A NEW CLASS OF MULTIVARIATE TESTS BASED
ON THE UNION-INTERSECTION PRINCIPLE

BY

JACK LEON TOMSKY

TECHNICAL REPORT NO. 83
AUGUST 1, 1974

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-32326X

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A NEW CLASS OF MULTIVARIATE TESTS BASED
ON THE UNION-INTERSECTION PRINCIPLE

by

Jack Leon Tomsky

Lockheed Palo Alto, Research Laboratory

and

Stanford University

TECHNICAL REPORT NO. 83

August 1, 1974

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT GP-32326X

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Professor Ingram Olkin for his guidance, encouragement, and patience. His enthusiasm for the subject is contagious and I have learned much from my association with him.

This work was supported by the Independent Research program of Lockheed Missiles and Space Company. In line with this, I would like to thank Dr. Ted Jenkins and Dr. Lee Stewart for their encouragement and for granting me the time necessary to bring this dissertation to a conclusion.

My appreciation goes to Carolyn Knutsen for a superb and accurate typing job.

Finally, I would like to thank my family for their understanding attitude which made this thesis possible.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1 Notation</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>SOME RESULTS ON MATRIX EXTREMAL PROBLEMS</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>UNION-INTERSECTION TESTS FOR SOME STANDARD MULTIVARIATE PROBLEMS</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>3.1 Testing that the Mean Vector is Zero</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>3.2 Testing Equality of Several Mean Vectors</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>3.3 Testing that the Covariance Matrix is the Identity Matrix</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>3.4 Testing Equality of Two Covariance Matrices</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>3.5 Testing for Independence Between Two Sets of Variates</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>3.6 Testing for Sphericity</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>GENERALIZED STEP-DOWN PROCEDURE</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>4.1 The General Linear Model</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>4.2 Equality of the Covariance Matrix to the Identity Matrix</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>4.3 Equality of Two Covariance Matrices</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>MONOTONICITY OF THE POWER FUNCTIONS OF SOME PROPOSED TESTS</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>5.1 The General Linear Model</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>5.2 Independence Between Two Sets of Variates</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>COMPARISON OF TESTS BY BAHADUR EFFICIENCY</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>6.1 Bahadur Efficiency</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>6.2 Some Theorems Regarding the Class D(a,t)</td>
<td>59</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>6.3 Comparison of Tests on the Equality of the Covariance Matrix to the Identity</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>6.4 Comparison Tests on the Equality of Two Covariance Matrices</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>6.5 Comparison of Tests of the General Linear Hypothesis</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>6.6 Comparison of Tests on the Independence Between Two Sets of Variates</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>7 SIMULTANEOUS CONFIDENCE BOUNDS ON POPULATION ELEMENTARY SYMMETRIC FUNCTIONS</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>7.1 Simultaneous Confidence Bounds on a Single Covariance Matrix</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>7.2 Simultaneous Confidence Bounds of Two Covariance Matrices</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>7.3 The General Linear Model</td>
<td>83</td>
<td></td>
</tr>
</tbody>
</table>

BIBLIOGRAPHY 86
A NEW CLASS OF MULTIVARIATE TESTS BASED
ON THE UNION-INTERSECTION PRINCIPLE

by

Jack Leon Tomskey

1. INTRODUCTION.

In a paper by Roy (1953), the union-intersection principle was introduced as a heuristic method of test construction which can be described as follows. Let \( \{ \omega_a^*, \omega_a^\prime, \epsilon_a \in \Gamma \} \) be a collection of sets in the parameter space. It is assumed that \( \bigcap_{a \in \Gamma} \omega_a \) is non-null and \( \Gamma \) is an arbitrary index set. We define the component hypotheses \( H_a \) and \( K_a \) as the hypotheses,

\[
H_a : \forall \omega_a^* \in \omega_a \quad , \quad K_a : \forall \theta \in \omega_a^* 
\]

Suppose that there exists a test of size \( \alpha_a \) in which one accepts \( H_a \) over \( K_a \) for sample points in the set \( A_a \) and rejects \( H_a \) otherwise. We assume that \( \alpha_a = \alpha^* \) is the same for all component tests.

The union-intersection testing problem is constructed from the component testing problems. Let us set

\[
H = \bigcap_{a \in \Gamma} H_a \quad , \quad K = \bigcup_{a \in \Gamma} K_a .
\]

The above symbolism means that the null hypothesis \( H \) is true if and only if every null component hypothesis \( H_a \) is true. Similarly, the alternative hypothesis \( K \) is true if and only if at least one alternative component hypothesis is true. Under the union-intersection principle, \( H \) is accepted over \( K \) if and only if each component test accepts \( H_a \) over the corresponding \( K_a \). That is, the acceptance region for a union-intersection test is given by \( A = \bigcap_{a \in \Gamma} A_a \).
In practice, one starts with a null hypothesis $H$, an alternative hypothesis $K$, and a significance level $\alpha$. $H$ and $K$ are then represented as an intersection of null component hypotheses and a union of alternative component hypotheses, respectively. The (common) size $\alpha^*$ of each component test is so determined that the size of $A$ under $H$ is equal to the preassigned $\alpha$.

Although the union-intersection principle has been applied to univariate problems (for example, analysis of variance and homogeneity of variances), its main applications have been in the area of multivariate statistics. The reason is that for most univariate problems, the union-intersection principle yields a test that coincides with the classical test. For some univariate examples, see Roy (1957, pp. 6-8).

Union-intersection tests are generally dependent upon the selection of the representation for $H$ and $K$. In most applications, the index set $\Gamma$ has been chosen as consisting of all nonzero $p$-dimensional vectors. These union-intersection tests are then constructed from well-known univariate tests.

In this study we investigate union-intersection tests which are generalized in two directions. First, the index set $\Gamma$ consists of matrices $A_i$ of order $k \times p$ and of full rank $k$, where $1 \leq k \leq p$. We next have a choice of multivariate tests. A convenient way of representing these tests is the application of elementary symmetric functions $E_m$ to the $k$ characteristic roots of the matrix appearing in the likelihood ratio component test. Note that $E_k$ is the determinant and $E_1$ is the trace of this matrix.

A remaining source of ambiguity is that if $W^n$, say, is a matrix which appears in a likelihood ratio component test, then $E_k$ of the
characteristic roots of any power of \( W \) yields an equivalent test. But for \( m < k \), the test does depend on the specific chosen power of \( W \). In applications, where possible, \( n \) has been chosen as equal to one.

By denoting \( T_{mk} \) as the union-intersection test statistic obtained by applying the \( m \)-th elementary symmetric function to component hypotheses of dimension \( k \), where \( 1 \leq m \leq k \leq p \), it can be seen that this class includes several standard multivariate tests. For example, \( T_{pp} \) coincides with the likelihood ratio test based on determinants, \( T_{lp} \) is the Hotelling-Lawley "trace" criterion (to be discussed later), and \( T_{ll} \) is the union-intersection test developed by Roy based on the extreme sample characteristic roots.

Moments and distributions of elementary symmetric functions have been studied by Mijares (1961), Pillai and Mijares (1959), Pillai (1964, 1965, 1966), and Pillai and Gupta (1967). However, in this study, we are not direction concerned with distributional problems.

Multivariate testing problems have been studied fairly extensively in the literature. Most of the classical tests were obtained by using the likelihood ratio criterion. One of the earliest multivariate tests is the \( T^2 \)-test proposed by Hotelling (1931) for testing the equality of two mean vectors. Hotelling's test turns out to be equivalent to the likelihood ratio test. Other important multivariate tests on the normal and Wishart distributions which are based on the likelihood ratio criterion were obtained by Wilks (1932, 1935, 1946).

Another criterion for multivariate tests is the trace criterion, introduced by Lawley (1958) and later by Hotelling (1947). It is sometimes denoted by \( T^2 \) and is defined as follows. If \( S_1 \) and \( S_2 \) estimate covariance matrices \( \Sigma_1 \) and \( \Sigma_2 \), respectively, then \( T^2 = tr(S_1S_2^{-1}) \) is used for testing \( \Sigma_1 = \Sigma_2 \).
Other principles which have been used for multivariate testing (when they exist) are uniformly most powerful tests, uniformly most powerful invariant tests, and most stringent tests. The reason for the proliferation of tests is that the class of admissible tests is usually very broad. For example, suppose that we restrict our attention to tests which are invariant under a group of linear transformations. These tests are based on the functions of the sample characteristic roots of a random matrix. In certain problems, such as tests on the mean vector or multiple correlation, there is only a single nonzero characteristic root. For these cases, there exists a uniformly most powerful invariant test which is based on this single characteristic root. In the more general case with more than one nonzero characteristic root, no uniformly most powerful invariant test exists (Lehmann (1959), p. 299). Most of the proposed tests have been based on various functions of the characteristic roots.

The union-intersection principle has been used to further increase the number of available multivariate tests. For particular problems, union-intersection tests are known to possess some "good" properties. One such property is "monotonicity". A test is said to have the monotonicity property if its power function is an increasing function of each ordered population characteristic root (noncentrality parameter). The standard union-intersection tests of Roy have been shown to satisfy the monotonicity property for the following problems: the general linear model (Roy and Mikhail (1961), Anderson, Das Gupta, and Mudkolkar (1964)), independence between two sets of variates (Roy and Mikhail (1961), Anderson and Das Gupta (1964a)), and equality of two covariance matrices against certain restricted alternatives (Mikhail (1962), Anderson and Das Gupta (1964b)).
Another type of monotonicity has been studied by Das Gupta and Perlman (1973, 1974). They consider the general linear model in which the rank of the population noncentrality matrix is one; that is, there is only a simple nonzero population characteristic root. They show that when this characteristic root is fixed, the power of the likelihood ratio test strictly decreases with the dimension $p$ and the hypothesis degrees of freedom and strictly increases with the error degrees of freedom.

A special type of union-intersection procedure, called the step-down procedure has been suggested by J. Roy (1958). Under this approach, the null hypothesis is expressed as a finite intersection of component hypotheses concerning the conditional distribution of a single variate, given all the previous variates. When the null hypothesis is true, these tests are independent. An advantage of the step-down procedure is the simplification of distribution problems. A disadvantage is that this procedure is dependent on the ordering of the variates.

The structure of union-intersection tests leads naturally into the construction of simultaneous confidence bounds. By inverting each component hypothesis into a confidence statement about a parametric function concerning that hypothesis, it is possible to compute the probability of the simultaneous truth of all confidence statements. Simultaneous confidence bounds constructed in this manner were first developed by Roy and Bose (1953).

In the application of the union-intersection principle to multivariate problems, we are usually led to solving extremal problems. For example, when the indexing set $\Gamma$ consists of $p$-dimensional vectors (i.e., $k = 1$), the construction of a union-intersection test usually involves minimizing
and/or maximizing a quadratic form with respect to a vector. As we consider component hypotheses indexed by \( k \times p \) matrices, the computations require us to minimize and/or maximize a function with respect to a matrix argument. Chapter 2 provided a discussion and solutions to these extremal problems.

In Chapter 3, we express several standard multivariate problems in terms of multivariate component hypotheses. The component tests are based on elementary symmetric functions \( E_m \) of the \( k \) characteristic roots associated with the component hypothesis in the likelihood ratio. For certain values of \( m \) and \( k \), the resulting union-intersection test coincides with a standard test. For other combinations of \( m \) and \( k \), the derived test is new.

In Chapter 4, the step-down procedure is generalized. The variates are partitioned into blocks and the null hypothesis \( H \) is represented as a finite intersection of component hypotheses. Each component hypothesis concerns the conditional distribution of the \( i \)-th block of variates given the first \( i-1 \) blocks. This generalized step-down procedure is applied to the three testing problems considered by J. Roy (1958); namely, (a) the general linear model, (b) equality of the covariance matrix to the identity matrix, and (c) equality of two covariance matrices.

In Chapter 5, we prove that for the problems of the general linear model and independence between two sets of variates, each of the proposed tests satisfies the property of monotonicity.

Chapter 6 compares the various tests with respect to indices \( m \) and \( k \). By using Bahadur efficiency as a method of comparison, we are able to compare the indices without using the non-null distributions of the test statistics.
In Chapter 7, simultaneous confidence bounds are derived for elementary symmetric functions of population parameters. The simultaneity is respect to the index matrices \( A \). These bounds are not direct inversions of the component tests and may tend to be rather conservative. That is, the simultaneous confidence level is greater than \( 1 - \alpha \).

