CHAPTER 10

Testing Hypotheses of Equality of Covariance Matrices and Equality of Mean Vectors and Covariance Matrices

10.1. INTRODUCTION

In this chapter we study the problems of testing hypotheses of equality of covariance matrices and equality of both covariance matrices and mean vectors. In each case (except one) the problem and tests considered are multivariate generalizations of a univariate problem and test. Many of the tests are likelihood ratio tests or modifications of likelihood ratio tests. Invariance considerations lead to other test procedures.

First, we consider equality of covariance matrices and equality of covariance matrices and mean vectors of several populations without specifying the common covariance matrix or the common covariance matrix and mean vector. The multivariate analysis of variance with random factors is considered in this context. Later we treat the equality of a covariance matrix to a given matrix and also simultaneous equality of a covariance matrix to a given matrix and equality of a mean vector to a given vector. One other hypothesis considered, the equality of a covariance matrix to a given matrix except for a proportionality factor, has only a trivial corresponding univariate hypothesis.

In each case the class of tests for a class of hypotheses leads to a confidence region. Families of simultaneous confidence intervals for covariances and for ratios of covariances are given.

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

The application of the tests for elliptically contoured distributions is treated in Section 10.11.

10.2. CRITERIA FOR TESTING EQUALITY OF SEVERAL COVARIANCE MATRICES

In this section we study several normal distributions and consider using a set of samples, one from each population, to test the hypothesis that the covariance matrices of these populations are equal. Let $x_\alpha^{(g)}$, $\alpha = 1, \dots, N_g$, $g = 1, \dots, q$, be an observation from the gth population $N(\mu^{(g)}, \Sigma_g)$. We wish to test the hypothesis

$$H_1 : \Sigma_1 = \cdots = \Sigma_q. \tag{1}$$

Let $\sum_{g=1}^q N_g = N$,

$$A_g = \sum_{\alpha=1}^{N_g} \left(x_\alpha^{(g)} - \bar{x}^{(g)}\right)\left(x_\alpha^{(g)} - \bar{x}^{(g)}\right)', \qquad g = 1, \dots, q, \tag{2}$$

$$A = \sum_{g=1}^q A_g.$$

First we shall obtain the likelihood ratio criterion. The likelihood function is

$$L = \prod_{g=1}^q \frac{1}{(2\pi)^{\frac{1}{2}pN_g} |\Sigma_g|^{\frac{1}{2}N_g}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{N_g} \left(x_\alpha^{(g)} - \mu^{(g)}\right)' \Sigma_g^{-1} \left(x_\alpha^{(g)} - \mu^{(g)}\right)\right]. \tag{3}$$

The space Ω is the parameter space in which each Σ_g is positive definite and $\mu^{(g)}$ any vector. The space ω is the parameter space in which $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_q$ (positive definite) and $\mu^{(g)}$ is any vector. The maximum likelihood estimators of $\mu^{(g)}$ and Σ_g in Ω are given by

$$\hat{\mu}_\Omega^{(g)} = \bar{x}^{(g)}, \qquad \hat{\Sigma}_{g\Omega} = \frac{1}{N_g} A_g, \qquad g = 1, \dots, q. \tag{4}$$

The maximum likelihood estimators of $\mu^{(g)}$ in ω are given by (4), $\hat{\mu}_\omega^{(g)} = \bar{x}^{(g)}$, since the maximizing values of $\mu^{(g)}$ are the same regardless of Σ_g. The function to be maximized with respect to $\Sigma_1 = \cdots = \Sigma_q = \Sigma$, say, is

$$\frac{1}{(2\pi)^{\frac{1}{2}pN} |\Sigma|^{\frac{1}{2}N}} \exp\left[-\frac{1}{2} \sum_{g=1}^q \sum_{\alpha=1}^{N_g} \left(x_\alpha^{(g)} - \bar{x}^{(g)}\right)' \Sigma^{-1} \left(x_\alpha^{(g)} - \bar{x}^{(g)}\right)\right]. \tag{5}$$

By Lemma 3.2.2, the maximizing value of Σ is

$$\hat{\Sigma}_\omega = \frac{1}{N} A, \tag{6}$$

and the maximum of the likelihood function is

$$\frac{1}{(2\pi)^{\frac{1}{2}pN} |\hat{\Sigma}_\omega|^{\frac{1}{2}N}} e^{-\frac{1}{2}pN}. \tag{7}$$

The likelihood ratio criterion for testing (1) is

$$\lambda_1 = \frac{\prod_{g=1}^{q} |\hat{\Sigma}_{g\Omega}|^{\frac{1}{2}N_g}}{|\hat{\Sigma}_\omega|^{\frac{1}{2}N}} = \frac{\prod_{g=1}^{q} |A_g|^{\frac{1}{2}N_g}}{|A|^{\frac{1}{2}N}} \cdot \frac{N^{\frac{1}{2}pN}}{\prod_{g=1}^{q} N_g^{\frac{1}{2}pN_g}}. \tag{8}$$

The critical region is

$$\lambda_1 \leq \lambda_1(\varepsilon), \tag{9}$$

where $\lambda_1(\varepsilon)$ is defined so that (9) holds with probability ε when (1) is true.

Bartlett (1937a) has suggested modifying λ_1 in the univariate case by replacing sample numbers by the numbers of degrees of freedom of the A_g. Except for a numerical constant, the statistic he proposes is

$$V_1 = \frac{\prod_{g=1}^{q} |A_g|^{\frac{1}{2}n_g}}{|A|^{\frac{1}{2}n}}, \tag{10}$$

where $n_g = N_g - 1$ and $n = \sum_{g=1}^{q} n_g = N - q$. The numerator is proportional to a power of a weighted geometric mean of the sample generalized variances, and the denominator is proportional to a power of the determinant of a weighted arithmetic mean of the sample covariance matrices.

In the scalar case ($p = 1$) of two samples the criterion (10) is

$$\frac{(n_1)^{\frac{1}{2}n_1}(n_2)^{\frac{1}{2}n_2}(s_1^2)^{\frac{1}{2}n_1}(s_2^2)^{\frac{1}{2}n_2}}{(n_1 s_1^2 + n_2 s_2^2)^{\frac{1}{2}(n_1+n_2)}} = \frac{(n_1)^{\frac{1}{2}n_1}(n_2)^{\frac{1}{2}n_2} F^{\frac{1}{2}n_1}}{(n_1 F + n_2)^{\frac{1}{2}(n_1+n_2)}}, \tag{11}$$

where s_1^2 and s_2^2 are the usual unbiased estimators of σ_1^2 and σ_2^2 (the two population variances) and

$$F = \frac{s_1^2}{s_2^2}. \tag{12}$$

Thus the critical region

$$V_1 \le V_1(\varepsilon) \tag{13}$$

is based on the F-statistic with n_1 and n_2 degrees of freedom, and the inequality (13) implies a particular method of choosing $F_1(\varepsilon)$ and $F_2(\varepsilon)$ for the critical region

$$F \le F_1(\varepsilon), \qquad F \ge F_2(\varepsilon). \tag{14}$$

Brown (1939) and Scheffé (1942) have shown that (14) yields an unbiased test.

Bartlett gave a more intuitive argument for the use of V_1 in place of λ_1. He argues that if N_1, say, is small, A_1 is given too much weight in λ_1, and other effects may be missed. Perlman (1980) has shown that the test based on V_1 is unbiased.

If one assumes

$$\mathscr{E} X_\alpha^{(g)} = \boldsymbol{\beta}_g z_\alpha^{(g)}, \tag{15}$$

where $z_\alpha^{(g)}$ consists of k_g components, and if one estimates the matrix $\boldsymbol{\beta}_g$, defining

$$A_g = \sum_{\alpha=1}^{N_g} \left(x_\alpha^{(g)} - \hat{\boldsymbol{\beta}}_g z_\alpha^{(g)}\right)\left(x_\alpha - \hat{\boldsymbol{\beta}}_g z_\alpha^{(g)}\right)', \tag{16}$$

one uses (10) with $n_g = N_g - k_g$.

The statistical problem (parameter space Ω and null hypothesis ω) is invariant with respect to changes of location within populations and a common linear transformation

$$X^{*(g)} = CX^{(g)} + \nu^{(g)}, \qquad g = 1, \dots, q, \tag{17}$$

where C is nonsingular. Each matrix A_g is invariant under change of location, and the modified criterion (10) is invariant:

$$V_1^* = \frac{\prod_{g=1}^q |A_g^*|^{\frac{1}{2}n_g}}{|A^*|^{\frac{1}{2}n}} = \frac{\prod_{g=1}^q |CA_gC'|^{\frac{1}{2}n_g}}{|CAC'|^{\frac{1}{2}n}} = \frac{\prod_{g=1}^q |A_g|^{\frac{1}{2}n_g}}{|A|^{\frac{1}{2}n}} = V_1. \tag{18}$$

Similarly, the likelihood ratio criterion (8) is invariant.

An alternative invariant test procedure [Nagao (1973a)] is based on the criterion

$$(19)\qquad \tfrac{1}{2}\sum_{g=1}^{q} n_g \operatorname{tr}\left(S_g S^{-1} - I\right)^2 = \tfrac{1}{2}\sum_{g=1}^{q} n_g \operatorname{tr}\left(S_g - S\right)S^{-1}\left(S_g - S\right)S^{-1},$$

where $S_g = (1/n_g)A_g$ and $S = (1/n)A$. (See Section 7.8.)

10.3. CRITERIA FOR TESTING THAT SEVERAL NORMAL DISTRIBUTIONS ARE IDENTICAL

In Section 8.8 we considered testing the equality of mean vectors when we assumed the covariance matrices were the same; that is, we tested

$$(1)\qquad H_2 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \cdots = \boldsymbol{\mu}^{(q)} \qquad \text{given} \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \cdots = \boldsymbol{\Sigma}_q.$$

The test of the assumption in H_2 was considered in Section 10.2. Now let us consider the hypothesis that both means and covariances are the same; this is a combination of H_1 and H_2. We test

$$(2)\qquad H : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \cdots = \boldsymbol{\mu}^{(q)}, \ \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_q.$$

As in Section 10.2, let $x_\alpha^{(g)}$, $\alpha = 1, \dots, N_g$, be an observation from $N(\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g)$, $g = 1, \dots, q$. Then Ω is the unrestricted parameter space of $\{\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g\}$, $g = 1, \dots, q$, where $\boldsymbol{\Sigma}_g$ is positive definite, and ω^* consists of the space restricted by (2).

The likelihood function is given by (3) of Section 10.2. The hypothesis H_1 of Section 10.2 is that the parameter point falls in ω; the hypothesis H_2 of Section 8.8 is that the parameter point falls in ω^* given it falls in $\omega \supset \omega^*$; and the hypothesis H here is that the parameter point falls in ω^* given that it is in Ω.

We use the following lemma:

Lemma 10.3.1. *Let y be an observation vector on a random vector with density $f(z, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector in a space Ω. Let H_a be the hypothesis $\boldsymbol{\theta} \in \Omega_a \subset \Omega$, let H_b be the hypothesis $\boldsymbol{\theta} \in \Omega_b$, $\subset \Omega_a$, given $\boldsymbol{\theta} \in \Omega_a$, and let H_{ab} be the hypothesis $\boldsymbol{\theta} \in \Omega_b$, given $\boldsymbol{\theta} \in \Omega$. If λ_a, the likelihood ratio criterion for testing H_a, λ_b for H_b, and λ_{ab} for H_{ab} are uniquely defined for the observation vector y, then*

$$(3)\qquad \lambda_{ab} = \lambda_a \lambda_b.$$

Proof. The lemma follows from the definitions:

$$\lambda_a = \frac{\max_{\theta \in \Omega_a} f(y, \theta)}{\max_{\theta \in \Omega} f(y, \theta)}, \tag{4}$$

$$\lambda_b = \frac{\max_{\theta \in \Omega_b} f(y, \theta)}{\max_{\theta \in \Omega_a} f(y, \theta)}, \tag{5}$$

$$\lambda_{ab} = \frac{\max_{\theta \in \Omega_b} f(y, \theta)}{\max_{\theta \in \Omega} f(y, \theta)}. \tag{6}$$ ■

Thus the likelihood ratio criterion for the hypothesis H is the product of the likelihood ratio criteria for H_1 and H_2,

$$\lambda = \lambda_1 \lambda_2 = \left(\prod_{g=1}^{q} \frac{|A_g|^{\frac{1}{2}N_g}}{N_g^{\frac{1}{2}pN_g}} \right) \frac{N^{\frac{1}{2}pN}}{|B|^{\frac{1}{2}N}}, \tag{7}$$

where

$$B = \sum_{g=1}^{q} \sum_{\alpha=1}^{N_g} \left(x_\alpha^{(g)} - \bar{x} \right) \left(x_\alpha^{(g)} - \bar{x} \right)' \tag{8}$$

$$= A + \sum_{g=1}^{q} N_g (\bar{x}^{(g)} - \bar{x})(\bar{x}^{(g)} - \bar{x})'.$$

The critical region is defined by

$$\lambda \leq \lambda(\varepsilon), \tag{9}$$

where $\lambda(\varepsilon)$ is chosen so that the probability of (9) under H is ε.

Let

$$V_2 = \frac{|A|^{\frac{1}{2}n}}{|B|^{\frac{1}{2}n}} = \lambda_2^{n/N}; \tag{10}$$

this is equivalent to λ_2 for testing H_2, which is λ of (12) of Section 8.8. We might consider

$$V = V_1 V_2 = \frac{\prod_{g=1}^{q} |A_g|^{\frac{1}{2}n_g}}{|B|^{\frac{1}{2}n}}. \tag{11}$$

However, Perlman (1980) has shown that the likelihood ratio test is unbiased.

10.4. DISTRIBUTIONS OF THE CRITERIA

10.4.1. Characterization of the Distributions

First let us consider V_1 given by (10) of Section 10.2. If

$$V_{1g} = \frac{|A_1 + \cdots + A_{g-1}|^{\frac{1}{2}(n_1 + \cdots + n_{g-1})} |A_g|^{\frac{1}{2}n_g}}{|A_1 + \cdots + A_g|^{\frac{1}{2}(n_1 + \cdots + n_g)}}, \qquad g = 2, \ldots, q, \tag{1}$$

then

$$V_1 = \prod_{g=2}^{q} V_{1g}. \tag{2}$$

Theorem 10.4.1. *$V_{12}, V_{13}, \ldots, V_{1q}$ defined by* (1) *are independent when $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q$ and $n_g \ge p$, $g = 1, \ldots, q$.*

The theorem is a consequence of the following lemma:

Lemma 10.4.1. *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are independently distributed according to $W(\boldsymbol{\Sigma}, m)$ and $W(\boldsymbol{\Sigma}, n)$, respectively, $n \ge p$, $m \ge p$, and $\boldsymbol{C}$ is such that $\boldsymbol{C}(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C}' = \boldsymbol{I}$, then $\boldsymbol{A} + \boldsymbol{B}$ and $\boldsymbol{CAC}'$ are independently distributed; $\boldsymbol{A} + \boldsymbol{B}$ has the Wishart distribution with $m + n$ degrees of freedom, and $\boldsymbol{CAC}'$ has the multivariate beta distribution with n and m degrees of freedom.*

Proof of Lemma. The density of $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{B}$ and $\boldsymbol{E} = \boldsymbol{CAC}'$ is found by replacing $\boldsymbol{A}$ and $\boldsymbol{B}$ in their joint density by $\boldsymbol{C}^{-1}\boldsymbol{E}\boldsymbol{C}'^{-1}$ and $\boldsymbol{D} - \boldsymbol{C}^{-1}\boldsymbol{E}\boldsymbol{C}'^{-1} = \boldsymbol{C}^{-1}(\boldsymbol{I} - \boldsymbol{E})\boldsymbol{C}'^{-1}$, respectively, and multiplying by the Jacobian, which is $\operatorname{mod}|\boldsymbol{C}|^{-(p+1)} = |\boldsymbol{D}|^{\frac{1}{2}(p+1)}$, to obtain

$$\begin{aligned}
&K(\boldsymbol{\Sigma}, m)K(\boldsymbol{\Sigma}, n)|\boldsymbol{C}^{-1}\boldsymbol{E}\boldsymbol{C}'^{-1}|^{\frac{1}{2}(m-p-1)} \\
&\qquad \cdot |\boldsymbol{C}^{-1}(\boldsymbol{I} - \boldsymbol{E})\boldsymbol{C}'^{-1}|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}} |\boldsymbol{D}|^{\frac{1}{2}(p+1)} \\
&= K(\boldsymbol{\Sigma}, m+n)|\boldsymbol{D}|^{\frac{1}{2}(m+n-p-1)} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}} \\
&\qquad \cdot \frac{\Gamma_p\left[\frac{1}{2}(m+n)\right]}{\Gamma_p\left(\frac{1}{2}m\right)\Gamma_p\left(\frac{1}{2}n\right)} |\boldsymbol{E}|^{\frac{1}{2}(m-p-1)} |\boldsymbol{I} - \boldsymbol{E}|^{\frac{1}{2}(n-p-1)}
\end{aligned} \tag{3}$$

for $\boldsymbol{D}$, $\boldsymbol{E}$, and $\boldsymbol{I} - \boldsymbol{E}$ positive definite. ■

Proof of Theorem. If we let $A_1 + \cdots + A_g = D_g$ and $C_g(A_1 + \cdots + A_{g-1})C_g' = E_g$, where $C_g D_g C_g' = I$, $g = 2, \dots, q$, then

$$V_{1g} = \frac{|C_g^{-1} E_g C_g'^{-1}|^{\frac{1}{2}(n_1 + \cdots + n_{g-1})} |C_g^{-1}(I - E_g) C_g'^{-1}|^{\frac{1}{2}n_g}}{|C_g^{-1} C_g'^{-1}|^{\frac{1}{2}(n_1 + \cdots + n_g)}} \tag{4}$$

$$= |E_g|^{\frac{1}{2}(n_1 + \cdots + n_{g-1})} |I - E_g|^{\frac{1}{2}n_g}, \qquad g = 2, \dots, q,$$

and $E_2, \dots, E_q$ are independent by Lemma 10.4.1. ∎

We shall now find a characterization of the distribution of V_{1g}. A statistic V_{1g} is of the form

$$\frac{|B|^b |C|^c}{|B + C|^{b+c}}. \tag{5}$$

Let B_i and C_i be the upper left-hand square submatrices of B and C, respectively, of order i. Define $b_{(i)}$ and $c_{(i)}$ by

$$B_i = \begin{pmatrix} B_{i-1} & b_{(i)} \\ b_{(i)}' & b_{ii} \end{pmatrix}, \qquad C_i = \begin{pmatrix} C_{i-1} & c_{(i)} \\ c_{(i)}' & c_{ii} \end{pmatrix}, \qquad i = 2, \dots, p. \tag{6}$$

Then (5) is ($B_0 = C_0 = I$, $b_{(1)} = c_{(1)} = 0$)

(7)

$$\frac{|B|^b |C|^c}{|B + C|^{b+c}} = \prod_{i=1}^p \frac{|B_i|^b |C_i|^c}{|B_{i-1}|^b |C_{i-1}|^c} \cdot \frac{|B_{i-1} + C_{i-1}|^{b+c}}{|B_i + C_i|^{b+c}}$$

$$= \prod_{i=1}^p \frac{\left(b_{ii} - b_{(i)}' B_{i-1}^{-1} b_{(i)}\right)^b \left(c_{ii} - c_{(i)}' C_{i-1}^{-1} c_{(i)}\right)^c}{\left[b_{ii} + c_{ii} - \left(b_{(i)} + c_{(i)}\right)'\left(B_{i-1} + C_{i-1}\right)^{-1}\left(b_{(i)} + c_{(i)}\right)\right]^{b+c}}$$

$$= \prod_{i=1}^p \left\{ \frac{b_{ii\cdot i-1}^b c_{ii\cdot i-1}^c}{\left(b_{ii\cdot i-1} + c_{ii\cdot i-1}\right)^{b+c}} \cdot \frac{\left(b_{ii\cdot i-1} + c_{ii\cdot i-1}\right)^{b+c}}{\begin{matrix}\left[b_{ii\cdot i-1} + c_{ii\cdot i-1} + b_{(i)}' B_{i-1}^{-1} b_{(i)} + c_{(i)}' C_{i-1}^{-1} c_{(i)}\right. \\ \left. - \left(b_{(i)} + c_{(i)}\right)'\left(B_{i-1} + C_{i-1}\right)^{-1}\left(b_{(i)} + c_{(i)}\right)\right]^{b+c}\end{matrix}} \right\},$$

where $b_{ii\cdot i-1} = b_{ii} - b'_{(i)}B_{i-1}^{-1}b_{(i)}$ and $c_{ii\cdot i-1} = c_{ii} - c'_{(i)}C_{i-1}^{-1}c_{(i)}$. The second term for $i = 1$ is defined as 1.

