned}$$

and hence

$$\log C_{\boldsymbol{V}_n}(\boldsymbol{t}) = -\sqrt{n}\,i\boldsymbol{t}'\boldsymbol{\xi} + n\log C_{\boldsymbol{Z}_1}((1/\sqrt{n})\boldsymbol{t}). \tag{5.1.4}$$

Since $\boldsymbol{Z}_1$ has the second moments, we have (see Theorem 1.4.1)

$$\log C_{\boldsymbol{Z}_1}(\boldsymbol{t}) = i\boldsymbol{t}'\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{t}'\Psi\boldsymbol{t} + \mathrm{o}(\|\boldsymbol{t}\|^2), \tag{5.1.5}$$

where $\|\boldsymbol{t}\| = \sqrt{t_1^2 + \cdots + t_m^2}$. Using (5.1.4) and (5.1.5), it is seen that

$$\lim_{n\to\infty} C_{\boldsymbol{V}_n}(\boldsymbol{t}) = \exp\left(-\frac{1}{2}\boldsymbol{t}'\Psi\boldsymbol{t}\right),$$

and hence from Theorem 1.1.5 we get the multivariate central limit theorem.

One important approach is to use an asymptotic expansion for the distribution of $\boldsymbol{V}_n$ in powers of $n^{-1/2}$. It is suggested that such an expansion will be obtained by expanding $C_{\boldsymbol{V}_n}(\boldsymbol{t})$ in powers of $n^{-1/2}$ and inverting. It is known that if $\boldsymbol{Z}_1$ has to sth moments with $s \geq 3$, then from Theorem 1.4.1 we can expand

$$\log C_{\boldsymbol{Z}_1}(\boldsymbol{t}) = i\boldsymbol{t}'\boldsymbol{\xi} + \frac{1}{2}i^2\boldsymbol{t}'\Psi\boldsymbol{t} + \sum_{r=3}^{s}\frac{i^r}{r!}\kappa_{j_1\cdots j_r}t_{j_1}\cdots t_{j_r} + \mathrm{o}(\|\boldsymbol{t}\|^s), \tag{5.1.6}$$

where $\kappa_{j_1\cdots j_r}$ is the cumulant of $\boldsymbol{Z}_1$. Substituting the expansion in (5.1.4), we obtain

$$\log C_{\boldsymbol{V}_n}(\boldsymbol{t}) = -\frac{1}{2}\boldsymbol{t}'\Psi\boldsymbol{t} + \sum_{r=3}^{s} n^{-(r-2)/2}\frac{i^r}{r!}\kappa_{j_1\cdots j_r}t_{j_1}\cdots t_{j_r} + \mathrm{o}(n^{(s-2)/2}) \tag{5.1.7}$$

and hence

$$C_{\boldsymbol{V}_n}(\boldsymbol{t}) = \exp\left(-\frac{1}{2}i^2\boldsymbol{t}'\Psi\boldsymbol{t}\right)\left[1 + \sum_{r=3}^{s} n^{-(r-2)/2}\frac{i^r}{r!}\tilde{\kappa}_{j_1\cdots j_r}t_{j_1}\cdots t_{j_r}\right] + \mathrm{o}(n^{-(s-2)/2}), \tag{5.1.8}$$

where $\tilde{\kappa}_{j_1j_2j_3} = \kappa_{j_1j_2j_3}$ and the remainder coefficients $\tilde{\kappa}_{j_1\cdots j_r}$ can be determined in terms of $\kappa_{j_1\cdots j_q}$ with $q \leq r$. Inverting (5.1.8), it is expected that there exists an asymptotic expansion of the density function of $\boldsymbol{V}_n$ given by

$$\phi_{s,n}(\boldsymbol{u}) = \phi(\boldsymbol{u})\left[1 + \sum_{j=1}^{s-2} a_j(\boldsymbol{u})n^{-j/2}\right], \tag{5.1.9}$$

where $\phi(\boldsymbol{u})$ is the density function of $\mathrm{N}_m(\boldsymbol{0}, \Psi)$.

The expansion $\phi_{s,n}(\boldsymbol{u})$ is called the Edgeworth expansion (E-expansion) of order $\mathrm{o}(n^{-(s-2)/2})$ for the distribution of $\boldsymbol{V}_n$. In general, the term *Edgeworth expansion* is used for a sequence $\{\boldsymbol{V}_n\}, n = 1, 2, \ldots$ of random m-vectors that possesses an expansion as in (5.1.9), with a general density function, not necessarily $\phi(\boldsymbol{u})$.

The E-expansion for the distribution of $\boldsymbol{V}_n$ defined by (5.1.3) has a long history. The first development was done by Edgeworth (1905), who gave a formal asymptotic expansion in the one-dimensional case. Cramér (1928) proved the validity of this formal expansion under certain regularity conditions. The following theorem on the existence of a valid asymptotic expansion for the distribution of $\boldsymbol{V}_n$ is due to Bhattacharya (1971).

Theorem 5.1.1 *Let $\{\boldsymbol{Z}_j\}$, $j = 1, 2, \ldots$ be a sequence of i.i.d. random p-vectors from a distribution with mean vector $\boldsymbol{\xi}$ and covariance matrix $\Psi >$ O. Suppose further that for some integer $s \geq 3$, $\boldsymbol{Z}_1$ satisfies the following assumptions:*

A1. $\mathrm{E}[\|\boldsymbol{Z}_1\|^s] < \infty$, *where* $\| * \|$ *is the Euclidean norm.*

A2. $\boldsymbol{Z}_1$ *(or its distribution) obeys Cramér's condition:*

$$\limsup_{\|\boldsymbol{t}\|\to\infty} \left| \mathrm{E}[\exp(i\boldsymbol{t}'\boldsymbol{Z}_1)] \right| < 1.$$

Then for the density function of $\boldsymbol{V}_n$ *defined by* (5.1.3),

$$\sup_{B\in\mathcal{B}} \left| \Pr\{\boldsymbol{V}_n \in B\} - \int_B \psi_{s,n}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \right| = \mathrm{o}(n^{-1/2(s-2)}),$$

where the supremum is over all convex Borel sets in R^m, $\phi_{s,n}$ *is defined by* (5.1.9), *and* $a_j(\boldsymbol{u})$ *is a polynomial in* $\boldsymbol{u}$ *whose coefficients are continuous functions of moments of* $\boldsymbol{Z}_1$ *up to the order* $j+2$.

It is more useful to consider a real-valued function $h(\bar{\boldsymbol{Z}}_n)$ of the sample mean $\bar{\boldsymbol{Z}}_n = (1/n)\sum_{j=1}^n \boldsymbol{Z}_j$, where $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ are i.i.d. random vectors with mean $\boldsymbol{\xi}$ and covariance matrix $\Psi > \mathrm{O}$. The random vectors may be defined as

$$\begin{aligned} \boldsymbol{Z}_j &= (Z_{j1}, \ldots, Z_{jm})' \\ &= (f_1(\boldsymbol{X}_j), \ldots, f_m(\boldsymbol{X}_j))', \; j = 1, 2, \ldots, \end{aligned} \tag{5.1.10}$$

where $f_1, \ldots, f_m$ are real-valued functions on R^p. In fact, many test statistics and estimators can be so expressed, since all functions of sample moments are of the form $h(\bar{\boldsymbol{Z}}_n)$. For example, we shall see that a statistic $T = g(\mathbf{S})$ is of the form $h(\bar{\boldsymbol{Z}}_n)$, where $\mathbf{S} = (S_{ij})$ is a $p \times p$ sample covariance matrix defined by

$$\mathbf{S} = \frac{1}{n}\sum_{j=1}^n (\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})' = \frac{1}{n}\sum_{j=1}^n \boldsymbol{X}_j\boldsymbol{X}_j' - \bar{\boldsymbol{X}}\bar{\boldsymbol{X}}'.$$

For simplicity, assume that $p = 2$, and let

$$\boldsymbol{X}_j = (X_{j1}, X_{j2})', \qquad j = 1, 2, \ldots.$$

We define $\boldsymbol{Z}_j = (Z_{j1}, \ldots, Z_{j5})'$, $j = 1, 2, \ldots$ as

$$Z_{j1} = X_{j1}, \quad Z_{j2} = X_{j2}, \quad Z_{j3} = X_{j1}^2, \quad Z_{j4} = X_{j2}^2, \quad Z_{j5} = X_{j1}X_{j2}.$$

Then, putting $\bar{Z}_{n\ell} = (1/n)\sum_{j=1}^n Z_{j\ell}$, we have

$$S_{11} = \bar{Z}_{n3} - \bar{Z}_{n1}^2, \quad S_{22} = \bar{Z}_{n4} - \bar{Z}_{n2}^2, \quad S_{12} = \bar{Z}_{n5} - \bar{Z}_{n1}\bar{Z}_{n2}.$$

These imply that T is a function of $\bar{\boldsymbol{Z}}_n = (\bar{Z}_{n1}, \ldots, \bar{Z}_{n5})'$.