1.1 NOTATION

The following notation is used throughout. The dimension of a matrix \( A \) is denoted by \( A: m \times n \). The characteristic roots of an \( n \times n \) matrix \( A \) are denoted by \( \lambda(A) \). When these roots are real, we order them by

\[
\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A).
\]

We write \( S \geq 0 \) when the symmetric matrix \( S \) is positive semi-definite. Similarly, \( S > 0 \) denotes that the symmetric matrix \( S \) is positive definite. When \( S_1 \) and \( S_2 \) are each symmetric matrices, \( S_1 \geq S_2 \) and \( S_1 > S_2 \) mean that \( S_1-S_2 \geq 0 \) and \( S_1-S_2 > 0 \), respectively.

Unless otherwise noted, all positive (semi-)definite matrices are of dimension \( p \times p \). The symbol \( \text{etr}(A) \) is used to denote the function \( \exp(\text{trace}(A)) \).

The \( m \)-th elementary symmetric function \( E_m \) of \( n \) variables \( x_1, x_2, \ldots, x_n \) with \( m \leq n \) is defined by

\[
E_m(x_1, x_2, \ldots, x_n) = \sum_{i_1 < \cdots < i_m} x_{i_1} x_{i_2} \ldots x_{i_m}.
\]

(1.3)

If \( \lambda_1, \ldots, \lambda_n \) are the characteristic roots of a symmetric matrix \( A \), the function \( \text{tr}_m(A) \) is defined by

\[
\text{tr}_m(A) = E_m(\lambda_1, \lambda_2, \ldots, \lambda_n) = \text{tr} A^{(m)},
\]

(1.4)

where \( A^{(m)} \) is the \( m \)-th compound of \( A \).
A random vector $\mathbf{x}$ has a $p$-dimensional normal distribution, $\mathcal{N}_p(\mu, \Sigma)$, if its density function is

$$f(\mathbf{x}) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \text{etr}(-\frac{1}{2} \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)')$$.

A random $p \times p$ matrix $S > 0$ has a Wishart distribution, $W(\Sigma, p, n)$, if its density function is given by

$$f(S) = c(\Sigma, p, n) |S|^{n-p-1} \text{etr}(-\frac{1}{2} \Sigma^{-1} S)$$

where

$$[c(\Sigma, p, n)]^{-1} = 2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma(\frac{n+i-1}{2})$$.
2. SOME RESULTS ON MATRIX EXTREMAL PROBLEMS.

This chapter contains results on extremal problems which are needed in the derivation of union-intersection tests in Chapter 3. Throughout this section we define

\[(2.1) \quad \mathcal{A}_{kp} = \{A: k \times p, \quad AA' = I_k\}\]

as the class of row-orthogonal matrices \(A\) of dimension \(k \times p\). Note that necessarily, \(1 \leq k \leq p\). In particular \(\mathcal{A}_{lp}\) is the class of unit vectors \(a\) and \(\mathcal{A}_{pp}\) is the class of orthogonal \(p \times p\) matrices \(A\). Because we deal with real matrices, the results are stated for real matrices; however, they can generally apply to complex matrices.

We use as a starting point a result which was obtained by Poincaré (1890) and is known as the Poincaré Separation Theorem.

**THEOREM 2.1 [Poincaré (1890)].** Let \(S \geq 0\) and \(A \in \mathcal{A}_{kp}\). Then

\[(2.2)\]

\[\text{(i) } \lambda_j(ASA') \leq \lambda_j(S) \quad \text{ and } \quad \lambda_{k-j+1}(ASA') \geq \lambda_{p-j+1}(S) \quad \text{ for } j = 1, 2, \ldots, k\]

From Theorem 2.1, we can deduce the following important min-max corollary.

**COROLLARY 2.2.** If \(S \geq 0\), then

\[\text{(i) } \min_{\mathcal{A}_{kp}} \lambda_1(ASA') = \lambda_{p-k+1}(S), \]

\[\text{(ii) } \max_{\mathcal{A}_{kp}} \lambda_1(ASA') = \lambda_1(S), \]

9
(iii) $\min_{\mathcal{A}_{kp}} \lambda_k(ASA') = \lambda_p(S)$,

(iv) $\max_{\mathcal{A}_{kp}} \lambda_k(ASA') = \lambda_k(S)$.

Proof. By Theorem 2.1,

$$\lambda_{p-k+1}(S) \leq \lambda_1(ASA') \leq \lambda_1(S),$$

(2.3)

$$\lambda_p(S) \leq \lambda_k(ASA') \leq \lambda_k(S).$$

It is well known [e.g., Bellman (1960, p. 54)] that since $S$ is symmetric, it can be factored according to the special decomposition, $S = \Gamma D_{\lambda} \Gamma'$, where $\Gamma$ is a $p \times p$ orthogonal matrix and $D_{\lambda}$ is a diagonal matrix whose diagonal elements are $\lambda_j(S)$, arranged in nonincreasing order.

It is easily verified that the upper limits in (2.3) are attained for $A_1 = (I_k, 0)\Gamma'$ and the lower limits are attained for $A_2 = (0, I_k)\Gamma'$. In both cases $A_1 A_1' = I_k$, so that $A_1 \in \mathcal{A}_{kp}$ for $i = 1, 2$. ||

The following corollary is also obtained from Theorem 2.1.

**Corollary 2.3.** If $S \geq 0$, then

(i) $\max_{\mathcal{A}_{kp}} \text{tr}(ASA') = \sum_{j=1}^{k} \lambda_j(S)$

(ii) $\min_{\mathcal{A}_{kp}} \text{tr}(ASA') = \sum_{j=1}^{k} \lambda_{p-j+1}(S)$

(iii) $\max_{\mathcal{A}_{kp}} |ASA'| = \prod_{j=1}^{k} \lambda_j(S)$

(iv) $\min_{\mathcal{A}_{kp}} |ASA'| = \prod_{j=1}^{k} \lambda_{p-j+1}(S)$.

Proof. By Theorem 2.1,
\[ 0 \leq \lambda_j'(\text{ASA}') \leq \lambda_j(S) \]

\[ 0 \leq \lambda_{p-j+1}(S) \leq \lambda_{k-j+1}(\text{ASA}') \]

for \( j = 1, 2, \ldots, k \). Taking the sum and product of both inequalities in (2.4) and noting that

\[ \text{tr}(\text{ASA}') = \sum_{j=1}^{k} \lambda_j(\text{ASA}') = \sum_{j=1}^{k} \lambda_{k-j+1}(\text{ASA}') \]

and

\[ |\text{ASA}'| = \prod_{j=1}^{k} \lambda_j(\text{ASA}') = \prod_{j=1}^{k} \lambda_{k-j+1}(\text{ASA}') , \]

we have

\[ \sum_{j=1}^{k} \lambda_{p-j+1}(S) \leq \text{tr}(\text{ASA}') \leq \sum_{j=1}^{k} \lambda_j(S) \]

\[ \prod_{j=1}^{k} \lambda_{p-j+1}(S) \leq |\text{ASA}'| \leq \prod_{j=1}^{k} \lambda_j(S) . \]

If \( S \) has the spectral decomposition, \( S = \lambda D \lambda', \) it is easily verified that the upper limits of (2.5) are attained for \( A = (I_k, 0) \lambda' \) and that the lower limits are attained for \( A = (0, I_A) \lambda' . \)

The first two parts of Corollary 2.3 were obtained by Fan (1951) and the last two parts were derived by Fan (1949). His derivation differs from that presented here and is more general in that it applies to complex matrices.

**DEFINITION.** Let \( A \) be a \( p \times q \) matrix and let \( m \leq \min(p, q) \). The \( m \)-th compound of a matrix \( A \), denoted by \( A^{(m)} \) is a matrix of dimension \( \binom{p}{m} \)
by \( \binom{a}{m} \) obtained by forming all minors of \( A \) and arranging them in lexicographic order.

Some of the properties of compound matrices are listed below and are found in Olkin (1966).

**Theorem 2.4.**

(i) if \( C = AB \), then \( C^{(m)} = A^{(m)}B^{(m)} \)

(ii) \( (A^{(m)})' = A^{(m)} \)

(iii) if \( A \) is (row-)orthogonal, then \( A^{(m)} \) is (row-)orthogonal

(iv) \( \text{tr}_m A = \text{tr} A^{(m)} \)

(v) if \( A \) is \( n \times n \) with characteristic roots \( \lambda_1, \ldots, \lambda_n \), then the characteristic roots of \( A^{(m)} \) are the \( \binom{n}{m} \) products of the form \( \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_m} \).

Part (i) of Theorem 2.4 is known as the Gauchy-Binet Theorem. Parts (iii) and (iv) are useful in extending certain extremal solutions of the trace to \( \text{tr}_m \).

**Theorem 2.5.** If \( S \geq 0 \) is a \( p \times p \) matrix, then for \( m = 1, 2, \ldots, k, \)

\[
\begin{align*}
(i) \max_{\mathcal{A}_{kp}} \text{tr}_m (ASA') &= E_m (\lambda_1(S), \ldots, \lambda_k(S)), \\
(ii) \min_{\mathcal{A}_{kp}} \text{tr}_m (ASA') &= E_m (\lambda_{p-k+1}(S), \ldots, \lambda_p(S)).
\end{align*}
\]

**Proof.** By Theorem 2.4 (iv), (i), and (ii),

\[
\begin{align*}
\text{tr}_m (ASA') &= \text{tr}(ASA')^{(m)} = \text{tr} A^{(m)} S^{(m)} A^{(m)}' \\
&= \text{tr} A^{(m)} S^{(m)} A^{(m)}'.
\end{align*}
\]

12
The matrix $A^{(m)}$ is row-orthogonal by Theorem 2.4(iii). Hence, by applying Corollary 2.3(i),(ii) and Theorem 2.4(v), we have

$$\max_{A_{kp}} \text{tr} A^{(m)} S^{(m)} A^{(m)\prime} = k \sum_{j=1}^{k} \lambda_j(S^{(m)}) = \text{E}_m(\lambda_1(S), \ldots, \lambda_k(S)),$$

$$\min_{A_{kp}} \text{tr} A^{(m)} S^{(m)} A^{(m)\prime} = k \sum_{j=1}^{k} \lambda_{p-j+1}(S^{(m)}) = \text{E}_m(\lambda_{p-k+1}(S), \ldots, \lambda_k(S)).$$

A recent paper by Berkowitz (1974) studies elementary symmetric functions of the form $\text{tr}_m(AS_1A^\prime + S_2)$, where $S_1$ and $S_2$ are positive-definite matrices of dimensions $p \times p$ and $k \times k$, respectively $A$ is a $k \times p$ row-orthogonal matrix. His result is that

$$\max_{A_{kp}} \text{tr}_m(AS_1A^\prime + S_2) = \text{E}_m(\theta_1, \theta_2, \ldots, \theta_k),$$

where $\theta_1 = \lambda_1(S_1) + \lambda_1(S_2)$.

All of the results presented thus far in this section concern a single positive semi-definite matrix $S$. Many applications, however, involve two matrices $S_1$ and $S_2$. These results can be easily extended to two matrices. Given two matrices $S_1$ and $S_2$ with $S_1 \succeq 0$ and $S_2 \succ 0$, consider the characteristic roots of the matrix $(AS_1A^\prime)(AS_2A^\prime)^{-1}$, where $A \in A_{kp}$.