Now we want to argue that the ratios on the right-hand side of (7) are statistically independent when B and C are independently distributed according to $W(\Sigma, m)$ and $W(\Sigma, n)$, respectively. It follows from Theorem 4.3.3 that for B_{i-1} fixed $b_{(i)}$ and $b_{ii\cdot i-1}$ are independently distributed according to $N(\beta_{(i)}, \sigma_{ii\cdot i-1}B_{i-1}^{-1})$ and $\sigma_{ii\cdot i-1}\chi^2$ with $m-(i-1)$ degrees of freedom, respectively. Lemma 10.4.1 implies that the first term (which is a function of $b_{ii\cdot i-1}/c_{ii\cdot i-1}$) is independent of $b_{ii\cdot i-1} + c_{ii\cdot i-1}$.

We apply the following lemma:

Lemma 10.4.2. *For B_{i-1} and C_{i-1} positive definite*

$$b'_{(i)}B_{i-1}^{-1}b_{(i)} + c'_{(i)}C_{i-1}^{-1}c_{(i)} - \left(b_{(i)} + c_{(i)}\right)'\left(B_{i-1} + C_{i-1}\right)^{-1}\left(b_{(i)} + c_{(i)}\right) \tag{8}$$
$$= \left(B_{i-1}^{-1}b_{(i)} - C_{i-1}^{-1}c_{(i)}\right)'\left(B_{i-1}^{-1} + C_{i-1}^{-1}\right)^{-1}\left(B_{i-1}^{-1}b_{(i)} - C_{i-1}^{-1}c_{(i)}\right).$$

Proof. Use of $(B^{-1} + C^{-1})^{-1} = [C^{-1}(B + C)B^{-1}]^{-1} = B(B + C)^{-1}C$ shows the left-hand side of (8) is (omitting i and $i-1$)

(9)

$$b'B^{-1}\left(B^{-1}+C^{-1}\right)^{-1}\left(B^{-1}+C^{-1}\right)b + c'\left(B^{-1}+C^{-1}\right)\left(B^{-1}+C^{-1}\right)^{-1}C^{-1}c$$
$$-(b+c)'B^{-1}\left(B^{-1}+C^{-1}\right)^{-1}C^{-1}(b+c)$$
$$= b'B^{-1}\left(B^{-1}+C^{-1}\right)^{-1}B^{-1}b + c'C^{-1}\left(B^{-1}+C^{-1}\right)^{-1}C^{-1}c$$
$$-b'B^{-1}\left(B^{-1}+C^{-1}\right)^{-1}C^{-1}c - c'C^{-1}\left(B^{-1}+C^{-1}\right)B^{-1}b,$$

which is the right-hand side of (8). ■

The denominator of the ith second term in (7) is the numerator plus (8). The conditional distribution of $B_{i-1}^{-1}b_{(i)} - C_{i-1}^{-1}c_{(i)}$ is normal with mean $B_{i-1}^{-1}\beta_{(i)} - C_{i-1}^{-1}\gamma_{(i)}$ and covariance matrix $\sigma_{ii\cdot i-1}(B_{i-1}^{-1} + C_{i-1}^{-1})$. The covariance matrix is $\sigma_{ii\cdot i-1}$ times the inverse of the second matrix on the right-hand side of (8). Thus (8) is distributed as $\sigma_{ii\cdot i-1}\chi^2$ with $i-1$ degrees of freedom, independent of B_{i-1}, C_{i-1}, $b_{ii\cdot i-1}$, and $c_{ii\cdot i-1}$.

Then

$$\frac{b_{ii\cdot i-1}^{b}c_{ii\cdot i-1}^{2}}{\left(b_{ii\cdot i-1} + c_{ii\cdot i-1}\right)^{b+c}} = \left(\frac{b_{ii\cdot i-1}}{b_{ii\cdot i-1} + c_{ii\cdot i-1}}\right)^{b}\left(\frac{c_{ii\cdot i-1}}{b_{ii\cdot i-1} + c_{ii\cdot i-1}}\right)^{c} \tag{10}$$

is distributed as $X_i^b(1-X_i)^c$, where X_i has the $\beta[\frac{1}{2}(m-i+1), \frac{1}{2}(n-i+1)]$

distribution, $i = 1, \ldots, p$. Also

$$\left[\frac{b_{ii\cdot i-1} + c_{ii\cdot i-1}}{b_{ii\cdot i-1} + c_{ii\cdot i-1} + (8)}\right]^{b+c}, \qquad i = 2, \ldots, p, \tag{11}$$

is distributed as Y_i^{b+c}, where Y_i has the $\beta[\frac{1}{2}(m+n) - i + 1, \frac{1}{2}(i-1)]$ distribution. Then (5) is distributed as $\prod_{i=1}^{p} X_i^b (1 - X_i)^c \prod_{i=2}^{p} Y_i^{b+c}$, and the factors are mutually independent.

Theorem 10.4.2.

$$V_1 = \prod_{g=2}^{q} \left\{ \prod_{i=1}^{p} X_{ig}^{\frac{1}{2}(n_1 + \cdots + n_{g-1})} (1 - X_{ig})^{\frac{1}{2}n_g} \cdot \prod_{i=2}^{p} Y_{ig}^{\frac{1}{2}(n_1 + \cdots + n_g)} \right\}, \tag{12}$$

where the X's and Y's are independent, X_{ig} has the $\beta[\frac{1}{2}(n_1 + \cdots + n_{g-1} - i + 1), \frac{1}{2}(n_g - i + 1)]$ distribution, and Y_{ig} has the $\beta[\frac{1}{2}(n_1 + \cdots + n_g) - i + 1, \frac{1}{2}(i-1)]$ distribution.

Proof. The factors $V_{12}, \ldots, V_{1q}$ are independent by Theorem 10.4.1. Each term V_{1g} is decomposed according to (7), and the factors are independent. ∎

The factors of V_1 can be interpreted as test criteria for subhypotheses. The term depending on X_{i2} is the criterion for testing the hypothesis that $\sigma_{ii\cdot i-1}^{(1)} = \sigma_{ii\cdot i-1}^{(2)}$, and the term depending on Y_{i2} is the criterion for testing $\boldsymbol{\sigma}_{(i)}^{(1)} = \boldsymbol{\sigma}_{(i)}^{(2)}$ given $\sigma_{ii\cdot i-1}^{(1)} = \sigma_{ii\cdot i-1}^{(2)}$, and $\boldsymbol{\Sigma}_{i-1,1} = \boldsymbol{\Sigma}_{i-1,2}$. The terms depending on X_{ig} and Y_{ig} similarly furnish criteria for testing $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_g$ given $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_{g-1}$.

Now consider the likelihood ratio criterion λ given by (7) of Section 10.3 for testing the hypothesis $\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(q)}$ and $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q$. It is equivalent to the criterion

$$W = \frac{\prod_{g=1}^{q} |A_g|^{\frac{1}{2}N_g}}{|A_1 + \cdots + A_q|^{\frac{1}{2}(N_1 + \cdots + N_g)}} \cdot \frac{|A_1 + \cdots + A_q|^{\frac{1}{2}N}}{|A_1 + \cdots + A_q + \sum_{g=1}^{q} N_g(\bar{\boldsymbol{x}}^{(g)} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)} - \bar{\boldsymbol{x}})'|^{\frac{1}{2}N}}. \tag{13}$$

The two factors of (13) are independent because the first factor is independent of $A_1 + \cdots + A_q$ (by Lemma 10.4.1 and the proof of Theorem 10.4.1) and of $\bar{\boldsymbol{x}}^{(1)}, \ldots, \bar{\boldsymbol{x}}^{(q)}$.

Theorem 10.4.3

$$(14)\qquad W=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}X_{ig}^{\frac{1}{2}(N_1+\cdots+N_{g-1})}(1-X_{ig})^{\frac{1}{2}N_g}\prod_{i=2}^{p}Y_{ig}^{\frac{1}{2}(N_1+\cdots+N_g)}\right\}\prod_{i=1}^{p}Z_i^{\frac{1}{2}N},$$

where the X's, Y's, and Z's are independent, X_{ig} has the $\beta[\frac{1}{2}(n_1+\cdots+n_{g-1}-i+1),\frac{1}{2}(n_g-i+1)]$ distribution, Y_{ig} has the $\beta[\frac{1}{2}(n_1+\cdots+n_g)-i+1,\frac{1}{2}(i-1)]$ distribution, and Z_i has the $\beta[\frac{1}{2}(n+1-i),\frac{1}{2}(q-1)]$ distribution.

Proof. The characterization of the first factor in (13) corresponds to that of V_1 with the exponents of X_{ig} and $1-X_{ig}$ modified by replacing n_g by N_g. The second term in $U_{p,q-1,n}^{\frac{1}{2}N}$, and its characterization follows from Theorem 8.4.1. ■

10.4.2. Moments of the Distributions

We now find the moments of V_1 and of W. Since $0\le V_1\le 1$ and $0\le W\le 1$, the moments determine the distributions uniquely. The hth moment of V_1 we find from the characterization of the distribution in Theorem 10.4.2:

(15)

$$\begin{aligned}
\mathscr{E}V_1^h &= \prod_{g=2}^{q}\left\{\prod_{i=1}^{p}\mathscr{E}X_{ig}^{\frac{1}{2}(n_1+\cdots+n_{g-1})h}(1-X_{ig})^{\frac{1}{2}n_g h}\prod_{i=2}^{p}\mathscr{E}Y_{ig}^{\frac{1}{2}(n_1+\cdots+n_g)h}\right\}\\
&=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1})(1+h)-\frac{1}{2}(i-1)\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1}-i+1)\right]}\right.\\
&\qquad\cdot\frac{\Gamma\left[\frac{1}{2}n_g(1+h)-\frac{1}{2}(i-1)\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1\right]}{\Gamma\left[\frac{1}{2}(n_g-i+1)\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)(1+h)-i+1\right]}\\
&\qquad\left.\cdot\prod_{i=2}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)(1+h)-i+1\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g-i+1)\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)(1+h)-\frac{1}{2}(i-1)\right]}\right\}\\
&=\prod_{i=1}^{p}\left\{\frac{\Gamma\left[\frac{1}{2}(n+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+hn+1-i)\right]}\prod_{g=1}^{q}\frac{\Gamma\left[\frac{1}{2}(n_g+hn_g+1-i)\right]}{\Gamma\left[\frac{1}{2}(n_g+1-i)\right]}\right\}\\
&=\frac{\Gamma_p(\frac{1}{2}n)}{\Gamma_p\left[\frac{1}{2}(n+hn)\right]}\prod_{g=1}^{q}\frac{\Gamma_p\left[\frac{1}{2}(n_g+hn_g)\right]}{\Gamma_p(\frac{1}{2}n_g)}.
\end{aligned}$$

The hth moment of W can be found from its representation in Theorem 10.4.3. We have

(16)

$$\begin{aligned}
\mathscr{E}W^h &= \prod_{g=2}^{q}\prod_{i=1}^{p}\mathscr{E}X_{ig}^{\frac{1}{2}(N_1+\cdots+N_{g-1})h}(1-X_{ig})^{\frac{1}{2}N_g h}\prod_{i=2}^{p}\mathscr{E}Y_{ig}^{\frac{1}{2}(N_1+\cdots+N_g)h}\,\mathscr{E}U_{p,q-1,n}^{\frac{1}{2}Nh}\\
&=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1}+1-i)+\frac{1}{2}h(N_1+\cdots+N_{g-1})\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1}+1-i)\right]\Gamma\left[\frac{1}{2}(n_g+1-i)\right]}\right.\\
&\qquad\cdot\frac{\Gamma\left[\frac{1}{2}(n_g+1-i+N_g h)\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)+\frac{1}{2}h(N_1+\cdots+N_g)+1-i\right]}\\
&\qquad\cdot\prod_{i=2}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)+\frac{1}{2}h(N_1+\cdots+N_g)+1-i\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)+1-i\right]}\\
&\qquad\left.\cdot\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g+1-i)\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g+1-i)+\frac{1}{2}h(N_1+\cdots+N_g)\right]}\right\}\\
&\qquad\cdot\prod_{i=1}^{p}\frac{\Gamma\left[\frac{1}{2}(n+1-i+hN)\right]\Gamma\left[\frac{1}{2}(N-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]\Gamma\left[\frac{1}{2}(N+hN-i)\right]}\\
&=\prod_{i=1}^{p}\left\{\prod_{g=1}^{q}\frac{\Gamma\left[\frac{1}{2}(N_g+hN_g-i)\right]}{\Gamma\left[\frac{1}{2}(N_g-i)\right]}\right\}\frac{\Gamma\left[\frac{1}{2}(N-i)\right]}{\Gamma\left[\frac{1}{2}(N+hN-i)\right]}\\
&=\frac{\Gamma_p(\frac{1}{2}n)}{\Gamma_p(\frac{1}{2}n+\frac{1}{2}hN)}\prod_{g=1}^{q}\frac{\Gamma_p\left[\frac{1}{2}(n_g+hN_g)\right]}{\Gamma_g(\frac{1}{2}n_g)}.
\end{aligned}$$

We summarize in the following theorem:

Theorem 10.4.4. *Let V_1 be the criterion defined by* (10) *of Section* 10.2 *for testing the hypothesis that $H_1:\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_q$, where A_g is n_g times the sample covariance matrix and n_g+1 is the size of the sample from the gth population; let W be the criterion defined by* (13) *for testing the hypothesis $H:\boldsymbol{\mu}_1=\cdots=\boldsymbol{\mu}_q$ and H_1, where $\boldsymbol{B}=\boldsymbol{A}+\sum_g N_g(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})'$. The hth moment of V_1 when H_1 is true is given by* (15). *The hth moment of W, the criterion for testing H, is given by* (16).

This theorem was first proved by Wilks (1932). See Problem 10.5 for an alternative approach.

If p is even, say $p = 2r$, we can use the duplication formula for the gamma function $[\Gamma(\alpha + \frac{1}{2})\Gamma(\alpha + 1) = \sqrt{\pi}\Gamma(2\alpha + 1)2^{-2\alpha}]$. Then

$$(17)\quad \mathscr{E}V_1^h = \prod_{j=1}^{r}\left\{\left[\prod_{g=1}^{q}\frac{\Gamma(n_g + hn_g + 1 - 2j)}{\Gamma(n_g + 1 - 2j)}\right]\frac{\Gamma(n + 1 - 2j)}{\Gamma(n + hn + 1 - 2j)}\right\}$$

and

$$(18)\quad \mathscr{E}W^h = \prod_{j=1}^{r}\left\{\left[\prod_{g=1}^{q}\frac{\Gamma(n_g + hN_g + 1 - 2j)}{\Gamma(n_g + 1 - 2j)}\right]\frac{\Gamma(N - 2j)}{\Gamma(N + hN - 2j)}\right\}.$$

In principle the distributions of the factors can be integrated to obtain the distributions of V_1 and W. In Section 10.6 we consider V_1 when $p = 2$, $q = 2$ (the case of $p = 1, q = 2$ being a function of an F-statistic). In other cases, the integrals become unmanageable. To find probabilities we use the asymptotic expansion given in the next section. Box (1949) has given some other approximate distributions.