Now we consider the distribution of

$$W_n = \sqrt{n}\{h(\bar{\boldsymbol{Z}}_n) - h(\boldsymbol{\xi})\}. \tag{5.1.11}$$

Assume that $h(\boldsymbol{z})$ is continuously differentiable in a neighborhood of $\boldsymbol{z} = \boldsymbol{\xi}$ and

$$\sigma^2 = \boldsymbol{b}'\Psi\boldsymbol{b} > 0,$$

where

$$\begin{aligned}\boldsymbol{b} &= (b_1, \ldots, b_p)' \\ &= (\partial h(\boldsymbol{z}_n)/\partial z_1, \ldots, \partial h(\boldsymbol{z}_n)/\partial z_p)'\Big|_{\boldsymbol{z}=\boldsymbol{\xi}} \equiv \partial h(\boldsymbol{z})/\partial \boldsymbol{z}\Big|_{\boldsymbol{z}=\boldsymbol{\xi}}.\end{aligned}$$

Then, according to Theorem 2.5.2, the limiting distribution of W_n is normal with mean zero and variance σ^2.

Precisely, the procedure is described as follows. Suppose that we have the characteristic function of W_n in the expanded form

$$C_n(t) = \mathrm{E}\big[\exp(i\,t\,W_n)\big] = C_{s,n}(t) + \mathrm{o}(n^{-(s-2)/2}),$$

where

$$C_{s,n}(t) = \exp\Big(-\frac{1}{2}\sigma^2 t^2\Big)\left[1 + \sum_{j=1}^{s-2} p_j(it)n^{-j/2}\right]; \tag{5.1.12}$$

$p_j(t)$ is a polynomial in t whose coefficients are independent of n. Then the E-expansion $\phi_{s,n}(u)$ is determined by inverting $C_{s,n}(it)$, that is,

$$\phi_{s,n}(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\,t\,u)C_{s,n}(t)dt.$$

The terms in the inversion are evaluated by using the formulas

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\,t\,u)\exp(-\sigma^2 t^2/2)\,\mathrm{d}t = \phi_{\sigma^2}(u), \\ &\frac{1}{2\pi}\int_{-\infty}^{\infty} (-i\,t)^j \exp(-i\,t\,u)\exp(-\sigma^2 t^2/2)\,\mathrm{d}t = \phi_{\sigma^2}^{(j)}(u),\end{aligned}$$

where $\phi_{\sigma^2}(u)$ is the density function of $\mathrm{N}(0, \sigma^2)$ and $\phi_{\sigma^2}^{(j)}(u)$ is the jth derivative of $\phi_{\sigma^2}(u)$. The inversion of (5.1.12) can be written as

$$\begin{aligned}\phi_{s,n}(u) &= \left[1 + \sum_{j=1}^{s-2} n^{-j/2} p_j\Big(-\frac{d}{du}\Big)\right]\phi_{\sigma^2}(u) \\ &= \phi_{\sigma^2}(u)\left[1 + \sum_{j=1}^{s-2} n^{-j/2} q_j(u)\right],\end{aligned} \tag{5.1.13}$$

where $\phi_{\sigma^2}(u)$ is the density function of $\mathrm{N}(0, \sigma^2)$ and $q_j(u)$ is a polynomial in u whose coefficients do not depend on n and is determined by the relation

$$\phi_{\sigma^2}(u)q_j(u) = p_j\Big(\frac{-\mathrm{d}}{\mathrm{d}u}\Big)\phi_{\sigma^2}(u).$$

Now we consider a general and powerful method for obtaining an asymptotic expansion of $C(t)$. First we use the approximation of W_n obtained by the Taylor expansion of $h(\bar{\boldsymbol{Z}}_n) = h(\boldsymbol{\xi} + \boldsymbol{V}_n/\sqrt{n})$; that is,

$$\hat{W}_{s,n}(\boldsymbol{V}_n) = \sum_j b_j V_{jn} + \frac{1}{2\sqrt{n}} \sum_{j_1,j_2} b_{j_1 j_2} V_{j_1 n} V_{j_2 n} + \cdots + \frac{1}{(s-1)!} n^{-(s-2)/2} \sum_{j_1,\ldots,j_{s-1}} b_{j_1 \cdots j_{s-1}} V_{j_1 n} \cdots V_{j_{s-1} n},$$

where $\boldsymbol{V}_n = (V_{n1}, \ldots, V_{np})'$, and

$$b_{j_1 \cdots j_\nu} = \partial^\nu h(\boldsymbol{z})/(\partial z_{j_1} \cdots \partial z_{j_\nu})|_{\boldsymbol{z}=\boldsymbol{\xi}}, \quad 1 \le j_1, \ldots, j_\nu \le p.$$

Next, using this approximation, we obtain the jth cumulant of $\tilde{W}_n$, which is given by

$$\kappa_{j,n} = \hat{\kappa}_{j,n} + \mathrm{o}\big(n^{-(s-2)/2}\big).$$

Then the expression

$$\exp\left\{ it\tilde{\kappa}_{1,n} + \frac{1}{2}(it)^2(\tilde{\kappa}_{2,n} - \sigma^2) + \sum_{j=3}^{s} \frac{1}{j!}(it)^j \hat{\kappa}_{j,n} \right\} \exp\left(-\frac{1}{2}\sigma^2 t^2 \right) \quad (5.1.14)$$

provides an approximation to the characteristic function of $\tilde{W}_n$ (or W_n). Expanding the first exponent, one may reduce (5.1.14) to (5.1.12).

The problem of the existence of a valid E-expansion of W_n was conjectured by Wallace (1958) and treated with limited success by several authors; see, for example, Chambers (1967), Sargan (1976), Phillips (1977, 1977a), Chibisov (1972), Bhattacharya (1985) and Bai and Rao (1991). The following theorem was developed by Bhattacharya and Ghosh (1978).

Theorem 5.1.2 *Suppose that for some integer* $s \ge 3$, $\boldsymbol{Z}_1$ *satisfies* A1 (Theorem 5.1.1), *while* $h(\boldsymbol{z})$ *satisfies* A3 *and* A4 *below.*

A3. *All the derivatives of* $h(\boldsymbol{z})$ *of order* s *and less are continuous in a neighborhood of* $\boldsymbol{z} = \boldsymbol{\xi}$.

A4. $\boldsymbol{a} = \frac{\partial h(\boldsymbol{z})}{\partial \boldsymbol{z}}\Big|_{\boldsymbol{z}=\boldsymbol{\xi}} \neq \boldsymbol{0}$.

If, in addition, (1) *the distribution of* $\boldsymbol{X}_1$ *in* (5.1.10) *has a nonzero absolutely continuous component* (*with respect to Lebesgue measure on* R^p) *and* (2) *the density of this component is strictly positive on some nonempty open set* U *on which* $f_1, \ldots, f_m$ *are continuously differentiable and* $1, f_1, \ldots, f_m$ *are* *linearly independent (as elements of the vector space of a continuous function on U) then*

$$\sup_B \left| P(W_n \in B) - \int_B \phi_{s,n}(u)\mathrm{d}u \right| = \mathrm{o}(n^{-(s-2)/2}), \tag{5.1.15}$$

where the supremum is over all measurable convex sets B in R^1 *and* $\phi_{s,n}(u)$ *is given by* (5.1.13).

(b) *If, instead of* (a), $\boldsymbol{Z}_1$ *satisfies* A2, *then the relation* (5.1.15) *holds uniformly over every class* $\mathcal{B}$ *of Borel sets satisfying*

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^\varepsilon} \phi_{\sigma^2}(v)\mathrm{d}v = \mathrm{O}(\varepsilon), \qquad \varepsilon \downarrow 0.$$

Here ∂B *is the boundary of* B *and* $(\partial B)^\varepsilon$ *is the* ε*-neighborhood of* B.