$S_1$ and $S_2$ can be simultaneously factored [e.g., Bellman (1960, p. 51)] according to $S_1 = WD_{\lambda}W^\prime$ and $S_2 = WW^\prime$, where $W$ is a non-singular matrix and $D_{\lambda}$ is a diagonal matrix whose diagonal elements consist of $\lambda_j(S_1^{-1}S_2^{-1})$, the ordered characteristic roots of $S_1S_2^{-1}$. Then
\[ \lambda[(A_1 A')(A_2 A')^{-1}] = \lambda[AWD_{W'} W' A' (AWW'A')^{-1}] = \lambda[BD_{\lambda} B'(BB')^{-1}] \]

where \( B = AW \). Now,

\[ \lambda[BD_{\lambda} B'(BB')^{-1}] = \lambda[(BB')^{-1/2} \lambda BD_{\lambda} B'(BB')^{-1/2}] = \lambda(CD_{\lambda} C') \]

where \( C = (BB')^{-1/2} B \) and \( (BB')^{-1/2} \) is the (unique) symmetric square root matrix of \( (BB')^{-1} \). We note that \( C \in \mathcal{A}_{kp} \) since \( CC' = I_k \). Thus, all the preceding results hold with \( ASA' \) replaced by \( (A_1 A')(A_2 A')^{-1} \) and \( \lambda(S) \) replaced by \( \lambda(S_1 S_2^{-1}) \). For future reference, these results are listed in the following theorem.

**Theorem 2.6.** Suppose \( S_1 \geq 0 \) and \( S_2 > 0 \). For \( A \in \mathcal{A}_{kp} \), the following results hold. (The maximizations and minimizations are over the set \( \mathcal{A}_{kp} \).)

(i) \[ \lambda_j[(A_1 A')(A_2 A')^{-1}] \leq \lambda_j(S_1 S_2^{-1}) \quad (j=1, \ldots, k) \]

(ii) \[ \lambda_{k-j+1}[(A_1 A')(A_2 A')^{-1}] \geq \lambda_{p-j+1}(S_1 S_2^{-1}) \quad (j=1, \ldots, k) \]

(iii) \[ \min \lambda_1[(A_1 A')(A_2 A')^{-1}] = \lambda_{p-k+1}(S_1 S_2^{-1}) \]

(iv) \[ \max \lambda_1[(A_1 A')(A_2 A')^{-1}] = \lambda_1(S_1 S_2^{-1}) \]

(v) \[ \min \lambda_k[(A_1 A')(A_2 A')^{-1}] = \lambda_k(S_1 S_2^{-1}) \]

(vi) \[ \max \lambda_k[(A_1 A')(A_2 A')^{-1}] = \lambda_k(S_1 S_2^{-1}) \]

(vii) \[ \max \text{tr}_m[(A_1 A')(A_2 A')^{-1}] = E_m(\lambda_1(S_1 S_2^{-1}), \ldots, \lambda_k(S_1 S_2^{-1})) \]

(viii) \[ \min \text{tr}_m[(A_1 A')(A_2 A')^{-1}] = E_m(\lambda_{p-k+1}(S_1 S_2^{-1}), \ldots, \lambda_p(S_1 S_2^{-1})) \]
Let $\mathcal{B}_{kp}$ be the set of all matrices $B$, each $k \times p$ and of rank $k$. We note that Theorem 2.6 remains true for $A \in \mathcal{B}_{kp}$. To see this, suppose $B \in \mathcal{B}_{kp}$ and consider the characteristic roots of $(BS_1S')(BS_2S')^{-1}$.

The matrix $B$ can be factored as $B = TA$ where $T$ is a $k \times k$ nonsingular matrix and $A \in \mathcal{A}_{kp}$ (see Roy (1957, p. 149)). Then,

$$\lambda(BS_1S')(BS_2S')^{-1} = \lambda(TAS_1A'T')(TAS_2A'T')^{-1} = \lambda(AS_1A')(AS_2A')^{-1}.$$ 

Thus, we need not restrict $A$ to $\mathcal{A}_{kp}$ as far as Theorem 2.6 is concerned. The conclusions are valid for $A \in \mathcal{B}_{kp}$.

In constructing union-intersection tests in Chapter 3, we are led to maximizing elementary symmetric functions with respect to a matrix $A$.

The following theorem is basic in solving this problem.

**Theorem 2.7.** Let $f(x)$ be a non-negative analytic function of a real variable $x$ over the interval $0 \leq x \leq R$. ($R$ may be infinity.) Suppose that for some $x_0 \in [0,R]$, the function $f(x)$ is decreasing for $0 \leq x \leq x_0$ and increasing for $x_0 \leq x \leq R$. Let $S \geq 0$ have characteristic roots $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p$. Let $\theta[1], \theta[2], \ldots, \theta[p]$ denote the $\theta$'s arranged in decreasing order according to the values of $f(\theta)$. That is, $f(\theta[1]) \geq f(\theta[2]) \geq \cdots \geq f(\theta[p])$. Then for $m = 1, 2, \ldots, k$

$$\max_{\mathcal{A}_{kp}} \mathbb{E}_m[f(\lambda_1(ASA')), f(\lambda_2(ASA')), \ldots, f(\lambda_k(ASA'))]$$

$$= \mathbb{E}_m[f(\theta[1]), f(\theta[2]), \ldots, f(\theta[k])] .$$
Proof. We will first show that for any $A \in A_{kp}$,

\[(2.11) \quad E_m[f(\lambda_1(ASA')), \ldots, f(\lambda_k(ASA'))] \leq E_m[f(\theta_{[1]}), \ldots, f(\theta_{[k]})].\]

Consider first $k = 1$. For any $A \in A_{kp}$, $A$ is a row vector of unit length. The single characteristic root $\lambda_1(ASA')$ is equal to the scalar $ASA'$. By the characterization of $\theta_{[1]}$ and $\theta_{[p]}$ as the extreme values of $x'Sx$ when $x'x = 1$, we have $\theta_{[p]} \leq ASA' \leq \theta_{[1]}$. Because of the assumptions on the function $f(x)$, the maximum in every interval occurs at one of its end-points. Thus, $f(ASA')$ cannot be greater than both $f(\theta_{[1]})$ and $f(\theta_{[p]})$. Therefore,

\[(2.12) \quad \max_{A_{kp}} f(ASA') \leq f(\theta_{[1]}).\]

When $k = 1$, we must have $m = 1$ and since $E_1(x) = x$, (2.11) holds for $k = 1$.

Consider now the case for general $k$. Let $\theta_{[1]}', \theta_{[2]}', \ldots, \theta_{[k]}'$ be the $k$ $\theta'$s which have the largest values of $f(\theta)$. In terms of the original ordering of $\theta$, these form sequences at either one or both ends of the $\theta'$s. Let us denote these $\theta'$s by $\{\theta_{[1]}', \ldots, \theta_{[k]}', \theta_{[p-k+m+1]}', \ldots, \theta_{[p]}', \}$, admitting the possibility that $m = 0$ or $k$. By the characterization of the function $f$, we must have $\theta_{[m]}' \leq x_0 \leq \theta_{[p-k+m+1]}'$, where $x_0$ minimizes $f$. By Theorem 2.1,

\[(2.13) \quad \lambda_j(ASA') \leq \theta_j, \quad j = 1, \ldots, m\]

\[\lambda_{k-j+1}(ASA') \geq \theta_{p-j+1}, \quad j = 1, \ldots, k-m\]
Because of the assumptions on \( f \),

\[
\begin{align*}
\lambda_j(A \Sigma A^\prime) & \leq f(\theta_j) \quad j = 1, \ldots, m \\
\lambda_{k-j+1}(A \Sigma A^\prime) & \leq f(\theta_{p-j+1}) \quad j = 1, \ldots, k-m
\end{align*}
\]

(2.14)

Now \( \mathbb{E}_m(x_1, \ldots, x_k) \) is a nondecreasing function of each argument when all \( x_i \geq 0 \). Therefore, by (2.15),

\[
\begin{align*}
\mathbb{E}_m[f(\lambda_1(A \Sigma A^\prime), \ldots, f(\lambda_k(A \Sigma A^\prime))] & \leq \\
\mathbb{E}_m[f(\theta_1), \ldots, f(\theta_m), f(\theta_{p-k+m+1}), \ldots, f(\theta_p)] \\
& = \mathbb{E}_m[f(\theta[1]), \ldots, f(\theta[k])].
\end{align*}
\]

(2.15)

All that remains to be shown is that this upper bound can be attained. Let \( S \) be factored according to \( S = D[\theta] \Gamma \), where \( \Gamma \) is a \( p \times p \) orthogonal matrix and \( D[\theta] \) is a \( p \times p \) diagonal matrix whose diagonal elements are \( \theta[j] \). Then the upper bound is attained for \( A = (I_k, 0) \Gamma \).

The possibility that \( x_0 \) may be equal to \( 0 \) or \( R \) admits monotone functions. From the discussion leading to Theorem 2.5, we can generalize Theorem 2.7 into a theorem concerning two matrices.

**Theorem 2.8.** Let \( f(x) \) be a function of a real variable \( x \) satisfying the conditions specified in Theorem 2.7. If \( S_1 \geq 0 \) and \( S_2 > 0 \), let \( \theta_1 \geq \theta_2 \geq \cdots \geq \theta_p \) be the characteristic roots of \( S_1 S_2^{-1} \). Let \( \theta[1], \theta[2], \ldots, \theta[p] \) denote the ordering of the \( \theta \)'s according to \( f(\theta) \).
That is, \( f(\hat{\theta}[1]) \geq f(\theta[2]) \geq \cdots \geq f(\theta[p]). \) Then for \( m = 1,2,\ldots,k \)

\[
(2.16) \max_{A_{kp}} \left\{ f(\lambda_1(AS_1A')(AS_2A')^{-1}), \ldots, f(\lambda_k(AS_1A')(AS_2A')^{-1}) \right\}
\]

\[= E_m(f(\theta[1]), \ldots, f(\theta[k])).\]

From the two preceding theorems, we can derive special results which will be needed in Chapter 3. As a temporary digression, we now review the definition of a matrix function.

Suppose \( S \) is an \( n \times n \) symmetric matrix which can be factored according to \( S = \Gamma D_\lambda \Gamma' \), where \( \Gamma \) is orthogonal and \( D_\lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \).

If \( g(x) \) is an analytic function such that \( g(\lambda_i) \) is defined for each \( i \), then \( g(S) \) is a symmetric matrix defined by

\[ g(S) = \Gamma D g(\lambda) \Gamma', \]

where \( D g(\lambda) = \text{diag}[g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)] \). For example, if \( g(x) = \exp(x) \) and \( S = \Gamma D_\lambda \Gamma' \), then \( \exp(S) = \Gamma D e^{\lambda} \Gamma' \). We note that if \( \lambda \) is a characteristic root of \( S \), then \( g(\lambda) \) is a characteristic root of \( g(S). \) The characteristic vectors of \( S \) and \( g(S) \) are identical.

**Corollary 2.9.** If \( S > 0 \), then for \( m = 1,2,\ldots,k, \)

\[
(2.17) \max_{A_{kp}} \left\{ \text{tr}_m[(AS_1A')^{-1}\exp(AS_1A')] \right\} = E_m(\theta^{-1}[1], \theta^{-1}[2], \ldots, \theta^{-1}[k]),
\]

where \( \theta[j] \) are the characteristic roots of \( S \) arranged according to decreasing values of \( f(\theta) = \theta^{-1}\exp(\theta). \)
Proof. If the characteristic roots of $ASA'$ are denoted by $\lambda = \lambda(ASA')$, then the characteristic roots of $(ASA')^{-1}\exp(ASA')$ are $\lambda^{-1}\exp(\lambda)$. Thus,

$$\text{tr}_m[(ASA')^{-1}\exp(ASA')] = E_m(\lambda_1^{-1}e_1, \ldots, \lambda_k^{-1}e_k).$$

Let $f(x) = x^{-1}e^x$. Since $f'(x) = x^{-2}(x-1)e^x$, the conditions of Theorem 2.7 on $f(x)$ are satisfied with $R = \infty$ and $x_0 = 1$. 

**COROLLARY 2.10.** If $S_1 \geq 0$ and $S_2 > 0$, then for $m = 1, 2, \ldots, k$,

$$\max_{\mathcal{A}_{kp}} \text{tr}_m\left[(AS_1A')^{-m_1}(A(S_1+S_2)A')^{n_1+n_2}(AS_2A')^{-n_2}\right]$$

$$= E_m(\theta_{[1]}(1-\theta_{[1]})^{-n_2}, \ldots, \theta_{[k]}(1-\theta_{[k]})^{-n_2}),$$

where $\theta_{[j]}$ are the characteristic roots of $S_1(S_1+S_2)^{-1}$ arranged according to decreasing values of $f(\theta) = \theta^{-n_1}(1-\theta)^{-n_2}$.