10.4.3. Step-down Tests

The characterizations of the distributions of the criteria in terms of independent factors suggests testing the hypotheses H_1 and H by testing component hypotheses sequentially. First, we consider testing $H_1 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ for $q = 2$. Let

$$(19)\quad X_{(i)}^{(g)} = \begin{pmatrix} X_{(i-1)}^{(g)} \\ X_i^{(g)} \end{pmatrix}, \qquad \boldsymbol{\mu}_{(i)}^{(g)} = \begin{pmatrix} \boldsymbol{\mu}_{(i-1)}^{(g)} \\ \mu_i^{(g)} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_i^{(g)} = \begin{bmatrix} \boldsymbol{\Sigma}_{(i-1)}^{(g)} & \boldsymbol{\sigma}_{(i)}^{(g)} \\ \boldsymbol{\sigma}_{(i)}^{(g)\prime} & \sigma_{ii}^{(g)} \end{bmatrix},$$
$$i = 2, \dots, p, \quad g = 1, 2.$$

The conditional distribution of $X_i^{(g)}$ given $X_{(i-1)}^{(g)} = x_{(i-1)}^{(g)}$ is

$$(20)\qquad N\left[\mu_i^{(g)} + \boldsymbol{\sigma}_{(i)}^{(g)\prime}\boldsymbol{\Sigma}_{i-1}^{-1}\left(x_{(i-1)}^{(g)} - \boldsymbol{\mu}_{(i-1)}^{(g)}\right), \sigma_{ii\cdot i-1}^{(g)}\right],$$

where $\sigma_{ii\cdot i-1}^{(g)} = \sigma_{ii}^{(g)} - \boldsymbol{\sigma}_{(i)}^{(g)\prime}\boldsymbol{\Sigma}_{i-1}^{(g)}\boldsymbol{\sigma}_{(i)}^{(g)}$. It is assumed that the components of X have been numbered in descending order of importance. At the ith step the component hypothesis $\sigma_{ii\cdot i-1}^{(1)} = \sigma_{ii\cdot i-1}^{(2)}$ is tested at significance level ε_i by means of an F-test based on $s_{ii\cdot i-1}^{(1)}/s_{ii\cdot i-1}^{(2)}$; S_1 and S_2 are partitioned like $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$. If that hypothesis is accepted, then the hypothesis $\boldsymbol{\sigma}_{(i)}^{(1)} = \boldsymbol{\sigma}_{(i)}^{(2)}$ (or $\boldsymbol{\Sigma}_{i-1}^{(1)-1}\boldsymbol{\sigma}_{(i)}^{(1)} = \boldsymbol{\Sigma}_{i-1}^{(2)-1}\boldsymbol{\sigma}_{(i)}^{(2)}$) is tested at significance level δ_i on the assumption that $\boldsymbol{\Sigma}_{i-1}^{(1)} = \boldsymbol{\Sigma}_{i-1}^{(2)}$ (a hypothesis previously accepted). The criterion is

$$(21)\quad \frac{\left(S_{i-1}^{(1)-1}s_{(i)}^{(1)} - S_{i-1}^{(2)-1}s_{(i)}^{(2)}\right)'\left(S_{i-1}^{(1)-1} + S_{i-1}^{(2)-1}\right)^{-1}\left(S_{i-1}^{(1)-1}s_{(i)}^{(1)} - S_{i-1}^{(2)-1}s_{(i)}^{(2)}\right)}{(i-1)s_{ii\cdot i-1}},$$

where $(n_1+n_2-2i+2)s_{ii\cdot i-1}=(n_1-i+1)s^{(1)}_{ii\cdot i-1}+(n_2-i+1)s^{(2)}_{ii\cdot i-1}$. Under the null hypothesis (21) has the F-distribution with $i-1$ and n_1+n_2-2i+2 degrees of freedom. If this hypothesis is accepted, the $(i+1)$st step is taken. The overall hypothesis $\mathbf{\Sigma}_1=\mathbf{\Sigma}_2$ is accepted if the $2p-1$ component hypotheses are accepted. (At the first step, $\boldsymbol{\sigma}^{(g)}_{(1)}$ is vacuous.) The overall significance level is

$$(22)\qquad 1-\prod_{i=1}^{p}(1-\varepsilon_i)\prod_{i=2}^{p}(1-\delta_i).$$

If any component null hypothesis is rejected, the overall hypothesis is rejected.

If $q>2$, the null hypotheses $H_1:\mathbf{\Sigma}_1=\cdots=\mathbf{\Sigma}_q$ is broken down into a sequence of hypotheses $[1/(g-1)](\mathbf{\Sigma}_1+\cdots+\mathbf{\Sigma}_{g-1})=\mathbf{\Sigma}_g$ and tested sequentially. Each such matrix hypothesis is tested as $\mathbf{\Sigma}_1=\mathbf{\Sigma}_2$ with S_2 replaced by S_g and S_1 replaced by $[1/(n_1+\cdots+n_{g-1})](A_1+\cdots+A_{g-1})$.

In the case of the hypothesis H, consider first $q=2$, $\mathbf{\Sigma}_1=\mathbf{\Sigma}_2$, and $\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)}$. One can test $\mathbf{\Sigma}_1=\mathbf{\Sigma}_2$. The steps for testing $\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)}$ consist of t-tests for $\mu_i^{(1)}=\mu_i^{(2)}$ based on the conditional distribution of $X_i^{(1)}$ and $X_i^{(2)}$ given $x^{(1)}_{(i-1)}$ and $x^{(2)}_{(i-1)}$. Alternatively one can test in sequence the equality of the conditional distributions of $X_i^{(1)}$ and $X_i^{(2)}$ given $x^{(1)}_{(i-1)}$ and $x^{(2)}_{(i-1)}$.

For $q>2$, the hypothesis $\mathbf{\Sigma}_1=\cdots=\mathbf{\Sigma}_q$ can be tested, and then $\boldsymbol{\mu}_1=\cdots=\boldsymbol{\mu}_q$. Alternatively, one can test $[1/(g-1)](\mathbf{\Sigma}_1+\cdots+\mathbf{\Sigma}_{g-1})=\mathbf{\Sigma}_g$ and $[1/(g-1)](\boldsymbol{\mu}^{(1)}+\cdots+\boldsymbol{\mu}^{(g-1)})=\boldsymbol{\mu}^{(g)}$.

10.5. ASYMPTOTIC EXPANSIONS OF THE DISTRIBUTIONS OF THE CRITERIA

Again we make use of Theorem 8.5.1 to obtain asymptotic expansions of the distributions of V_1 and of λ. We assume that $n_g=k_g n$, where $\sum_{g=1}^q k_g=1$. The asymptotic expansion is in terms of n increasing with $k_1,\dots,k_q$ fixed. (We could assume only $\lim n_g/n=k_g>0$.)

The hth moment of

$$(1)\qquad \lambda_1^*=V_1\cdot\frac{n^{\frac12 pn}}{\prod_{g=1}^q n_g^{\frac12 pn_g}}=V_1\cdot\prod_{g=1}^q\left(\frac{n}{n_g}\right)^{\frac12 pn_g}=\left[\prod_{g=1}^q\left(\frac{1}{k_g}\right)^{k_g}\right]^{\frac12 pn}V_1$$

is

$$(2)\qquad \mathscr{E}\lambda_1^{*h}=K\left(\frac{\prod_{j=1}^p\left(\frac12 n\right)^{\frac12 n}}{\prod_{g=1}^q\prod_{i=1}^p\left(\frac12 n_g\right)^{\frac12 n_g}}\right)^h\frac{\prod_{g=1}^q\prod_{i=1}^p\Gamma\left[\frac12 n_g(1+h)+\frac12(1-i)\right]}{\prod_{j=1}^p\Gamma\left[\frac12 n(1+h)+\frac12(1-j)\right]}.$$

This is of the form of (1) of Section 8.6 with

$$
\begin{array}{llll}
b = p, & y_j = \tfrac{1}{2}n, & \eta_j = \tfrac{1}{2}(1-j), & j = 1, \ldots, p, \\
a = pq, & x_k = \tfrac{1}{2}n_g, & k = (g-1)p+1, \ldots, gp, & g = 1, \ldots, q, \\
\xi_k = \tfrac{1}{2}(1-i), & & k = i, p+i, \ldots, (q-1)p+i, & i = 1, \ldots, p.
\end{array} \tag{3}
$$

Then

$$
\begin{aligned}
f &= -2\left[\sum \xi_k - \sum \eta_j - \tfrac{1}{2}(a-b)\right] \\
&= -\left[q\sum_{i=1}^{p}(1-i) - \sum_{j=1}^{p}(1-j) - (qp-p)\right] \\
&= -\left[-q\tfrac{1}{2}p(p-1) + \tfrac{1}{2}p(p-1) - (q-1)p\right] \\
&= \tfrac{1}{2}(q-1)p(p+1),
\end{aligned} \tag{4}
$$

$\varepsilon_j = \frac{1}{2}(1-\rho)n$, $j = 1, \ldots, p$, and $\beta_k = \frac{1}{2}(1-\rho)n_g = \frac{1}{2}(1-\rho)k_g n$, $k = (g-1)p + 1, \ldots, gp$.

In order to make the second term in the expansion vanish, we take ρ as

$$
\rho = 1 - \left(\sum_{g=1}^{q}\frac{1}{n_g} - \frac{1}{n}\right)\frac{2p^2+3p-1}{6(p+1)(q-1)}. \tag{5}
$$

Then

$$
\omega_2 = \frac{p(p+1)\left[(p-1)(p+2)\left(\displaystyle\sum_{g=1}^{q}\frac{1}{n_g^2} - \frac{1}{n^2}\right) - 6(q-1)(1-\rho)^2\right]}{48\rho^2}. \tag{6}
$$

Thus

$$
\begin{aligned}
&\Pr\{-2\rho\log\lambda_1^* \le z\} \\
&\quad = \Pr\{\chi_f^2 \le z\} + \omega_2\left[\Pr\{\chi_{f+4}^2 \le z\} - \Pr\{\chi_f^2 \le z\}\right] + O(n^{-3}).
\end{aligned} \tag{7}
$$

Let $\lambda = WN^{\frac{1}{2}pN}\prod_{g=1}^{q}N_g^{-\frac{1}{2}pN_g}$. The hth moment is

$$
\mathscr{E}\lambda^h = K\left[\frac{\prod_{j=1}^{p}\left(\frac{1}{2}N\right)^{\frac{1}{2}N}}{\prod_{g=1}^{q}\prod_{i=1}^{p}\left(\frac{1}{2}N_g\right)^{\frac{1}{2}N_g}}\right]^h \frac{\prod_{g=1}^{q}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}N_g(1+h) - \frac{1}{2}i\right]}{\prod_{j=1}^{p}\Gamma\left[\frac{1}{2}N(1+h) - \frac{1}{2}j\right]}. \tag{8}
$$

This is the form (1) of Section 8.5 with

$$
\begin{aligned}
&b = p, && y_j = \tfrac{1}{2}N = \tfrac{1}{2}\sum_{g=1}^{q} N_g, && n_j = -\tfrac{1}{2}j, && j = 1, \dots, p,\\
(9)\qquad &a = pq, && x_k = \tfrac{1}{2}N_g, && k = (g-1)p+1, \dots, gp, && g = 1, \dots, q,\\
&\xi_k = -\tfrac{1}{2}i, && && k = i, p+i, \dots, (q-1)p+i, && i = 1, \dots, p.
\end{aligned}
$$

The basic number of degrees of freedom is $f = \frac{1}{2}p(p+3)(q-1)$. We use (11) of Section 8.5 with $\beta_k = (1-\rho)x_k$ and $\varepsilon_j = (1-\rho)y_j$. To make $\omega_1 = 0$, we take

$$
\rho = 1 - \left(\sum_{g=1}^{q} \frac{1}{N_g} - \frac{1}{N} \right) \frac{2p^2 + 9p + 11}{6(q-1)(p+3)}. \tag{10}
$$

Then

$$
\omega_2 = \frac{p(p+3)}{48\rho^2} \left[\sum_{g=1}^{q} \left(\frac{1}{N_g^2} - \frac{1}{N^2} \right)(p+1)(p+2) - 6(1-\rho)^2(q-1) \right]. \tag{11}
$$

The asymptotic expansion of the distribution of $-2\rho \log \lambda$ is

$$
\begin{aligned}
(12)\quad &\Pr\{-2\rho \log \lambda \le z\}\\
&= \Pr\{\chi_f^2 \le z\} + \omega_2 \left[\Pr\{\chi_{f+4}^2 \le z\} - \Pr\{\chi_f^2 \le z\} \right] + O(n^{-3}).
\end{aligned}
$$

Box (1949) considered the case of λ_1^* in considerable detail. In addition to this expansion he considered the use of (13) of Section 8.6. He also gave an F-approximation.

As an example, we use one given by E. S. Pearson and Wilks (1933). The measurements are made on tensile strength (X_1) and hardness (X_2) of aluminum die castings. There are 12 observations in each of five samples. The observed sums of squares and cross-products in the five samples are

$$
\begin{aligned}
A_1 &= \begin{pmatrix} 78.948 & 214.18 \\ 214.18 & 1247.18 \end{pmatrix},\\
A_2 &= \begin{pmatrix} 223.695 & 657.62 \\ 657.62 & 2519.31 \end{pmatrix},\\
A_3 &= \begin{pmatrix} 57.448 & 190.63 \\ 190.63 & 1241.78 \end{pmatrix},\\
A_4 &= \begin{pmatrix} 187.618 & 375.91 \\ 375.91 & 1473.44 \end{pmatrix},\\
A_5 &= \begin{pmatrix} 88.456 & 259.18 \\ 259.18 & 1171.73 \end{pmatrix},
\end{aligned} \tag{13}
$$

and the sum of these is

$$\sum A_i = \begin{pmatrix} 636.165 & 1697.52 \\ 1697.52 & 7653.44 \end{pmatrix}. \tag{14}$$

The $-\log \lambda_1^*$ is 5.399. To use the asymptotic expansion we find $\rho = 152/165 = 0.9212$ and $\omega_2 = 0.0022$. Since ω_2 is small, we can consider $-2\rho \log \lambda_1^*$ as χ^2 with 12 degrees of freedom. Our observed criterion, therefore, is clearly not significant.

Table B.5 [due to Korin (1969)] gives 5% significance points for $-2\log \lambda_1^*$ for $N_1 = \cdots = N_q$ for various q, small values of N_g, and $p = 2(1)6$.

The limiting distribution of the criterion (19) of Section 10.1 is also χ_f^2. An asymptotic expansion of the distribution was given by Nagao (1973b) to terms of order $1/n$ involving χ^2-distributions with f, $f+2$, $f+4$, and $f+6$ degrees of freedom.

10.6. THE CASE OF TWO POPULATIONS

10.6.1. Invariant Tests

When $q = 2$, the null hypothesis H_1 is $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. It is invariant with respect to transformations

$$\boldsymbol{x}^{*(1)} = \boldsymbol{C}\boldsymbol{x}^{(1)} + \boldsymbol{\nu}^{(1)}, \qquad \boldsymbol{x}^{*(2)} = \boldsymbol{C}\boldsymbol{x}^{(2)} + \boldsymbol{\nu}^{(2)}, \tag{1}$$

where $\boldsymbol{C}$ is nonsingular. The maximal invariant of the parameters under the transformation of locations ($\boldsymbol{C} = \boldsymbol{I}$) is the pair of covariance matrices $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$, and the maximal invariant of the sufficient statistics $\bar{\boldsymbol{x}}^{(1)}, \boldsymbol{S}_1, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S}_2$ is the pair of matrices $\boldsymbol{S}_1, \boldsymbol{S}_2$ (or equivalently $\boldsymbol{A}_1, \boldsymbol{A}_2$). The transformation (1) induces the transformations $\boldsymbol{\Sigma}_1^* = \boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}'$, $\boldsymbol{\Sigma}_2^* = \boldsymbol{C}\boldsymbol{\Sigma}_2\boldsymbol{C}'$, $\boldsymbol{S}_1^* = \boldsymbol{C}\boldsymbol{S}_1\boldsymbol{C}'$, and $\boldsymbol{S}_2^* = \boldsymbol{C}\boldsymbol{S}_2\boldsymbol{C}'$. The roots $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ of

$$|\boldsymbol{\Sigma}_1 - \lambda\boldsymbol{\Sigma}_2| = 0 \tag{2}$$

are invariant under these transformations since

$$|\boldsymbol{\Sigma}_1^* - \lambda\boldsymbol{\Sigma}_2^*| = |\boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}' - \lambda\boldsymbol{C}\boldsymbol{\Sigma}_2\boldsymbol{C}'| = |\boldsymbol{C}\boldsymbol{C}'| \cdot |\boldsymbol{\Sigma}_1 - \lambda\boldsymbol{\Sigma}_2|. \tag{3}$$

Moreover, the roots are the only invariants because there exists a nonsingular matrix $\boldsymbol{C}$ such that

$$\boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}' = \boldsymbol{\Lambda}, \qquad \boldsymbol{C}\boldsymbol{\Sigma}_2\boldsymbol{C}' = \boldsymbol{I}, \tag{4}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix with λ_i as the ith diagonal element, $i = 1, \ldots, p$. (See Theorem A.2.2 of the Appendix.) Similarly, the maximal invariants of S_1 and S_2 are the roots $l_1 \geq l_2 \geq \cdots \geq l_p$ of

$$|S_1 - lS_2| = 0. \tag{5}$$

Theorem 10.6.1. *The maximal invariant of the parameters of $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_1)$ and $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_2)$ under the transformation* (1) *is the set of roots $\lambda_1 \geq \cdots \geq \lambda_p$ of* (2). *The maximal invariant of the sufficient statistics $\bar{\mathbf{x}}^{(1)}, \boldsymbol{S}_1, \bar{\mathbf{x}}^{(2)}, \boldsymbol{S}_2$ is the set of roots $l_1 \geq \cdots \geq l_p$ of* (5).

Any invariant test criterion can be expressed in terms of the roots $l_1, \ldots, l_p$. The criterion V_1 is $n_1^{\frac{1}{2}pn_1} n_2^{\frac{1}{2}pn_2}$ times

$$\frac{|S_1|^{\frac{1}{2}n_1}|S_2|^{\frac{1}{2}n_2}}{|n_1 S_1 + n_2 S_2|^{\frac{1}{2}n}} = \frac{|L|^{\frac{1}{2}n_1}|I|^{\frac{1}{2}n_2}}{|n_1 L + n_2 I|^{\frac{1}{2}n}} = \prod_{i=1}^{p} \frac{l_i^{\frac{1}{2}n_1}}{(n_1 l_i + n_2)^{\frac{1}{2}n}}, \tag{6}$$

where $\boldsymbol{L}$ is the diagonal matrix with l_i as the ith diagonal element. The null hypothesis is rejected if the smaller roots are too small or if the larger roots are too large, or both.

The null hypothesis is that $\lambda_1 = \cdots = \lambda_p = 1$. Any useful invariant test of the null hypothesis has a rejection region in the space of $l_1, \ldots, l_p$ that includes the points that in some sense are far from $l_1 = \cdots = l_p = 1$. The power of an invariant test depends on the parameters through the roots $\lambda_1, \ldots, \lambda_p$.

