CHAPTER 7

The Distribution of the Sample Covariance Matrix and the Sample Generalized Variance

7.1. INTRODUCTION

The sample covariance matrix, $S = [1/(N-1)]\Sigma_{\alpha=1}^{n}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'$, is an unbiased estimator of the population covariance matrix $\boldsymbol{\Sigma}$. In Section 4.2 we found the density of $\boldsymbol{A} = (N-1)\boldsymbol{S}$ in the case of a 2×2 matrix. In Section 7.2 this result will be generalized to the case of a matrix $\boldsymbol{A}$ of any order. When $\boldsymbol{\Sigma} = \boldsymbol{I}$, this distribution is in a sense a generalization of the χ^2-distribution. The distribution of $\boldsymbol{A}$ (or $\boldsymbol{S}$), often called the Wishart distribution, is fundamental to multivariate statistical analysis. In Sections 7.3 and 7.4 we discuss some properties of the Wishart distribution.

The generalized variance of the sample is defined as $|\boldsymbol{S}|$ in Section 7.5; it is a measure of the scatter of the sample. Its distribution is characterized. The density of the set of all correlation coefficients when the components of the observed vector are independent is obtained in Section 7.6.

The inverted Wishart distribution is introduced in Section 7.7 and is used as an a priori distribution of $\boldsymbol{\Sigma}$ to obtain a Bayes estimator of the covariance matrix. In Section 7.8 we consider improving on $\boldsymbol{S}$ as an estimator of $\boldsymbol{\Sigma}$ with respect to two loss functions. Section 7.9 treats the distributions for sampling from elliptically contoured distributions.

An Introduction to Multivariate Statistical Analysis, Third Edition. By T. W. Anderson
ISBN 0-471-36091-0

7.2. THE WISHART DISTRIBUTION

We shall obtain the distribution of $A = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'$, where $X_1, \ldots, X_N$ $(N > p)$ are independent, each with the distribution $N(\mu, \Sigma)$. As was shown in Section 3.3, A is distributed as $\sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$, where $n = N - 1$ and $Z_1, \ldots, Z_n$ are independent, each with the distribution $N(\mathbf{0}, \Sigma)$. We shall show that the density of A for A positive definite is

$$(1) \qquad \frac{|A|^{\frac{1}{2}(n-p-1)} \exp\left(-\frac{1}{2}\operatorname{tr} \Sigma^{-1} A\right)}{2^{\frac{1}{2}np} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(n+1-i)\right]}.$$

We shall first consider the case of $\Sigma = I$. Let

$$(2) \qquad (Z_1, \ldots, Z_n) = \begin{pmatrix} v_1' \\ \vdots \\ v_p' \end{pmatrix}.$$

Then the elements of $A = (a_{ij})$ are inner products of these n-component vectors, $a_{ij} = v_i' v_j$. The vectors $v_1, \ldots, v_p$ are independently distributed, each according to $N(\mathbf{0}, I_n)$. It will be convenient to transform to new coordinates according to the Gram–Schmidt orthogonalization. Let $w_1 = v_1$,

$$(3) \qquad w_i = v_i - \sum_{j=1}^{i-1} w_j \frac{w_j' v_i}{w_j' w_j}, \qquad i = 2, \ldots, p.$$

We prove by induction that w_k is orthogonal to w_i, $k < i$. Assume $w_k' w_h = 0$, $k \neq h$, $k, h = 1, \ldots, i-1$; then take the inner product of w_k and (3) to obtain $w_k' w_i = 0$, $k = 1, \ldots, i-1$. (Note that $\Pr\{\|w_i\| = 0\} = 0$.)

Define $t_{ii} = \|w_i\| = \sqrt{w_i' w_i}$, $i = 1, \ldots, p$, and $t_{ij} = v_i' w_j / \|w_j\|$, $j = 1, \ldots, i-1$, $i = 2, \ldots, p$. Since $v_i = \sum_{j=1}^{i} (t_{ij}/\|w_j\|) w_j$,

$$(4) \qquad a_{hi} = v_h' v_i = \sum_{j=1}^{\min(h,i)} t_{hj} t_{ij}.$$

If we define the lower triangular matrix $T = (t_{ij})$ with $t_{ii} > 0$, $i = 1, \ldots, p$, and $t_{ij} = 0$, $i < j$, then

$$(5) \qquad A = TT'.$$

Note that t_{ij}, $j = 1, \ldots, i-1$, are the first $i-1$ coordinates of v_i in the coordinate system with $w_1, \ldots, w_{i-1}$ as the first $i-1$ coordinate axes. (See Figure 7.1.) The sum of the other $n - i + 1$ coordinates squared is $\|v_i\|^2 - \sum_{j=1}^{i-1} t_{ij}^2 = t_{ii}^2 = \|w_i\|^2$; w_i is the vector from v_i to its projection on $w_1, \ldots, w_{i-1}$ (or equivalently on $v_1, \ldots, v_{i-1}$).

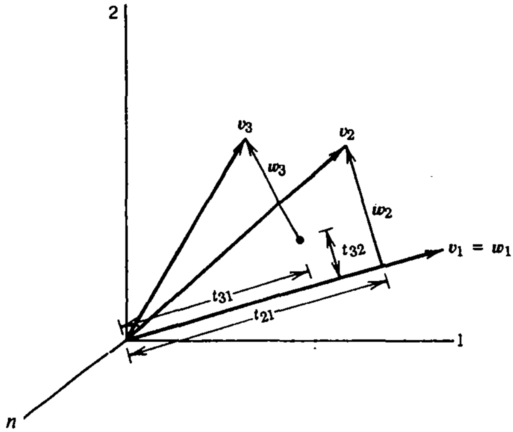

Figure 7.1. Transformation of coordinates.

Lemma 7.2.1. *Conditional on $w_1, \dots, w_{i-1}$ (or equivalently on $v_1, \dots, v_{i-1}$), $t_{i1}, \dots, t_{i,i-1}$ and t_{ii}^2 are independently distributed; t_{ij} is distributed according to $N(0,1)$, $i > j$; and t_{ii}^2 has the χ^2-distribution with $n-i+1$ degrees of freedom.*

Proof. The coordinates of v_i referred to the new orthogonal coordinates with $v_1, \dots, v_{i-1}$ defining the first coordinate axes are independently normally distributed with means 0 and variances 1 (Theorem 3.3.1). t_{ii}^2 is the sum of the coordinates squared omitting the first $i-1$. ■

Since the conditional distribution of $t_{i1}, \dots, t_{ii}$ does not depend on $v_1, \dots, v_{i-1}$, they are distributed independently of $t_{11}, t_{21}, t_{22}, \dots, t_{i-1,i-1}$.