Note: Theorem 5.1.2 holds true for a vector-valued function $\boldsymbol{h}(\boldsymbol{z}) = (h_1(\boldsymbol{z}), \ldots, h_k(\boldsymbol{z}))'$ provided that $\mathrm{M} = \mathrm{G}'\Psi\mathrm{G}$ is nonsingular, where

$$\mathrm{G}: p \times k = \left(\left.\frac{\partial h_1(\boldsymbol{z})}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{\xi}}, \ \ldots, \ \left.\frac{\partial h_k(\boldsymbol{z})}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{\xi}} \right).$$

Next we consider the distribution of W_n when $\boldsymbol{a} = \partial h(\boldsymbol{z})/\partial \boldsymbol{z}|_{\boldsymbol{z}=\boldsymbol{\xi}} = 0$. This occurs when we are dealing with the null distributions of certain test statistics. For this case, let

$$Y_n = 2n\{h(\bar{\boldsymbol{Z}}_n) - h(\boldsymbol{\xi})\}.$$

Then again using the Taylor series of $h(\bar{z}_n)$ around the point $\bar{z}_n = \boldsymbol{\xi}$ up to $(s-1)$th order, we obtain an approximation to y_n as follows:

$$\begin{aligned}\hat{Y}_{s-1,n}(\boldsymbol{V}_n) = &\sum_{j_1,j_2} b_{j_1j_2}V_{j_1n}V_{j_2n} \\ &+ \frac{2}{3!}n^{-1/2} \sum_{j_1,j_2,j_3} b_{j_1j_2j_3}V_{j_1n}V_{j_2n}V_{j_3n} + \cdots \\ &+ \frac{2}{(s-1)!}n^{-1/2(s-3)} \sum_{j_1,\ldots,j_{s-1}} b_{j_1\cdots j_{s-1}}V_{j_1n}\cdots V_{j_{s-1}n}.\end{aligned}$$

We observe that since the leading term of Y_n is quadratic in V_{jn}, the limiting distribution of Y_n is a central shi-square distribution with q degrees of freedom ($\chi^2(q)$), where q is the rank of $\mathrm{B} = (b_{ij})$ under the following assumption:

A5. $\boldsymbol{b} = \mathbf{0}$ and $\mathrm{B}\Psi\mathrm{B} = \mathrm{B}$.

Chandra and Ghosh (1979) have established the existence of a valid asymptotic expansion for the distribution of Y_n under appropriate conditions. Thus in this case, the asymptotic expansion for the distribution of Y_n is obtained by formally inverting the exact characteristic function $\mathrm{E}[\exp(itY_n)]$ or the approximate characteristic function; that is,

$$\mathrm{E}[\exp\{it\hat{Y}_{s-1,n}(\boldsymbol{V}_n)\}], \quad \int \exp\{it\hat{y}_{s-1,n}(\boldsymbol{u})\phi_{s-1,n}(\boldsymbol{u})\}\,\mathrm{d}\boldsymbol{u},$$

and so on.

5.2 Sample Mean Vector and Covariance Matrix

Many multivariate tests are based on the sample mean vector $\bar{\boldsymbol{X}}$ and the sample covariance matrix $\mathbf{S}$, based on a sample of size n drawn from a population with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Asymptotic results on the distribution of $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ are basic for asymptotic expansions of various statistics. Under the assumption that the samples have been drawn for a normal population,

$$\bar{\boldsymbol{X}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma), \ (n-1)\mathbf{S} \sim \mathrm{W}_p(n-1, \Sigma),$$

and they are independent. In this section we consider the case when the population is not necessarily normal. We consider asymptotic expansions of the distributions of the standardized statistics

$$\boldsymbol{Z} = \sqrt{n}\,\Sigma^{-1/2}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}), \quad \mathbf{V} = \sqrt{n}\,\Sigma^{-1/2}(\mathbf{S} - \mathrm{I}_p)\Sigma^{-1/2}.$$

Let

$$\boldsymbol{Y} = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}), \ \boldsymbol{Y}_i = \Sigma^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\mu}), \qquad i = 1, \ldots, n. \tag{5.2.1}$$

Then the standardized statistics are expressed as

$$\boldsymbol{Z} = \sqrt{n}\,\bar{\boldsymbol{Y}}, \quad \mathbf{V} = \sqrt{n}(\mathbf{S} - \mathrm{I}_p), \tag{5.2.2}$$

where $\mathbf{S} = (1/(n-1))\sum_{i=1}^{n}(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})'$. In this section we denote the moments of $\boldsymbol{Y}$ by

$$\mathrm{E}\left[Y_{i_1} \cdots Y_{i_k}\right] = \mu_{i_1 \cdots i_k},$$

and the corresponding cumulants as $\kappa_{i_1 \cdots i_k}$. Then $\mathrm{E}(\boldsymbol{Y}) = \mathbf{0}, \mathrm{Var}(\boldsymbol{Y}) = \mathrm{I}_p$, and

$$\begin{aligned} &\kappa_a = 0, \ \kappa_{ab} = \delta_{ab}, \ \kappa_{abc} = \mu_{abc}, \\ &\kappa_{abcd} = \mu_{abcd} - (\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}). \end{aligned}$$

We use the following moment notation:

$$\begin{aligned} m_{j;k;\ell} &= m_{j;k\ell} = \kappa_{jk\ell}, \\ m_{ab;cd} &= \kappa_{abcd} + \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}, \\ m_{j;ab;cd} &= \kappa_{jabcd} + \sum_{[4]} \kappa_{jac}\sigma_{bd}, \\ m_{ab;cd;ef} &= \kappa_{abcdef} + \sum_{[12]} \kappa_{abce}\sigma_{bd} + \sum_{[4]} \kappa_{ace}\kappa_{bdf} + \sum_{[8]} \sigma_{ac}\sigma_{be}\sigma_{df}. \end{aligned} \tag{5.2.3}$$

The multivariate skewness and kurtosis are expressed as

$$\begin{aligned} \kappa_{[3]}^{(1)} &= \sum_{a,b,c} (\kappa_{abc})^2, \\ \kappa_{[3]}^{(2)} &= \sum_{a} (\sum_{b} \kappa_{abb})^2 = \sum_{a,b,c} \kappa_{aab}\kappa_{bcc}, \\ \kappa_{[4]}^{(1)} &= \sum_{a,b} \kappa_{aabb}. \end{aligned}$$

Asymptotic expansion for the distribution of $\boldsymbol{Z}$ has been studied extensively. The formula is obtained from the following lemma.

Lemma 5.2.1 *Suppose that $\boldsymbol{Y}$ has a fourth moment. Then the characteristic function of $\boldsymbol{Z}$ can be expanded as*

$$\begin{aligned} C_{\boldsymbol{Z}}(\boldsymbol{t}) = \exp\left(\frac{i^2}{2}\boldsymbol{t}'\boldsymbol{t}\right) \Bigg[1 &+ \frac{i^3}{6\sqrt{n}} \sum_{a,b,c} \kappa_{abc} t_a t_b t_c \\ &+ \frac{i^4}{24n} \sum_{a,b,c,d} \kappa_{abcd} t_a t_b t_c t_d \\ &+ \frac{i^6}{72n} \sum_{a,b,c,d,e,f} \kappa_{abc}\kappa_{def} t_a t_b t_c t_d t_e t_f \Bigg] + \mathrm{o}(n^{-1}), \end{aligned} \tag{5.2.4}$$

where $\boldsymbol{t} = (t_1, \ldots, t_p)'$ and the t_j are real numbers.

Proof. The result is obtained from (5.1.6) and (5.1.7) with $\Psi = \mathrm{I}_p$. □

Inverting the characteristic function above and using Theorem 5.1.2 we have the following asymptotic expansion.

Theorem 5.2.1 *Suppose that the distribution of $\boldsymbol{Y}$ satisfies* Cramér*'s condition with a fourth order moment. Let $\boldsymbol{Z} = \sqrt{n}\,\bar{\boldsymbol{Y}}$ be defined as in* (5.2.2)*. Then we have*

$$\sup_{B} \left| \Pr\{\boldsymbol{Z} \in B\} - \int_B \phi_{4,n}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right| = \mathrm{o}(n^{-1}),$$

where the supremum is over all Borel sets B, and

$$\phi_{4,n}(\boldsymbol{z}) = \phi_p(\boldsymbol{z}) \left[1 + \frac{1}{6\sqrt{n}} \sum_{a,b,c} \kappa_{abc} H_{abc}(\boldsymbol{z}) \right.$$
$$\left. + \frac{1}{24n} \sum_{a,b,c,d} \kappa_{abcd} H_{abcd}(\boldsymbol{z}) + \frac{1}{72n} \sum_{a,b,c,d,e,f} \kappa_{abc}\kappa_{def} H_{abcdef}(\boldsymbol{z}) \right]. \tag{5.2.5}$$

Here $\phi_p(\boldsymbol{z})$ is the density function of $\mathrm{N}_p(\mathbf{0}, \mathrm{I}_p)$*:*

$$\phi_p(\boldsymbol{z}) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \exp\left(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right),$$

and $H_a(\boldsymbol{z}), H_{ab}(\boldsymbol{z}), \ldots$ are multivariate Hermite polynomials defined by

$$H_{a_1\cdots a_j}(\boldsymbol{z})\phi_p(\boldsymbol{z}) = (-1)^{a_1+\cdots+a_j} \frac{\partial^j}{\partial z_{a_1} \cdots \partial z_{a_j}} \phi_p(\boldsymbol{z}). \tag{5.2.6}$$

Next we consider asymptotic expansions of the joint distribution as well as the joint characteristic function of $\boldsymbol{Z}$ and $\mathbf{V}$ up to the order $\mathrm{O}(n^{-1/2})$. Let

$$\tilde{\mathbf{S}} = n^{-1} \sum_{j=1}^{n} \boldsymbol{X}_j \boldsymbol{X}_j', \quad \tilde{\mathbf{V}} = \sqrt{n}(\tilde{\mathbf{S}} - \mathrm{I}_p). \tag{5.2.7}$$