The criterion (19) of Section 10.2 is (with $nS = n_1 S_1 + n_2 S_2$)

$$\begin{aligned}
(7) \qquad & \tfrac{1}{2}n_1 \operatorname{tr}\left[(S_1 - S)S^{-1}\right]^2 + \tfrac{1}{2}n_2 \operatorname{tr}\left[(S_2 - S)S^{-1}\right]^2 \\
&= \tfrac{1}{2}n_1 \operatorname{tr}\left[C(S_1 - S)C'(CSC')^{-1}\right]^2 \\
&\quad + \tfrac{1}{2}n_2 \operatorname{tr}\left[C(S_2 - S)C'(CSC')^{-1}\right]^2 \\
&= \tfrac{1}{2}n_1 \operatorname{tr}\left[\left\{L - \left(\frac{n_1}{n}L + \frac{n_2}{n}I\right)\right\}\left(\frac{n_1}{n}L + \frac{n_2}{n}I\right)^{-1}\right]^2 \\
&\quad + \tfrac{1}{2}n_2 \operatorname{tr}\left[\left\{I - \left(\frac{n_1}{n}L + \frac{n_2}{n}I\right)\right\}\left(\frac{n_1}{n}L + \frac{n_2}{n}I\right)^{-1}\right]^2 \\
&= \tfrac{1}{2}n_1 n_2 n \sum_{i=1}^{p} \frac{(l_i - 1)^2}{(n_1 l_i + n_2)^2}.
\end{aligned}$$

This criterion is a measure of how close $l_1, \ldots, l_p$ are to 1; the hypothesis is rejected if the measure is too large. Under the null hypothesis, (7) has the χ^2-distribution with $f = \frac{1}{2}p(p+1)$ degrees of freedom as $n_1 \to \infty$, $n_2 \to \infty$, and n_1/n_2 approaches a positive constant. Nagao (1973b) gives an asymptotic expansion of this distribution to terms of order $1/n$.

Roy (1953) suggested a test based on the largest and smallest roots, l_1 and l_p. The procedure is to reject the null hypothesis if $l_1 > k_1$ or if $l_p < k_p$, where k_1 and k_p are chosen so that the probability of rejection when $\Lambda = I$ is the desired significance level. Roy (1957) proposed determining k_1 and k_p so that the test is locally unbiased, that is, that the power functions have a relative minimum at $\Lambda = I$. Since it is hard to determine k_1 and k_p on this basis, other proposals have been made. The limit k_1 can be determined so that $\Pr\{l_1 > k_1 | H_1\}$ is one-half the significance level, or $\Pr\{l_p < k_p | H_1\}$ is one-half of the significance level, or $k_1 + k_p = 2$, or $k_1 k_p = 1$. In principle k_1 and k_p can be determined from the distribution of the roots, given in Section 13.2. Schuurmann, Waikar, and Krishnaiah (1975) and Chu and Pillai (1979) give some exact values of k_1 and k_p for small values of p. Chu and Pillai (1979) also make some power comparisons of several test procedures.

In the case of $p = 1$ the only invariant of the sufficient statistics is S_1/S_2, which is the usual F-statistic with n_1 and n_2 degrees of freedom. The criterion V_1 is $(A_1/A_2)^{\frac{1}{2}n_1}[1 + A_1/A_2)]^{-\frac{1}{2}n}$; the critical region V_1 less than a constant is equivalent to a two-tailed critical region for the F-statistic. The quantity $n(B - A)/A$ has an independent F-distribution with 1 and n degrees of freedom. (See Section 10.3.)

In the case of $p = 2$, the hth moment of V_1 is, from (15) of Section 10.4,

$$\begin{aligned}(8)\qquad \mathscr{E} V_1^h &= \frac{\Gamma(n_1 + hn_1 - 1)\Gamma(n_2 + hn_2 - 1)\Gamma(n - 1)}{\Gamma(n_1 - 1)\Gamma(n_2 - 1)\Gamma(n + hn - 1)} \\ &= \mathscr{E}\left[X_1^{n_1}(1 - X_1)^{n_2} X_2^{n_1 + n_2}\right]^h,\end{aligned}$$

where X_1 and X_2 are independently distributed according to $\beta(x | n_1 - 1, n_2 - 1)$ and $\beta(x | n_1 + n_2 - 2, 1)$, respectively. Then $\Pr\{V_1 \le v\}$ can be found by integration. (See Problems 10.8 and 10.9.)

Anderson (1965a) has shown that a confidence interval for $a'\Sigma_1 a / a'\Sigma_2 a$ for all a with confidence coefficient ε is given by $(l_p/U, l_1/L)$, where $\Pr\{(n_2 - p + 1)L \le n_2 F_{n_1, n_2 - p + 1}\} \Pr\{(n_1 - p + 1)F_{n_1 - p + 1, n_2} \le n_1 U\} = 1 - \varepsilon$.

10.6.2. Components of Variance

In Section 8.8 we considered what is equivalent to the one-way analysis of variance with fixed effects. We can write the model in the balanced case $(N_1 = N_2 = \cdots = N_q)$ as

$$\begin{aligned}(9)\qquad X_\alpha^{(g)} &= \boldsymbol{\mu}^{(g)} + U_\alpha^{(g)} \\ &= \boldsymbol{\mu} + \boldsymbol{\nu}_g + U_\alpha^{(g)}, \qquad \alpha = 1, \dots, M, \quad g = 1, \dots, q.\end{aligned}$$

where $\mathscr{E}U^{(g)} = \mathbf{0}$ and $\mathscr{E}U^{(g)}U^{(g)\prime} = \mathbf{\Sigma}$, $\boldsymbol{\nu}_g = \boldsymbol{\mu}^{(g)} - \boldsymbol{\mu}$, and $\boldsymbol{\mu} = (1/q)\sum_{g=1}^q \boldsymbol{\mu}^{(g)}$ $(\sum_{g=1}^q \boldsymbol{\nu}_g = \mathbf{0})$. The null hypothesis of no effect is $\boldsymbol{\nu}_1 = \cdots = \boldsymbol{\nu}_q = \mathbf{0}$. Let $\bar{x}^{(g)} = (1/M)\sum_{\alpha=1}^M x_\alpha^{(g)}$ and $\bar{x} = (1/q)\sum_{g=1}^q \bar{x}^{(g)}$. The analysis of variance table is

Source	Sum of Squares	Degrees of Freedom
Effect	$H = M \sum_{g=1}^{q} (\bar{x}^{(g)} - \bar{x})(\bar{x}^{(g)} - \bar{x})'$	$q - 1$
Error	$G = \sum_{g=1}^{q} \sum_{\alpha=1}^{M} (x_\alpha^{(g)} - \bar{x}^{(g)})(x_\alpha^{(g)} - \bar{x}^{(g)})'$	$q(M-1)$
Total	$\sum_{g=1}^{q} \sum_{\alpha=1}^{M} (x_\alpha^{(g)} - \bar{x})(x_\alpha^{(g)} - \bar{x})'$	$qM - 1$

Invariant tests of the null hypothesis of no effect are based on the roots of $|H - mG| = 0$ or of $|S_h - lS_e| = 0$, where $S_h = [1/(q-1)]H$ and $S_e = [1/q(M-1)]G$. The null hypothesis is rejected if one or more of the roots is too large. The error matrix G has the distribution $W(\mathbf{\Sigma}, q(M-1))$. The effects matrix H has the distribution $W(\mathbf{\Sigma}, q-1)$ when the null hypothesis is true and has the noncentral Wishart distribution when the null hypothesis is not true; its expected value is

$$\mathscr{E}H = (q-1)\mathbf{\Sigma} + M \sum_{g=1}^{q} (\boldsymbol{\mu}^{(g)} - \boldsymbol{\mu})(\boldsymbol{\mu}^{(g)} - \boldsymbol{\mu})' \tag{10}$$
$$= (q-1)\mathbf{\Sigma} + M \sum_{g=1}^{q} \boldsymbol{\nu}_g \boldsymbol{\nu}_g'.$$

The MANOVA model with random effects is

$$X_\alpha^{(g)} = \boldsymbol{\mu} + V_g + U_\alpha^{(g)}, \qquad \alpha = 1, \dots, M, \quad g = 1, \dots, q, \tag{11}$$

where V_g has the distribution $N(\mathbf{0}, \mathbf{\Theta})$. Then $X_\alpha^{(g)}$ has the distribution $N(\boldsymbol{\mu}, \mathbf{\Sigma} + \mathbf{\Theta})$. The null hypothesis of no effect is

$$\mathbf{\Theta} = \mathbf{0}. \tag{12}$$

In this model G again has the distribution $W(\mathbf{\Sigma}, q(M-1))$. Since $\bar{X}^{(g)} = \boldsymbol{\mu} + V_g + \bar{U}^{(g)}$ has the distribution $N(\boldsymbol{\mu}, (1/M)\mathbf{\Sigma} + \mathbf{\Theta})$, H has the distribution $W(\mathbf{\Sigma} + M\mathbf{\Theta}, q-1)$. The null hypothesis (12) is equivalent to the equality o the covariance matrices in these two Wishart distributions; that is, $\Sigma = \Sigma + M\Theta$. The matrices G and H correspond to A_1 and A_2 in Section 10.6.1. However, here the alternative to the null hypothesis is that $(\Sigma + M\Theta) - \Sigma$ is positive semidefinite, rather than $\Sigma_1 \neq \Sigma_2$. The null hypothesis is to be rejected if H is too large relative to G. Any of the criteria presented in Section 10.2 can be used to test the null hypothesis here, and its distribution under the null hypothesis is the same as given there.

The likelihood ratio criterion for testing $\Theta = 0$ must take into account the fact that Θ is positive semidefinite; that is, the maximum likelihood estimators of Σ and $\Sigma + M\Theta$ under Ω must be such that the estimator of Θ is positive semidefinite. Let $l_1 > l_2 > \cdots > l_p$ be the roots of

$$\left| H - l\frac{1}{M-1}G \right| = 0. \tag{13}$$

(Note $\{1/[q(M-1)]\}G$ and $(1/q)H$ maximize the likelihood without regard to Θ being positive definite.) Let $l_1^* = l_i$ if $l_i > 1$, and let $l_i^* = 1$ if $l_i \leq 1$. Then the likelihood ratio criterion for testing the hypothesis $\Theta = 0$ against the alternative Θ positive semidefinite and $\Theta \neq 0$ is

$$M^{\frac{1}{2}qMp} \prod_{i=1}^{p} \frac{l_i^{*\frac{1}{2}q}}{(l_i^* + M - 1)^{\frac{1}{2}qM}} = M^{\frac{1}{2}qMk} \prod_{i=1}^{k} \frac{l_i^{\frac{1}{2}q}}{(l_i + M - 1)^{\frac{1}{2}qM}}, \tag{14}$$

where k is the number of roots of (13) greater than 1. [See Anderson (1946b), (1984a), (1989a), Morris and Olkin (1964), and Klotz and Putter (1969).]

10.7. TESTING THE HYPOTHESIS THAT A COVARIANCE MATRIX IS PROPORTIONAL TO A GIVEN MATRIX; THE SPHERICITY TEST

10.7.1. The Hypothesis

In many statistical analyses that are considered univariate, the assumption is made that a set of random variables are independent and have a common variance. In this section we consider a test of these assumptions based on repeated sets of observations.

More precisely, we use a sample of p-component vectors $x_1, \ldots, x_N$ from $N(\mu, \Sigma)$ to test the hypothesis $H: \Sigma = \sigma^2 I$, where σ^2 is not specified. The hypothesis can be given an algebraic interpretation in terms of the characteristic roots of Σ, that is, the roots of

$$|\Sigma - \phi I| = 0. \tag{1}$$

The hypothesis is true if and only if all the roots of (1) are equal.† Another way of putting it is that the arithmetic mean of roots $\phi_1, \ldots, \phi_p$ is equal to the geometric mean, that is,

$$\frac{\prod_{i=1}^{p} \phi_i^{1/p}}{\sum_{i=1}^{p} \phi_i / p} = \frac{|\Sigma|^{1/p}}{\operatorname{tr} \Sigma / p} = 1. \tag{2}$$

The lengths squared of the principal axes of the ellipsoids of constant density are proportional to the roots ϕ_i (see Chapter 11); the hypothesis specifies that these are equal, that is, that the ellipsoids are spheres.

The hypothesis H is equivalent to the more general form $\boldsymbol{\Psi} = \sigma^2 \boldsymbol{\Psi}_0$, with $\boldsymbol{\Psi}_0$ specified, having observation vectors $y_1, \ldots, y_N$ from $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$. Let C be a matrix such that

$$C \boldsymbol{\Psi}_0 C' = I, \tag{3}$$

and let $\boldsymbol{\mu}^* = C\boldsymbol{\nu}$, $\boldsymbol{\Sigma}^* = C \boldsymbol{\Psi} C'$, $x_\alpha^* = C y_\alpha$. Then $x_1^*, \ldots, x_N^*$ are observations from $N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, and the hypothesis is transformed into $H: \boldsymbol{\Sigma}^* = \sigma^2 I$.

10.7.2. The Criterion

In the canonical form the hypothesis H is a combination of the hypothesis $H_1: \boldsymbol{\Sigma}$ is diagonal or the components of X are independent and H_2: the diagonal elements of $\boldsymbol{\Sigma}$ are equal given that $\boldsymbol{\Sigma}$ is diagonal or the variances of the components of X are equal given that the components are independent. Thus by Lemma 10.3.1 the likelihood ratio criterion λ for H is the product of the criterion λ_1 for H_1 and λ_2 for H_2. From Section 9.2 we see that the criterion for H_1 is

$$\lambda_1 = \frac{|A|^{\frac{1}{2}N}}{\prod a_{ii}^{\frac{1}{2}N}} = |r_{ij}|^{\frac{1}{2}N}, \tag{4}$$

where

$$A = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' = (a_{ij}) \tag{5}$$

and $r_{ij} = a_{ij} / \sqrt{a_{ii} a_{jj}}$. We use the results of Section 10.2 to obtain λ_2 by considering the ith component of x_α as the αth observation from the ith population. (p here is q in Section 10.2; N here is N_g there; pN here is N

†This follows from the fact that $\boldsymbol{\Sigma} = O' \boldsymbol{\Phi} O$, where $\boldsymbol{\Phi}$ is a diagonal matrix with roots as diagonal elements and O is an orthogonal matrix.

there.) Thus

$$\lambda_2 = \frac{\prod_i \left[\sum_\alpha (x_{i\alpha} - \bar{x}_i)^2\right]^{\frac{1}{2}N}}{\left[\sum_{i,\alpha}(x_{i\alpha} - \bar{x}_i)^2/p\right]^{\frac{1}{2}pN}} \tag{6}$$

$$= \frac{\prod a_{ii}^{\frac{1}{2}N}}{(\operatorname{tr} A/p)^{\frac{1}{2}pN}}.$$

Thus the criterion for H is

$$\lambda = \lambda_1 \lambda_2 = \frac{|A|^{\frac{1}{2}N}}{(\operatorname{tr} A/p)^{\frac{1}{2}pN}}. \tag{7}$$

It will be observed that λ resembles (2). If $l_1, \ldots, l_p$ are the roots of

$$|S - lI| = 0, \tag{8}$$

where $S = (1/n)A$, the criterion is a power of the ratio of the geometric mean to the arithmetic mean,

$$\lambda = \left(\frac{\prod l_i^{1/p}}{\sum l_i/p}\right)^{\frac{1}{2}pN}. \tag{9}$$

Now let us go back to the hypothesis $\Psi = \sigma^2 \Psi_0$, given observation vectors $y_1, \ldots, y_N$ from $N(\nu, \Psi)$. In the transformed variables $\{x_\alpha^*\}$ the criterion is $|A^*|^{\frac{1}{2}N}(\operatorname{tr} A^*/p)^{-\frac{1}{2}pN}$, where

$$A^* = \sum_{\alpha=1}^{N} (x_\alpha^* - \bar{x}^*)(x_\alpha^* - \bar{x}^*)' \tag{10}$$

$$= C \sum_{\alpha=1}^{N} (y_\alpha - \bar{y})(y_\alpha - \bar{y})' C'$$

$$= CBC',$$

where

$$B = \sum_{\alpha=1}^{N} (y_\alpha - \bar{y})(y_\alpha - \bar{y})'. \tag{11}$$

From (3) we have $\boldsymbol{\Psi}_0 = \boldsymbol{C}^{-1}(\boldsymbol{C}')^{-1} = (\boldsymbol{C}'\boldsymbol{C})^{-1}$. Thus

$$|A^*| = \frac{|\boldsymbol{B}|}{|\boldsymbol{\Psi}_0|} = |\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}|,$$

$$\operatorname{tr} A^* = \operatorname{tr} \boldsymbol{CBC}' = \operatorname{tr} \boldsymbol{BC}'\boldsymbol{C} \tag{12}$$

$$= \operatorname{tr} \boldsymbol{B}\boldsymbol{\Psi}_0^{-1}.$$

The results can be summarized.

Theorem 10.7.1. *Given a set of p-component observation vectors* $\boldsymbol{y}_1, \dots, \boldsymbol{y}_N$ *from* $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$, *the likelihood ratio criterion for testing the hypothesis* $H: \boldsymbol{\Psi} = \sigma^2 \boldsymbol{\Psi}_0$, *where* $\boldsymbol{\Psi}_0$ *is specified and* σ^2 *is not specified, is*

$$\frac{|\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}|^{\frac{1}{2}N}}{\left(\operatorname{tr} \boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/p\right)^{\frac{1}{2}pN}}. \tag{13}$$

Mauchly (1940) gave this criterion and its moments under the null hypothesis.

The maximum likelihood estimator of σ^2 under the null hypothesis is $\operatorname{tr} \boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/(pN)$, which is $\operatorname{tr} A/(pN)$ in canonical form; an unbiased estimator is $\operatorname{tr} \boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/[p(N-1)]$ or $\operatorname{tr} A/[p(N-1)]$ in canonical form [Hotelling (1951)]. Then $\operatorname{tr} \boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/\sigma^2$ has the χ^2-distribution with $p(N-1)$ degrees of freedom.