Corollary 7.2.1. *Let $Z_1, \dots, Z_n$ $(n \ge p)$ be independently distributed, each according to $N(\mathbf{0}, \mathbf{I})$; let $A = \sum_{\alpha=1}^n Z_\alpha Z_\alpha' = TT'$, where $t_{ij} = 0$, $i < j$, and $t_{ii} > 0$, $i = 1, \dots, p$. Then $t_{11}, t_{21}, \dots, t_{pp}$ are independently distributed; t_{ij} is distributed according to $N(0,1)$, $i > j$; and t_{ii}^2 has the χ^2-distribution with $n-i+1$ degrees of freedom.*

Since t_{ii} has density $2^{-\frac{1}{2}(n-i-1)} t^{n-i} e^{-\frac{1}{2}t^2} / \Gamma[\frac{1}{2}(n+1-i)]$, the joint density of t_{ij}, $j = 1, \dots, i$, $i = 1, \dots, p$, is

$$(6) \qquad \prod_{i=1}^{p} \frac{t_{ii}^{n-i} \exp\left(-\frac{1}{2}\sum_{j=1}^{i} t_{ij}^2\right)}{\pi^{\frac{1}{2}(i-1)} 2^{\frac{1}{2}n-1} \Gamma\left[\frac{1}{2}(n+1-i)\right]}$$

$$= \frac{\prod_{i=1}^{p} t_{ii}^{n-i} \exp\left(-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{i} t_{ij}^2\right)}{2^{\frac{1}{2}p(n-2)} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(n+1-i)\right]}.$$

Let C be a lower triangular matrix ($c_{ij}=0$, $i<j$) such that $\Sigma = CC'$ and $c_{ii}>0$. The linear transformation $T^* = CT$, that is,

$$(7)\qquad t_{ij}^* = \sum_{k=j}^{i} c_{ik} t_{kj}, \qquad i \ge j,$$
$$= 0, \qquad i<j,$$

can be written

$$(8)\qquad \begin{bmatrix} t_{11}^* \\ t_{21}^* \\ t_{22}^* \\ \vdots \\ t_{p1}^* \\ \vdots \\ t_{pp}^* \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ x & c_{22} & 0 & \cdots & 0 & \cdots & 0 \\ x & x & c_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ x & x & x & \cdots & c_{pp} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ x & x & x & \cdots & x & \cdots & c_{pp} \end{bmatrix} \begin{bmatrix} t_{11} \\ t_{21} \\ t_{22} \\ \vdots \\ t_{p1} \\ \vdots \\ t_{pp} \end{bmatrix},$$

where x denotes an element, possibly nonzero. Since the matrix of the transformation is triangular, its determinant is the product of the diagonal elements, namely, $\prod_{i=1}^p c_{ii}^i$. The Jacobian of the transformation from T to T^* is the reciprocal of the determinant. The density of T^* is obtained by substituting into (6) $t_{ii} = t_{ii}^*/c_{ii}$ and

$$(9)\qquad \sum_{i=1}^{p} \sum_{j=1}^{i} t_{ij}^2 = \operatorname{tr} TT'$$
$$= \operatorname{tr} C^{-1} T^* T^{*\prime} (C^{-1})'$$
$$= \operatorname{tr} T^* T^{*\prime} C'^{-1} C^{-1}$$
$$= \operatorname{tr} T^* T^{*\prime} \Sigma^{-1} = \operatorname{tr} T^{*\prime} \Sigma^{-1} T^*,$$

and using $\prod_{i=1}^p c_{ii}^2 = |C||C'| = |\Sigma|$.

Theorem 7.2.1. *Let $Z_1, \dots, Z_n$ $(n \ge p)$ be independently distributed, each according to $N(\mathbf{0}, \Sigma)$; let $A = \sum_{\alpha=1}^n Z_\alpha Z_\alpha'$; and let $A = T^* T^{*\prime}$, where $t_{ij}^* = 0$, $i<j$, and $t_{ii}^* > 0$. Then the density of T^* is*

$$(10)\qquad \frac{\prod_{i=1}^p t_{ii}^{*\,n-i} e^{-\frac{1}{2}\operatorname{tr}\Sigma^{-1} T^* T^{*\prime}}}{2^{\frac{1}{2}p(n-2)} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]}.$$

We can write (4) as $a_{hi} = \sum_{j=1}^{i} t_{hj}^* t_{ij}^*$ for $h \geq i$. Then

$$(11) \qquad \frac{\partial a_{hi}}{\partial t_{kl}^*} = 0, \qquad k > h,$$

$$= 0, \qquad k = h, \quad l > i;$$

that is, $\partial a_{hi}/\partial t_{kl}^* = 0$ if k, l is beyond h, i in the lexicographic ordering. The Jacobian of the transformation from A to T^* is the determinant of the lower triangular matrix with diagonal elements

$$(12) \qquad \frac{\partial a_{hh}}{\partial t_{hh}^*} = 2t_{hh}^*,$$

$$(13) \qquad \frac{\partial a_{hi}}{\partial t_{hi}^*} = t_{ii}^*, \qquad h > i,$$

The Jacobian is therefore $2^p \prod_{i=1}^{p} t_{ii}^{*\,p+1-i}$. The Jacobian of the transformation from T^* to A is the reciprocal.

Theorem 7.2.2. *Let $Z_1, \dots, Z_n$ be independently distributed, each according to $N(\mathbf{0}, \Sigma)$. The density of $A = \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$ is*

$$(14) \qquad \frac{|A|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}A}}{2^{\frac{1}{2}pn} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(n+1-i)\right]}$$

for A positive definite, and 0 *otherwise.*

Corollary 7.2.2. *Let $X_1, \dots, X_N$ $(N > p)$ be independently distributed, each according to $N(\boldsymbol{\mu}, \Sigma)$. Then the density of $A = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ is* (14) *for $n = N - 1$.*

The density (14) will be denoted by $w(A|\Sigma, n)$, and the associated distribution will be termed $W(\Sigma, n)$. If $n < p$, then A does not have a density, but its distribution is nevertheless defined, and we shall refer to it as $W(\Sigma, n)$.

Corollary 7.2.3. *Let $X_1, \dots, X_N$ $(N > p)$ be independently distributed, each according to $N(\boldsymbol{\mu}, \Sigma)$. The distribution of $S = (1/n)\sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ is $W[(1/n)\Sigma, n]$, where $n = N - 1$.*

Proof. S has the distribution of $\sum_{\alpha=1}^{n}[(1/\sqrt{n})Z_\alpha][(1/\sqrt{n})Z_\alpha]'$, where $(1/\sqrt{n})Z_1, \dots, (1/\sqrt{n})Z_N$ are independently distributed, each according to $N(\mathbf{0}, (1/n)\Sigma)$. Theorem 7.2.2 implies this corollary. ■

The Wishart distribution for $p = 2$ as given in Section 4.2.1 was derived by Fisher (1915). The distribution for arbitrary p was obtained by Wishart (1928) by a geometric argument using $v_1, \dots, v_p$ defined above. As noted in Section 3.2, the ith diagonal element of A is the squared length of the ith vector, $a_{ii} = v_i'v_i = \|v_i\|^2$, and the i,jth off-diagonal element of A is the product of the lengths of v_i and v_j and the cosine of the angle between them. The matrix A specifies the lengths and configuration of the vectors.