Then it is easy to obtain an expansion formula of $C_{(\boldsymbol{Z},\tilde{\mathbf{V}})}(\boldsymbol{t}, \mathrm{T})$. On the other hand, note that

$$\mathbf{V} = \tilde{\mathbf{V}} + n^{-1/2}(\mathrm{I}_p - \boldsymbol{Z}\boldsymbol{Z}') + \mathrm{O}_p(n^{-1}).$$

This implies that

$$\begin{aligned} C_{(\boldsymbol{Z},\mathbf{V})}(\boldsymbol{t}, \mathrm{T}) &= \mathrm{E}\left[\exp\left(\sum_j it_j Z_j + \sum_{a \le b} it_{ab} \tilde{V}_{ab} \right) \right. \\ &\quad \left. \times \left\{ 1 + \frac{i}{\sqrt{n}} \left(\sum_{a \le b} t_{ab}\delta_{ab} - \sum_{a \le b} t_{ab} Z_a Z_b \right) + \mathrm{o}(n^{-1/2}) \right\} \right] \\ &= \left[1 + \frac{i}{\sqrt{n}} \sum_{a \le b} t_{ab} \left(\delta_{ab} + \frac{\partial^2}{\partial t_a \partial t_b} \right) + \mathrm{o}(n^{-1/2}) \right] C_{(\boldsymbol{Z},\tilde{\mathbf{V}})}(\boldsymbol{t}, \mathrm{T}). \end{aligned}$$

Using this relation we have the following expression for the joint characteristic function of $\boldsymbol{Z}$ and $\mathbf{V}$:

Lemma 5.2.2 *Suppose that $\boldsymbol{x}$ has a sixth moment. Then the characteristic function of $(\boldsymbol{Z}, \mathbf{V})$ can be expanded as*

$$C_{(\boldsymbol{Z},\mathbf{V})}(\boldsymbol{t}, \mathrm{T}) = \exp\left(\frac{i^2}{2}A_0\right)\left[1 + \frac{i^3}{6\sqrt{n}}A_1 + \mathrm{o}(n^{-1/2})\right], \tag{5.2.8}$$

where $\mathrm{T} = (\frac{1}{2}(1+\delta_{ab})t_{ab})$, $t_{ab} = t_{ba}$,

$$\begin{aligned} A_0 &= \sum_j t_j^2 + 2\sum_{j,a\le b}\kappa_{jab}t_j t_{ab} + \sum_{a\le b, c\le d} m_{ab;cd}t_{ab}t_{cd}, \\ A_1 &= \sum_{j,k,\ell}\kappa_{jk\ell}t_j t_k t_\ell + 3\sum_{j,k,a\le b}\kappa_{jkab}t_j t_k t_{ab} \\ &\quad + 3\sum_{j,a\le b,c\le d} m_{j;ab;cd}t_j t_{ab}t_{cd} + \sum_{a\le b,c\le d,e\le f} m_{ab;cd;ef}t_{ab}t_{cd}t_{ef}, \end{aligned}$$

and the coefficients $m_{j:k:\ell}, \ldots$ are as given in (5.2.3).

From Lemma 5.2.2 we have that $\boldsymbol{Z}$ and $\mathbf{V}$ are asymptotically independent if and only if $\kappa_{jab} = 0$ for all j and $a \le b$. The characteristic function of $\mathbf{V}$ is obtained from Lemma 5.2.2 with $\boldsymbol{t} = \mathbf{0}$ as

$$\begin{aligned} C_{(\boldsymbol{Z},\mathbf{V})}(\boldsymbol{t}, \mathrm{T}) &= \exp\left(\frac{i^2}{2}\sum_{a\le b,c\le d} m_{ab;cd}t_{ab}t_{cd}\right) \\ &\quad \times \left[1 + \frac{i^3}{6\sqrt{n}}\sum_{a\le b,c\le d,e\le f} m_{ab;cd;ef}t_{ab}t_{cd}t_{ef} + \mathrm{o}(n^{-1/2})\right]. \end{aligned} \tag{5.2.9}$$

The density function of the limiting distribution of $\mathbf{V}$ is given by

$$\begin{aligned} &\left(\frac{1}{2\pi}\right)^{p(p+1)/2}\int \exp\left(-i\sum_{a\le b}t_{ab}v_{ab} - \frac{1}{2}\sum_{a\le b,c\le d} m_{ab;cd}t_{ab}t_{cd}\right)\mathrm{dT} \\ &= \left(\frac{1}{2\pi}\right)^{p(p+1)/4}|\Lambda|^{1/2}\exp\left(-\frac{1}{2}\sum_{a\le b,c\le d}\lambda_{ab;cd}v_{ab}v_{cd}\right), \end{aligned} \tag{5.2.10}$$

where $\Lambda = (\lambda_{ab;cd})$ is the $\frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1)$ matrix defined by

$$\Lambda^{-1} = (m_{ab;cd}), \qquad a \le b, c \le d. \tag{5.2.11}$$

We can obtain an asymptotic expansion of the density function of $\mathbf{V}$ by inverting the term of $\mathrm{O}(n^{-1/2})$. The inversion is computed by differentiating both sides of (5.2.10) with respect to v_{ab}.

Now, we consider the case when Σ is not I_p, assuming normality. Let $\mathbf{S}$ be the sample covariance matrix based on a sample of size $N = n+1$ from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Then $n\mathbf{S} \sim \mathrm{W}_p(n, \Sigma)$, and let

$$\mathbf{U} = (U_{ij}) = \sqrt{n}(\mathbf{S} - \Sigma). \tag{5.2.12}$$

Theorem 5.2.2 *Let* $\mathbf{U}$ *be a random matrix defined by* (5.2.12). *Then the density function of* $\mathbf{U}$ *can be expanded as*

$$\begin{aligned} f(\mathrm{U}) &= [\pi^{p(p+1)/4} 2^{p(p+3)/4} |\Sigma|^{p(p+1)/2}]^{-1} \operatorname{etr}\left\{-\frac{1}{4}(\Sigma^{-1}\mathrm{U})^2\right\} \\ &\quad \times \left[1 + \frac{1}{\sqrt{n}} q_1(\mathrm{U}) + \frac{1}{n} q_2(\mathrm{U}) + \mathrm{o}(n^{-1})\right], \end{aligned}$$

where

$$\begin{aligned} q_1(\mathrm{U}) &= -\frac{1}{2}(p+1)\operatorname{tr}(\Sigma^{-1}\mathrm{U}) + \frac{1}{6}\operatorname{tr}(\Sigma^{-1}\mathrm{U})^3, \\ q_2(\mathrm{U}) &= \frac{1}{2}\{q_1(\mathrm{U})\}^2 - \frac{1}{24}p(2p^2+3p-1) \\ &\quad + \frac{1}{4}(p+1)\operatorname{tr}(\Sigma^{-1}\mathrm{U})^2 - \frac{1}{8}\operatorname{tr}(\Sigma^{-1}\mathrm{U})^4. \end{aligned}$$

Proof. Let $\mathbf{V} = (1/\sqrt{2})\Sigma^{-1/2}\mathbf{U}\Sigma^{-1/2}$, and $f(\mathrm{V})$ be the density function of $\mathbf{V}$. Since $J(\mathrm{V} \to \mathrm{U}) = 2^{-p(p+1)/2}|\Sigma|^{-(p+1)/2}$ (see [A.3.5]), the density function of $\mathbf{U}$ is

$$2^{-p(p+1)/2}|\Sigma|^{-(p+1)/2} g((1/\sqrt{2})\Sigma^{-1/2}\mathrm{U}\Sigma^{-1/2}),$$

so we obtain an asymptotic expansion of the density function of $\mathbf{V}$. Let $\mathrm{T} = ((1+\delta_{ij})t_{ij}/2)$, where $t_{ij} = t_{ji}$ and δ_{ij} is the Kronecker delta. Then the characteristic function of $\mathbf{V}$ is

$$\begin{aligned} C_n(\mathrm{T}) &= \mathrm{E}[\operatorname{etr}(i\mathrm{T}\mathbf{V})] \\ &= \operatorname{etr}\left(-i\sqrt{n/2}\mathrm{T}\right)\mathrm{E}\left[\operatorname{etr}(i\sqrt{n/2}\Sigma^{-1/2}\mathrm{T}\Sigma^{-1/2}\mathbf{S})\right] \\ &= \left|\mathrm{I}_p - \sqrt{n/2}\, i\mathrm{T}\right|^{-n/2} \operatorname{etr}\left(-i\sqrt{n/2}\mathrm{T}\right). \end{aligned}$$

Therefore, for any fixed T, we can expand in the form

$$C_n(\mathrm{T}) = \hat{C}_n(\mathrm{T}) + \mathrm{o}(n^{-1}),$$

where

$$\hat{C}_n(\mathrm{T}) = \operatorname{etr}\left(-\frac{1}{2}\mathrm{T}^2\right)\left[1 + \frac{1}{3}\sqrt{\frac{2}{n}}i^3 \operatorname{tr}\mathrm{T}^3 + \frac{1}{n}\left\{\frac{1}{2}i^4 \operatorname{tr}\mathrm{T}^4 + \frac{1}{9}i^6(\operatorname{tr}\mathrm{T}^3)^2\right\}\right].$$

Inverting $\hat{C}_n(\mathrm{T})$, we obtain an asymptotic expansion of the density function $g(\mathrm{V})$ of $\mathbf{V}$, which is given by

$$g(\mathrm{V}) = (2\pi)^{-p(p+1)/2} \int \operatorname{etr}(i\mathrm{TV})\hat{C}_n(\mathrm{T})\,\mathrm{dT}.$$

The inversion can be calculated by the formula in Problem 5.9.