10.7.3. The Distribution and Moments of the Criterion

The distribution of the likelihood ratio criterion under the null hypothesis can be characterized by the facts that $\lambda = \lambda_1\lambda_2$ and λ_1 and λ_2 are independent and by the characterizations of λ_1 and λ_2. As was observed in Section 7.6, when $\boldsymbol{\Sigma}$ is diagonal the correlation coefficients $\{r_{ij}\}$ are distributed independently of the variances $\{a_{ii}/(N-1)\}$. Since λ_1 depends only on $\{r_{ij}\}$ and λ_2 depends only on $\{a_{ii}\}$, they are independently distributed when the null hypothesis is true. Let $W = \lambda^{2/N}$, $W_1 = \lambda_1^{2/N}$, $W_2 = \lambda_2^{2/N}$. From Theorem 9.3.3, we see that W_1 is distributed as $\prod_{i=2}^p X_i$, where $X_2, \dots, X_p$ are independent and X_i has the density $\beta[x|\frac{1}{2}(n-i+1), \frac{1}{2}(i-1)]$, where $n = N-1$. From Theorem 10.4.2 with $W_2 = p^p V_1^{2/n}$, we find that W_2 is distributed as $p^p \prod_{j=2}^p Y_j^{j-1}(1-Y_j)$, where $Y_2, \dots, Y_p$ are independent and Y_j has the density $\beta(y|\frac{1}{2}n(j-1), \frac{1}{2}n)$. Then W is distributed as W_1W_2, where W_1 and W_2 are independent.

The moments of W can be found from this characterization or from Theorems 9.3.4 and 10.4.4. We have

(14)
$$\mathscr{E} W_1^h = \frac{\Gamma^p(\tfrac{1}{2}n)}{\Gamma^p(\tfrac{1}{2}n+h)} \frac{\Gamma_p(\tfrac{1}{2}n+h)}{\Gamma_p(\tfrac{1}{2}n)},$$

(15)
$$\mathscr{E} W_2^h = p^{hp} \frac{\Gamma^p(\tfrac{1}{2}n+h)\Gamma(\tfrac{1}{2}pn)}{\Gamma^p(\tfrac{1}{2}n)\Gamma(\tfrac{1}{2}pn+ph)}.$$

It follows that

(16)
$$\mathscr{E} W^h = p^{hp} \frac{\Gamma(\tfrac{1}{2}pn)}{\Gamma(\tfrac{1}{2}pn+ph)} \frac{\Gamma_p(\tfrac{1}{2}n+h)}{\Gamma_p(\tfrac{1}{2}n)}.$$

For $p=2$ we have

(17)
$$\begin{aligned}\mathscr{E} W^h &= 4^h \frac{\Gamma(n)}{\Gamma(n+2h)} \prod_{i=1}^{2} \frac{\Gamma[\tfrac{1}{2}(n+1-i)+h]}{\Gamma[\tfrac{1}{2}(n+1-i)]} \\ &= \frac{\Gamma(n)\Gamma(n-1+2h)}{\Gamma(n+2h)\Gamma(n-1)} = \frac{n-1}{n-1+2h} \\ &= (n-1)\int_0^1 z^{n-2+2h}\,dz,\end{aligned}$$

by use of the duplication formula for the gamma function. Thus W is distributed as Z^2, where Z has the density $(n-1)z^{n-2}$, and W has the density $\frac{1}{2}(n-1)w^{\frac{1}{2}(n-3)}$. The cdf is

(18)
$$\Pr\{W \le w\} = F(w) = w^{\frac{1}{2}(n-1)}.$$

This result can also be found from the joint distribution of l_1, l_2, the roots of (8). The density for $p = 3$, 4, and 6 has been obtained by Consul (1967b). See also Pillai and Nagarsenkar (1971).

10.7.4. Asymptotic Expansion of the Distribution

From (16) we see that the rth moment of $W^{\frac{1}{2}n} = Z$, say, is

(19)
$$\mathscr{E} Z^r = K p^{\frac{1}{2}npr} \frac{\prod_{i=1}^p \Gamma[\tfrac{1}{2}n(1+r)+\tfrac{1}{2}(1-i)]}{\Gamma[\tfrac{1}{2}pn(1+r)]}.$$

This is of the form of (1), Section 8.5, with

$$\text{(20)}\qquad \begin{aligned} &a = p, && x_k = \tfrac{1}{2}n, && \xi_k = \tfrac{1}{2}(1-k), && k = 1, \dots, p, \\ &b = 1, && y_1 = \tfrac{1}{2}np, && \eta_1 = 0. \end{aligned}$$

Thus the expansion of Section 8.5 is valid with $f = \frac{1}{2}p(p+1) - 1$. To make the second term in the expansion zero we take ρ so

$$\text{(21)}\qquad 1 - \rho = \frac{2p^2 + p + 2}{6pn}.$$

Then

$$\text{(22)}\qquad \omega_2 = \frac{(p+2)(p-1)(p-2)(2p^3 + 6p^2 + 3p + 2)}{288 p^2 n^2 \rho^2}.$$

Thus the cdf of W is found from

$$\begin{aligned} \text{(23)}\quad \Pr\{-2\rho \log Z \le z\} &= \Pr\{-n\rho \log W \le z\} \\ &= \Pr\{\chi_f^2 \le z\} + \omega_2\left(\Pr\{\chi_{f+4}^2 \le z\} - \Pr\{\chi_f^2 \le z\}\right) + O(n^{-3}). \end{aligned}$$

Factors $c(n, p, \varepsilon)$ have been tabulated in Table B.6 such that

$$\text{(24)}\qquad \Pr\left\{-n\rho \log W \le c(n, p, \varepsilon)\, \chi^2_{\frac{1}{2}p(n+1)-1}(\varepsilon)\right\} = \varepsilon.$$

Nagarsenkar and Pillai (1973a) have tables for W.

10.7.5. Invariant Tests

The null hypothesis $H: \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ is invariant with respect to transformations $\boldsymbol{X}^* = c\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{\nu}$, where c is a scalar and $\boldsymbol{Q}$ is an orthogonal matrix. The invariant of the sufficient statistic under shift of location is $\boldsymbol{A}$, the invariants of $\boldsymbol{A}$ under orthogonal transformations are the characteristic roots $l_1, \dots, l_p$, and the invariants of the roots under scale transformations are functions that are homogeneous of degree 0, such as the ratios of roots, say $l_1/l_2, \dots, l_{p-1}/l_p$. Invariant tests are based on such functions; the likelihood ratio criterion is such a function.

Nagao (1973a) proposed the criterion

$$
\begin{aligned}
(25)\qquad & \tfrac{1}{2}n\,\mathrm{tr}\left(S-\frac{\mathrm{tr}\,S}{p}I\right)\frac{p}{\mathrm{tr}\,S}\left(S-\frac{\mathrm{tr}\,S}{p}I\right)\frac{p}{\mathrm{tr}\,S}\\
&=\tfrac{1}{2}n\,\mathrm{tr}\left(\frac{p}{\mathrm{tr}\,S}S-I\right)^2=\tfrac{1}{2}n\left[\frac{p^2}{(\mathrm{tr}\,S)^2}\mathrm{tr}\,S^2-p\right]\\
&=\tfrac{1}{2}n\left[\frac{p}{\left(\sum_{i=1}^p l_i\right)^2}\sum_{i=1}^p l_i^2-p\right]=\tfrac{1}{2}n\frac{\sum_{i=1}^p\left(l_i-\bar{l}\right)^2}{\bar{l}^2},
\end{aligned}
$$

where $\bar{l}=\sum_{i=1}^p l_i/p$. The left-hand side of (25) is based on the loss function $L_q(\mathbf{\Sigma}, G)$ of Section 7.8; the right-hand side shows it is proportional to the square of the coefficient of variation of the characteristic roots of the sample covariance matrix S. Another criterion is l_1/l_p. Percentage points have been given by Krishnaiah and Schuurmann (1974).

10.7.6. Confidence Regions

Given observations $y_1, \dots, y_N$ from $N(\boldsymbol{\nu}, \mathbf{\Psi})$, we can test $\mathbf{\Psi}=\sigma^2\mathbf{\Psi}_0$ for any specified $\mathbf{\Psi}_0$. From this family of tests we can set up a confidence region for $\mathbf{\Psi}$. If any matrix is in the confidence region, all multiples of it are. This kind of confidence region is of interest if all components of y_α are measured in the same unit, but the investigator wants a region independent of this common unit. The confidence region of confidence $1-\varepsilon$ consists of all matrices $\mathbf{\Psi}^*$ satisfying

$$
(26)\qquad \frac{|B\mathbf{\Psi}^{*-1}|}{\left[(\mathrm{tr}\,B\mathbf{\Psi}^{*-1})/p\right]^p}\ge\lambda^{2/N}(\varepsilon),
$$

where $\lambda(\varepsilon)$ is the ε significance level for the criterion.

Consider the case of $p=2$. If the common unit of measurement is irrelevant, the investigator is interested in $\tau=\psi_{11}/\psi_{22}$ and $\rho=\psi_{12}/\sqrt{\psi_{11}\psi_{22}}$. In this case

$$
\begin{aligned}
(27)\qquad \mathbf{\Psi}^{-1}&=\frac{1}{\psi_{11}\psi_{22}(1-\rho^2)}\begin{pmatrix}\psi_{22} & -\rho\sqrt{\psi_{11}\psi_{22}}\\ -\rho\sqrt{\psi_{11}\psi_{22}} & \psi_{11}\end{pmatrix}\\
&=\frac{1}{\psi_{11}(1-\rho^2)}\begin{pmatrix}1 & -\rho\sqrt{\tau}\\ -\rho\sqrt{\tau} & \tau\end{pmatrix}.
\end{aligned}
$$

The region in terms of τ and ρ is

$$4\frac{\left(b_{11}b_{22}-b_{12}^2\right)\left(1-\rho^2\right)^{\tau}}{\left(b_{11}+\tau b_{22}-2\rho\sqrt{\tau}\,b_{12}\right)^2} \geq \lambda^{2/N}(\varepsilon). \tag{28}$$

Hickman (1953) has given an example of such a confidence region.

10.8. TESTING THE HYPOTHESIS THAT A COVARIANCE MATRIX IS EQUAL TO A GIVEN MATRIX

10.8.1. The Criteria

If Y is distributed according to $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$, we wish to test H_1 that $\boldsymbol{\Psi} = \boldsymbol{\Psi}_0$, where $\boldsymbol{\Psi}_0$ is a given positive definite matrix. By the argument of the preceding section we see that this is equivalent to testing the hypothesis $H_1 : \boldsymbol{\Sigma} = \boldsymbol{I}$, where $\boldsymbol{\Sigma}$ is the covariance matrix of a vector $\boldsymbol{X}$ distributed according to $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Given a sample $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$, the likelihood ratio criterion is

$$\lambda_1 = \frac{\max_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \boldsymbol{I})}{\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}, \tag{1}$$

where the likelihood function is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{1}{2}pN} |\boldsymbol{\Sigma}|^{-\frac{1}{2}N} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right]. \tag{2}$$

Results in Chapter 3 show that

$$\begin{aligned} \lambda_1 &= \frac{(2\pi)^{-\frac{1}{2}pN}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})\right]}{(2\pi)^{-\frac{1}{2}pN}|(1/N)\boldsymbol{A}|^{-\frac{1}{2}N} e^{-\frac{1}{2}pN}} \\ &= \left(\frac{e}{N}\right)^{\frac{1}{2}pN} |\boldsymbol{A}|^{\frac{1}{2}N} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}}, \end{aligned} \tag{3}$$

where

$$\boldsymbol{A} = \sum_{\alpha}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'. \tag{4}$$

Sugiura and Nagao (1968) have shown that the likelihood ratio test is biased, but the modified likelihood ratio test based on

$$\lambda_1^* = \left(\frac{e}{n}\right)^{\frac{1}{2}pn} |\boldsymbol{A}|^{\frac{1}{2}n} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}} = e^{\frac{1}{2}pn}\left(|\boldsymbol{S}|e^{-\operatorname{tr}\boldsymbol{S}}\right)^{\frac{1}{2}n}, \tag{5}$$

where $S = (1/n)A$, is unbiased. Note that

$$(6) \qquad -\frac{2}{n}\log \lambda_1^* = \operatorname{tr} S - \log|S| - p = L_l(I, S),$$

where $L_l(I, S)$ is the loss function for estimating I by S defined in (2) of Section 7.8. In terms of the characteristic roots of S the criterion (6) is a constant plus

$$(7) \qquad \sum_{i=1}^{p} l_i - \log \prod_{i=1}^{p} l_i - p = \sum_{i=1}^{p} (l_i - \log l_i - 1);$$

for each i the minimum of (7) is at $l_i = 1$.

Using the algebra of the preceding section, we see that given $y_1, \dots, y_N$ as observation vectors of p components from $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$, the modified likelihood ratio criterion for testing the hypothesis $H_1: \boldsymbol{\Psi} = \boldsymbol{\Psi}_0$, where $\boldsymbol{\Psi}_0$ is specified, is

$$(8) \qquad \lambda_1^* = \left(\frac{e}{n}\right)^{\frac{1}{2}pn} |B\boldsymbol{\Psi}_0^{-1}|^{\frac{1}{2}n} e^{-\frac{1}{2}\operatorname{tr} B\boldsymbol{\Psi}_0^{-1}},$$

where

$$(9) \qquad B = \sum_{\alpha=1}^{N} (y_\alpha - \bar{y})(y_\alpha - \bar{y})'.$$

10.8.2. The Distribution and Moments of the Modified Likelihood Ratio Criterion

The null hypothesis $H_1: \Sigma = I$ is the intersection of the null hypothesis of Section 10.7, $H: \Sigma = \sigma^2 I$, and the null hypothesis $\sigma^2 = 1$ given $\Sigma = \sigma^2 I$. The likelihood ratio criterion for H_1 given by (3) is the product of (7) of Section 10.7 and

$$(10) \qquad \left(\frac{\operatorname{tr} A}{pN}\right)^{\frac{1}{2}pN} e^{-\frac{1}{2}\operatorname{tr} A + \frac{1}{2}pN},$$

which is the likelihood ratio criterion for testing the hypothesis $\sigma^2 = 1$ given $\Sigma = \sigma^2 I$. The modified criterion λ_1^* is the product of $|A|^{\frac{1}{2}n}/(\operatorname{tr} A/p)^{\frac{1}{2}pn}$ and

$$(11) \qquad \left(\frac{\operatorname{tr} A}{pn}\right)^{\frac{1}{2}pn} e^{-\frac{1}{2}\operatorname{tr} A + \frac{1}{2}pn};$$

these two factors are independent (Lemma 10.4.1). The characterization of the distribution of the modified criterion can be obtained from Section

10.7.3. The quantity $\operatorname{tr} A$ has the χ^2-distribution with np degrees of freedom under the null hypothesis.

Instead of obtaining the moments and characteristic function of λ_1^* [defined by (5)] from the preceding characterization, we shall find them by use of the fact that A has the distribution $W(\boldsymbol{\Sigma}, n)$. We shall calculate

$$
\begin{aligned}
(12) \qquad \mathscr{E}\lambda_1^{*h} &= \int \cdots \int \left(\frac{e^{\frac{1}{2}pn}}{n^{\frac{1}{2}pn}} |A|^{\frac{1}{2}n} e^{-\frac{1}{2}\operatorname{tr} A} \right)^h w(A|\boldsymbol{\Sigma}, n)\, dA \\
&= \frac{e^{\frac{1}{2}pnh}}{n^{\frac{1}{2}pnh}} \int \cdots \int |A|^{\frac{1}{2}nh} e^{-\frac{1}{2}h \operatorname{tr} A} w(A|\boldsymbol{\Sigma}, n)\, dA.
\end{aligned}
$$

Since

(13)

$$
\begin{aligned}
|A|^{\frac{1}{2}nh} e^{-\frac{1}{2}h\operatorname{tr} A} w(A|\boldsymbol{\Sigma}, n) &= \frac{|A|^{\frac{1}{2}(n+nh-p-1)} e^{-\frac{1}{2}(\operatorname{tr}\boldsymbol{\Sigma}^{-1}A + \operatorname{tr} hA)}}{2^{\frac{1}{2}pn}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\Gamma_p(\frac{1}{2}n)} \\
&= \frac{2^{\frac{1}{2}pnh}\Gamma_p\left[\frac{1}{2}n(1+h)\right]}{|\boldsymbol{\Sigma}^{-1} + hI|^{\frac{1}{2}(n+nh)}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\Gamma_p(\frac{1}{2}n)} \\
&\quad \cdot \frac{|\boldsymbol{\Sigma}^{-1} + hI|^{\frac{1}{2}(n+nh)}|A|^{\frac{1}{2}(n+nh-p-1)} e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}+hI)A}}{2^{\frac{1}{2}p(n+nh)}\Gamma_p\left[\frac{1}{2}n(1+h)\right]} \\
&= \frac{2^{\frac{1}{2}pnh}|\boldsymbol{\Sigma}|^{\frac{1}{2}nh}\Gamma_p\left[\frac{1}{2}n(1+h)\right]}{|I + h\boldsymbol{\Sigma}|^{\frac{1}{2}(n+nh)}\Gamma_p(\frac{1}{2}n)} \\
&\quad \cdot w\left(A|(\boldsymbol{\Sigma}^{-1} + hI)^{-1}, n + nh\right),
\end{aligned}
$$

the hth moment of λ_1^* is

$$
(14) \qquad \mathscr{E}\lambda_1^{*h} = \left(\frac{2e}{n}\right)^{\frac{1}{2}phn} \frac{|\boldsymbol{\Sigma}|^{\frac{1}{2}nh}\prod_{j=1}^p \Gamma\left[\frac{1}{2}(n+nh+1-j)\right]}{|I + h\boldsymbol{\Sigma}|^{\frac{1}{2}(n+nh)}\prod_{j=1}^p \Gamma\left[\frac{1}{2}(n+1-j)\right]}.
$$

Then the characteristic function of $-2\log\lambda^*$ is

$$
\begin{aligned}
(15) \qquad \mathscr{E}e^{-2it\log\lambda_1^*} &= \mathscr{E}\lambda_1^{*-2it} \\
&= \left(\frac{2e}{n}\right)^{-ipnt} \frac{|\boldsymbol{\Sigma}|^{-int}}{|I - 2it\boldsymbol{\Sigma}|^{\frac{1}{2}n-int}} \cdot \prod_{j=1}^p \frac{\Gamma\left[\frac{1}{2}(n+1-j) - int\right]}{\Gamma\left[\frac{1}{2}(n+1-j)\right]}.
\end{aligned}
$$