We shall give a geometric interpretation† of the derivation of the density of the rectangular coordinates t_{ij}, $i \ge j$, when $\Sigma = I$. The probability element of t_{11} is approximately the probability that $\|v_1\|$ lies in the interval $t_{11} < \|v_1\| < t_{11} + dt_{11}$. This is the probability that v_1 falls in a spherical shell in n dimensions with inner radius t_{11} and thickness dt_{11}. In this region, the density $(2\pi)^{-\frac{1}{2}n} \exp(-\frac{1}{2}v_1'v_1)$ is approximately constant, namely, $(2\pi)^{-\frac{1}{2}n}\exp(-\frac{1}{2}t_{11}^2)$. The surface area of the unit sphere in n dimensions is $C(n) = 2\pi^{\frac{1}{2}n}/\Gamma(\frac{1}{2}n)$ (Problems 7.1–7.3), and the volume of the spherical shell is approximately $C(n)t_{11}^{n-1}\,dt_{11}$. The probability element is the product of the volume and approximate density, namely,

$$2^{-(\frac{1}{2}n-1)} t_{11}^{n-1} \exp\left(-\tfrac{1}{2}t_{11}^2\right) dt_{11}/\Gamma\left(\tfrac{1}{2}n\right). \tag{15}$$

The probability element of $t_{i1}, \dots, t_{i,i-1}, t_{ii}$ given $v_1, \dots, v_{i-1}$ (i.e., given $w_1, \dots, w_{i-1}$) is approximately the probability that v_i falls in the region for which $t_{i1} < v_i'w_1/\|w_1\| < t_{i1} + dt_{i1}, \dots, t_{i,i-1} < v_i'w_{i-1}/\|w_{i-1}\| < t_{i,i-1} + dt_{i,i-1}$, and $t_{ii} < \|w_i\| < t_{ii} - dt_{ii}$, where w_i is the projection of v_i on the $(n-i+1)$-dimensional space orthogonal to $w_1, \dots, w_{i-1}$. Each of the first $i-1$ pairs of inequalities defines the region between two hyperplanes (the different pairs being orthogonal). The last pair of inequalities defines a cylindrical shell whose intersection with the $(i-1)$-dimensional hyperplane spanned by $v_1, \dots, v_{i-1}$ is a spherical shell in $n-i+1$ dimensions with inner radius t_{ii}. In this region the density $(2\pi)^{-\frac{1}{2}n}\exp(-\frac{1}{2}v_i'v_i)$ is approximately constant, namely, $(2\pi)^{-\frac{1}{2}n}\exp(-\frac{1}{2}\sum_{j=1}^{i} t_{ij}^2)$. The volume of the region is approximately $dt_{i1} \cdots dt_{i,i-1}\, C(n-i+1)t_{ii}^{n-i}\,dt_{ii}$. The probability element is

$$\frac{2^{-(\frac{1}{2}n-1)}\pi^{-\frac{1}{2}(i-1)} t_{ii}^{n-i} \exp\left(-\frac{1}{2}\sum_{j=1}^{i} t_{ij}^2\right)}{\Gamma\left[\frac{1}{2}(n+1-i)\right]} dt_{i1} \cdots dt_{ii}. \tag{16}$$

Then the product of (15) and (16) for $i = 2, \dots, p$ is (6) times $dt_{11} \cdots dt_{pp}$.

This analysis, which exactly parallels the geometric derivation by Wishart [and later by Mahalanobis, Bose, and Roy (1937)], was given by Sverdrup

†In the first edition of this book, the derivation of the Wishart distribution and its geometric interpretation were in terms of the nonorthogonal vectors $v_1, \dots, v_p$.

(1947) [and by Fog (1948) for $p = 3$]. Another method was used by Madow (1938), who drew on the distribution of correlation coefficients (for $\Sigma = I$) obtained by Hotelling by considering certain partial correlation coefficients. Hsu (1939b) gave an inductive proof, and Rasch (1948) gave a method involving the use of a functional equation. A different method is to obtain the characteristic function and invert it, as was done by Ingham (1933) and by Wishart and Bartlett (1933).

Cramér (1946) verified that the Wishart distribution has the characteristic function of A. By means of alternative matrix transformations Elfving (1947), Mauldon (1955), and Olkin and Roy (1954) derived the Wishart distribution via the *Bartlett decomposition*; Kshirsagar (1959) based his derivation on random orthogonal transformations. Narain (1948), (1950) and Ogawa (1953) used a regression approach. James (1954), Khatri and Ramachandran (1958), and Khatri (1963) applied different methods. Giri (1977) used invariance. Wishart (1948) surveyed the derivations up to that date. Some of these methods are indicated in the problems.

The relation $A = TT'$ is known as the *Bartlett decomposition* [Bartlett (1939)], and the (nonzero) elements of T were termed *rectangular coordinates* by Mahalanobis, Bose, and Roy (1937).

Corollary 7.2.4

$$\int_{B>0}\cdots\int |B|^{t-\frac{1}{2}(p+1)} e^{-\operatorname{tr} B}\, dB = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left[t - \tfrac{1}{2}(i-1)\right]. \tag{17}$$

Proof. Here $B > \mathbf{0}$ denotes B positive definite. Since (14) is a density, its integral for $A > \mathbf{0}$ is 1. Let $\Sigma = I$, $A = 2B$ ($dA = 2\, dB$), and $n = 2t$. Then the fact that the integral is 1 is identical to (17) for t a half integer. However, if we derive (14) from (6), we can let n be any real number greater than $p - 1$. In fact (17) holds for complex t such that $\mathscr{R}t > p - 1$. ($\mathscr{R}t$ means the real part of t.) ■

Definition 7.2.1. *The multivariate gamma function is*

$$\Gamma_p(t) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left[t - \tfrac{1}{2}(i-1)\right]. \tag{18}$$

The Wishart density can be written

$$w(A|\Sigma, n) = \frac{|A|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\operatorname{tr} \Sigma^{-1} A}}{2^{\frac{1}{2}pn} |\Sigma|^{\frac{1}{2}n} \Gamma_p\left(\frac{1}{2}n\right)}. \tag{19}$$

7.3. SOME PROPERTIES OF THE WISHART DISTRIBUTION

7.3.1. The Characteristic Function

The characteristic function of the Wishart distribution can be obtained directly from the distribution of the observations. Suppose $Z_1, \dots, Z_n$ are distributed independently, each with density

$$(1) \qquad \frac{1}{(2\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}z'\Sigma^{-1}z\right).$$