**Proof.** For any $A \in \mathcal{A}_{kp}$, we first apply a simultaneous factorization to $AS_1A'$ and $A(S_1+S_2)A'$.

$$A(S_1+S_2)A' = WW'$$

(2.18)

$$AS_1A' = WD_\lambda W'$$

$$AS_2A' = W(I_k-D_\lambda)W' .$$

Here, $W$ is a $k \times k$ nonsingular matrix and $D_\lambda$ is a diagonal matrix whose diagonal elements are the $k$ characteristic roots of $A(S_1+S_2)A'$. By direct substitution,
(2.19) \[(AS_1A')^{-n_1}(A(S_1+S_2)A')^{-n_2} = \]
\[(WD_\lambda W')^{-n_1}(W(W(I_k-D_\lambda)W')^{-n_2} \cdot \]

We next make use of the fact that the characteristic roots of the product of square matrices are invariant under commutation; that is, \(\lambda(AB) = \lambda(BA)\). Applying this fact repeatedly, we find that the characteristic roots of \((WD_\lambda W')^{-n_1}(W(W(I_k-D_\lambda)W')^{-n_2}\) are the same as the characteristic roots of \(D_\lambda^{-n_1}(I_k-D_\lambda)^{-n_2}\). These characteristic roots are equal to \(\lambda^{-n_1(1-\lambda)^{-n_2}}\). Therefore,

(2.20) \[
\text{tr}_m[(AS_1A')^{-n_1}(A(S_1+S_2)A')^{-n_2} = \text{E}_m[\lambda_1^{-n_1(1-\lambda)^{-n_2}}, \ldots, \lambda_k^{-n_k(1-\lambda_k)^{-n_2}}],
\]

where \(\lambda_j\) are the characteristic roots of \((AS_1A')(A(S_1+S_2)A')^{-1}\).

Let \(f(x) = x^{-n_1(1-x)^{-n_2}}\). Since \(f'(x) = x^{-(n_1+1)}(1-x)^{-(n_2+1)}[(n_1+n_2)x-n_1]\), the conditions of Theorem 2.8 on \(f(x)\) are satisfied with \(R = 1\) and \(x_0 = n_1/(n_1+n_2)\).

The following theorem is required in the construction of a test for the independence between two sets of variates.

**Theorem 2.11.** Let \(D_r\) be a \(p \times p\) diagonal matrix whose diagonal elements are \(r_j\) with \(1 > r_1 > r_2 \geq \cdots \geq r_p > 0\). Let \(G = (D_r, 0)\) then be a \(p \times q\) matrix with \(p \leq q\). If \(A = a_{k_1 p} \times a_{k_2 q} = (A_1, A_2; A_1: k_1 \times p), A_2: k_2 \times q, A_1A_1' = I_{k_1}, A_2A_2' = I_{k_2}\),
where \( k_1 \leq p, k_2 \leq q, \) and \( k = \min(k_1, k_2), \) then for \( m = 1, 2, \ldots, k \)

\[
\begin{align*}
\max \mathcal{A} & \left( I_{k_1} - (A_1^* A_2^2)(A_1^* A_2^2)' \right)^{-1} \\
& = E_m \left[ (1-r_1^2)^{-1}, (1-r_2^2)^{-1}, \ldots, (1-r_k^2)^{-1}, 1, \ldots, 1 \right]
\end{align*}
\]

where there are \( k_1 - k \) "ones". \( (k_1 - k) \) may be zero.

\textbf{Proof.} If the characteristic roots of \( (A_1^* A_2^2)(A_1^* A_2^2)' \) are denoted by \( \lambda_j \), then

\[
\begin{align*}
\text{tr}_m \left[ (I_{k_1} - (A_1^* A_2^2)(A_1^* A_2^2)')^{-1} \right] & = \\
E_m \left[ (1-\lambda_1)^{-1}, (1-\lambda_2)^{-1}, \ldots, (1-\lambda_{k_1})^{-1} \right].
\end{align*}
\]

Since the nonzero characteristic roots of a product are invariant under commutation of matrices, for \( j = 1, 2, \ldots, k, \)

\[
\lambda_j (A_1^* A_2^2)(A_1^* A_2^2)' = \lambda_j (GA_1^2 A_2 GA_1^1 A_1^1).
\]

The matrix \( A_2^1 A_2^2 \) is a symmetric indempotent matrix of rank \( k_2 \), and hence can be factored:

\[
(2.24) \quad A_2^1 A_2^2 = L \begin{pmatrix} I_{k_2} & 0 \\ 0 & 0 \end{pmatrix} L'.
\]

where \( L \) is a \( q \times q \) orthogonal matrix. Next partition \( L \) according to
\[(2.25) \quad L = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}, \]

where \(L_{11}\) is \(p \times k_2\) and \(L_{22}\) is \((q-p) \times (q-k_2)\). By substituting (2.24) and (2.25) in (2.23), we obtain

\[
\lambda_j (GA_2 A_2 G^t A_1) = \lambda_j [CL \begin{pmatrix} I_{k_2} & 0 \\ 0 & 0 \end{pmatrix} L^t G A^t A_1] \\
= \lambda_j (D_r L_{11} L_{11}^t D_r A_1) = \lambda_j (A_1 D_r L_{11} L_{11}^t D_r A_1). \tag{2.26}
\]

By Theorem 2.1,

\[
\lambda_j (A_1 D_r L_{11} L_{11}^t D_r A_1) \leq \lambda_j (D_r L_{11} L_{11}^t D_r). \tag{2.27}
\]

Since \(L_{11} L_{11}^t = I_{p_1} - L_{12} L_{12}^t\), we have \(L_{11} L_{11}^t \leq I_{p_1}\), from which

\(D_r L_{11} L_{11}^t D_r \leq D_r^2\). But if \(A \preceq B\), then \(\lambda_j (A) \leq \lambda_j (B)\). Thus,

\[
\lambda_j (D_r L_{11} L_{11}^t D_r) \leq \lambda_j (D_r^2) = r_j^2. \tag{2.28}
\]

We therefore have for \(j = 1, \ldots, k\), that

\[
\lambda_j = \lambda_j [(A_1 G A_2^t) (A_1 G A_2^t)^t] \leq r_j^2. \tag{2.29}
\]

Since \(\lambda_j = 0\) for \(k+1 \leq j \leq k_1\) and \(E_m [(1-\lambda_1)^{-1}, (1-\lambda_2)^{-1}, \ldots, (1-\lambda_{k_1})^{-1}]\)

is an increasing function of each \(\lambda_j\), it follows that
\[(2.30) \quad E_m[(1-\lambda_1)^{-1}, (1-\lambda_2)^{-1}, \ldots, (1-\lambda_k)^{-1}] \leq E_m[(1-r_1^2)^{-1}, (1-r_2^2)^{-1}, \ldots, (1-r_k^2)^{-1}, 1, \ldots, 1].\]

The upper bound in (2.29) is attained for

\[(2.31) \quad L_{11} = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}, \quad A_1 = (I_k, 0)\]

as can be readily checked by direct substitution in (2.27) and (2.28).

The following series of lemmas lead to Theorem 2.15, which is applied in the construction of tests for sphericity.

**Lemma 2.12.** Suppose $\Delta_1$ and $\Delta_2$ are two $n \times n$ diagonal matrices, each of whose diagonal elements are positive and arranged in nonincreasing order. If $\Gamma_1$ and $\Gamma_2$ are any two orthogonal matrices, then for $m = 1, 2, \ldots, n,$

\[(2.32) \quad \text{tr}_m(\Gamma_1 \Delta_1 \Gamma_2 \Delta_2) \leq \text{tr}_m(\Delta_1 \Delta_2).\]

**Proof.** By Theorem 2.4(iv),

\[(2.33) \quad \text{tr}_m(\Gamma_1 \Delta_1 \Gamma_2 \Delta_2) = \text{tr}(\Gamma_1 \Delta_1 \Gamma_2 \Delta_2)^{(m)}.\]

By Theorem 2.4(i),

\[(2.34) \quad \text{tr}(\Gamma_1 \Delta_1 \Gamma_2 \Delta_2)^{(m)} = \text{tr}(\Gamma_1^{(m)} \Delta_1 \Gamma_2^{(m)} \Delta_2^{(m)}).\]

Since $\Gamma_1^{(m)}$ and $\Gamma_2^{(m)}$ are each orthogonal by Theorem 2.4(iii), we can
apply a result due to von Neumann (1937) which states that if $\psi_1$ and $\psi_2$ are orthogonal and $\Delta_1$ and $\Delta_2$ are as assumed in this lemma, then

\begin{equation}
(2.35) \quad \text{tr}(\psi_1 \Delta_1 \psi_2 \Delta_2) \leq \text{tr}(\Delta_1 \Delta_2).
\end{equation}

Applying (2.35) to (2.34), we obtain,

\begin{equation}
(2.36) \quad \text{tr}(\Gamma_1^{(m)} \Delta_1^{(m)} \Gamma_2^{(m)} \Delta_2^{(m)}) \leq \text{tr}(\Delta_1^{(m)} \Delta_2^{(m)}) = \text{tr}_m(\Delta_1 \Delta_2). \quad \|\end{equation}

Note that if $\Gamma$ is an orthogonal matrix, then

\begin{equation}
(2.37) \quad \text{tr}_m(\Gamma \Delta_1 \Gamma^t \Delta_2) \leq \max_{\Gamma_1, \Gamma_2} \text{tr}_m(\Gamma_1 \Delta_1 \Gamma_2 \Delta_2) \leq \text{tr}_m(\Delta_1 \Delta_2).
\end{equation}

Therefore, Lemma 2.12 holds even if we restrict $\Gamma_2 = \Gamma_1^t$. This is stated in Corollary 2.13.

**Corollary 2.13.** If $\Delta_1$ and $\Delta_2$ are two $n \times n$ diagonal matrices whose diagonal elements are positive and arranged in nonincreasing order and $\Gamma$ is any orthogonal matrix, then for $m = 1, 2, \ldots, n$,

\[ \text{tr}_m(\Gamma \Delta_1 \Gamma^t \Delta_2) \leq \text{tr}_m(\Delta_1 \Delta_2). \]

Equality is achieved for $\Gamma = I_n$.

The following lemma relates the extremal solution of $\text{tr}_m(W)$ to $\text{tr}_m(W + W^{-1})$.

**Lemma 2.14.** Let $Q_n$ be a closed and bounded set in $R_n$ in which each
\[ z_n = (x_1, \ldots, x_n) \in Q_n \] has the property that \( x_1 \geq \cdots \geq x_n \geq 1 \). Suppose that if \( z^*_{n-1} = (x^*_1, \ldots, x^*_{n-1}) \) maximizes

\[ f_{n-1}(z_n) = E_m(x_1, \ldots, x_{n-1}) \]

for \( z_{n-1} \in Q_{n-1} \), then \( z^*_{n-1} \) also maximizes \( E_{m-1}(x_1, \ldots, x_n) \) and

\[ z^*_{n} = (z^*_{n-1}, x^*_n) \] maximizes \( f_n(z_n) = E_m(x_1, \ldots, x_{n-1}, x_n) \) for some \( x_n = x^*_n \).

Then \( z^*_{n} \) also maximizes

\[ g_n(z_n) = E_m(x_1 + x^{-1}_1, \ldots, x_n + x^{-1}_n) \]

**Proof.** The proof is by induction on \( n \). For \( n = 1 \), the lemma is trivially true.

Assume that \( z^*_j = (x^*_1, \ldots, x^*_j) \) maximizes both \( f_j(z_j) \) and \( g_j(z_j) \) for \( z_j \in Q_j \) and \( j = 1, 2, \ldots, n \). Now \( E_m(x_1, \ldots, x_n, x_{n+1}) \) can be expressed as

\[
E_m(x_1, \ldots, x_n, x_{n+1}) = x_{n+1} E_{m-1}(x_1, \ldots, x_n) + \\
E_m(x_1, \ldots, x_n)
\]

Similarly,

\[
E_m(x_1 + x^{-1}_1, \ldots, x_n + x^{-1}_n, x_{n+1} + x^{-1}_{n+1}) = (x_{n+1} + x^{-1}_{n+1})^* \]

\[
E_{m-1}(x_1 + x^{-1}_1, \ldots, x_n + x^{-1}_n) + E_m(x_1 + x^{-1}_1, \ldots, x_n + x^{-1}_n).
\]

By the assumptions stated in the lemma and the induction hypothesis, \( z^*_n \)
maximizes not only \( E_m(x_1, \ldots, x_n) \), but \( E_{m-1}(x_1, \ldots, x_n), E_{m-1}(x_1 x_n^{-1}, \ldots, x_n x_n^{-1}) \), and \( E_m(x_1 x_n^{-1}, \ldots, x_n x_n^{-1}) \) as well. The proof is completed by noting that since \( x_n+1 \geq 1 \), the maximum of \( x_n+1 \) also maximizes \( x_n+1 x_n^{-1} \).