When the null hypothesis is true, $\boldsymbol{\Sigma} = \boldsymbol{I}$, and

(16)
$$\mathscr{E} e^{-2it \log \lambda_1^*} = \left(\frac{2e}{n}\right)^{-ipnt} (1-2it)^{-\frac{1}{2}p(n-2int)} \prod_{j=1}^{p} \frac{\Gamma\left[\frac{1}{2}(n+1-j) - int\right]}{\Gamma\left[\frac{1}{2}(n+1-j)\right]}.$$

This characteristic function is the product of p terms such as

(17)
$$\phi_j(t) = \left(\frac{2e}{n}\right)^{-int} (1-2it)^{-\frac{1}{2}(n-2int)} \frac{\Gamma\left[\frac{1}{2}(n+1-j) - int\right]}{\Gamma\left[\frac{1}{2}(n+1-j)\right]}.$$

Thus $-2\log\lambda_1^*$ is distributed as the sum of p independent variates, the characteristic function of the jth being (17). Using Stirling's approximation for the gamma function, we have

(18)
$$\begin{aligned}\phi_j(t) &\sim 2^{-int} e^{-int} n^{int} (1-2it)^{\frac{1}{2}(2int-n)} \\ &\quad \cdot \frac{e^{-[\frac{1}{2}(n+1-j)-int]}\left[\frac{1}{2}(n+1-j) - int\right]^{\frac{1}{2}(n-j)-int}}{e^{-[\frac{1}{2}(n+1-j)]}\left[\frac{1}{2}(n-j+1)\right]^{\frac{1}{2}(n-j)}} \\ &= (1-2it)^{-\frac{1}{2}j}\left(1 - \frac{it(j-1)}{\frac{1}{2}(n-j+1)(1-2it)}\right)^{\frac{1}{2}(n+1-j)-\frac{1}{2}} \\ &\quad \cdot \left(1 - \frac{2j-1}{n(1-2it)}\right)^{-int}.\end{aligned}$$

As $n \to \infty$, $\phi_j(t) \to (1-2it)^{-\frac{1}{2}j}$, which is the characteristic function of χ_j^2 (χ^2 with j degrees of freedom). Thus $-2\log\lambda_1^*$ is asymptotically distributed as $\sum_{j=1}^{p}\chi_j^2$, which is χ^2 with $\sum_{j=1}^{p} j = \frac{1}{2}p(p+1)$ degrees of freedom. The distribution of λ_1^* can be further expanded [Korin (1968), Davis (1971)] as

(19)
$$\begin{aligned}&\Pr\{-2\rho\log\lambda_1^* \le z\} \\ &= \Pr\{\chi_f^2 \le z\} + \frac{\gamma_2}{\rho^2 N^2}\left(\Pr\{\chi_{f+4}^2 \le z\} - \Pr\{\chi_f^2 \le z\}\right) + O(N^{-3}),\end{aligned}$$

where

(20)
$$\rho = 1 - \frac{2p^2+3p-1}{6N(p+1)},$$

(21)
$$\gamma_2 = \frac{p(2p^4+6p^3+p^2-12p-13)}{288(p+1)}.$$

Nagarsenker and Pillai (1973b) found exact distributions and tabulated 5% and 1% significant points, as did Davis and Field (1971), for $p = 2(1)10$ and $n = 6(1)30(5)50, 60, 120$. Table B.7 [due to Korin (1968)] gives some 5% and 1% significance points of $-2\log \lambda_1^*$ for small values of n and $p = 2(1)10$.

10.8.3. Invariant Tests

The null hypothesis $H: \boldsymbol{\Sigma} = \boldsymbol{I}$ is invariant with respect to transformations $\boldsymbol{X}^* = \boldsymbol{Q}\boldsymbol{X} + \boldsymbol{\nu}$, where $\boldsymbol{Q}$ is an orthogonal matrix. The invariants of the sufficient statistics are the characteristic roots $l_1, \dots, l_p$ of $\boldsymbol{S}$, and the invariants of the parameters are the characteristic roots of $\boldsymbol{\Sigma}$. Invariant tests are based on the roots of $\boldsymbol{S}$; the modified likelihood ratio criterion is one of them. Nagao (1973a) suggested the criterion

$$\tag{22} \tfrac{1}{2} n \operatorname{tr}(\boldsymbol{S} - \boldsymbol{I})^2 = \tfrac{1}{2} n \sum_{i=1}^{p} (l_i - 1)^2.$$

Under the null hypothesis this criterion has a limiting χ^2-distribution with $\frac{1}{2}p(p+1)$ degrees of freedom.

Roy (1957), Section 6.4, proposed a test based on the largest and smallest characteristic roots l_1 and l_p: Reject the null hypothesis if

$$\tag{23} l_p < l \quad \text{or} \quad l_1 > u,$$

where

$$\tag{24} \Pr\{l < l_p, l_1 < u \mid \boldsymbol{\Sigma} = \boldsymbol{I}\} = 1 - \varepsilon$$

and ε is the significance level. Clemm, Krishnaiah, and Waikar (1973) give tables of $u = 1/l$. See also Schuurman and Waikar (1973).

10.8.4. Confidence Bounds for Quadratic Forms

The test procedure based on the smallest and largest characteristic roots can be inverted to give confidence bounds on qudaratic forms in $\boldsymbol{\Sigma}$. Suppose $n\boldsymbol{S}$ has the distribution $W(\boldsymbol{\Sigma}, n)$. Let $\boldsymbol{C}$ be a nonsingular matrix such that $\boldsymbol{\Sigma} = \boldsymbol{C}'\boldsymbol{C}$. Then $n\boldsymbol{S}^* = n\boldsymbol{C}'^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}$ has the distribution $W(\boldsymbol{I}, n)$. Since $l_p^* \le \boldsymbol{a}'\boldsymbol{S}^*\boldsymbol{a}/\boldsymbol{a}'\boldsymbol{a} < l_1^*$ for all $\boldsymbol{a}$, where l_p^* and l_1^* are the smallest and largest characteristic roots of $\boldsymbol{S}^*$ (Sections 11.2 and A.2),

$$\tag{25} \Pr\left\{l \le \frac{\boldsymbol{a}'\boldsymbol{S}^*\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}} \le u \ \ \forall \boldsymbol{a} \ne \boldsymbol{0}\right\} = 1 - \varepsilon,$$

where

$$\tag{26} \Pr\{l \le l_p^* \le l_1^* \le u\} = 1 - \varepsilon.$$

Let $a = Cb$. Then $a'a = b'C'Cb = b'\Sigma b$ and $a'S^*a = b'C'S^*Cb = b'Sb$. Thus (25) is

$$(27)\qquad 1-\varepsilon = \Pr\left\{l \le \frac{b'Sb}{b'\Sigma b} \le u \quad \forall b \ne 0\right\}$$
$$= \Pr\left\{\frac{b'Sb}{u} \le b'\Sigma b \le \frac{b'Sb}{l} \quad \forall b\right\}.$$

Given an observed S, one can assert

$$(28)\qquad \frac{b'Sb}{u} \le b'\Sigma b \le \frac{b'Sb}{l} \qquad \forall b$$

with confidence $1-\varepsilon$.

If b has 1 in the ith position and 0's elsewhere, (28) is $s_{ii}/u \le \sigma_{ii} \le s_{ii}/l$. If b has 1 in the ith position, -1 in the jth position, $i \ne j$, and 0's elsewhere, then (28) is

$$(29)\qquad \frac{s_{ii}+s_{jj}-2s_{ij}}{u} \le \sigma_{ii}+\sigma_{jj}-2\sigma_{ij} \le \frac{s_{ii}+s_{jj}-2s_{ij}}{l}.$$

Manipulation of these inequalities yields

$$(30)\qquad \frac{s_{ij}}{l} - \frac{s_{ii}+s_{jj}}{2}\left(\frac{1}{l}-\frac{1}{u}\right) \le \sigma_{ij} \le \frac{s_{ij}}{u} + \frac{s_{ii}+s_{jj}}{2}\left(\frac{1}{l}-\frac{1}{u}\right), \qquad i \ne j.$$

We can obtain simultaneously confidence intervals on all elements of Σ.

From (27) we can obtain

$$(31)\qquad 1-\varepsilon = \Pr\left\{\frac{1}{u}\frac{b'Sb}{b'b} \le \frac{b'\Sigma b}{b'b} \le \frac{1}{l}\frac{b'Sb}{b'b} \quad \forall b\right\}$$
$$\le \Pr\left\{\frac{1}{u}\min_a \frac{a'Sa}{a'a} \le \frac{b'\Sigma b}{b'b} \le \frac{1}{l}\max_a \frac{a'Sa}{a'a} \quad \forall b\right\}$$
$$= \Pr\left\{\frac{1}{u}l_p \le \lambda_p \le \lambda_1 \le \frac{1}{l}l_1\right\},$$

where l_1 and l_p are the largest and smallest characteristic roots of S and λ_1 and λ_p are the largest and smallest characteristic roots of Σ. Then

$$(32)\qquad \frac{1}{u}l_p \le \lambda(\Sigma) \le \frac{1}{l}l_1$$

is a confidence interval for all characteristic roots of Σ with confidence at least $1-\varepsilon$. In Section 11.6 we give tighter bounds on $\lambda(\Sigma)$ with exact confidence.

10.9. TESTING THE HYPOTHESIS THAT A MEAN VECTOR AND A COVARIANCE MATRIX ARE EQUAL TO A GIVEN VECTOR AND MATRIX

In Chapter 3 we pointed out that if $\boldsymbol{\Psi}$ is known, $(\bar{y}-\boldsymbol{\nu}_0)'\boldsymbol{\Psi}_0^{-1}(\bar{y}-\boldsymbol{\nu}_0)$ is suitable for testing

$$(1) \qquad H_2: \boldsymbol{\nu}=\boldsymbol{\nu}_0, \qquad \text{given} \quad \boldsymbol{\Psi}\equiv\boldsymbol{\Psi}_0.$$

Now let us combine H_1 of Section 10.8 and H_2, and test

$$(2) \qquad H: \boldsymbol{\nu}=\boldsymbol{\nu}_0, \qquad \boldsymbol{\Psi}=\boldsymbol{\Psi}_0,$$

on the basis of a sample $y_1,\ldots,y_N$ from $N(\boldsymbol{\nu},\boldsymbol{\Psi})$.

Let

$$(3) \qquad X=C(Y-\boldsymbol{\nu}_0),$$

where

$$(4) \qquad C\boldsymbol{\Psi}_0C'=I.$$

Then $x_1,\ldots,x_N$ constitutes a sample from $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, and the hypothesis is

$$(5) \qquad H: \boldsymbol{\mu}=\mathbf{0}, \qquad \boldsymbol{\Sigma}=I.$$

The likelihood ratio criterion for $H_2: \boldsymbol{\mu}=\mathbf{0}$, given $\boldsymbol{\Sigma}=I$, is

$$(6) \qquad \lambda_2=e^{-\frac{1}{2}N\bar{x}'\bar{x}}.$$

The likelihood ratio criterion for H is (by Lemma 10.3.1)

$$(7) \qquad \begin{aligned}\lambda=\lambda_1\lambda_2&=\left(\frac{e}{N}\right)^{\frac{1}{2}pN}|A|^{\frac{1}{2}N}e^{-\frac{1}{2}\operatorname{tr}A}e^{-\frac{1}{2}N\bar{x}'\bar{x}}\\ &=\left(\frac{e}{N}\right)^{\frac{1}{2}pN}|A|^{\frac{1}{2}N}e^{-\frac{1}{2}\operatorname{tr}(A+N\bar{x}\bar{x}')}\\ &=\left(\frac{e}{N}\right)^{\frac{1}{2}pN}|A|^{\frac{1}{2}N}e^{-\frac{1}{2}\Sigma x_\alpha'x_\alpha}.\end{aligned}$$

The likelihood ratio test (rejecting H if λ is less than a suitable constant) is unbiased [Srivastava and Khatri (1979), Theorem 10.4.5]. The two factors λ_1 and λ_2 are independent because λ_1 is a function of A and λ_2 is a function of $\bar{x}$, and A and $\bar{x}$ are independent. Since

$$(8) \qquad \mathscr{E}\lambda_2^h=\mathscr{E}e^{-\frac{1}{2}hN\Sigma\bar{x}_i^2}=\mathscr{E}e^{-\frac{1}{2}h\chi_p^2}=(1+h)^{-\frac{1}{2}p},$$

the hth moment of λ is

$$\mathscr{E}\lambda^h = \mathscr{E}\lambda_1^h \mathscr{E}\lambda_2^h = \left(\frac{2e}{N}\right)^{\frac{1}{2}pNh} \frac{1}{(1+h)^{\frac{1}{2}pN(1+h)}} \frac{\Gamma_p\left[\frac{1}{2}(n+Nh)\right]}{\Gamma_p\left(\frac{1}{2}n\right)} \tag{9}$$

under the null hypothesis. Then

$$-2\log\lambda = -2\log\lambda_1 - 2\log\lambda_2 \tag{10}$$

has asymptotically the χ^2-distribution with $f = p(p+1)/2 + p$ degrees of freedom. In fact, an asymptotic expansion of the distribution [Davis (1971)] of $-2\rho\log\lambda$ is

$$\begin{aligned}(11)\quad &\Pr\{-2\rho\log\lambda \le z\} \\ &= \Pr\left\{\chi_f^2 \le z\right\} + \frac{\gamma_2}{\rho^2 N^2}\left(\Pr\left\{\chi_{f+4}^2 \le z\right\} - \Pr\left\{\chi_f^2 \le z\right\}\right) + O(N^{-3}),\end{aligned}$$

where

$$\rho = 1 - \frac{2p^2 + 9p - 11}{6N(p+3)}, \tag{12}$$

$$\gamma_2 = \frac{p(2p^4 + 18p^3 + 49p^2 + 36p - 13)}{288(p-3)}. \tag{13}$$

Nagarsenker and Pillai (1974) used the moments to derive exact distributions and tabulated the 5% and 1% significance points for $p = 2(1)6$ and $N = 4(1)20(2)40(5)100$.

Now let us return to the observations $y_1, \ldots, y_N$. Then

$$\begin{aligned}(14)\quad \sum_\alpha x'_\alpha x_\alpha &= \sum_\alpha (y_\alpha - \nu_0)' C' C (y_\alpha - \nu_0) \\ &= \sum_\alpha (y_\alpha - \nu_0)' \Psi_0^{-1} (y_\alpha - \nu_0) \\ &= \operatorname{tr} A + N\bar{x}'\bar{x} \\ &= \operatorname{tr}\left(B\Psi_0^{-1}\right) + N(\bar{y} - \nu_0)'\Psi_0^{-1}(\bar{y} - \nu_0)\end{aligned}$$

and

$$|A| = |B\Psi_0^{-1}|. \tag{15}$$

Theorem 10.9.1. *Given the p-component observation vectors $y_1, \ldots, y_N$ from $N(\nu, \Psi)$, the likelihood ratio criterion for testing the hypothesis $H: \nu = \nu_0$,*

$\boldsymbol{\Psi} = \boldsymbol{\Psi}_0$, *is*

$$\lambda = \left(\frac{e}{N}\right)^{\frac{1}{2}pN} |\mathbf{B}\boldsymbol{\Psi}_0^{-1}|^{\frac{1}{2}N} e^{-\frac{1}{2}[\operatorname{tr} \mathbf{B}\boldsymbol{\Psi}_0^{-1} + N(\bar{\mathbf{y}} - \boldsymbol{\nu}_0)'\boldsymbol{\Psi}_0^{-1}(\bar{\mathbf{y}} - \boldsymbol{\nu}_0)]}. \tag{16}$$

When the null hypothesis is true, $-2\log\lambda$ is asymptotically distributed as χ^2 with $\frac{1}{2}p(p+1)+p$ degrees of freedom.