Let

$$(2) \qquad A = \sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha .$$

Introduce the $p \times p$ matrix $\Theta = (\theta_{ij})$ with $\theta_{ij} = \theta_{ji}$. The characteristic function of $A_{11}, A_{22}, \dots, A_{pp}, 2A_{12} 2A_{13}, \dots, 2A_{p-1,p}$ is

$$(3) \qquad \begin{aligned} \mathscr{E} \exp[i \operatorname{tr}(A\Theta)] &= \mathscr{E} \exp\left(i \operatorname{tr} \sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha \Theta\right) \\ &= \mathscr{E} \exp\left(i \operatorname{tr} \sum_{\alpha=1}^{n} Z'_\alpha \Theta Z_\alpha\right) \\ &= \mathscr{E} \exp\left(i \sum_{\alpha=1}^{n} Z'_\alpha \Theta Z_\alpha\right). \end{aligned}$$

It follows from Lemma 2.6.1 that

$$(4) \qquad \mathscr{E} \exp\left(i \sum_{\alpha=1}^{n} Z'_\alpha \Theta Z_\alpha\right) = \prod_{\alpha=1}^{n} \mathscr{E} \exp(iZ'_\alpha \Theta Z_\alpha) = [\mathscr{E} \exp(iZ'\Theta Z)]^n,$$

where Z has the density (1). For Θ real, there is a real nonsingular matrix B such that

$$(5) \qquad B'\Sigma^{-1}B = I,$$

$$(6) \qquad B'\Theta B = D,$$

where D is a real diagonal matrix (Theorem A.1.2 of the Appendix). If we let $z = By$, then

$$(7) \qquad \begin{aligned} \mathscr{E} \exp(iZ'\Theta Z) &= \mathscr{E} \exp(iY'DY) \\ &= \mathscr{E} \prod_{j=1}^{p} \exp\left(id_{jj}Y_j^2\right) \\ &= \prod_{j=1}^{p} \mathscr{E} \exp\left(id_{jj}Y_j^2\right) \end{aligned}$$

by Lemma 2.6.2. The jth factor in the second product is $\mathscr{E}\exp(id_{jj}Y_j^2)$, where Y_j has the distribution $N(0,1)$; this is the characteristic function of the χ^2-distribution with one degree of freedom, namely $(1-2id_{jj})^{-\frac{1}{2}}$ [as can be proved by expanding $\exp(id_{jj}y_j^2)$ in a power series and integrating term by term]. Thus

$$\mathscr{E}\exp(i\boldsymbol{Z}'\boldsymbol{\Theta}\boldsymbol{Z}) = \prod_{j=1}^{p}(1-2id_{jj})^{-\frac{1}{2}} = |\boldsymbol{I}-2i\boldsymbol{D}|^{-\frac{1}{2}} \tag{8}$$

since $\boldsymbol{I}-2i\boldsymbol{D}$ is a diagonal matrix. From (5) and (6) we see that

$$\begin{aligned} |\boldsymbol{I}-2i\boldsymbol{D}| &= |\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{B}-2i\boldsymbol{B}'\boldsymbol{\Theta}\boldsymbol{B}| \\ &= |\boldsymbol{B}'(\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{\Theta})\boldsymbol{B}| \\ &= |\boldsymbol{B}'|\cdot|\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{\Theta}|\cdot|\boldsymbol{B}| \\ &= |\boldsymbol{B}|^2\cdot|\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{\Theta}|, \end{aligned} \tag{9}$$

$|\boldsymbol{B}'|\cdot|\boldsymbol{\Sigma}^{-1}|\cdot|\boldsymbol{B}| = |\boldsymbol{I}| = 1$, and $|\boldsymbol{B}|^2 = 1/|\boldsymbol{\Sigma}^{-1}|$. Combining the above results, we obtain

$$\mathscr{E}\exp[i\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Theta})] = \frac{|\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}n}}{|\boldsymbol{\Sigma}^{-1}-2i\boldsymbol{\Theta}|^{\frac{1}{2}n}} = |\boldsymbol{I}-2i\boldsymbol{\Theta}\boldsymbol{\Sigma}|^{-\frac{1}{2}n}. \tag{10}$$

It can be shown that the result is valid provided $(\mathscr{R}(\sigma^{ik}-2i\theta_{jk}))$ is positive definite. In particular, it is true for all real $\boldsymbol{\Theta}$. It also holds for $\boldsymbol{\Sigma}$ singular.

Theorem 7.3.1. *If $\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n$ are independent, each with distribution $N(\boldsymbol{0},\boldsymbol{\Sigma})$, then the characteristic function of $A_{11},\dots,A_{pp},2A_{12},\dots,2A_{p-1,p}$, where $(A_{ij})=\boldsymbol{A}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$, is given by* (10).

7.3.2. The Sum of Wishart Matrices

Suppose the $\boldsymbol{A}_i$, $i=1,2$, are distributed independently according to $W(\boldsymbol{\Sigma},n_i)$, respectively. Then $\boldsymbol{A}_1$ is distributed as $\sum_{\alpha=1}^{n_1}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$, and $\boldsymbol{A}_2$ is distributed as $\sum_{\alpha=n_1+1}^{n_1+n_2}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$, where $\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_{n_1+n_2}$ are independent, each with distribution $N(\boldsymbol{0},\boldsymbol{\Sigma})$. Then $\boldsymbol{A}=\boldsymbol{A}_1+\boldsymbol{A}_2$ is distributed as $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$, where $n=n_1+n_2$. Thus $\boldsymbol{A}$ is distributed according to $W(\boldsymbol{\Sigma},n)$. Similarly, the sum of q matrices distributed independently, each according to a Wishart distribution with covariance $\boldsymbol{\Sigma}$, has a Wishart distribution with covariance matrix $\boldsymbol{\Sigma}$ and number of degrees of freedom equal to the sum of the numbers of degrees of freedom of the component matrices.

Theorem 7.3.2. *If $A_1, \dots, A_q$ are independently distributed with A_i distributed according to $W(\Sigma, n_i)$, then*

$$A = \sum_{i=1}^{q} A_i \tag{11}$$

is distributed according to $W(\Sigma, \sum_{i=1}^{q} n_i)$.