**Theorem 2.15.** Let \( S: p \times p \) be positive definite with distinct characteristic roots and let

\[
\mathcal{A} = \{ A, B; A: k \times p, B: k \times p, AA' = BB' = I_k, AB' = 0, k \leq p/2 \}.
\]

If \((ASA')^{1/2}\) and \((BSB')^{1/2}\) are the symmetric square root matrices of \(ASA'\) and \(BSB'\), respectively, and if

\[
\tau_j = \sqrt{\frac{\lambda_j(S)}{\lambda_{p-j+1}(S)}} + \sqrt{\frac{\lambda_{p-j+1}(S)}{\lambda_j(S)}}
\]

then

\[
\max_{\mathcal{A}} \max_{m} \text{tr} [(ASA')^{1/2}(BSB')^{-1/2} + (BSB')^{1/2}(ASA')^{-1/2}]
\]

\[
= E_m[\tau_1, \tau_2, \ldots, \tau_k].
\]

**Proof.** In the course of this proof, we use a few well-known facts about ordered matrices which are stated here. If \( 0 \leq X \leq Y \), then

\[
(i) \quad X^{-1} \geq Y^{-1}
\]

\[
(ii) \quad \lambda_j(X) \leq \lambda_j(Y)
\]

\[
(iii) \quad TXT' \leq TYT'
\]

for any matrix \( T \).
Let \( \theta_j = \lambda_j(S) \) by the \( j \)-th largest characteristic root of \( S \) and let \( D_1 = \text{diag}[\theta_1^{1/2}, \theta_2^{1/2}, \ldots, \theta_k^{1/2}] \) and \( D_2 = [\theta_p^{1/2}, \theta_{p-1}^{1/2}, \ldots, \theta_{p-k+1}^{1/2}] \).

By Theorem 2.1,

\[
\lambda_j(\text{ASA}')^{1/2} = \lambda_j^{1/2}(\text{ASA}') \leq \lambda_j(D_1)
\]

for \( j = 1, 2, \ldots, k \). Let \( \Gamma_1 \) be a \( k \times k \) orthogonal matrix such that \( \Gamma_1(\text{ASA}')^{1/2}\Gamma_1' \) is a diagonal matrix with decreasing diagonal elements.

The diagonal terms are, of course, the square roots of the characteristic roots of \( \text{ASA}' \). Then \( \Gamma_1(\text{ASA}')^{1/2}\Gamma_1' \leq D_1 \), or, equivalently, \( (\text{ASA}')^{1/2} \leq \Gamma_1' D_1 \Gamma_1 \).

Similarly, for any \( B \) which is row-orthogonal, there is an orthogonal matrix \( \Gamma_2 \) such that \( \Gamma_2(\text{BSB}')^{1/2}\Gamma_2' \geq D_2 \), or, equivalently, \( (\text{BSB}')^{-1/2} \leq \Gamma_2' D_2^{-1} \Gamma_2 \). Now because of the invariance of characteristic roots under commutation of matrices, for \( j = 1, 2, \ldots, k \),

\[
(2.42) \quad \lambda_j(\text{ASA}')^{1/2}(\text{BSB}')^{-1/2} = \lambda_j(\text{BSB}')^{-1/4}(\text{ASA}')^{1/2}(\text{BSB}')^{-1/4}.
\]

Since by \( (2.41(iii)) \)

\[
(\text{BSB}')^{-1/4}(\text{ASA}')^{1/2}(\text{BSB}')^{-1/4} \leq (\text{BSB}')^{-1/4}(\Gamma_1' D_1 \Gamma_1')(\text{BSB}')^{-1/4},
\]

it follows that

\[
(2.43) \quad \lambda_j[(\text{BSB}')^{-1/4}(\text{ASA}')^{1/2}(\text{BSB}')^{-1/4}] \leq \lambda_j[(\text{BSB}')^{-1/4}(\Gamma_1' D_1 \Gamma_1')(\text{BSB}')^{-1/4}].
\]
Furthermore, by matrix commutation, \((2.41) (ii), (iii)\),

\[
(2.44) \quad \lambda_j[(B, B')^{-1/4} (\Gamma_1 \Gamma_1', \Gamma_1') (B, B')^{-1/4}] = \lambda_j[(\Gamma_1^{1/2} \Gamma_1') (B, B')^{-1/2} (\Gamma_1^{1/2} \Gamma_1')] \\
\leq \lambda_j[(\Gamma_1^{1/2} \Gamma_1') (\Gamma_2^{-1} \Gamma_2')(\Gamma_1^{1/2} \Gamma_1')] .
\]

Letting \( \Gamma = \Gamma_1' \Gamma_2' \), \((2.44)\) becomes

\[
(2.45) \quad \lambda_j(D_1^{1/2} \Gamma_2^{-1} \Gamma_1^{1/2}) = \lambda_j(D_1 \Gamma_2^{-1} \Gamma_1') = \lambda_j(\Gamma' D_1 \Gamma_2^{-1}) .
\]

Note that \( \Gamma \) is a \( k \times k \) orthogonal matrix. Thus, by \((2.42) -(2.45)\)
there exists an orthogonal matrix \( \Gamma \) depending on \( A \) and \( B \) such that

\[
\lambda_j[(A, A')^{1/2} (B, B')^{-1/2}] \leq \lambda_j(\Gamma' D_1 \Gamma_2^{-1}) .
\]

Therefore, since \( 1 \leq \lambda_j(X) \leq \lambda_j(Y) \) implies \( \text{tr}_m(XX^{-1}) \leq \text{tr}_m(YY^{-1}) \),
we have

\[
(2.46) \quad \text{tr}_m[(A, A')^{1/2} (B, B')^{-1/2} + (B, B')^{1/2} (A, A')^{-1/2}] \leq \\
\quad \quad \text{tr}_m[\Gamma' D_1 \Gamma_2^{-1} + D_2 \Gamma' D_1^{-1} \Gamma] .
\]

We now seek to maximize the right-hand side of \((2.46)\). Let
\( x_1 \geq \cdots \geq x_k \) be the ordered characteristic roots of \( \Gamma' D_1 \Gamma_2^{-1} \). Then

\[
(2.47) \quad \text{tr}_m[\Gamma' D_1 \Gamma_2^{-1} + D_2 \Gamma' D_1^{-1} \Gamma] = E_m(x_1 + x_1^{-1}, \ldots, x_k + x_k^{-1}) = g_k(z_k) ,
\]

28
where \( g_k(z_k) \) is defined in Lemma 2.13. Applying Lemma 2.14 to (2.47), if \( z_k^* = (x_1^*, \ldots, x_k^*) \) maximizes

\[
(2.48) \quad f_k(z_k) = E_m(x_1, \ldots, x_k) = \text{tr}_m(G D_1 D_2^{-1}),
\]

then \( z_k^* \) maximizes \( g_k(z_k) \). But the maximum of (2.48) is given by Corollary 2.13, by setting \( \Delta_1 = D_1 \) and \( \Delta_2 = D_2^{-1} \), as \( \text{tr}_m(D_1 D_2^{-1}) \).

Since \( D_1 = \text{diag}[\theta_1^{1/2}, \theta_2^{1/2}, \ldots, \theta_p^{1/2}] \) we have \( D_1 D_2^{-1} = \text{diag}[(\theta_1/\theta_p)^{1/2}, (\theta_2/\theta_{p-1})^{1/2}, \ldots, (\theta_k/\theta_{p-k+1})^{1/2}] \). Therefore,

\[
(2.49) \quad \text{tr}_m(D_1 D_2^{-1}) = E_m[(\theta_1/\theta_p)^{1/2}, (\theta_2/\theta_{p-1})^{1/2}, \ldots, (\theta_k/\theta_{p-k+1})^{1/2}].
\]

The maximum of (2.47) is thus given by

\[
(2.50) \quad \max_{\Gamma} \text{tr}_m(G D_1 D_2^{-1} + D_2 \Gamma D_1^{-1}) = \text{tr}_m(D_1 D_2^{-1} + D_2 D_1^{-1})
\]

\[
= E_m(\tau_1, \tau_2, \ldots, \tau_k)
\]

where \( \tau_j \) is defined by (2.39).
3. **UNION-INTERSECTION TESTS FOR SOME STANDARD MULTIVARIATE PROBLEMS.**

In this chapter, we obtain some union-intersection tests for a number of standard problems in multivariate statistics. In the usual application of the union-intersection principle, the unions and intersections have been performed over univariate tests. This has led in most cases to a test based on the extreme sample characteristic roots, or as in the case of the step-down procedure, on a finite number of independent statistics.

The idea here is to perform the unions and intersections over multivariate tests, of lower dimensionality \( k \), which are indexed by a \( k \times p \) matrix \( A \). A matrix \( W_A \) is said to be a likelihood ratio component matrix if the likelihood ratio test for the component hypothesis is equivalent to rejection if \( |W_A| > c \). In general, there are many matrices satisfying this property.

The test criterion for each component hypothesis is to reject the component hypothesis \( H_A \) if \( \text{tr}(W_A) > c \). The null hypothesis \( H \) is rejected if at least one \( H_A \) is rejected. Thus, for each \( k \), the union-intersection test is based on the maximum of \( \text{tr}(W_A) \) with respect to the index matrix \( A \). For \( m < k \), the resulting union-intersection test generally depends on the rule for selecting \( W_A \).

By denoting this test statistic as \( T_{mk} \), where \( 1 \leq m \leq k \leq p \), it can be seen that several standard multivariate tests are included in this class. For example, \( T_{pp} \) coincides with the likelihood ratio test based on determinants, \( T_{1p} \) is a Hotelling-Lawley "trace" criterion, and \( T_{11} \) is the union-intersection test of Roy based on the extreme sample characteristic roots.
This research is not directly concerned with distribution problems. The constant \( c \) which appears in the rejection regions is a generic constant and may differ for different statistics.

3.1 TESTING THAT THE MEAN VECTOR IS ZERO.

Suppose that \( x_1, x_2, \ldots, x_N \) is a sample of size \( N \) from a \( p \)-variate normal distribution, \( N_p(\mu, \Sigma) \), and we wish to test the null hypothesis \( H \) against the alternative hypothesis \( K \), where \( H \) and \( K \) are

\[
\begin{align*}
H &: \mu = 0, \\
K &: \mu \neq 0.
\end{align*}
\]

Let \( \mathcal{G}_{kp} \) be the set of all \( k \times p \) matrices \( A \) of rank \( k \). The null component hypotheses are taken as

\[
H_A: A\mu = 0, \quad K_A: A\mu \neq 0
\]

for \( A \in \mathcal{G}_{kp} \). Define

\[
\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha, \quad S = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})'(N-1).
\]

A likelihood ratio component test matrix [see Anderson (1958, p. 103)] for \( H_A \) against \( K_A \) is

\[
[A((N-1)S+N\bar{x}\bar{x}')A'][(N-1)(ASA')]^{-1} = I_k + \frac{N}{N-1} (A\bar{x}\bar{x}'A')(ASA')^{-1}
\]

The component hypothesis \( H_A \) is rejected against \( K_A \) if

\[
\text{tr}_m[I_k + \frac{N}{N-1} (A\bar{x}\bar{x}'A')(ASA')^{-1}] > c,
\]

31
where \( c \) is a generic constant to be determined later. The characteristic roots of
\[
\frac{N}{N-1} (A - x' A')(ASA')^{-1}
\]
are equal to unity plus the characteristic roots of
\[
\frac{N}{N-1} (A - x' A')(ASA')^{-1}
\]
Since the rank of the latter matrix is one, it has \( k-1 \) characteristic roots equal to zero and one non-zero root,
\[
T_A^2/(N-1),
\]
where
\[
T_A^2 = N x' A'(ASA')^{-1} A x .
\]
Thus, the component rejection criterion given by (3.4) is equivalent to
\[
E_m (1 + T_A^2/(N-1), 1, \ldots, 1) =
\]
\[
\binom{k-1}{m} + \binom{k-1}{m-1} (1 + T_A^2/(N-1)) > c ,
\]
where \( \binom{a}{b} \) is the standard binomial.