10.10. ADMISSIBILITY OF TESTS

We shall consider some Bayes solutions to the problem of testing the hypothesis

$$\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q \tag{1}$$

as in Section 10.2. Under the alternative hypothesis, let

$$\left[\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g\right] = \left[(\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\boldsymbol{y}^{(g)}, (\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g)^{-1}\right], \qquad g = 1, \dots, q, \tag{2}$$

where the $p \times r_g$ matrix $\boldsymbol{C}_g$ has density proportional to $|\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}$, $n_g = N_g - 1$, the r_g-component random vector $\boldsymbol{y}^{(g)}$ has the conditional normal distribution with mean $\mathbf{0}$ and covariance matrix $(1/N_g)[\boldsymbol{I}_{r_g} - \boldsymbol{C}_g'(\boldsymbol{I}_p + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g]^{-1}$ given $\boldsymbol{C}_g$, and $(\boldsymbol{C}_1, \boldsymbol{y}^{(1)}), \dots, (\boldsymbol{C}_q, \boldsymbol{y}^{(q)})$ are independently distributed. As we shall see, we need to choose suitable integers $r_1, \dots, r_q$. Note that the integral of $|\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}$ is finite if $n_g \geq p + r_g$. Then the numerator of the Bayes ratio is

$$\begin{aligned}
\text{(3)} \quad \text{const} \prod_{g=1}^{q} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} &|\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g'|^{\frac{1}{2}N_g} \\
&\cdot \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N_g}\left[\boldsymbol{x}_\alpha^{(g)} - (\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\boldsymbol{y}^{(g)}\right]'\right. \\
&\qquad \left.\cdot(\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g')\left[\boldsymbol{x}_\alpha^{(g)} - (\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\boldsymbol{y}^{(g)}\right]\right\} \\
&\cdot |\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}\left|\boldsymbol{I} - \boldsymbol{C}_g'(\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\right|^{\frac{1}{2}} \\
&\qquad \cdot \exp\left\{-\tfrac{1}{2}N_g\boldsymbol{y}^{(g)\prime}\left[\boldsymbol{I} - \boldsymbol{C}_g'(\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g^{(g)}\right]\boldsymbol{y}^{(g)}\right\} d\boldsymbol{y}^{(g)}\, d\boldsymbol{C}_g
\end{aligned}$$

$$= \text{const} \prod_{g=1}^{q} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2} \left[\sum_{\alpha=1}^{N_g} x_\alpha^{(g)\prime}(I + C_g C_g') x_\alpha^{(g)} \right.\right.$$

$$\left.\left. -2 y^{(g)\prime} C_g' \sum_{\alpha=1}^{N_g} x_\alpha^{(g)} + N_g y^{(g)\prime} y^{(g)} \right] \right\} dy^{(g)}\, dC_g$$

$$= \text{const} \prod_{g=1}^{q} \exp\left\{ -\frac{1}{2} \sum_{\alpha=1}^{N_g} x_\alpha^{(g)\prime} x_\alpha^{(g)} \right\} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}$$

$$\cdot \exp\left\{ -\tfrac{1}{2} N_g \left(y^{(g)} - C_g' \bar{x}^{(g)} \right)' \left(y^{(g)} - C_g' \bar{x}^{(g)} \right) - \tfrac{1}{2} \operatorname{tr} C_g' A_g C_g \right\} dy^{(g)}\, dC_g$$

$$= \text{const} \prod_{g=1}^{q} \exp\left\{ -\tfrac{1}{2} \left[\operatorname{tr} A_g + N_g \bar{x}^{(g)\prime} \bar{x}^{(g)} \right] \right\} |A_g|^{-\frac{1}{2} r_g}.$$

Under the null hypothesis let

$$\left[\mu^{(g)}, \Sigma_g \right] = \left[(I + CC')^{-1} C y^{(g)}, (I + CC')^{-1} \right], \tag{4}$$

where the $p \times r$ matrix C has density proportional to $|I + CC'|^{-\frac{1}{2}n}$, $n = \sum_{g=1}^{q} n_g$, the r-component vector $y^{(g)}$ has the conditional normal distribution with mean $\mathbf{0}$ and covariance matrix $(1/N_g)[I_r - C'(I_p + CC')^{-1} C]^{-1}$ given C, and $y^{(1)}, \ldots, y^{(g)}$ are conditionally independent. Note that the integral of $|I + CC'|^{-\frac{1}{2}n}$ is finite if $n \geq p + r$. The denominator of the Bayes ratio is

(5)

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{g=1}^{q} \left[|I + CC'|^{\frac{1}{2} N_g} \exp\left\{ -\frac{1}{2} \sum_{\alpha=1}^{N_g} \left[x_\alpha^{(g)} - (I + CC')^{-1} C y^{(g)} \right]' \right.\right.$$

$$\left. \cdot (I + CC') \left[x_\alpha^{(g)} - (I + CC')^{-1} C y^{(g)} \right] \right\}$$

$$\cdot |I + CC'|^{-\frac{1}{2} n_g} \left| I - C'(I + CC')^{-1} C \right|^{\frac{1}{2}}$$

$$\left. \cdot \exp\left\{ -\tfrac{1}{2} N_g y^{(g)\prime} \left[I - C'(I + CC')^{-1} C \right] y^{(g)} \right\} dy^{(g)} \right] dC$$

$$= \text{const} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{g=1}^{q} \exp\left\{-\frac{1}{2}\left[\sum_{\alpha=1}^{N_g} x_\alpha^{(g)\prime}(I + CC')x_\alpha^{(g)} - 2y^{(g)\prime}C' \sum_{\alpha=1}^{N_g} x_\alpha^{(g)} + N_g y^{(g)\prime}y^{(g)}\right]\right\} dy^{(g)}\, dC$$

$$= \text{const} \exp\left\{-\frac{1}{2}\sum_{g=1}^{q}\sum_{\alpha=1}^{N_g} x_\alpha^{(g)\prime}x_\alpha^{(g)}\right\}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}$$

$$\cdot \exp\left\{-\frac{1}{2}\sum_{g=1}^{q} N_g\left(y^{(g)} - C'\bar{x}^{(g)}\right)'\left(y^{(g)} - C'\bar{x}^{(g)}\right) - \tfrac{1}{2}\operatorname{tr} C'AC\right\}\prod_{g=1}^{q} dy^{(g)}\, dC$$

$$= \text{const} \exp\left\{-\frac{1}{2}\left(\operatorname{tr} A + \sum_{g=1}^{q} N_g \bar{x}^{(g)\prime}\bar{x}^{(g)}\right)\right\}|A|^{-\frac{1}{2}r}.$$

The Bayes test procedure is to reject the hypothesis if

$$\frac{|A|^{\frac{1}{2}r}}{\prod_{g=1}^{q}|A_g|^{\frac{1}{2}r_g}} \geq c. \tag{6}$$

For invariance we want $\sum_{g=1}^{q} r_g = r$.

The binding constraint on the choice of $r_1,\ldots,r_q$ is $r_g \leq n_g - p$, $g = 1,\ldots,q$. It is possible in some special cases to choose $r_1,\ldots,r_q$ so that $(r_1,\ldots,r_q)$ is proportional to $(N_1,\ldots,N_q)$ and hence yield the likelihood ratio test or proportional to $(n_1,\ldots,n_q)$ and hence yield the modified likelihood ratio test, but since $r_1,\ldots,r_q$ have to be integers, it may not be possible to choose them in either such way. Next we consider an extension of this approach that involves the choice of numbers $t_1,\ldots,t_q$, and t as well as $r_1,\ldots,r_q$, and r.

Suppose $2(p-1) < n_g$, $g = 1,\ldots,q$, and take $r_g \geq p$. Let t_g be a real number such that $2p - 1 < r_g + t_g + p < n_g + 1$, and let t be a real number such that $2p - 1 < r + t + p < n + 1$. Under the alternative hypothesis let the marginal density of C_g be proportional to $|C_g C_g'|^{\frac{1}{2}t_g}|I + C_g C_g'|^{-\frac{1}{2}n_g}$, $g = 1,\ldots,q$, and under the null hypothesis let the marginal density of C be proportional to $|CC'|^{\frac{1}{2}t}|I + CC'|^{-\frac{1}{2}n}$. (The conditions on $t_1,\ldots,t_q$, and t ensure that the purported densities have finite integrals; see Problem 10.18.) Then the Bayes procedure is to reject the null hypothesis if

$$\frac{|A|^{\frac{1}{2}(r+t)}}{\prod_{g=1}^{q}|A_g|^{\frac{1}{2}(r_g+t_g)}} \geq c. \tag{7}$$

For invariance we want $t = \Sigma_{g=1}^{q} t_g$. If $t_1, \dots, t_q$ are taken so $r_g + t_g = kN_g$ and $p - 1 < kN_g < N_g - p$, $g = 1, \dots, q$, for some k, then (7) is the likelihood ratio test; if $r_g + t_g = kn_g$ and $p - 1 < kn_g < n_g + 1 - p$, $g = 1, \dots, q$, for some k, then (7) is the modified test [i.e., $(p-1)/\min_g N_g < k < 1 - p/\min_g N_g$].

Theorem 10.10.1. *If $2p < N_g + 1$, $g = 1, \dots, q$, then the likelihood ratio test and the modified likelihood ratio test of the null hypothesis* (1) *are admissible.*

Now consider the hypothesis

$$\mu^{(1)} = \cdots = \mu^{(q)}, \qquad \Sigma_1 = \cdots = \Sigma_q. \tag{8}$$

The alternative hypothesis has been treated before. For the null hypothesis let

$$\left[\mu^{(g)}, \Sigma_g\right] = \left[(I + CC')Cy, (I + CC')^{-1}\right], \tag{9}$$

where the $p \times r$ matrix C has the density proportional to $|I + CC'|^{-\frac{1}{2}(N-1)}$ and the r-component vector y has the conditional normal distribution with mean $\mathbf{0}$ and covariance matrix $(1/N)[I - C'(I + CC')^{-1}C]^{-1}$ given C. Then the Bayes procedure is to reject the null hypothesis (8) if

$$\frac{\left|\Sigma_{g=1}^{q} A_g + \Sigma_{g=1}^{q} N_g\left(\bar{y}^{(g)} - \bar{y}\right)\left(\bar{y}^{(g)} - \bar{y}\right)'\right|^{\frac{1}{2}r}}{\prod_{g=1}^{q} |A_g|^{\frac{1}{2}r_g}} \geq c. \tag{10}$$

If $2p < N_g + 1$, $g = 1, \dots, q$, the prior distribution can be modified as before to obtain the likelihood ratio test and modified likelihood ratio test.

Theorem 10.10.2. *If $2p < N_g + 1$, $g = 1, \dots, q$, the likelihood ratio test and modified likelihood ratio test of the null hypothesis* (8) *are admissible.*

For more details see Kiefer and Schwartz (1965).

10.11. ELLIPTICALLY CONTOURED DISTRIBUTIONS

10.11.1. Observations Elliptically Contoured

Let $x_\alpha^{(g)}$, $\alpha = 1, \dots, N_g$, be N_g observations on $X^{(g)}$ having the density

$$|\Lambda_g|^{-\frac{1}{2}} g\left[(x - \nu^{(g)})'\Lambda_g^{-1}(x - \nu^{(g)})\right], \tag{1}$$

where $\mathscr{E}[(X - \nu^{(g)})'\Lambda_g^{-1}(X - \nu^{(g)})]^2 = \mathscr{E}R_g^4 < \infty$, $g = 1, \dots, q$. Note that the same function $g(\cdot)$ is used for the density in all q populations. Define N, A_g,

$g = 1, \ldots, q$, and A by (1) of Section 10.2. Let $S_g = (1/n_g)A_g$, where $n_g = N_g - 1$, and $S = (1/n)A$, where $n = \sum_{g=1}^q n_g$.

Since the likelihood ratio criterion λ_1 is invariant under the transformation $X^{(g)} = CX^{(g)} + \nu^{(g)}$, under the null hypothesis we can take $\Sigma_1 = \cdots = \Sigma_q = I$ and $\nu^{(1)} = \cdots = \nu^{(g)} = 0$. Then

$$
\begin{aligned}
(2) \quad -2\log\lambda_1 &= -\left[\sum_{g=1}^{q} N_g \log|\hat{\Sigma}_{g\Omega}| - N\log|\hat{\Sigma}_\omega|\right] \\
&= -\left[\sum_{g=1}^{g} N_g \log\left|I + \left(\hat{\Sigma}_{g\Omega} - I\right)\right| - N\log\left|I + \left(\hat{\Sigma}_\omega - I\right)\right|\right] \\
&= -\left\{\sum_{g=1}^{q} N_g\left[\operatorname{tr}\left(\hat{\Sigma}_{g\Omega} - I\right) - \tfrac{1}{2}\operatorname{tr}\left(\hat{\Sigma}_{g\Omega} - I\right)^2 + O_p\left(N_g^{-3}\right)\right]\right. \\
&\qquad \left. - N\left[\operatorname{tr}\left(\hat{\Sigma}_\omega - I\right) - \tfrac{1}{2}\operatorname{tr}\left(\hat{\Sigma}_\omega - I\right)^2 + O_p\left(N^{-3}\right)\right]\right\} \\
&= \tfrac{1}{2}\sum_{g=1}^{q} N_g \operatorname{tr}\left(\hat{\Sigma}_{g\Omega} - I\right)^2 - \tfrac{1}{2}N\operatorname{tr}\left(\hat{\Sigma}_\omega - I\right)^2 + O_p(N^{-3}) \\
&= \tfrac{1}{2}\sum_{g=1}^{q} N_g\left[\operatorname{vec}\left(\hat{\Sigma}_{g\Omega} - I\right)\right]'\operatorname{vec}\left(\hat{\Sigma}_{g\Omega} - I\right) \\
&\qquad - \tfrac{1}{2}N\left[\operatorname{vec}\left(\hat{\Sigma}_\omega - I\right)\right]'\operatorname{vec}\left(\hat{\Sigma}_\omega - I\right) + O_p(N^{-3}).
\end{aligned}
$$

By Theorem 3.6.2

$$
(3) \quad \sqrt{N_g}\operatorname{vec}(S_g - I_p) \xrightarrow{d} N\left[0, (\kappa+1)(I_{p^2} + K_{pp}) + \kappa \operatorname{vec} I_p (\operatorname{vec} I_p)'\right],
$$

and $n_g S_g = N_g \hat{\Sigma}_{g\Omega}$, $g = 1, \ldots, q$, are independent. Let $N_g = k_g N$, $g = 1, \ldots, q$, $\sum_{g=1}^q k_g = 1$, and let $N \to \infty$. In terms of this asymptotic theory the limiting distribution of $\operatorname{vec}(S_1 - I), \ldots, \operatorname{vec}(S_q - I)$ is the same as the distribution of $\bar{y}^{(1)}, \ldots, \bar{y}^{(q)}$ of Section 8.8, with Σ of Section 8.8 replaced by $(\kappa+1)(I_{p^2} + K_{pp}) + \kappa \operatorname{vec} I_p (\operatorname{vec} I_p)'$.

When $\Sigma = I$, the variance of the limiting distribution of $\sqrt{N_g}(s_{ii}^{(g)} - 1)$ is $3\kappa + 2$; the covariance of the limiting distribution of $\sqrt{N_g}(s_{ii}^{(g)} - 1)$ and $\sqrt{N_g}(s_{jj}^{(g)} - 1)$, $i \neq j$, is κ; the variance of $s_{ij}^{(g)}$, $i \neq j$, is $\kappa + 1$; the set $\sqrt{N}(s_{11}^{(g)} - 1), \ldots, \sqrt{N}(s_{pp}^{(g)} - 1)$ is independent of the set $(s_{ij}^{(g)})$, $i \neq j$; and the $s_{ij}^{(g)}$, $i < j$, are mutually uncorrelated (as in Section 7.9.1).

Let $\bar{y}_g = \operatorname{vec}(\hat{\Sigma}_{g\Omega} - I)$ and $\bar{y} = \operatorname{vec}(\hat{\Sigma}_\omega - I)$. Then $\bar{y} = \sum_{g=1}^q (N_g/N)\bar{y}_g$ and

(4)
$$\begin{aligned}-2\log\lambda &= \tfrac{1}{2}\sum_{g=1}^q N_g(\bar{y}_g - \bar{y})'(\bar{y}_g - \bar{y}) \\ &= \operatorname{tr}\tfrac{1}{2}\sum_{g=1}^q N_g(\bar{y}_g - \bar{y})(\bar{y}_g - \bar{y})' \\ &= \tfrac{1}{2}\operatorname{tr}\left(\sum_{g=1}^q N_g \bar{y}_g \bar{y}_g' - N\bar{y}\bar{y}'\right).\end{aligned}$$

Let Q be a $q \times q$ orthogonal matrix with last column $(\sqrt{N_1/N}, \dots, \sqrt{N_q/N})'$. Define

(5)
$$(w_1, \dots, w_q) = \left(\sqrt{N_1}\,\bar{y}_1, \dots, \sqrt{N_g}\,\bar{y}_g\right)Q.$$

Then $w_q = \sqrt{N}\bar{y}$ and

(6)
$$\sum_{g=1}^q N_g \bar{y}_g \bar{y}_g' - N\bar{y}\bar{y} = \sum_{g=1}^{q-1} w_g w_g'.$$

In these terms

(7)
$$-2\log\lambda_1 = \tfrac{1}{2}\sum_{g=1}^{q-1} w_g' w_g + O_p(N^{-3}),$$

and $w_1, \dots, w_{q-1}$ are asymptotically independent, w_g having the covariance matrix of $\sqrt{N}\bar{y}_g$; that is, $(\kappa+1)(I_{p^2} + K_{pp}) + \kappa \operatorname{vec} I_p (\operatorname{vec} I_p)$. Then $w_g' w_g = \sum_{i,j=1}^p (w_{ij}^{(g)})^2 = \sum_{i=1}^p (w_{ii}^{(g)})^2 + 2\sum_{i<j}(w_{ij}^{(g)})^2$. The covariance matrix of $w_{11}^{(g)}, \dots, w_{pp}^{(g)}$ is $2(\kappa+1)I_p + \kappa\varepsilon\varepsilon$, where $\varepsilon = (1, \dots, 1)'$. The characteristic roots of this matrix are $2(\kappa+1)$ of multiplicity $p-1$ and a single root of $2(\kappa+1) + p\kappa$. Thus $\sum_{i=1}^p (w_{ii}^{(g)})^2$ has the distribution of $2(\kappa+1)\chi_{p-1}^2 + [2(\kappa+1) + p\kappa]\chi_1^2$. The distribution of $2\sum_{i<j}(w_{ij}^{(g)})^2$ is $2(\kappa+1)\chi_{p(p-1)/2}^2$.

Theorem 10.11.1. *When sampling from* (1) *and the null hypothesis is true,*

(8)
$$-2\log\lambda_1 \xrightarrow{d} (\kappa+1)\chi^2_{(q-1)(p-1)(p+2)/2} + [(\kappa+1) + p\kappa/2]\,\chi^2_{q-1}.$$

When $\kappa = 0$, $-2\log\lambda_1 \xrightarrow{d} \chi^2_{(q-1)p(p+1)/2}$ is in agreement with (12) of Section 10.5. The validity of the distributions derived in Section 10.4 depend on the observations being normally distributed; Theorem 10.11.1 shows that even the asymptotic theory depends on nonnormality.

The likelihood criteria for testing the null hypothesis (2) of Section 10.3 is the product $\lambda_1\lambda_2$ or V_1V_2. Lemma 10.4.1 states that under normality V_1 and V_2 (or equivalently λ_1 and λ_2) are independent. In the elliptically contoured case we want to show that $\log V_1$ and $\log V_2$ are asymptotically independent.

Lemma 10.11.1. *Let $A_1 = n_1S_1$ and $A_2 = n_2S_2$ be defined by* (2) *of Section* 10.2 *with $\Sigma_1 = \Sigma_2 = I$. Then $A_1(A_1 + A_2)^{-1}$ and $A_1 + A_2$ are asymptotically independent.*

Proof. Let $(1\sqrt{n_g})(A_g - n_gI) = W_g$, $g = 1,2$. Then

(9)
$$\sqrt{n_1}\left[A_1(A_1+A_2)^{-1} - \frac{n_1}{n_1+n_2}I\right] = \frac{n_1n_2}{(n_1+n_2)^2}W_1 - \frac{n_1\sqrt{n_1n_2}}{(n_1+n_2)^2}W_2 + O_p(1),$$

(10)
$$\sqrt{n_1+n_2}\left[A_1+A_2-(n_1+n_2)I\right] = \sqrt{\frac{n_1}{n_1+n_2}}\,W_1 + \sqrt{\frac{n_2}{n_1+n_2}}\,W_2 + O_p(1).$$

Then

$$\mathscr{E}\operatorname{vec}\left(\frac{n_1n_2}{(n_1+n_2)^2}W_1 - \frac{n_1\sqrt{n_1n_2}}{(n_1+n_2)^2}W_2\right)\left[\operatorname{vec}\left(\sqrt{\frac{n_1}{n_1+n_2}}\,W_1 + \sqrt{\frac{n_2}{n_1+n_2}}\,W_2\right)\right]' = \mathbf{0}. \tag{11}$$

■

By application of Lemma 10.11.1 in succession to A_1 and $A_1 + A_2$, to $A_1 + A_2$ and $A_1 + A_2 + A_3$, etc., we establish that $A_1A^{-1}, A_2A^{-1}, \ldots, A_qA^{-1}$ are independent of $A = A_1 + \cdots + A_q$. It follows that V_1 and V_2 are asymptotically independent.