Next, we consider the distribution of

$$W_n = \sqrt{n}\{h(\mathbf{S}) - h(\Sigma)\}, \tag{5.2.13}$$

which possesses a valid asymptotic expansion under assumptions A3 and A4. Following the general theory discussed in Section 5.2, consider a stochastic approximation to W_n up to the order n^{-1}, which is obtained by a Taylor series for $f(\mathbf{S}) = h(\Sigma + (1/\sqrt{n})\mathbf{U})$ about $\mathrm{S} = \Sigma$,

$$\hat{W}_n = \operatorname{tr}(\mathrm{A}\mathbf{U}) + n^{-1/2}q_1(\mathbf{U}) + n^{-1}q_2(\mathbf{U}),$$

where $\mathbf{U}$ is given by (5.2.12), and

$$\begin{aligned}
\mathrm{A}: p \times p &= \left(\frac{1}{2}(1+\delta_{ab})\partial h(\mathrm{S})/\partial s_{ab}\Big|_{\mathrm{S}=\Sigma}\right),\\
q_1(\mathrm{U}) &= \frac{1}{2}\sum_{a\geq b}\sum_{c\geq d} u_{ab}u_{cd}\frac{\partial^2 h(\mathrm{S})}{\partial s_{ab}\partial s_{cd}}\Big|_{\mathrm{S}=\Sigma},\\
q_2(\mathrm{U}) &= \frac{1}{6}\sum_{a\geq b}\sum_{c\geq d}\sum_{e\geq f} u_{ab}u_{cd}u_{ef}\frac{\partial^3 h(\mathrm{S})}{\partial s_{ab}\partial s_{cd}\partial s_{ef}}\Big|_{\mathrm{S}=\Sigma}.
\end{aligned}$$

To evaluate the characteristic function of W_n, we calculate the first three cumulants of $\hat{W}_n$ by using formulas on moments of $\mathbf{U}$ in Theorem 2.2.5 and 2.2.6; that is,

$$\begin{aligned}
\kappa_1(\hat{W}_n) &= \mathrm{E}[\hat{W}_n] = n^{-1/2}g_1 + \mathrm{O}(n^{-3/2}),\\
\kappa_2(\hat{W}_n) &= \mathrm{E}[\hat{W}_n^2] - \{\mathrm{E}[\hat{W}_n]\}^2 = \sigma^2 + \mathrm{O}(n^{-1}),\\
\kappa_3(\hat{W}_n) &= \mathrm{E}[\hat{W}_n^3] - 3\,\mathrm{E}[\hat{W}_n^2]\,\mathrm{E}[\hat{W}_n] + 2\{\mathrm{E}(\hat{W}_n)\}^3\\
&= 6n^{-1/2}g_3 + \mathrm{O}(n^{-3/2}),
\end{aligned}$$

where

$$\begin{aligned}
g_1 &= \frac{1}{2}\sum_{a\geq b}\sum_{c\geq d}(\sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc})\frac{\partial^2 h(\mathrm{S})}{\partial s_{ab}\partial s_{cd}}\Big|_{\mathrm{S}=\Sigma},\\
\sigma^2 &= 2\operatorname{tr}(\mathrm{A}\Sigma)^2,\\
g_3 &= \frac{4}{3}\operatorname{tr}(\mathrm{A}\Sigma)^3 + \sum_{a\geq b}\sum_{c\geq d}(\Sigma\mathrm{A}\Sigma)_{ab}(\Sigma\mathrm{A}\Sigma)_{cd}\frac{\partial^2 h(\mathrm{S})}{\partial s_{ab}\partial s_{cd}}\Big|_{\mathrm{S}=\Sigma},
\end{aligned}$$

and $(\mathrm{A})_{ab}$ denotes the (a, b) element of a matrix A. Hence the characteristic function of W_n is obtained in the form

$$\begin{aligned} C_n(t) &= \mathrm{E}\big[\exp\big(itW_n\big)\big] \\ &= \exp\Big\{it\,\kappa_1(W_n) + \frac{1}{2}(it)^2\kappa_2(W_n) + \frac{1}{6}(it)^3\kappa_3(W_n) + \cdots\Big\} \\ &= \exp\Big(-\frac{1}{2}\sigma^2t^2\Big)\Big[1 + n^{-1/2}\{itg_1 + (it)^3g_3\} + \mathrm{o}(n^{-1/2})\Big], \end{aligned}$$

and hence the characteristic function of W_n/σ is given by

$$\exp\Big(-\frac{1}{2}t^2\Big)\Big[1 + n^{-1/2}\big\{it\,\sigma^{-1}g_1 + (it)^3\sigma^{-3}g_3\big\} + \mathrm{o}(n^{-1/2})\Big]. \tag{5.2.14}$$

Inverting (5.2.14), the density function $f(x)$ and the distribution function $F(x)$ of W_n/σ are expanded as

$$f(x) = \phi(x)\Big[1 + n^{-1/2}\{\sigma^{-1}g_1H_1(x) + \sigma^{-3}g_3H_3(x)\}\Big] + \mathrm{o}(n^{-1/2}), \tag{5.2.15}$$

$$F(x) = \Phi(x) + n^{-1/2}\phi(x)\{\sigma^{-1}g_1 + \sigma^{-3}g_3H_2(x)\} + \mathrm{o}(n^{-1/2}), \tag{5.2.16}$$

where H_j are the Hermite polynomials defined by

$$H_j(x)\phi(x) = (-1)^j(d/dx)^j\phi(x), \qquad j = 1, 2, \ldots. \tag{5.2.17}$$

5.3 T^2 Statistic

5.3.1 Outlines of Two Methods

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a p-dimensional population with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Then we consider the distribution of T^2 statistic defined by

$$T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}), \tag{5.3.1}$$

where

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_j, \quad \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})'.$$

By considering the transformation $\boldsymbol{X}_j \to \Sigma^{-1/2}(\boldsymbol{X}_j - \boldsymbol{\mu})$, without loss of generality we may consider the distribution of

$$T^2 = n\bar{\boldsymbol{X}}'\mathbf{S}^{-1}\bar{\boldsymbol{X}}, \tag{5.3.2}$$

assuming that the moments and cumulants of $\boldsymbol{X}$ are those of $\Sigma^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu})$. So $\mathrm{E}(\boldsymbol{X})=\mathbf{0}, \mathrm{Var}(\boldsymbol{X})=\mathrm{I}_p$, and the cumulants are as given in (5.2.1). Let

$$\boldsymbol{Z}=\sqrt{n}\,\bar{\boldsymbol{X}}, \quad \mathbf{V}=\sqrt{n}(\mathbf{S}-\mathrm{I}_p).$$

In this section we give outlines of two methods for deriving an asymptotic expansion of T^2.