7.3.3. A Certain Linear Transformation

We shall frequently make the transformation

$$A = CBC', \tag{12}$$

where C is a nonsingular $p \times p$ matrix. If A is distributed according to $W(\Sigma, n)$, then B is distributed according to $W(\Phi, n)$ where

$$\Phi = C^{-1} \Sigma C'^{-1}. \tag{13}$$

This is proved by the following argument: Let $A = \sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha$, where $Z_1, \dots, Z_n$ are independently distributed, each according to $N(\mathbf{0}, \Sigma)$. Then $Y_\alpha = C^{-1} Z_\alpha$ is distributed according to $N(\mathbf{0}, \Phi)$. However,

$$B = \sum_{\alpha=1}^{n} Y_\alpha Y'_\alpha = C^{-1} \sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha C'^{-1} = C^{-1} A C'^{-1} \tag{14}$$

is distributed according to $W(\Phi, n)$. Finally, $|\partial(A)/\partial(B)|$, the Jacobian of the transformation (12), is

$$\left| \frac{\partial(A)}{\partial(B)} \right| = \frac{w(B, \Phi, n)}{w(A, \Sigma, n)} = \frac{|B|^{\frac{1}{2}(n-p-1)} |\Sigma|^{\frac{1}{2}n}}{|A|^{\frac{1}{2}(n-p-1)} |\Phi|^{\frac{1}{2}n}} = \operatorname{mod}|C|^{p+1}. \tag{15}$$

Theorem 7.3.3. *The Jacobian of the transformation* (12) *from A to B, where A and B are symmetric, is* $\operatorname{mod}|C|^{p+1}$.

7.3.4. Marginal Distribntions

If A is distributed according to $W(\Sigma, n)$, the marginal distribution of any arbitrary set of the elements of A may be awkward to obtain. However, the marginal distribution of some sets of elements can be found easily. We give some of these in the following two theorems.

Theorem 7.3.4. *Let A and Σ be partitioned into q and $p-q$ rows and columns,*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{16}$$

If A is distributed according to $W(\Sigma, n)$, then A_{11} is distributed according to $W(\Sigma_{11}, n)$.

Proof. A is distributed as $\sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha$, where the Z_α are independent, each with the distribution $N(\mathbf{0}, \Sigma)$. Partition Z_α into subvectors of q and $p-q$ components, $Z_\alpha = (Z_\alpha^{(1)\prime}, Z_\alpha^{(2)\prime})'$. Then $Z_1^{(1)}, \dots, Z_n^{(1)}$ are independent, each with the distribution $N(\mathbf{0}, \Sigma_{11})$, and A_{11} is distributed as $\sum_{\alpha=1}^{n} Z_\alpha^{(1)} Z_\alpha^{(1)\prime}$, which has the distribution $W(\Sigma_{11}, n)$. ∎

Theorem 7.3.5. *Let A and Σ be partitioned into $p_1, p_2, \dots, p_q$ rows and columns $(p_1 + \cdots + p_q = p)$,*

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \vdots \\ A_{q1} & \cdots & A_{qq} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1q} \\ \vdots & & \vdots \\ \Sigma_{q1} & \cdots & \Sigma_{qq} \end{pmatrix}. \tag{17}$$

If $\Sigma_{ij} = \mathbf{0}$ for $i \neq j$ and if A is distributed according to $W(\Sigma, n)$, then $A_{11}, A_{22}, \dots, A_{qq}$ are independently distributed and A_{jj} is distributed according to $W(\Sigma_{jj}, n)$.

Proof. A is distributed as $\sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha$, where $Z_1, \dots, Z_n$ are independently distributed, each according to $N(\mathbf{0}, \Sigma)$. Let Z_α be partitioned

$$Z_\alpha = \begin{pmatrix} Z_\alpha^{(1)} \\ \vdots \\ Z_\alpha^{(q)} \end{pmatrix} \tag{18}$$

as A and Σ have been partitioned. Since $\Sigma_{ij} = \mathbf{0}$, the sets $Z_1^{(1)}, \dots, Z_n^{(1)}, \dots, Z_1^{(q)}, \dots, Z_n^{(q)}$ are independent. Then $A_{11} = \sum_{\alpha=1}^{n} Z_\alpha^{(1)} Z_\alpha^{(1)\prime}, \dots, A_{qq} = \sum_{\alpha=1}^{n} Z_\alpha^{(q)} Z_\alpha^{(q)\prime}$ are independent. The rest of Theorem 7.3.5 follows from Theorem 7.3.4. ∎

7.3.5. Conditional Distributions

In Section 4.3 we considered estimation of the parameters of the conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$. Application of Theorem 7.2.2 to Theorem 4.3.3 yields the following theorem:

Theorem 7.3.6. *Let A and Σ be partitioned into q and $p - q$ rows and columns as in* (16). *If A is distributed according to $W(\Sigma, n)$, the distribution of $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ is $W(\Sigma_{11\cdot 2}, n - p + q)$, $n \geq p - q$.*

Note that Theorem 7.3.6 implies that $A_{11\cdot 2}$ is independent of A_{22} and $A_{12}A_{22}^{-1}$ regardless of Σ.

7.4. COCHRAN'S THEOREM

Cochran's theorem [Cochran (1934)] is useful in proving that certain *vector quadratic forms* are distributed as sums of *vector squares*. It is a statistical statement of an algebraic theorem, which we shall give as a lemma.

Lemma 7.4.1. *If the $N \times N$ symmetric matrix C_i has rank r_i, $i = 1, \dots, m$, and*

$$\sum_{i=1}^{m} C_i = I_N, \tag{1}$$

then

$$\sum_{i=1}^{m} r_i = N \tag{2}$$

is a necessary and sufficient condition for there to exist an $N \times N$ orthogonal matrix P such that for $i = 1, \dots, m$

$$PC_iP' = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{3}$$

where I is of order r_i, the upper left-hand $\mathbf{0}$ is square of order $\sum_{j=1}^{i-1} r_j$ (which is vacuous for $i = 1$), and the lower-right hand $\mathbf{0}$ is square of order $\sum_{j=i+1}^{m} r_j$ (which is vacuous for $i = m$).

Proof. The necessity follows from the fact that (1) implies that the sum of (3) over $i = 1, \dots, m$ is I_N. Now let us prove the sufficiency; we assume (2).

There exists an orthogonal matrix P_j such that $P_jC_jP_j'$ is diagonal with diagonal elements the characteristic roots of C_j. The number of nonzero roots is r_j, the rank of C_j, and the number of 0 roots is $N - r_j$. We write

$$P_jC_jP_j' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Delta_j & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4}$$

where the partitioning is according to (3), and Δ_j is diagonal of order r_j. This is possible in view of (2). Then

$$P_j \sum_{\substack{i=1 \\ i \neq j}}^{m} C_iP_j' = P_j(I - C_j)P_j' = \begin{pmatrix} I & 0 & 0 \\ 0 & I - \Delta_j & 0 \\ 0 & 0 & I \end{pmatrix}. \tag{5}$$

Since the rank of (5) is not greater than $\sum_{i=1}^{m} r_i - r_j = N - r_j$, which is the sum of the orders of the upper left-hand and lower right-hand I's in (5), the rank of $I - \Delta_j$ is 0 and $\Delta_j = I$. (Thus the r_j nonzero roots of C_j are 1, and C_j is positive semidefinite.) From (4) we obtain

$$C_j = P_j' \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix} P_j = B_jB_j', \tag{6}$$

where B_j consists of the r_j columns of P_j' corresponding to I in (6). From (1) we obtain

$$I = \sum_{j=1}^{m} B_jB_j' = (B_1, B_2, \ldots, B_m) \begin{pmatrix} B_1' \\ B_2' \\ \vdots \\ B_m' \end{pmatrix} = P'P, \tag{7}$$

where $P = (B_1, B_2, \ldots, B_m)'$. ■

We now state a multivariate analog to Cochran's theorem.