The null hypothesis \( H \) is rejected against \( K \) if and only if at least one \( H_A \) is rejected against \( K_A \). This is equivalent to the maximum of \( E_m \) with respect \( A \in \Theta_{kp} \) being greater than \( c \), which in turn is equivalent to \( \max_{\Theta_{kp}} T_A^2 > c \). Now
\[
T_A^2 = N x' A'(ASA')^{-1} A x = \lambda_1 (N(A x' A')(ASA')^{-1})
\]
because of the invariance of the nonzero characteristic roots under commutation of matrix products and the fact that the rank of a scalar is one. Let \( S_1 = N x' x' \) and \( S_2 = S \). Then from Theorem 2.6(iv) and the subsequent comments,
\[
\max_{\Theta_{kp}} \lambda_1 (N(A x' A')(ASA')^{-1}) = \max_{\Theta_{kp}} \lambda_1 ((A S_A')(A S_A')^{-1})
\]
\[
= \lambda_1 (S_1 S_2^{-1}) = \lambda_1 (N x' x' S^{-1}) = N x' S^{-1} x .
\]

32
Thus, the rejection region for testing $H$ against $K$ is the same for all $m$ and $k$. They are each equivalent to the rejection region,

$$T^2 = N x' S^{-1} x > c.$$  \hspace{1cm} (3.8)

This test, which is known as Hotelling's $T^2$-test (1931) has been shown by Stein (1956) to be admissible with a loss function of 1 for an incorrect decision and 0 for a correct decision. That is, its power function is not dominated by the power function of any other test. It also has been shown to be UMP invariant under the group of all non-singular linear transformations by Hunt and Stein (1946).

### 3.2 Testing Equality of Several Mean Vectors.

This problem is a particular case of the general linear model. Suppose $x_{j1}', x_{j2}', \ldots, x_{jN_j}'$ is a sample of size $N_j$ from a $p$-variate normal distribution, $N_p[\mu_j, \Sigma]$, $j = 1, 2, \ldots, r$. We wish to test the null hypothesis $H$ against the alternative hypothesis $K$: not $H$, where $H$ is

$$H: \mu_1 = \mu_2 = \cdots = \mu_r.$$  \hspace{1cm} (3.9)

Let $\mathcal{B}_{kp}$ be the set of all $k \times p$ matrices $A$ of rank $k$. The null component hypotheses and the alternative component hypotheses are taken as

$$H_A = A \mu_1 = A \mu_2 = \cdots = A \mu_r,$$

$$K_A: \text{not } H_A,$$

for $A \in \mathcal{B}_{kp}$. Define the statistics:
\[
\bar{x}_j = \frac{\sum_{\alpha=1}^{N_j} x_{j\alpha}}{N_j}, \quad \bar{x} = \frac{\sum_{j=1}^{r} \sum_{\alpha=1}^{N_j} x_{j\alpha}}{\sum_{j=1}^{r} N_j},
\]

\[
S_H = \sum_{j=1}^{r} N_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})', \quad S_E = \sum_{j=1}^{r} \sum_{\alpha=1}^{N_j} (x_{j\alpha} - \bar{x}_j)(x_{j\alpha} - \bar{x}_j)'.
\]

A likelihood ratio component test matrix [see Anderson (1958, p. 217)] for \(H_A\) against \(K_A\) is

\[
A(S_E + S_H)A' \left((AS_E A')^{-1}\right) = I_k + (AS_H A')(AS_E A')^{-1}.
\]

The component hypothesis \(H_A\) is rejected against \(K_A\) if

\[
tr_m[I_k + (AS_H A')(AS_E A')^{-1}] > c. \tag{3.13}
\]

Denote the characteristic roots of \((AS_H A')(AS_E A')^{-1}\) by \(\lambda_j, j = 1, 2, \ldots, k\).

Then (3.13) is equivalent to

\[
E_m[1+\lambda_1, 1+\lambda_2, \ldots, 1+\lambda_k] > c. \tag{3.14}
\]

The null hypothesis \(H\) is rejected if and only if \(E_m > c\) for some \(A \in \mathcal{B}_{kp}\). To find the maximum of \(E_m\), we use Theorem 2.8 with \(f(x) = 1+x\), \(x_0 = 0\) and \(R = \infty\). This maximum is

\[
\max_{\mathcal{B}_{kp}} E_m[1+\lambda_1, 1+\lambda_2, \ldots, 1+\lambda_k] = E_m[1+\lambda_1(S_H S_E^{-1}), 1+\lambda_2(S_H S_E^{-1}), \ldots, 1+\lambda_k(S_H S_E^{-1})]. \tag{3.15}
\]

The resulting test for \(H\) against \(K\) is to reject \(H\) if
\begin{equation}
\tag{3.16}
T_{mk} = E_m \left[ 1 + \lambda_1(S_HS_E^{-1}), 1 + \lambda_2(S_HS_E^{-1}), \ldots, 1 + \lambda_k(S_HS_E^{-1}) \right] > c.
\end{equation}

It is shown in Section 5.1 that each of the tests based on $T_{mk}$ satisfy the monotonicity property; that is, the power function is an increasing function of each characteristic root of the matrix $\Sigma^{-1}\mu$, where $\mu = (\mu_1, \mu_2, \ldots, \mu_r)$.

For certain combinations of $m$ and $k$, we can obtain some standard multivariate tests. We note that

\begin{equation}
\tag{3.17}
T_{pp} = \prod_{j=1}^{P} (1 + \lambda_j(S_HS_E^{-1})) = |I_p + S_HS_E^{-1}|
= \frac{|S_E + S_H|}{|S_E|}
\end{equation}

is equivalent to the likelihood ratio statistic.

\begin{equation}
\tag{3.18}
T_{lp} = \sum_{j=1}^{P} (1 + \lambda_j(S_HS_E^{-1})) = p + \sum_{j=1}^{P} \lambda_j(S_HS_E^{-1}) = p + \text{tr}(S_HS_E^{-1})
\end{equation}

is equivalent to the Hotelling-Lawley "trace" statistic.

\begin{equation}
\tag{3.19}
T_{ll} = 1 + \lambda_1(S_HS_E^{-1})
\end{equation}

is equivalent to the Roy "maximum root" statistic.

3.3 TESTING THAT THE COVARIANCE MATRIX IS THE IDENTITY MATRIX.

Suppose that $nS > 0$ has a Wishart distribution, $W(\Sigma, p, n)$ and we wish to test the null hypothesis $H$ against the alternative hypothesis $K$, where
(3.20) \[ H: \Sigma = I_p , \quad K: \Sigma \neq I_p . \]

Let

\[ a_{kp} = \{ A: k \times p, AA' = I_k \} . \]

The null component hypotheses \( H_A \) and the alternative component hypotheses \( K_A \) are

(3.21) \[ H_A: A A^\prime = I_k , \quad K_A: A A^\prime \neq I_k , \]

for \( A \in a_{kp} \).

A likelihood ratio component test matrix [see Anderson (1958, p. 265)] is \((ASA')^{-1} \exp (ASA')\). The component test for \( H_A \) against \( K_A \) is to reject \( H_A \) if

(3.22) \[ \text{tr}_m [(ASA')^{-1} \exp (ASA')] > c . \]

The corresponding union-intersection test is to reject \( H \) against \( K \) if

(3.23) \[ \max_{a_{kp}} \text{tr}_m [(ASA')^{-1} \exp (ASA')] > c . \]

By Corollary 2.9, (3.23) is equivalent to

(3.24) \[ T_{mk} = E_m (\lambda_{[1]}^{-1} e^{\lambda_{[1]}}, \lambda_{[2]}^{-1} e^{\lambda_{[2]}}, \ldots, \lambda_{[k]}^{-1} e^{\lambda_{[k]}}) > c , \]

where \( \lambda_{[j]} \) are the characteristic roots of \( S \) arranged according to decreasing values of \( f(\lambda) = \lambda^{-1} \exp(\lambda) \).
(3.25) \[ T_{pp} = \prod_{j=1}^{p} \lambda_j^{-1} e_j = e_j^{j=1}^{p} \frac{\sum_{j=1}^{p} \lambda_j}{j=1} = \frac{\text{etr}(S)}{|S|} \]

is equivalent to the likelihood ratio statistic. A Hotelling-Lawley type statistic is given by

(3.26) \[ T_{lp} = \sum_{j=1}^{p} \lambda_j^{-1} e_j = \text{tr}(S^{-1}\exp(S)) \]

For \( m = k = 1 \),

(3.27) \[ T_{ll} = \lambda_1^{-1} e_1^{[1]} \]

which is identical to a test based on \( \lambda_1 \) and/or \( \lambda_p \). This test differs slightly from Roy's test, which is based on "equal tails" for \( \lambda_1 \) and \( \lambda_p \).

3.4 TESTING EQUALITY OF TWO COVARIANCE MATRICES.

Suppose that \( S_1 \) and \( S_2 \) have Wishart distributions \( W(\Sigma_1, p_n_1) \) and \( W(\Sigma_2, p_n_2) \), respectively. We wish to test \( H \) against \( K \), where

(3.28) \[ H: \Sigma_1 = \Sigma_2 \quad K: \Sigma_1 \neq \Sigma_2 \]

For \( A \in A_{kp} \), we define the null component hypotheses \( H_A \) and the alternative component hypotheses \( K_A \) by

(3.29) \[ H_A: \Sigma_1 A' = \Sigma_2 A' \quad K_A: \Sigma_1 A' \neq \Sigma_2 A' \]
A test matrix for the likelihood ratio component test [see Anderson (1958, p. 256)] is

\[(3.30)\quad (A S_1 A')^{-1} (A (S_1 + S_2) A')^{-1} (A S_2 A')^{-2}.\]

The test for \( H_A \) against \( K_A \) is to reject \( H_A \) if

\[(3.31)\quad \text{tr}_m [(A S_1 A')^{-1} (A (S_1 + S_2) A')^{-1} (A S_2 A')^{-2}] > c.\]

The union-intersection test is to reject \( H \) against \( K \) if

\[(3.32)\quad \max_{kp} \text{tr}_m [(A S_1 A')^{-1} (A (S_1 + S_2) A')^{-1} (A S_2 A')^{-2}] > c.\]

By Corollary 2.10, (3.32) is equivalent to

\[(3.33)\quad T_{mk} = \sum_{i=1}^{n_1} \text{tr}_m \left[ \lambda_{[i]} (1 - \lambda_{[i]})^{-n_2} \right] > c,\]

where \( \lambda_{[j]} \) are characteristic roots of \( S_1 (S_1 + S_2)^{-1} \) arranged according to decreasing values of \( f(\theta) = \theta^{-1} (1 - \theta)^{-n_2}. \)

The likelihood ratio test statistic is equivalent to

\[(3.34)\quad T_{pp} = \prod_{j=1}^{p} \lambda_j^{-1} (1 - \lambda_j^{-1})^{-n_2} = \frac{\left| S_1 + S_2 \right|^{n_1 + n_2}}{\left| S_1 \right|^{n_1} \left| S_2 \right|^{n_2}};\]

a Hotelling-Lawley type statistic is given by

\[(3.35)\quad T_{lp} = \sum_{j=1}^{p} \lambda_j^{-1} (1 - \lambda_j^{-1})^{-n_2} = \text{tr}((I_p + S_2 S_1^{-1})^{-1} (I_p + S_1 S_2^{-1})^{-n_2});\]
Roy's union-intersection statistic is equivalent to

\[(3.37) \quad T_{11} = \lambda_1^{1-(\lambda_1 - 1)} \lambda_2^{n_2},\]

which is identical to a test based on \(\lambda_1\) and/or \(\lambda_p\).