Method I. First we expand T^2 as

$$\begin{aligned}\mathrm{T}^2 &= \boldsymbol{Z}'(\mathrm{I}_p+n^{-1/2}\mathbf{V})^{-1}\boldsymbol{Z} \\ &= \boldsymbol{Z}\boldsymbol{Z}'-\frac{1}{\sqrt{n}}\boldsymbol{Z}'\mathbf{V}\boldsymbol{Z}+\frac{1}{n}\boldsymbol{Z}'\mathbf{V}^2\boldsymbol{Z}+\mathrm{o}(n^{-1}).\end{aligned} \tag{5.3.3}$$

Then the characteristic function of T^2 can be expanded as

$$C_{T^2}(t)=C_0(t)+\frac{1}{\sqrt{n}}C_1(t)+\frac{1}{n}C_2(t)+\mathrm{o}(n^{-1}), \tag{5.3.4}$$

where

$$\begin{aligned}C_0(t) &= \mathrm{E}\left[\exp(it\boldsymbol{Z}'\boldsymbol{Z})\right], \\ C_1(t) &= \mathrm{E}\left[-it\boldsymbol{Z}'\mathbf{V}\boldsymbol{Z}\exp(it\boldsymbol{Z}'\boldsymbol{Z})\right], \\ C_2(t) &= \mathrm{E}\left[\left\{it\boldsymbol{Z}'\mathbf{V}^2\boldsymbol{Z}+\frac{1}{2}(-it\boldsymbol{Z}'\mathbf{V}\boldsymbol{Z})^2\right\}\exp(it\boldsymbol{Z}'\boldsymbol{Z})\right].\end{aligned} \tag{5.3.5}$$

After much computation, the characteristic function of T^2 can be expressed as

$$C_{T^2}(t)=\varphi^{p/2}\left[1+\frac{1}{n}\sum_{j=0}^{3}b_j\varphi^j+\mathrm{o}(n^{-1})\right], \tag{5.3.6}$$

whose inversion gives a desired result, where $\varphi=(1-2it)^{-1}$. The computation of $C_0(t)$ is carried out by using an asymptotic expansion of the distribution of $\boldsymbol{Z}$ up to the order $\mathrm{O}(n^{-1})$. Similarly, $C_1(t)$ and $C_2(t)$ can be computed by using asymptotic expansions of the joint distribution of $\boldsymbol{Z}$ and $\mathbf{V}$ up to the orders $\mathrm{O}(n^{-1/2})$ and $\mathrm{O}(1)$, respectively.

The validity of the method can be checked from the results in Section 5.2. In fact, let

$$\begin{aligned}\boldsymbol{Y} &= (X_1,\cdots,X_p,X_1^2,X_1X_2,\cdots,X_p^2)' \\ &= (Y_1,\ldots,Y_m)', \quad m=\frac{1}{2}p(p+3),\end{aligned}$$

and $\boldsymbol{Y}_j, j=1,\ldots,n$, be the corresponding ones to $\boldsymbol{X}_j$. Then we can see that T^2 is of the form

$$T^2=n\{h(\bar{\boldsymbol{Y}})-h(\boldsymbol{\nu})\}(1-n^{-1})^{1/2}, \tag{5.3.7}$$

where $\bar{\boldsymbol{Y}} = (1/n)\sum_{j=1}^{n} \boldsymbol{Y}_j, \boldsymbol{\nu} = \mathrm{E}(\boldsymbol{Y})$ and $h(\boldsymbol{y})$ is a smooth vector function in a neighborhood of $\boldsymbol{y} = \boldsymbol{\nu}$. In this case,

$$\mathrm{grad}(h(\boldsymbol{\nu})) \equiv ((\partial/\partial x_1)h(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{\nu}}, \ldots, (\partial/\partial x_m)h(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{\nu}})' = 0,$$

and we need to apply the validity condition as in Chandra and Ghosh (1979).

Method II. This method uses a decomposition of $T^2 = \boldsymbol{U}'\boldsymbol{U}$, where $\boldsymbol{U}$ is a multivariate t statistic defined by

$$\boldsymbol{U} = \sqrt{n}\,\mathbf{S}^{-1/2}\bar{\boldsymbol{X}}. \tag{5.3.8}$$

Similarly, it is seen that $\boldsymbol{U}$ is of the form

$$\boldsymbol{U} = \sqrt{n}\{\boldsymbol{h}(\bar{\boldsymbol{Y}}) - \boldsymbol{h}(\boldsymbol{\nu})\}(1-n^{-1})^{1/2}. \tag{5.3.9}$$

In this case the gradient matrix $(\partial H_i/\partial y_j)$ at $\boldsymbol{y} = \boldsymbol{\nu}$ is nonsingular. Using a multivariate version of Theorem 5.1.2, we have a valid expansion for $\boldsymbol{U}$ under appropriate regularity conditions. To obtain an asymptotic expansion up to the order $\mathrm{O}(n^{-1})$, it is sufficient to satisfy:

C1. $\mathrm{E}(\|\ \boldsymbol{Y}\ \|^4) < \infty$, [i.e., $\mathrm{E}(\|\ \boldsymbol{X}\ \|^8) < \infty$].

C2-1. $\boldsymbol{X}$ has an absolutely continuous component.

The condition C2-1 may be changed to:

C2-2. $\boldsymbol{Y}$ satisfies Cramér's condition.

The distribution of $\boldsymbol{U}$ can be expanded as

$$\begin{aligned} g(\boldsymbol{u}) &= \phi_p(\boldsymbol{u})\left\{1 + \frac{1}{\sqrt{n}}p_1(\boldsymbol{u}) + \frac{1}{n}p_2(\boldsymbol{u})\right\} + \mathrm{o}(n^{-1}) \\ &= \phi_{4,n}(\boldsymbol{u}) + \mathrm{o}(n^{-1}), \end{aligned} \tag{5.3.10}$$

where $\phi_p(\boldsymbol{u})$ is the density of $\mathrm{N}_p(0, I_p)$. The derivation is based on a perturbation expansion of $\boldsymbol{U}$ given by

$$\begin{aligned} \boldsymbol{U} &= \left(\mathrm{I}_p + \frac{1}{\sqrt{n}}\mathbf{V}\right)^{-1/2}\boldsymbol{Z} \\ &= \left(\mathrm{I}_p - \frac{1}{2\sqrt{n}}\mathbf{V} + \frac{3}{8n}\mathbf{V}^2 + \mathrm{O}_p(n^{-3/2})\right)\boldsymbol{Z}. \end{aligned} \tag{5.3.11}$$

Next the characteristic function of T^2 can be computed as

$$\begin{aligned} C_{T^2}(t) &= \mathrm{E}\{\exp(itT^2)\} \\ &= \int \exp(it\boldsymbol{u}'\boldsymbol{u})\phi_{4,n}(\boldsymbol{u})d\boldsymbol{u} + \mathrm{o}(n^{-1}), \end{aligned} \tag{5.3.12}$$

which has an expression as in (5.3.4).

5.3.2 Multivariate t-Statistic

We consider an asymptotic expansion of the distribution function of a multivariate t-statistic $\boldsymbol{U}$ in (5.3.9) with a remainder term $\mathrm{o}(n^{-1})$. Let $\nu_{i_1 \cdots i_k} = \mathrm{E}(U_{i_1} \cdots U_{i_k})$. Similarly, the corresponding cumulants of $\boldsymbol{U}$ are denoted by $\gamma_{i_1 \cdots i_k}$. Then in Section 5.2 and Bhattacharya and Ghosh (1978) it is known that the cumulants of $\boldsymbol{U}$ take the following form:

$$\begin{aligned}
\gamma_i &= \frac{1}{\sqrt{n}} q_i + \mathrm{O}(n^{-3/2}), \\
\gamma_{ij} &= \delta_{ij} + \frac{2}{n} q_{ij} + \mathrm{O}(n^{-2}), \\
\gamma_{ijk} &= \frac{6}{\sqrt{n}} q_{ijk} + \mathrm{O}(n^{-3/2}), \\
\gamma_{ijk\ell} &= \frac{24}{n} q_{ijk\ell} + \mathrm{O}(n^{-2}),
\end{aligned}$$

and in general, for $j \geq 2$,

$$\gamma_{i_1 \cdots i_j} = \mathrm{O}(n^{-(j-2)/2}).$$

The explicit expressions for q_i, q_{ij}, $\ldots$ can be computed from the definition by using the perturbation expansion of $\boldsymbol{U}$ in (5.3.11). They are expressed as

$$\begin{aligned}
q_i &= -\frac{1}{2} \sum_a \kappa_{iaa}, \\
q_{ij} &= \frac{3}{8} \sum_{a,b} \kappa_{ija}\kappa_{abb} + \frac{1}{2} \sum_{a,b} \kappa_{iab}\kappa_{jab} + \frac{1}{2}(p+1)\delta_{ij}, \\
q_{ijk} &= -\frac{1}{3}\kappa_{ijk}, \\
q_{ijkl} &= -\frac{1}{12}\kappa_{ijkl} + \frac{1}{2} \sum_a \kappa_{aij}\kappa_{akl} + \frac{1}{4}\delta_{ij}\delta_{kj}.
\end{aligned} \tag{5.3.13}$$

Theorem 5.3.1 *Under assumptions C1 and C2-1 of Section 5.3.1 the distribution of a multivariate t-statistic* $\boldsymbol{U}$ *can be expanded as*

$$\sup_B \left| \Pr\{\boldsymbol{U} \in B\} - \int_B \phi_{4,n}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \right| = \mathrm{o}(n^{-1}), \tag{5.3.14}$$

the supremum is over all measurable convex sets B *in* R^m *for C2-1 and every class* $\mathcal{B}$ *of Borel sets satisfying a boundary condition (see Theorem 5.1.2).*

The density function $\phi_{4,n}(\boldsymbol{u})$ is given by

$$\phi_{2,n}(\boldsymbol{u}) = \phi_p(\boldsymbol{u}) \left[1 + \frac{1}{\sqrt{n}} \left\{ \sum_i q_i H_i(\boldsymbol{x}) + \sum_{i,j,k} q_{ijk} H_{ijk}(\boldsymbol{x}) \right\} \right.$$
$$+ \frac{1}{n} \left\{ \sum_{i,j} (q_{ij} + \frac{1}{2} q_i q_j) H_{ij}(\boldsymbol{u}) + \sum_{i,j,k,\ell} (q_{ijk\ell} + q_i q_{jk\ell}) H_{ijk\ell}(\boldsymbol{u}) \right.$$
$$\left. \left. + \frac{1}{2} \sum_{i,j,k,\ell,e,f} q_{ijk} q_{\ell e f} H_{ijk\ell e f}(\boldsymbol{u}) \right\} \right]. \tag{5.3.15}$$

Here the coefficients q_i, q_{ij}, q_{ijk}, and $q_{ijk\ell}$ are as given in (5.3.13).