Theorem 7.4.1. *Suppose $Y_1, \ldots, Y_N$ are independently distributed, each according to $N(\mathbf{0}, \Sigma)$. Suppose the matrix $(c_{\alpha\beta}^i) = C_i$ used in forming*

$$Q_i = \sum_{\alpha, \beta=1}^{N} c_{\alpha\beta}^i Y_\alpha Y_\beta', \qquad i = 1, \ldots, m, \tag{8}$$

is of rank r_i, and suppose

$$\sum_{i=1}^{m} Q_i = \sum_{\alpha=1}^{N} Y_\alpha Y_\alpha'. \tag{9}$$

Then (2) *is a necessary and sufficient condition for $Q_1, \dots, Q_m$ to be independently distributed with Q_i having the distribution $W(\Sigma, r_i)$.*

It follows from (3) that C_i is idempotent. See Section A.2 of the Appendix.

This theorem is useful in generalizing results from the univariate analysis of variance. (See Chapter 8.) As an example of the use of this theorem, let us prove that the mean of a sample of size N times its transpose and a multiple of the sample covariance matrix are independently distributed with a singular and a nonsingular Wishart distribution, respectively. Let $Y_1, \dots, Y_N$ be independently distributed, each according to $N(\mathbf{0}, \Sigma)$. We shall use the matrices $C_1 = (c_{\alpha\beta}^{(1)}) = (1/N)$ and $C_2 = (c_{\alpha\beta}^{(2)}) = [\delta_{\alpha\beta} - (1/N)]$. Then

$$Q_1 = \sum_{\alpha, \beta=1}^{N} \frac{1}{N} Y_\alpha Y_\beta' = N\bar{Y}\bar{Y}', \tag{10}$$

$$\begin{aligned} Q_2 &= \sum_{\alpha, \beta=1}^{N} \left(\delta_{\alpha\beta} - \frac{1}{N}\right) Y_\alpha Y_\beta' \\ &= \sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' - N\bar{Y}\bar{Y}' \\ &= \sum_{\alpha=1}^{N} (Y_\alpha - \bar{Y})(Y_\alpha - \bar{Y})', \end{aligned} \tag{11}$$

and (9) is satisfied. The matrix C_1 is of rank 1; the matrix C_2 is of rank $N-1$ (since the rank of the sum of two matrices is less than or equal to the sum of the ranks of the matrices and the rank of the second matrix is less than N). The conditions of the theorem are satisfied; therefore Q_1 is distributed as ZZ', where Z is distributed according to $N(\mathbf{0}, \Sigma)$, and Q_2 is distributed independently according to $W(\Sigma, N-1)$.

Anderson and Styan (1982) have given a survey of proofs and extensions of Cochran's theorem.

7.5. THE GENERALIZED VARIANCE

7.5.1. Definition of the Generalized Variance

One multivariate analog of the variance σ^2 of a univariate distribution is the covariance matrix Σ. Another multivariate analog is the scalar $|\Sigma|$, which is called the *generalized variance* of the multivariate distribution [Wilks (1932); see also Frisch (1929)]. Similarly, the generalized variance of the sample of vectors $x_1, \dots, x_N$ is

$$|S| = \left| \frac{1}{N-1} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \right|. \tag{1}$$

In some sense each of these is a measure of spread. We consider them here because the sample generalized variance will recur in many likelihood ratio criteria for testing hypotheses.

A geometric interpretation of the sample generalized variance comes from considering the p rows of $X = (x_1, \dots, x_N)$ as p vectors in N-dimensional space. In Section 3.2 it was shown that the rows of

$$(x_1 - \bar{x}, \dots, x_N - \bar{x}) = X - \bar{x}\varepsilon', \tag{2}$$

where $\varepsilon = (1, \dots, 1)'$, are orthogonal to the equiangular line (through the origin and ε); see Figure 3.2. Then the entries of

$$A = (X - \bar{x}\varepsilon')(X - \bar{x}\varepsilon')' \tag{3}$$

are the inner products of rows of $X - \bar{x}\varepsilon'$.

We now define a *parallelotope* determined by p vectors $v_1, \dots, v_p$ in an n-dimensional space ($n \geq p$). If $p = 1$, the parallelotope is the line segment v_1. If $p = 2$, the parallelotope is the parallelogram with v_1 and v_2 as principal edges; that is, its sides are v_1, v_2, v_1 translated so its initial endpoint is at v_2, and v_2 translated so its initial endpoint is at v_1. See Figure 7.2. If $p = 3$, the parallelotope is the conventional parallelepided with v_1, v_2, and v_3 as

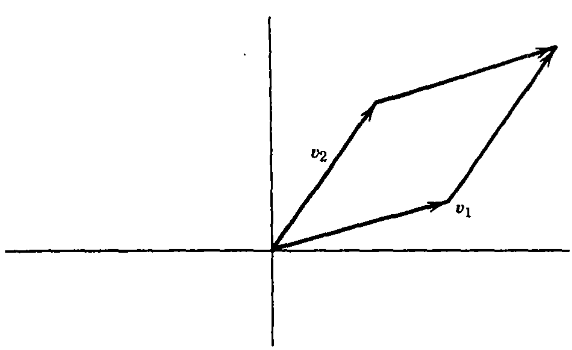

Figure 7.2. A parallelogram.

principal edges. In general, the parallelotope is the figure defined by the principal edges $v_1, \dots, v_p$. It is cut out by p pairs of parallel $(p-1)$-dimensional hyperplanes, one hyperplane of a pair being spanned by $p-1$ of $v_1, \dots, v_p$ and the other hyperplane going through the endpoint of the remaining vector.