3.5 TESTING FOR INDEPENDENCE BETWEEN TWO SETS OF VARIATES.

Let \(S\) be a sample covariance matrix partitioned according to

\[
S = \begin{pmatrix}
S_{11} & S_{12} \\
S_{12}' & S_{22}
\end{pmatrix},
\]

where \(S_{11}\) is \(p \times p\) and \(S_{22}\) is \(q \times q\). Without loss of generality we assume that \(p \leq q\). Suppose that \(S\) has a Wishart distribution, \(W(\Sigma, p, n)\), where \(\Sigma\) is partitioned as

\[
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{12}' & \Sigma_{22}
\end{pmatrix}
\]

in conformity to the partition of \(S\). We wish to test the null hypothesis \(H\) against the alternative hypothesis \(K\), where

\[(3.38) \quad H: \Sigma_{12} = 0, \quad K: \Sigma_{12} \neq 0.\]

For \(p = 1\), this is equivalent to testing that the multiple correlation between the first variate and the remaining \(q\) variates is zero. Let
(3.39) \[ A = a_{k_1 p} \times a_{k_2 q} \]

\[ \{ A_1, A_2; A_1: k_1 \times p, A_2: k_2 \times q, A_1 A'_1 = I_{k_1}, A_2 A'_2 = I_{k_2} \} , \]

where \( k_1 \leq p \) and \( k_2 \leq q \). Let \( k = \min(k_1, k_2) \).

The component hypotheses are

(3.40) \[ H_{A_1 A_2} : A_1 \Sigma_{12} A'_2 = 0, \quad K_{A_1 A_2} : A_1 \Sigma_{12} A'_2 \neq 0 \]

for \((A_1, A_2) \in \Lambda\). The component rejection regions are

(3.41) \[ \text{tr}_m[(I_{k_1} - (A_1 G A_2')(A_1 G A_2'))^{-1}] > c \]

where \( G = (D_r, 0) \) is a \( p \times q \) matrix and \( D_r = \text{diag}(r_1, r_2, \ldots, r_p) \). The \( r_j \) are given by

(3.42) \[ r_j = \lambda_j^{1/2}(S_{11}^{-1}S_{12}^{-1}S_{22}^{-1}S_{22}^{t}) . \]

The rejection region for the null hypotheses \( H \) is given by the union of (3.41) over the set \( \Lambda \). Applying Theorem 2.10 to (3.41), this rejection region becomes

(3.43) \[ T_{mk_1 k_2} = E_m[(1-r_1^2)^{-1}, (1-r_2^2)^{-1}, \ldots, (1-r_k^2)^{-1}, 1, \ldots, 1] > c \]

where there are \( k_1-k \) ones in the argument of \( E_m \).

It is shown in Section 5.2 that each of the test statistics \( T_{mk_1 k_2} \) defines a test which satisfies the property of monotonicity. Special
cases of $T_{mk_1k_2}$ are of particular interest.

The likelihood ratio test statistic is equivalent to

$$(3.44) \quad T_{ppq} = \prod_{j=1}^{p} (1-r_j^2)^{-1} = |I_p - S_{11}^{-1} S_{12} S_{22}^{-1} S_{22}'|^{-1};$$

a Hotelling-Lawley type statistic is given by

$$(3.45) \quad T_{lpq} = \sum_{j=1}^{p} (1-r_j^2)^{-1} = \text{tr}(I_p - S_{11}^{-1} S_{12} S_{22}^{-1} S_{22}')^{-1};$$

Roy's union-intersection statistic is equivalent to

$$(3.46) \quad T_{lll} = (1-r_1^2)^{-1},$$

which is identical to a test based on $r_1^2$.

3.6 TESTING FOR SPHERICITY.

Suppose that $S > 0$ has a Wishart distribution, $W(\Sigma, p, n)$ and we wish to test $H$ against $K$, where

$$H: \Sigma = \sigma^2 I_p$$

for some unknown constant $\sigma^2$,

$$K: \Sigma \neq \sigma^2 I_p.$$

Let

$$\mathcal{B} = \{A, B; A: k \times p, B: k \times p, \text{rank}(A) = kp, \text{rank}(B) = k, AA' = BB', AB' = 0\}.$$
Since $A'B' = 0$, it follows that the common rank $k \leq p/2$. The component hypotheses are

$$H_{AB}: AA' = BB', \quad K_{AB}: AA' \neq BB'$$

for $(A,B) \in \mathcal{P}_{kp}$. A likelihood ratio component test matrix for (3.47) [see Anderson (1958, p. 256)] is

$$(3.48) \quad (ASA')^{-1/2}(ASA' + BSB')(BSB')^{-1/2}.$$  

Denote the characteristic roots of $(ASA')(BSB')^{-1}$ by $\lambda_j$, $j = 1, \ldots, k$. Then the characteristic roots of the matrix in (3.48) are $\lambda_j^{1/2} + \lambda_j^{-1/2}$. The component rejection regions are given by

$$tr_m[(ASA')^{-1/2}(ASA' + BSB')(BSB')^{-1/2}] > c,$$

which is equivalent to

$$E_m[\lambda_1^{1/2} + \lambda_1^{-1/2}, \lambda_2^{1/2} + \lambda_2^{-1/2}, \ldots, \lambda_k^{1/2} + \lambda_k^{-1/2}] > c.$$  

Note that in the intersection of (3.50) over $\mathcal{P}_{kp}$, we can restrict attention to the class $\mathcal{P}_{kp}^{(1)}$ for which $AA' = BB' = I_k$. To see this, suppose $(A,B) \in \mathcal{P}_{kp}$ and $AA' = BB' = G \neq I_k$. Let $G^{-1/2}$ be the symmetric square root matrix of $G^{-1}$. Let $A_1 = G^{-1/2}A$ and $B_1 = G^{-1/2}B$. Since $A_1A_1' = B_1B_1' = I_k$ and $A_1B_1' = 0$, it follows that $(A_1,B_1) \in \mathcal{P}_{kp}^{(1)}$.

Consider now the characteristic roots of $(ASA')(BSB')^{-1}$. We have $\lambda(ASA')(BSB')^{-1} = \lambda(A_1SA_1')(B_1SB_1')^{-1}$. Since the region described by (3.50) is based on the characteristic roots of $(ASA')(BSB')^{-1}$, we can therefore restrict the intersections to the class $\mathcal{P}_{kp}^{(1)}$.

42
By Theorem 2.15, the union-intersection test for $H$ against $K$ is to reject $H$ if

$$T_{mk} = E\left[\theta_1(s), \theta_2(s), \ldots, \theta_k(s)\right] > c,$$

where,

$$\theta_i(s) = \left[\frac{\lambda_i(s)}{\lambda_{p-i+1}(s)}\right]^{1/2} + \left[\frac{\lambda_{p-i+1}(s)}{\lambda_i(s)}\right]^{1/2}.$$

As far as can be determined, none of the tests described by (3.51) have previously appeared in the literature. An interesting special case is when $m = k = 1$. The test based on $T_{11}$ is equivalent to that based on $\lambda_1(s)/\lambda_p(s)$.

The likelihood ratio test, which is not included in (3.51), was obtained by Mauchly (1940) it has the rejection region,

$$\frac{\text{tr} S}{|S|} > c.$$

The moments of this likelihood ratio statistic were computed by Mauchly when $H$ is true. Wijsman (1959) showed that the likelihood ratio test is UMP invariant under the group of orthogonal transformations.
4. **GENERALIZED STEP-DOWN PROCEDURE.**

The step-down procedure was introduced by Roy and Bargmann (1958) and by J. Roy (1958) as a special type of union-intersection procedure for constructing multivariate tests. Under the step-down procedure, the null hypothesis $H$ to be tested is represented as a finite intersection of univariate component hypotheses. Each component hypothesis concerns the parameters of the conditional distribution of a single variate, given all the previous variates. When $H$ is true, it turns out that the component tests are independent, thus making it easy to determine the significance level of $H$. A disadvantage of the step-down procedure is that the resulting test is dependent upon the ordering of the variates.

Under a generalized step-down procedure which is proposed here, the $p$ variates are partitioned into $k$ blocks. The $i$-th block contains $p_i$ variates, with $\sum_{i=1}^{k} p_i = p$. The null hypothesis $H$ is represented as an intersection of $k$ multivariate component hypotheses. The $i$-th component hypothesis concerns the parameters of the conditional distribution of the $i$-th block of variates, given all the previous variates. The component tests are based on elementary symmetric functions of the characteristic roots corresponding to each component hypothesis. When $H$ is true, all these tests are independent.

This generalized step-down procedure is applied to the three problems considered by J. Roy: (a) the general linear model, (b) equality of a covariance matrix to the identity matrix, and (c) equality of two covariance matrices.
4.1 THE GENERAL LINEAR MODEL.

Let \( Y: n \times p \) be a matrix of observed random variables. It is assumed that the rows of \( Y \) are independently and normally distributed with a common covariance matrix \( \Sigma \); the expectation of \( Y \) is given by

\[
(4.1) \quad EY = A\theta ,
\]

where \( A \) is an \( n \times m \) "design matrix" of known constants, and \( \theta \) is an \( m \times p \) matrix of unknown parameters. It is assumed that \( A \) is of rank \( r \) (\( \leq n-p \)). Let \( \Phi = G\theta \) be a matrix of parametric functions, where \( \Phi \) is \( t \times p \) and \( G \) is a \( t \times m \) specified matrix of rank \( t \leq r \). It is assumed that \( \Phi \) is an estimable function in the sense that there exists an unbiased estimate of \( \Phi \), linear in \( Y \). The hypothesis to be tested is

\[
(4.2) \quad H: \Phi = 0 .
\]

Partition \( Y \) into

\[
(4.3) \quad Y = (Y_1, Y_2, \ldots, Y_k) ,
\]

where each \( Y_i \) is \( n \times p_i \), and define

\[
(4.4) \quad Y[i] = (Y_1, Y_2, \ldots, Y_i)
\]

for \( i = 1, \ldots, k \). Note that \( Y[k] = Y \). The dimension of \( Y[i] \) is \( n \times p[i] \), where \( p[i] = \sum_{j=1}^{i} p_j \). Next, partition the columns of \( \theta \) in conformity with the partition of \( Y \).

\[
(4.5) \quad \theta = (\theta_1, \theta_2, \ldots, \theta_k) ,
\]
(4.6) \( \theta_{[i]} = (\theta_1, \theta_2, \ldots, \theta_i), \ i = 1, \ldots, k \).

Each \( \theta_i \) is \( n \times p_i \) and the dimension of \( \theta_{[i]} \) is \( n \times p_{[i]} \). Now, partition the covariance matrix \( \Sigma \) in conformity with the partition of \( Y \) and \( \theta \): \( \Sigma = (\Sigma_{ij}); \ i, j = 1, 2, \ldots, k \), with each \( \Sigma_{ij} \) being a \( p_i \times p_j \) matrix of covariances between the \( i \)-th and \( j \)-th blocks of variates.

Denote the upper left-hand block of \( \Sigma \) by \( \Sigma_{[m]} = (\Sigma_{ij}); \ i, j = 1, 2, \ldots, m \).

Thus \( \Sigma_{[i]} \) is a \( p_{[i]} \) by \( p_{[i]} \) covariance matrix of the first \( p_{[i]} \) variates.

Now let \( \Sigma_{[i,i+1]} \) be a \( p_{[i]} \times p_{i+1} \) matrix defined by

(4.7) \[
\Sigma_{[i,i+1]} = \begin{bmatrix}
\Sigma_{1,i+1} \\
\Sigma_{2,i+1} \\
\vdots \\
\Sigma_{i,i+1}
\end{bmatrix}.
\]

Note that

\[
\Sigma_{[i+1]} = \begin{pmatrix}
\Sigma_{[i]} & \Sigma_{[i,i+1]} \\
\Sigma_{[i+1]} & \Sigma_{i+1,i+1}
\end{pmatrix}.
\]

Consider the conditional distribution of the \( n \times p_{i+1} \) matrix \( Y_{i+1} \) given the \( n \times p_{[i]} \) matrix \( Y_{[i]} \). The rows of \( Y_{i+1} \) are conditionally independent and normally distributed with covariance matrix

(4.8) \[
\Sigma_{i+1} \cdot [i] = \Sigma_{i+1,i+1} - \Sigma_{[i,i+1]}^t \Sigma_{[i]}^{-1} \Sigma_{[i,i+1]}.
\]

and expectations.