In the special case $p = 1$, the density function (5.3.14) can be simplified as

$$\phi_{4,n}(u) = \phi_1(u) \left[1 - \frac{1}{6\sqrt{n}} \kappa_3 (2u^3 - 3u) \right.$$
$$+ \frac{1}{n} \left\{ \frac{1}{18} \kappa_3^2 (u^6 - 3u^4 - 9u^2 + 3) \right.$$
$$\left. \left. - \frac{1}{12} \kappa_4 (u^4 - 6u^2 + 3) + \frac{1}{4} (u^4 - 2u^2 - 1) \right\} \right],$$

where $\kappa_3 = \mathrm{E}(Y^3)$ and $\kappa_4 = \mathrm{E}(Y^4) - 3$. This result is essentially the same as a previous result [see, e.g., Hall, 1987] for the distribution function of U.

Note that the characteristic function of $\boldsymbol{U}$ can be expressed as

$$C_{\boldsymbol{U}}(\boldsymbol{t}) = \mathrm{E} \left[\exp(i\boldsymbol{t}'\boldsymbol{Z}) \left\{ 1 + \frac{-i}{2\sqrt{n}} \boldsymbol{t}' \mathbf{V} \boldsymbol{Z} \right. \right.$$
$$\left. \left. + \frac{3i}{8n} \boldsymbol{t}' \mathbf{V}^2 \boldsymbol{Z} + \frac{1}{2} \left(\frac{-i}{2\sqrt{n}} \boldsymbol{t}' \mathbf{V} \boldsymbol{Z} \right)^2 \right\} \right] + \mathrm{o}(n^{-1}), \tag{5.3.16}$$

which is evaluated based on an Edgeworth expansion of $(\boldsymbol{U}, V)$. By evaluating each term and inverting the resulting expansion, we can obtain Theorem 5.3.1.

It may be noted that the final results depend on fourth and less than the fourth moments of $\boldsymbol{X}$. So it is expected that condition C1 can be relaxed as

C1′. $\mathrm{E}(\|\boldsymbol{X}\|^2) < \infty$ [i.e., $\mathrm{E}(\|\boldsymbol{Y}\|^4) < \infty$].

For $p = 1$, the relaxation has been proved in Hall (1979). For related discussions, see Bhattacharya and Ghosh (1988) and Babu and Bai (1993).

5.3.3 Asymptotic Expansions

It is easily seen that T^2 has a valid expansion from Theorem 5.3.1 . In fact, taking $B = B_x = \{\boldsymbol{u}; \boldsymbol{u}'\boldsymbol{u} \le x\}$ in Theorem 5.3.1, the validity follows from

$$P(T^2 \le x) = \int_{B_x} \phi_{4,n}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \mathrm{o}(n^{-1}), \tag{5.3.17}$$

uniformly for all positive real numbers x. A powerful method for getting a simplified asymptotic expansion is to expand the characteristic function in power series up to the order n^{-1} and then invert it term by term. Following method II, we can write

$$\begin{aligned} C_{T^2}(t) &= \mathrm{E}\{\exp(itT^2)\} \\ &= \int \exp(it\boldsymbol{u}'\boldsymbol{u})\phi_{4,n}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \mathrm{o}(n^{-1}). \end{aligned}$$

Considering the transformation from $\boldsymbol{u}$ to $\boldsymbol{x} = \varphi^{-1}\boldsymbol{u}$, we have

$$\begin{aligned} C_{T^2}(t) = \varphi^p \mathrm{E}_{\boldsymbol{X}} \Bigg[1 &+ \frac{1}{\sqrt{n}} \left\{ \sum_i q_i H_i(\varphi \boldsymbol{X}) + \sum_{i,j,k} q_{ijk} H_{ijk}(\varphi \boldsymbol{X}) \right\} \\ &+ \frac{1}{n} \Bigg\{ \sum_{i,j} (q_{ij} + \frac{1}{2} q_i q_j H_{ij}(\varphi \boldsymbol{X}) \\ &+ \sum_{i,j,k,\ell} (q_{ijk\ell} + q_i q_{jk\ell} H_{ijk\ell}(\varphi \boldsymbol{X}) \\ &+ \frac{1}{2} \sum_{i,j,k,l,e,f} q_{ijk} q_{\ell e f} H_{ijk\ell e f}(\varphi \boldsymbol{X}) \Bigg\} \Bigg] + \mathrm{o}(n^{-1}), \end{aligned} \tag{5.3.18}$$

where $\varphi = (1 - 2it)^{-1/2}$, and the expectation is taken under the normal random vector $\boldsymbol{X}$ whose elements are independently distributed as $\mathrm{N}(0, 1)$.

Lemma 5.3.1 *Suppose that* $\boldsymbol{X} = (X_1, \dots, X_p)'$ *and the* X_i *are independently distributed as* $\mathrm{N}(0, 1)$. *Then*

$$\begin{aligned} \mathrm{E}[H_i(\varphi \boldsymbol{X})] &= 0, \\ \mathrm{E}[H_{ijk}(\varphi \boldsymbol{X})] &= 0, \\ \mathrm{E}[H_{ij}(\varphi \boldsymbol{X})] &= (\varphi^2 - 1)\delta_{ij}, \\ \mathrm{E}[H_{ijkl}(\varphi \boldsymbol{X})] &= (\varphi^2 - 1)^2 \sum_{[3]} \delta_{ij}\delta_{kl}, \\ \mathrm{E}[H_{ijklef}(\varphi \boldsymbol{X})] &= (\varphi^2 - 1)^3 \sum_{[15]} \delta_{ij}\delta_{kl}\delta_{ef}, \end{aligned}$$

where $\sum_{[m]}$ *denotes a sum of* m *similar terms determined by suitable permutations of the indices.*

Substituting the results in Lemma 5.3.1 in (5.3.18), we have

$$C_{T^2}(t) = \varphi^p \left[1 + \frac{1}{n} \{ a_1(\varphi^2 - 1) + a_2(\varphi^2 - 1)^2 + a_3(\varphi^2 - 1)^3 \} \right] + \mathrm{o}(n^{-1}), \tag{5.3.19}$$

where

$$\begin{aligned} a_1 &= \sum_i q_{ii} + \frac{1}{2} \sum_i q_i^2 \\ &= \frac{1}{2} p(p+1) + \frac{1}{2} (\kappa_3^{(1)})^2 + \frac{1}{2} (\kappa_3^{(2)})^2, \\ a_2 &= 3 \sum_{i,j} q_{iijj} + 3 \sum_{i,j} q_i q_{ijj} \\ &= \frac{1}{4} p(p+2) + (\kappa_3^{(1)})^2 + (\kappa_3^{(2)})^2 - \frac{1}{4} \kappa_4^{(1)}, \\ a_3 &= \frac{9}{2} \sum_{i,j,k} q_{iij} q_{jkk} + 3 \sum_{i,j,k} (q_{ijk})^2 \\ &= \frac{1}{3} (\kappa_3^{(1)})^2 + \frac{1}{2} (\kappa_3^{(2)})^2. \end{aligned}$$

Inverting the characteristic function above formally, we obtain the following theorem.