Theorem 7.5.1. *If $V = (v_1, \dots, v_p)$, then the square of the p-dimensional volume of the parallelotope with $v_1, \dots, v_p$ as principal edges is $|V'V|$.*

Proof. If $p = 1$, then $|V'V| = v_1'v_1 = \|v_1\|^2$, which is the square of the one-dimensional volume of v_1. If two k-dimensional parallelotopes have bases consisting of $(k-1)$-dimensional parallelotopes of equal $(k-1)$-dimensional volumes and equal altitudes, their k-dimensional volumes are equal [since the k-dimensional volume is the integral of the $(k-1)$-dimensional volumes]. In particular, the volume of a k-dimensional parallelotope is equal to the volume of a parallelotope with the same base (in $k-1$ dimensions) and same altitude with sides in the kth direction orthogonal to the first $k-1$ directions. Thus the volume of the parallelotope with principal edges $v_1, \dots, v_k$, say P_k, is equal to the volume of the parallelotope with principal edges $v_1, \dots, v_{k-1}$, say P_{k-1}, times the altitude of P_k over P_{k-1}'; that is,

$$\operatorname{Vol}(P_k) = \operatorname{Vol}(P_{k-1}) \times \operatorname{Alt}(P_k | P_{k-1}). \tag{4}$$

It follows (by induction) that

$$\operatorname{Vol}(P_p) = \operatorname{Vol}(P_1) \times \operatorname{Alt}(P_2 | P_1) \times \cdots \times \operatorname{Alt}(P_p | P_{p-1}). \tag{5}$$

By the construction in Section 7.2 the altitude of P_k over P_{k-1} is $t_{kk} = \|w_k\|$; that is, t_{kk} is the distance of v_k from the $(k-1)$-dimensional space spanned by $v_1, \dots, v_{k-1}$ (or $w_1, \dots, w_{k-1}$). Hence $\operatorname{Vol}(P_p) = \prod_{k=1}^p t_{kk}$. Since $|V'V| = |TT'| = \prod_{i=1}^p t_{ii}^2$, the theorem is proved. ∎

We now apply this theorem to the parallelotope having the rows of (2) as principal edges. The dimensionality in Theorem 7.5.1 is arbitrary (but at least p).

Corollary 7.5.1. *The square of the p-dimensional volume of the parallelotope with the rows of* (2) *as principal edges is $|A|$, where A is given by* (3).

We shall see later that many multivariate statistics can be given an interpretation in terms of these volumes. These volumes are analogous to distances that arise in special cases when $p = 1$.

We now consider a geometric interpretation of $|A|$ in terms of N points in p-space. Let the columns of the matrix (2) be $y_1, \ldots, y_N$, representing N points in p-space. When $p = 1$, $|A| = \Sigma_\alpha y_{1\alpha}^2$, which is the sum of squares of the distances from the points to the origin. In general $|A|$ is the sum of squares of the volumes of all parallelotopes formed by taking as principal edges p vectors from the set $y_1, \ldots, y_N$.

We see that

$$(6) \qquad |A| = \begin{vmatrix} \sum_\alpha y_{1\alpha}^2 & \cdots & \sum_\alpha y_{1\alpha} y_{p-1,\alpha} & \sum_\beta y_{1\beta} y_{p\beta} \\ \vdots & & \vdots & \vdots \\ \sum_\alpha y_{p-1,\alpha} y_{1\alpha} & \cdots & \sum_\alpha y_{p-1,\alpha}^2 & \sum_\beta y_{p-1,\beta} y_{p\beta} \\ \sum_\alpha y_{p\alpha} y_{1\alpha} & \cdots & \sum_\alpha y_{p\alpha} y_{p-1,\alpha} & \sum_\beta y_{p\beta}^2 \end{vmatrix}$$

$$= \sum_\beta \begin{vmatrix} \sum_\alpha y_{1\alpha}^2 & \cdots & \sum_\alpha y_{1\alpha} y_{p-1,\alpha} & y_{1\beta} y_{p\beta} \\ \vdots & & \vdots & \vdots \\ \sum_\alpha y_{p-1,\alpha} y_{1\alpha} & \cdots & \sum_\alpha y_{p-1,\alpha}^2 & y_{p-1,\beta} y_{p\beta} \\ \sum_\alpha y_{p\alpha} y_{1\alpha} & \cdots & \sum_\alpha y_{p\alpha} y_{p-1,\alpha} & y_{p\beta}^2 \end{vmatrix}$$

by the rule for expanding determinants. [See (24) of Section A.1 of the Appendix.] In (6) the matrix A has been partitioned into $p-1$ and 1 columns. Applying the rule successively to the columns, we find

$$(7) \qquad |A| = \sum_{\alpha_1, \ldots, \alpha_p = 1}^{N} |y_{i\alpha_j} y_{j\alpha_j}|.$$

By Theorem 7.5.1 the square of the volume of the parallelotope with $y_{\gamma_1}, \ldots, y_{\gamma_p}$, $\gamma_1 < \cdots < \gamma_p$, as principal edges is

$$(8) \qquad V_{\gamma_1, \ldots, \gamma_p}^2 = \left| \sum_\beta y_{i\beta} y_{j\beta} \right|,$$

where the sum on β is over $(\gamma_1, \ldots, \gamma_p)$. If we now expand this determinant in the manner used for $|A|$, we obtain

$$(9) \qquad V_{\gamma_1, \ldots, \gamma_p}^2 = \sum |y_{i\beta_j} y_{j\beta_j}|,$$

where the sum is for each β_j over the range $(\gamma_1, \dots, \gamma_p)$. Summing (9) over all different sets $(\gamma_1 < \cdots \gamma_p)$, we obtain (7). ($|y_{i\beta_j} y_{j\beta_j}| = 0$ if two or more β_j are equal.) Thus $|A|$ is the sum of volumes squared of all different parallelotopes formed by sets of p of the vectors y_α as principal edges. If we replace y_α by $x_\alpha - \bar{x}$, we can state the following theorem:

Theorem 7.5.2. *Let $|S|$ be defined by* (1), *where $x_1, \dots, x_N$ are the N vectors of a sample. Then $|S|$ is proportional to the sum of squares of the volumes of all the different parallelotopes formed by using as principal edges p vectors with p of $x_1, \dots, x_N$ as one set of endpoints and $\bar{x}$ as the other, and the factor of proportionality is $1/(N-1)^p$.*

The population analog of $|S|$ is $|\Sigma|$, which can also be given a geometric interpretation. From Section 3.3 we know that

$$\Pr\left\{X'\Sigma^{-1}X \le \chi_p^2(\alpha)\right\} = 1 - \alpha \tag{10}$$

if X is distributed according to $N(\mathbf{0}, \Sigma)$; that is, the probability is $1 - \alpha$ that X fall inside the ellipsoid

$$x'\Sigma^{-1}x = \chi_p^2(\alpha). \tag{11}$$

The volume of this ellipsoid is $C(p)|\Sigma|^{\frac{1}{2}}[\chi_p^2(\alpha)]^{\frac{1}{2}p}/p$, where $C(p)$ is defined in Problem 7.3.