Theorem 10.11.2. *When $\Sigma_1 = \cdots = \Sigma_g$ and $\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(g)}$,*

$$\begin{aligned} -2\log\lambda_1\lambda_2 &= -2\log\lambda_1 - 2\log\lambda_2 \\ &\xrightarrow{d} (\kappa+1)\chi^2_{(q-1)(p-1)(p+2)/2} + [(\kappa+1)+p\kappa/2]\,\chi^2_{q-1} + \chi^2_{p(q-1)}. \end{aligned} \tag{12}$$

The hypothesis of sphericity is that $\Sigma = \sigma^2 I$ (or $\Lambda = \lambda I$). The criteron is $\lambda_1 \lambda_2$, where

$$\lambda_1 = \left(\frac{|A|}{\prod_{i=1}^p a_{ii}} \right)^{N/2}, \qquad \lambda_2 = \left[\frac{\prod_{i=1}^p a_{ii}}{\left(\frac{\operatorname{tr} A}{p} \right)^p} \right]^{N/2}. \tag{13}$$

The first factor is the criteron for independence of the components of X, and the second is that the variances of the components are equal. For the first we set $q = p$ and $p_i = 1$ in Theorem 9.10, and for the second we set $q = p$ and $p = 1$. Thus

$$-2 \log(\lambda_1 \lambda_2) \xrightarrow{d} (1 + \kappa) \chi^2_{p-1} + \tfrac{1}{2}(3\kappa + 2) \chi^2_{p-1}. \tag{14}$$

10.11.2. Elliptically Contoured Matrix Distributions

Consider the density

$$\begin{aligned} \prod_{g=1}^q |\Lambda_g|^{-N_g/2} g \left[\operatorname{tr} \sum_{g=1}^q \Lambda_g^{-1} \left(X^{(g)} - \nu_g \varepsilon'_{N_g} \right) \left(X^{(g)} - \nu_g \varepsilon'_{N_g} \right)' \right] \\ = \prod_{g=1}^q |\Lambda_g|^{-N_g/2} g \left[\operatorname{tr} \sum_{g=1}^q \Lambda_g^{-1} A_g + \sum_{g=1}^q N_g (\bar{x}^{(g)} - \nu^{(g)})' \Lambda_g^{-1} (\bar{x}^{(g)} - \nu^{(g)}) \right]. \end{aligned} \tag{15}$$

In this density $(A_g, \bar{x}_g)$, $g = 1, \ldots, q$, is a sufficient set of statistics, and the likelihood ratio criterion is (8) of Section 10.2, the same as for normality [Anderson and Fang (1990b)].

Theorem 10.11.3. *Let $f(X)$ be a vector-valued function of $X = (X^{(1)}, \ldots, X^{(q)})$ $(p \times N)$ such that*

$$f\left(X^{(1)} + \nu^{(1)} \varepsilon'_{N_1}, \ldots, X^{(q)} + \nu^{(q)} \varepsilon'_{N_g} \right) = f(X^{(1)}, \ldots, X^{(q)}) \tag{16}$$

for every $(\nu^{(1)}, \ldots, \nu^{(q)})$ and

$$f(CX^{(1)}, \ldots, CX^{(q)}) = f(X^{(1)}, \ldots, X^{(q)}) \tag{17}$$

for every nonsingular C. Then the distribution of $f(X)$ where X has the arbitrary density (15) with $\Lambda_1 = \cdots = \Lambda_q$ is the same as the distribution of $f(X)$ where X has the normal density (15).

The proof of Theorem 10.11.3 is similar to the proof of Theorem 4.5.4. The theorem implies that the distribution of the criterion V_1 of (10) of Section 10.2 when the density of X is (15) with $\Lambda_1 = \cdots = \Lambda_q$ is the same as for normality. Hence the distributions and their asymptotic expansions are those discussed in Sections 10.4 and 10.5.

Corollary 10.11.1. *Let $f(X)$ be a vector-valued function of X $(p \times N)$ such that*

$$f(X + \boldsymbol{\nu}\boldsymbol{\varepsilon}_N') = f(X) \tag{18}$$

for every $\boldsymbol{\nu}$ and (17) holds. Then the distribution of $f(X)$, where X has the arbitrary density (15) with $\Lambda_1 = \cdots = \Lambda_q$ and $\boldsymbol{\nu}^{(1)} = \cdots = \boldsymbol{\nu}^{(q)}$, is the same as the distribution of $f(X)$, where X has the normal density (15).

If follows that the distribution of the criterion λ of (7) or V of (11) of Section 10.3 is the same for the density (15) as for X being normally distributed.

Let X $(p \times N)$ have the density

$$|\Lambda|^{-N/2} g\left[\operatorname{tr} \Lambda^{-1}(X - \boldsymbol{\nu}\boldsymbol{\varepsilon}_N')(X - \boldsymbol{\nu}\boldsymbol{\varepsilon}_N')'\right]. \tag{19}$$

Then the likelihood ratio criterion for testing the null hypothesis $\Lambda = \lambda I$ for some $\lambda > 0$ is (7) of Section 10.7, and its distribution under the null hypothesis is the same as for X being normally distributed.

For more detail see Anderson and Fang (1990b) and Fang and Zhang (1990).

PROBLEMS

10.1. (Sec. 10.2) Sums of squares and cross-products of deviations from the means of four measurements are given below (from Table 3.4). The populations are *Iris versicolor* (1), *Iris setosa* (2), and *Iris virginica* (3); each sample consists of 50 observations:

$$A_1 = \begin{pmatrix} 13.0552 & 4.1740 & 8.9620 & 2.7332 \\ 4.1740 & 4.8250 & 4.0500 & 2.0190 \\ 8.9620 & 4.0500 & 10.8200 & 3.5820 \\ 2.7332 & 2.0190 & 3.5820 & 1.9162 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 6.0882 & 4.8616 & 0.8014 & 0.5062 \\ 4.8616 & 7.0408 & 0.5732 & 0.4556 \\ 0.8014 & 0.5732 & 1.4778 & 0.2974 \\ 0.5062 & 0.4556 & 0.2974 & 0.5442 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 19.8128 & 4.5944 & 14.8612 & 2.4056 \\ 4.5944 & 5.0962 & 3.4976 & 2.3338 \\ 14.8612 & 3.4976 & 14.9248 & 2.3924 \\ 2.4056 & 2.3338 & 2.3924 & 3.6962 \end{pmatrix}.$$

(a) Test the hypothesis $\Sigma_1 = \Sigma_2$ at the 5% significance level.
(b) Test the hypothesis $\Sigma_1 = \Sigma_2 = \Sigma_3$ at the 5% significance level.

10.2. (Sec. 10.2)

(a) Let $Y^{(g)}$, $g = 1, \ldots, q$, be a set of random vectors each with p components. Suppose

$$\mathscr{E} Y^{(g)} = \mathbf{0}, \qquad \mathscr{E} Y^{(g)} Y^{(h)\prime} = \delta_{gh} \Sigma_g.$$

Let C be an orthogonal matrix of order q such that each element of the last row is

$$c_{qh} = 1/\sqrt{q}.$$

Define

$$Z^{(g)} = \sum_{h=1}^{q} c_{gh} Y^{(h)}, \qquad g = 1, \ldots, q.$$

Show that

$$\mathscr{E} Z^{(q)} Z^{(g)\prime} = \mathbf{0}, \qquad g = 1, \ldots, q-1,$$

if and only if

$$\Sigma_1 = \Sigma_2 = \cdots = \Sigma_q.$$

(b) Let $X_\alpha^{(g)}$, $\alpha = 1, \ldots, N$, be a random sample from $N(\mu^{(g)}, \Sigma_g)$, $g = 1, \ldots, q$. Use the result from (a) to construct a test of the hypothesis

$$H: \Sigma_1 = \cdots = \Sigma_q,$$

based on a test of independence of $Z^{(q)}$ and the set $Z^{(1)}, \ldots, Z^{(q-1)}$. Find the exact distribution of the criterion for the case $p = 2$.

10.3. (Sec. 10.2) *Unbiasedness of the modified likelihood ratio test of $\sigma_1^2 = \sigma_2^2$.* Show that (14) is unbiased. [*Hint:* Let $G = n_1 F/n_2$, $r = \sigma_1^2/\sigma_2^2$, and $c_1 < c_2$ be the solutions to $G^{\frac{1}{2}n_1}(1+G)^{-\frac{1}{2}(n_1+n_2)} = k$, the critical value for the modified likelihood ratio criterion. Then

$$\Pr\{\text{Acceptance} \mid \sigma_1^2/\sigma_2^2 = r\} = \text{const} \int_{c_1}^{c_2} r^{\frac{1}{2}n_1} G^{\frac{1}{2}n_1 - 1}(1 + rG)^{-\frac{1}{2}(n_1+n_2)}\, dG$$

$$= \text{const} \int_{rc_1}^{rc_2} H^{\frac{1}{2}n_1 - 1}(1 + H)^{-\frac{1}{2}(n_1+n_2)}\, dH.$$

Show that the derivative of the above with respect to r is positive for $0 < r < 1$, 0 for $r = 1$, and negative for $r > 1$.]

10.4. (Sec. 10.2) Prove that the limiting distribution of (19) is χ_f^2, where $f = \frac{1}{2}p(p+1)(q-1)$. [*Hint:* Let $\Sigma = I$. Show that the limiting distribution of (19) is the limiting distribution of

$$\frac{1}{2}\sum_{i=1}^{p}\sum_{g=1}^{q} n_g\left(s_{ii}^{(g)} - s_{ii}\right)^2 + \sum_{i<j}\sum_{g=1}^{q} n_g\left(s_{ij}^{(g)} - s_{ij}\right)^2,$$

where $S^{(g)} = (s_{ij}^{(g)})$, $S = (s_{ij})$, and the $\sqrt{n_g}\,(s_{ij}^{(g)} - \delta_{ij})$, $i \le j$, are independent in the limiting distribution, the limiting distribution of $\sqrt{n_g}\,(s_{ii}^{(g)} - 1)$ is $N(0,2)$, and the limiting distribution of $\sqrt{n_g}\,s_{ij}^{(g)}$, $i<j$, is $N(0,1)$.]

10.5. (Sec. 10.4) Prove (15) by integration of Wishart densities. [*Hint:* $\mathscr{E}V_1^h = \mathscr{E}\prod_{g=1}^q |A_g|^{\frac{1}{2}n_g}|A|^{-\frac{1}{2}n}$ can be written as the integral of a constant times $|A|^{-\frac{1}{2}n}\prod_{g=1}^q w(A_g|\Sigma, n_g + hn_g)$. Integration over $\sum_{g=1}^q A_g = A$ gives a constant times $w(A|\Sigma, n)$.]

10.6. (Sec. 10.4) Prove (16) by integration of Wishart and normal densities. [*Hint:* $\sum_{g=1}^q N_g(\bar{x}^{(g)} - \bar{x})(\bar{x}^{(g)} - \bar{x})'$ is distributed as $\sum_{f=1}^{q-1} y_f y_f'$. Use the hint of Problem 10.5.]

10.7. (Sec. 10.6) Let $x_1^{(\nu)}, \dots, x_N^{(\nu)}$ be observations from $N(\mu^{(\nu)}, \Sigma_\nu)$, $\nu = 1,2$, and let $A_\nu = \sum(x_\alpha^{(\nu)} - \bar{x}^{(\nu)})(x_\alpha^{(\nu)} - \bar{x}^{(\nu)})'$.

(a) Prove that the likelihood ratio test for $H: \Sigma_1 = \Sigma_2$ is equivalent to rejecting H if

$$T = \frac{|A_1| \cdot |A_2|}{|A_1 + A_2|^2} \le C.$$

(b) Let $d_1^2, d_2^2, \dots, d_p^2$ be the roots of $|\Sigma_1 - \lambda\Sigma_2| = 0$, and let

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_p \end{pmatrix}.$$

Show that T is distributed as $|B_1| \cdot |B_2| / |B_1 + B_2|^2$, where B_1 is distributed according to $W(D^2, N-1)$ and B_2 is distributed according to $W(I, N-1)$. Show that T is distributed as $|DC_1D| \cdot |C_2| / |DC_1D + C_2|^2$, where C_i is distributed according to $W(I, N-1)$.

10.8. (Sec. 10.6) For $p = 2$ show

$$\begin{aligned}\Pr\{V_1 \le v\} = {} & I_a(n_1 - 1, n_2 - 1) \\ & + B^{-1}(n_1 - 1, n_2 - 1)v^{(n_1+n_2-2)/n}\int_a^b x^{-2n_2/n}(1 - x_1)^{-n_1/n}\,dx_1 \\ & + 1 - I_b(n_1 - 1, n_2 - 1),\end{aligned}$$

where $a < b$ are the two roots of $x_1^{n_1}(1-x_1)^{n_2} = v \le n_1^{n_1} n_2^{n_2}/n^n$. [*Hint:* This follows from integrating the density defined by (8).]

10.9. (Sec. 10.6) For $p = 2$ and $n_1 = n_2 = m$, say, show

$$\Pr\{V_1 \le v\}$$

$$= 2I_a(m-1, m-1) + 2B^{-1}(m-1, m-1)v^{1-(1/m)} \log \frac{1+\sqrt{1-4v^{1/m}}}{1-\sqrt{1-4v^{1/m}}}.$$

where $a = \frac{1}{2}[1 - \sqrt{1-4v^{1/m}}]$.

10.10. (Sec. 10.7) Find the distribution of W for $p = 2$ under the null hypothesis (**a**) directly from the distribution of A and (**b**) from the distribution of the characteristic roots (Chapter 13).

10.11. (Sec. 10.7) Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be a sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. What is the likelihood ratio criterion for testing the hypothesis $\boldsymbol{\mu} = k\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0$, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are specified and k is unspecified?

10.12. (Sec. 10.7) Let $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_{N_1}^{(1)}$ be a sample from $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_1)$, and $\boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_{N_2}^{(2)}$ be a sample from $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_2)$. What is the likelihood ratio criterion for testing the hypothesis that $\boldsymbol{\Sigma}_1 = k^2\boldsymbol{\Sigma}_2$, where k is unspecified? What is the likelihood ratio criterion for testing the hypothesis that $\boldsymbol{\mu}^{(1)} = k\boldsymbol{\mu}^{(2)}$ and $\boldsymbol{\Sigma}_1 = k^2\boldsymbol{\Sigma}_2$, where k is unspecified?

10.13. (Sec. 10.7) Let $\boldsymbol{x}_\alpha$ of p components, $\alpha = 1, \ldots, N$, be observations from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We define the following hypotheses:

$$H: \boldsymbol{\mu} = \mathbf{0}, \ \boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0,$$

$$H_1: \boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0,$$

$$H_2: \boldsymbol{\mu} = \mathbf{0}, \qquad \text{given that} \quad \boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0.$$

In each case k^2 is unspecified, but $\boldsymbol{\Sigma}_0$ is specified. Find the likelihood ratio criterion λ_2 for testing H_2. Give the asymptotic distribution of $-2\log\lambda_2$ under H_2. Obtain the exact distribution of a suitable monotonic function of λ_2 under H_2.

10.14. (Sec. 10.7) Find the likelihood ratio criterion λ for testing H of Problem 10.13 (given $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$). What is the asymptotic distribution of $-2\log\lambda$ under H?

10.15. (Sec. 10.7) Show that $\lambda = \lambda_1\lambda_2$, where λ is defined in Problem 10.14, λ_2 is defined in Problem 10.13, and λ_1 is the likelihood ratio criterion for H_1 in Problem 10.13. Are λ_1 and λ_2 independently distributed under H? Prove your answer.

10.16. (Sec. 10.7) Verify that $\operatorname{tr} \boldsymbol{B}\boldsymbol{\Psi}_0^{-1}$ has the χ^2-distribution with $p(N-1)$ degrees of freedom.

10.17. (Sec. 10.7.1) *Admissibility of sphericity test.* Prove that the likelihood ratio test of sphericity is admissible. [*Hint:* Under the null hypothesis let $\boldsymbol{\Sigma} = [1/(1+\eta^2)]\boldsymbol{I}$, and let η have the density $(1+\eta^2)^{-\frac{1}{2}np}(\eta^2)^{p-\frac{1}{2}}$.]

10.18. (Sec. 10.10.1) Show that for $r \geq p$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{i=1}^{r} \boldsymbol{x}_i \boldsymbol{x}_i' \right|^{\frac{1}{2}t} \left| \boldsymbol{I} + \sum_{i=1}^{r} \boldsymbol{x}_i \boldsymbol{x}_i' \right|^{-\frac{1}{2}n} \prod_{i=1}^{r} d\boldsymbol{x}_i < \infty$$

if $2p - 1 < t + r + p < n + 1$. [*Hint:* $|\boldsymbol{A}|/|\boldsymbol{I}+\boldsymbol{A}| \leq 1$ if $\boldsymbol{A}$ is positive semidefinite. Also, $|\sum_{i=1}^{r} \boldsymbol{x}_i \boldsymbol{x}_i'|$ has the distribution of $\chi_r^2 \chi_{r-1}^2 \cdots \chi_{r-p+1}^2$ if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ are independently distributed according to $N(\boldsymbol{0}, \boldsymbol{I})$.]

10.19. (Sec. 10.10.1) Show

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\boldsymbol{C}\boldsymbol{C}'|^{\frac{1}{2}t} e^{-\frac{1}{2}\operatorname{tr} \boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}} \, d\boldsymbol{C} = \text{const} |\boldsymbol{A}|^{-\frac{1}{2}(t+r)},$$

where $\boldsymbol{C}$ is $p \times r$. [*Hint:* $\boldsymbol{C}\boldsymbol{C}'$ has the distribution $W(\boldsymbol{A}^{-1}, r)$ if $\boldsymbol{C}$ has a density proportional to $e^{-\frac{1}{2}\operatorname{tr} \boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}}$.]

10.20. (Sec. 10.10.1) Using Problem 10.18, complete the proof of Theorem 10.10.1.