Theorem 5.3.2 *Under the conditions C1 and C2-1 (or C2-2) the distribution function of* T^2 *can be expanded as*

$$\begin{aligned} \Pr\{T^2 \le x\} &= G_p(x) + \frac{1}{n} \sum_{j=0}^{3} b_j G_{p+2j}(x) + \mathrm{o}(n^{-1}) \\ &= G_p(x) - \frac{2x}{np} g_p(x) \left\{ b_1 + b_2 + b_3 + \frac{(b_2 + b_3)x}{p+2} + \frac{b_3 x^2}{(p+2)(p+4)} \right\} + \mathrm{o}(n^{-1}), \end{aligned} \tag{5.3.20}$$

uniformly for all positive real numbers x, *where* $G_p(x)$ *and* $g_p(x)$ *are the distribution and density functions of a chi-square variate with* p *degrees of* *freedom, respectively. The coefficients b_j are given by*

$$\begin{aligned}
b_0 &= -\frac{1}{4}p^2 + \frac{1}{6}(\kappa_3^{(1)})^2 - \frac{1}{4}\kappa_4^{(1)}, \\
b_1 &= -\frac{1}{2}p - \frac{1}{2}(\kappa_3^{(1)})^2 + \frac{1}{2}\kappa_4^{(1)}, \\
b_2 &= \frac{1}{4}p(p+2) - \frac{1}{2}(\kappa_3^{(2)})^2 - \frac{1}{4}\kappa_4^{(1)}, \\
b_3 &= \frac{1}{3}(\kappa_3^{(1)})^2 + \frac{1}{2}(\kappa_3^{(2)})^2.
\end{aligned}$$

The result was obtained independently by Kano (1995) and Fujikoshi (1997a) based on Methods I and II, respectively.

In the special case $p = 1$, the coefficients b_i are given by

$$\begin{aligned}
b_0 &= \frac{1}{6}\kappa_3^2 - \frac{1}{4}\kappa_4 - \frac{1}{4}, \\
b_1 &= -\frac{1}{2}\kappa_3^2 + \frac{1}{2}\kappa_4, \\
b_2 &= -\frac{1}{2}\kappa_3^2 - \frac{1}{4}\kappa_4 + \frac{3}{4}, \\
b_3 &= \frac{5}{6}\kappa_3^2,
\end{aligned}$$

where $\kappa_3 = \mathrm{E}(Y^3)$ and $\kappa_4 = \mathrm{E}(Y^4) - 3$.

If $\boldsymbol{Y}$ has an elliptical distribution $E_p(\mu, \Lambda)$ with the characteristic function $\exp(i\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'\Lambda\boldsymbol{t})$, the coefficients are expressed as

$$\begin{aligned}
b_0 &= -\frac{1}{4}p\{p + (p+2)\kappa\}, \\
b_1 &= -\frac{1}{2}p\{1 - (p+2)\kappa\}, \\
b_2 &= \frac{1}{4}p(p+2)(1-\kappa), \\
b_3 &= 0,
\end{aligned}$$

where κ is the kurtosis parameter defined by $\kappa = \psi^{(2)}(0)/(\psi'(0))^2 - 1$. This result was obtained by Iwashita (1997).

5.4 Statistics with a Class of Moments

5.4.1 Large-Sample Expansions

Consider the Lambda distribution defined by

$$\Lambda = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_h|} \sim \Lambda_p(q, n), \tag{5.4.1}$$

where $\mathbf{S}_h$ and $\mathbf{S}_e$ are independently distributed as $\mathrm{W}_p(q, \Sigma)$ and $\mathrm{W}_p(n, \Sigma)$, respectively. The Λ appears, for example, as a likelihood ratio test for testing the equality of mean vectors $\boldsymbol{\mu}_i$, $i = 1, \ldots, q+1$, based on an N_i sample from $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. In this case, $n = N - (q+1)$ and $N = N_1 + \cdots + N_{q+1}$. The likelihood ratio criterion is $\lambda = |\Lambda|^{N/2}$. From Theorem 3.3.3 the distribution of Λ is the same as

$$\Lambda = \beta_1 \cdots \beta_p, \tag{5.4.2}$$

where $\beta_1, \ldots, \beta_p$ are independent and $\beta_j \sim \mathrm{Be}(\frac{1}{2}(n-j+1), \frac{1}{2}q)$, and hence the hth moment of Λ is given by

$$\begin{aligned} \mathrm{E}[\Lambda^h] &= \prod_{j=1}^{p} \frac{\Gamma[\frac{1}{2}(n-j+1)+h]\Gamma[\frac{1}{2}(n+q-j+1)]}{\Gamma[\frac{1}{2}(n-j+1)]\Gamma[\frac{1}{2}(n+q-j+1)+h]} \\ &= K \prod_{j=1}^{p} \frac{\Gamma[\frac{1}{2}(n-j+1)+h]}{\Gamma[\frac{1}{2}(n+q-j+1)+h]}, \end{aligned} \tag{5.4.3}$$

where $K = \prod_{j=1}^{p} \Gamma[\frac{1}{2}(n+q-j+1)]/\Gamma[\frac{1}{2}(n-j+1)]$. Related to the distribution of $\lambda = \Lambda^{N/2}$, we consider a statistic W $(0 \le W \le 1)$ which has moments of the form

$$\mathrm{E}(W^h) = K \left(\frac{\prod_{k=1}^{b} y_k^{y_k}}{\prod_{j=1}^{a} x_j^{x_j}} \right)^h \frac{\prod_{j=1}^{a} \Gamma[x_j(1+h)+\xi_j]}{\prod_{k=1}^{b} \Gamma[y_k(1+h)+\eta_k]}, \quad h = 0, 1, \ldots, \tag{5.4.4}$$

where K is a constant such that $\mathrm{E}(W^0) = 1$ and

$$\sum_{j=1}^{a} x_j = \sum_{k=1}^{b} y_k. \tag{5.4.5}$$

The moments are expressed as

$$\mathrm{E}(W^h) = \left(\frac{\prod_{k=1}^{b} y_k^{y_k}}{\prod_{j=1}^{a} x_j^{x_j}} \right)^h \frac{\prod_{j=1}^{a} \Gamma[x_j(1+h)+\xi_j] \prod_{k=1}^{b} \Gamma[y_k+\eta_k]}{\prod_{j=1}^{a} \Gamma[x_j+\xi_j] \prod_{k=1}^{b} \Gamma[y_k(1+h)+\eta_k]}. \tag{5.4.6}$$

Let

$$V = -2\rho \log W, \tag{5.4.7}$$

where ρ is a constant such that $0 < \rho < 1$ and $\alpha_j = (1-\rho)x_j$, $\beta_k = (1-\rho)y_k$. Assume that:

C0: α_j and β_k have limits as x_j and y_k tend to infinity.

In most applications,

$$\text{C1: } x_j = c_j N, \quad i = 1, \ldots, a; \quad y_k = d_k N, \quad k = 1, \ldots, b, \tag{5.4.8}$$

where c_j and d_k are nonzero constants such that $\sum_{j=1}^{a} c_j = \sum_{k=1}^{b} d_k$, and N is a sample size or a number growing with the sample size.

Let $C(t)$ be the characteristic function of V. Then the cumulant generating function $\log C(t)$ is

$$\log C(t) = g(t) - g(0), \tag{5.4.9}$$

where

$$\begin{aligned} g(t) = (2it)\left(\sum_{j=1}^{a} x_j \log x_j - \sum_{k=1}^{b} y_k \log y_k\right) & \\ + \sum_{j=1}^{a} \log \Gamma[\rho x_j(1-2it) + \alpha_j + \xi_j] & \\ - \sum_{k=1}^{b} \log \Gamma[\rho y_k(1-2it) + \beta_k + \eta_k]. & \end{aligned}$$

To obtain an expansion for $\log C(t)$, we use the following generalized version of Stirling's formula for the gamma function, which is due to Barnes (1899):

$$\begin{aligned} \log \Gamma(z+h) = \log \sqrt{2\pi} + \left(z + h - \frac{1}{2}\right) \log z - z & \\ - \sum_{r=1}^{m} (-1)^r \frac{B_{r+1}}{r(r+1)z^r} + R_{m+1}(z), & \end{aligned} \tag{5.4.10}$$

where $R_{m+1}(z) = \mathrm{O}\big(z^{-(m+1)}\big)$ and $B_r(h)$ is the Bernoulli polynomial of degree r defined by

$$\frac{\tau e^{h\tau}}{e^{\tau}} = \sum_{r=0}^{\infty} \frac{\tau^r}{r!} B_r(h). \tag{5.4.11}$$

Some of them are

$$\begin{aligned} & B_0(h) = 1, \quad B_1(h) = h - \frac{1}{2}, \quad B_2(h) = h^2 - h + \frac{1}{6}, \\ & B_3(h) = h^3 - \frac{3}{2}h^2 + \frac{1}{2}h, \quad B_4(h) = h^4 - 2h^3 + h^2 - \frac{1}{30}, \\ & B_5(h) = h^5 - \frac{5}{2}h^4 + \frac{5}{3}h^3 - \frac{1}{6}h. \end{aligned} \tag{5.4.12}$$

Expanding each of the gamma functions in (5.4.10) by taking $z = \rho x_j(1 - 2it), \rho y_k(1-2it)$, and $h = \alpha_j + \xi_j, \beta_k + \eta_k$, we obtain

$$\log C(t) = -\frac{f}{2} \log(1-2it) + \sum_{\ell=1}^{r} \omega_r \left\{(1-2it)^{-\ell} - 1\right\} + R_{r+1}, \tag{5.4.13}$$