7.5.2. Distribution of the Sample Generalized Variance

The distribution of $|S|$ is the same as the distribution of $|A|/(N-1)^p$, where $A = \sum_{\alpha=1}^n Z_\alpha Z_\alpha'$ and $Z_1, \dots, Z_n$ are distributed independently, each according to $N(\mathbf{0}, \Sigma)$, and $n = N - 1$. Let $Z_\alpha = CY_\alpha$, $\alpha = 1, \dots, n$, where $CC' = \Sigma$. Then $Y_1, \dots, Y_n$ are independently distributed, each with distribution $N(\mathbf{0}, I)$. Let

$$B = \sum_{\alpha=1}^{n} Y_\alpha Y_\alpha' = \sum_{\alpha=1}^{n} C^{-1} Z_\alpha Z_\alpha' (C^{-1})' = C^{-1}A(C^{-1}); \tag{12}$$

then $|A| = |C| \cdot |B| \cdot |C'| = |B| \cdot |\Sigma|$. By the development in Section 7.2 we see that $|B|$ has the distribution of $\prod_{i=1}^p t_{ii}^2$ and that $t_{11}^2, \dots, t_{pp}^2$ are independently distributed with χ^2-distributions.

Theorem 7.5.3. *The distribution of the generalized variance $|S|$ of a sample $X_1, \dots, X_N$ from $N(\boldsymbol{\mu}, \Sigma)$ is the same as the distribution of $|\Sigma|/(N-1)^p$ times the product of p independent factors, the distribution of the* ith *factor being the χ^2-distribution with $N - i$ degrees of freedom.*

If $p = 1$, $|S|$ has the distribution of $|\Sigma| \cdot \chi^2_{N-1}/(N-1)$. If $p = 2$, $|S|$ has the distribution of $|\Sigma| \chi^2_{N-1} \cdot \chi^2_{N-2}/(N-1)^2$. It follows from Problem 7.15 or 7.37 that when $p = 2$, $|S|$ has the distribution of $|\Sigma|(\chi^2_{2N-4})^2/(2N-2)^2$. We can write

$$|A| = |\Sigma| \times \chi^2_{N-1} \times \chi^2_{N-2} \times \cdots \times \chi^2_{N-p}. \tag{13}$$

If $p = 2r$, then $|A|$ is distributed as

$$|\Sigma| \left(\chi^2_{2N-4} \times \chi^2_{2N-8} \times \cdots \times \chi^2_{2N-4r} \right)^2 / 2^{2r}. \tag{14}$$

Since the hth moment of a χ^2-variable with m degrees of freedom is $2^h \Gamma(\frac{1}{2}m + h)/\Gamma(\frac{1}{2}m)$ and the moment of a product of independent variables is the product of the moments of the variables, the hth moment of $|A|$ is

$$\begin{aligned} (15)\quad |\Sigma|^h \prod_{i=1}^{p} \left\{ 2^h \frac{\Gamma\left[\frac{1}{2}(N-i) + h\right]}{\Gamma\left[\frac{1}{2}(N-i)\right]} \right\} &= 2^{hp} |\Sigma|^h \frac{\prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(N-i) + h\right]}{\prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(N-i)\right]} \\ &= 2^{hp} |\Sigma|^h \frac{\Gamma_p\left[\frac{1}{2}(N-1) + h\right]}{\Gamma_p\left[\frac{1}{2}(N-1)\right]}. \end{aligned}$$

Thus

$$\mathscr{E}|A| = |\Sigma| \prod_{i=1}^{p} (N-i). \tag{16}$$

$$\mathscr{V}(|A|) = |\Sigma|^2 \prod_{i=1}^{p} (N-i) \left[\prod_{j=1}^{p} \Gamma(N-j+2) - \prod_{j=1}^{p} \Gamma(N-j) \right], \tag{17}$$

where $\mathscr{V}(|A|)$ is the variance of $|A|$.

7.5.3. The Asymptotic Distribution of the Sample Generalized Variance

Let $|B|/n^p = V_1(n) \times V_2(n) \times \cdots \times V_p(n)$, where the V's are independently distributed and $nV_i(n) = \chi^2_{n-p+i}$. Since χ^2_{n-p+i} is distributed as $\sum_{\alpha=1}^{n-p+i} W_\alpha^2$, where the W_α are independent, each with distribution $N(0,1)$, the central limit theorem (applied to W_α^2) states that

$$\frac{nV_i(n) - (n-p+i)}{\sqrt{2(n-p+i)}} = \sqrt{n} \frac{V_i(n) - 1 + \frac{p-i}{n}}{\sqrt{2} \sqrt{1 - \frac{p-i}{n}}} \tag{18}$$

is asymptotically distributed according to $N(0,1)$. Then $\sqrt{n}[V_i(n) - 1]$ is asymptotically distributed according to $N(0,2)$. We now apply Theorem 4.2.3.

We have

$$(19) \qquad U(n) = \begin{pmatrix} V_1(n) \\ \vdots \\ V_p(n) \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

$|B|/n^p = w = f(u_1, \ldots, u_p) = u_1 u_2 \cdots u_p$, $\mathbf{T} = 2I$, $\partial f/\partial u_i|_{u=b} = 1$, *and* $\phi_b' \mathbf{T} \phi_b = 2p$. *Thus*

$$(20) \qquad \sqrt{n}\left(\frac{|B|}{n^p} - 1\right)$$

is asymptotically distributed according to $N(0, 2p)$.

Theorem 7.5.4. *Let* S *be a* $p \times p$ *sample covariance matrix with* n *degrees of freedom. Then* $\sqrt{n}\,(|S|/|\Sigma| - 1)$ *is asymptotically normally distributed with mean* 0 *and variance* $2p$.

7.6. DISTRIBUTION OF THE SET OF CORRELATION COEFFICIENTS WHEN THE POPULATION COVARIANCE MATRIX IS DIAGONAL

In Section 4.2.1 we found the distribution of a single sample correlation when the corresponding population correlation was zero. Here we shall find the density of the set r_{ij}, $i<j$, $i, j = 1, \ldots, p$, when $\rho_{ij} = 0$, $i<j$.

We start with the distribution of A when Σ is diagonal; that is, $W[(\sigma_{ii}\delta_{ij}), n]$. The density of A is

$$(1) \qquad \frac{|a_{ij}|^{\frac{1}{2}(n-p-1)} \exp\left(-\frac{1}{2}\sum_{i=1}^p a_{ii}/\sigma_{ii}\right)}{2^{\frac{1}{2}np} \prod_{i=1}^p \sigma_{ii}^{\frac{1}{2}n} \Gamma_p\left(\frac{1}{2}n\right)},$$

since

$$(2) \qquad |\Sigma| = \begin{vmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{vmatrix} = \prod_{i=1}^p \sigma_{ii}.$$

We make the transformation

$$(3) \qquad a_{ij} = \sqrt{a_{ii}}\sqrt{a_{jj}}\, r_{ij}, \qquad i \neq j,$$

$$(4) \qquad a_{ii} = a_{ii}.$$