46
(4.9) \[ E(Y_{i+1}|Y_i) = A\xi_{i+1} + Y_i\beta_i, \]

where \( \xi_{i+1} \) and \( \beta_i \) are \( m \times p_{i+1} \) and \( p_i \times p_{i+1} \) matrices of unknown parameters. They are defined by

\[
\begin{align*}
\beta_0 &= 0, \\
\beta_i &= \Sigma_{i}^{-1} \Sigma_{i,i+1}, \quad i = 1, \ldots, k-1, \\
\xi_{i+1} &= \theta_{i+1} - \theta_{i} \beta_i, \quad i = 0, \ldots, k-1.
\end{align*}
\]

The matrix \( \Sigma_{i+1,i} \) is called the \( i \)-th order step-down residual covariance matrix and \( \beta_i \) is the \( i \)-th order step-down matrix of regression coefficients.

Let

(4.11) \[ \xi = (\xi_1, \xi_2, \ldots, \xi_k). \]

The hypothesis \( H: G\theta = 0 \) is equivalent to the hypothesis \( H: G\xi = 0 \), which, in turn is equivalent to the intersection representation \( \bigcap_{i=1}^{k} H_i \), where

(4.12) \[ H_i: G\xi_i = 0. \]

Thus, we can sequentially test (4.12) for \( i = 1, 2, \ldots, k \), using the conditional model (4.8) and (4.9).

Recalling that \( A \) is \( n \times m \) and of rank \( r \), partition \( A \) into

(4.13) \[ A = (A_1, A_2), \]

where \( A_1 \) is \( n \times r \) and of rank \( r \). \( A_1 \) is said to form a basis for \( A \). Similarly partition \( G: t \times m \) in conformity with the partition of \( A \):
\[ G = (G_1, G_2), \]

where \( G_1 \) is \( t \times r \). We test the conditional component hypothesis \( H_1 \) as follows. Let

\[ \Psi_i = G_{i1}, \quad i = 1, 2, \ldots, k, \]

(4.14)

\[ \Psi = (\Psi_1, \Psi_2, \ldots, \Psi_k) = G\xi; \]

\( \Psi_i \) is estimated by

\[ \hat{\Psi}_i = G_1(A_1\Lambda_1)^{-1}A'_1Y_1. \]

The matrix due to hypothesis, \( S_{H_1} : p_1 \times p_1 \) is given by

(4.15)

\[ S_{H_1} = \hat{\Psi}_i [G_1(A_1\Lambda_1)^{-1}G_1'] \hat{\Psi}_i. \]

The matrix due to error, \( S : p \times p \) is given by \( S = (S_{ij}); \ i, j = 1, \ldots, k, \)

where,

(4.16)

\[ S_{ij} = Y_i'[(I - A_1(A_1\Lambda_1)^{-1}A_1')]Y_j \]

is a \( p_1 \times p_1 \) matrix of covariances between the \( i \)-th and \( j \)-th blocks of variables. Define the conditional matrices due to error as

(4.17)

\[ S_{i+1.i-1} = S_{i+1.i+1} - S'[i,i+1]S^{-1}[i,i+1]S'[i,i+1], \]

where these quantities are the sample analogues of the population matrices defined by (4.9).

When \( H \) is true and \( Y_{i-1} \) is fixed, \( S_{H_1} \) and \( S_{i+1.i-1} \) are independently distributed according to \( W(S_{i+1.i-1}, p_1, t) \) and

48
\(W(\Sigma_{1:[i-1]}, p_1, n-r-i+1), \) respectively. In addition, the matrices
\[
\begin{align*}
S_{H_1}^{-1}, S_{H_2}^{-1}, S_{H_3}^{-1}, \ldots, S_{H_k}^{-1}
\end{align*}
\]
are independently distributed.

The next step is to select integers \(m_1, m_2, \ldots, m_k\) such that \(m_i \leq p_i\). The generalized step-down test is to accept \(H\) if
\[
(4.18) \quad \text{tr}_{m_i} \left[ I_{p_i} + S_{H_i}^{-1} \right] < c_i,
\]
for all \(i = 1, 2, \ldots, k\).

The \(c_i\) are chosen so that
\[
(4.19) \quad P\left( \text{tr}_{m_i} \left[ I_{p_i} + S_{H_i}^{-1} \right] < c_i \right) = (1-\alpha)^{1/k}
\]
for each \(i\).

### 4.2 Equality of the Covariance Matrix to the Identity Matrix

Suppose that \(S > 0\) has a Wishart distribution \(W(\Sigma, p, n)\). Partition \(\Sigma\) as \((\Sigma_{ij})\); \(i, j = 1, \ldots, k\), where \(\Sigma_{ij}\) is a \(p_i \times p_j\) matrix of covariances between the \(i\)-th and \(j\)-th blocks of variates. Partition \(S\) similar to the partition of \(\Sigma\). Define \(\Sigma_{i+1} [i]\) as in \((4.8)\) and \(\beta_i\) as in \((4.10)\).

The sample analogues are denoted by \(S_{i+1} [i]\) and \(b_i\), respectively. The hypothesis \(H: \Sigma = I_{p_i}\) is equivalent to \(H = \bigcap_{i=1}^{k-1} H_i\), where
\[
(4.20) \quad H_i: (\beta_i = 0) \cap (\Sigma_{i+1} [i] = I_{p_i}).
\]
Consider the matrices \( b_i S[i] b_i \) and \( S_{i+1}^{-1}[i] \). Under \( H_{i+1} \), when \( Y_i \) is fixed, they are independently distributed as \( W(I_{p_i+1}, p_{i+1}, p_i) \) and \( W(I_{p_{i+1}}, p_{i+1}, n-p_i) \), respectively. In addition, all the matrices \( b_i S[i] b_i \) and \( S_{i+1}^{-1}[i] \) for \( i = 1, \ldots, k-1 \) are independently distributed when \( H \) is true.

Select integers \( m_1, m_2, \ldots, m_{k-1} \) such that \( m_1 \leq p_{i+1} \). The generalized step-down test is to accept \( H \) if

\[
\text{tr}_{m_1} (I_{p_{i+1}} + b_i S[i] b_i) < c_{1i},
\]

\[
\text{tr}_{m_1} (S_{i+1}^{-1}[i] \exp (S_{i+1}^{-1}[i])) < c_{2i},
\]

for all \( i = 1, 2, \ldots, k-1 \). The constants \( c_{1i} \) and \( c_{2i} \) are chosen so that

\[
(1-\alpha)^{1/2(k-1)} = P[\text{tr}_{m_1} (I_{p_i} + b_i S[i] b_i) < c_{1i}]
\]

\[
= P[\text{tr}_{m_1} (S_{i+1}^{-1}[i] \exp (S_{i+1}^{-1}[i])) < c_{2i}] .
\]

4.3 **EQUALITY OF TWO COVARIANCE MATRICES.**

This problem is the two-sample analogue to that discussed in the preceding section. Suppose that \( S_1 \) and \( S_2 \) have the Wishart distributions, \( W(\Sigma_1, p, n_1) \) and \( W(\Sigma_2, p, n_2) \), respectively. Using definitions previously given, let \( \beta_{1i} \) and \( \beta_{2i} \) be the i-th order step-down matrices of regression coefficients for the two population covariance matrices \( \Sigma_1 \) and \( \Sigma_2 \). Let \( \Sigma_{1,i+1}^{-1}[i] \) and \( \Sigma_{2,i+1}^{-1}[i] \) be the respective residual matrices of i-th order for \( \Sigma_1 \) and \( \Sigma_2 \). The hypothesis to be tested is \( H: \Sigma_1 = \Sigma_2 \).
This is equivalent to $H = \bigcap_{i=1}^{k-1} H_i$, where

\begin{equation}
H_i: (\beta_{1i} = \beta_{2i}) \land (\Sigma_{1,i+1\cdot[i]} = \Sigma_{2,i+1\cdot[i]}).
\end{equation}

Let $b_{1i}, b_{2i}, S_{1,i+1\cdot[i]}$ and $S_{2,i+1\cdot[i]}$ be the sample analogues to $\beta_{1i}, \beta_{2i}, \Sigma_{1,i+1\cdot[i]}$ and $\Sigma_{2,i+1\cdot[i]}$, respectively. Let

\begin{equation}
S_i = (S_{1[i]} + S_{2[i]})^{-1}.
\end{equation}

Employing a similar approach to that used in the one-sample problem, we obtain the generalized step-down test which accepts $H$ if

\begin{equation}
\text{tr}_{m_1} \left( I_{P_{i+1}} + (b_{1i} - b_{2i})'S_i(b_{1i} - b_{2i}) \right) < c_{1i},
\end{equation}

\begin{equation}
\text{tr}_{m_1} \begin{pmatrix}
-(n_{1} - p_{[i]}), & n_{1} + n_{2} - 2p_{[i]} \end{pmatrix}^{-1}
\begin{pmatrix}
S_{1,i+1\cdot[i]}(S_{1,i+1\cdot[i]} + S_{2,i+1\cdot[i]}), & S_{2,i+1\cdot[i]} \end{pmatrix} < c_{2i}
\end{equation}

$c_{1i}$ and $c_{2i}$ are chosen so that

\begin{equation}
(1-\alpha)^{1/2(k-1)} = \mathbb{P}\left\{ \text{tr}_{m_1} \left( I_{P_{i+1}} + (b_{1i} - b_{2i})'S_i(b_{1i} - b_{2i}) \right) < c_{1i} \right\}
\end{equation}

\begin{equation}
= \mathbb{P}\left\{ \text{tr}_{m_1} \begin{pmatrix}
-(n-p_{[i]}), & n_{1} + n_{2} - 2p_{[i]} \end{pmatrix}^{-1}
\begin{pmatrix}
S_{1,i+1\cdot[i]}(S_{1,i+1\cdot[i]} + S_{2,i+1\cdot[i]}), & S_{2,i+1\cdot[i]} \end{pmatrix} < c_{2i} \right\}
\end{equation}
5. MONOTONICITY OF THE POWER FUNCTIONS OF SOME PROPOSED TESTS.

In many testing problems in multivariate analysis, tests which are invariant under a group of transformations depend upon the sample characteristic roots. The power of such tests is a function of the population characteristic roots, which can be thought of as noncentrality parameters. A test is said to have the monotonicity property if its power function is a monotonically increasing function of each population characteristic root.

The purpose of this chapter is to show that the two new classes of tests proposed each for the general linear model problem and the independence between two sets of variates satisfy the monotonicity property.

5.1 THE GENERAL LINEAR MODEL.

The canonical form for the general linear model is given by Roy (1957). Let \( U \) be a \( p \times s \) matrix and \( V \) be a \( p \times n-r \) matrix, where \( p \) is the number of variates, \( s \) is the degrees of freedom for the hypothesis, and \( n-r \) is the degrees of freedom for the error. The joint distribution of \( U \) and \( V \) is

\[
\begin{align*}
    f(U,V) &= (2\pi)^{-p(s+n-r)/2} \exp\left[-\frac{1}{2} \text{tr}(VV')\right] + \sum_{i=1}^{r} (u_{ii} - \theta_i)^2 \\
    &\quad + \sum_{i=t+1}^{p} u_{ii}^2 \quad + \sum_{i=1}^{r} \sum_{j \neq 1}^{s} u_{ij}^2 \biggr]\,.
\end{align*}
\]

The hypothesis to be tested is \( H: \theta_1 = \cdots = \theta_t = 0 \), against the alternative hypothesis \( K: \theta_1 > 0 \). (The \( \theta_i \)'s are arranged in nonincreasing order.)
A sufficient condition for a test to satisfy the monotonicity property (for testing \( H \) against \( K \)) has been given by Anderson, Das Gupta, and Mudholkar (1964) and is stated below.

**Theorem 5.1.** For each \( i \) \((i = 1, \ldots, s)\) and for each set of fixed values of \( u_j \)'s \((j \neq i)\) and \( V \), suppose there exists an orthogonal transformation:

\[
u_i \rightarrow M u_i = u_i^* = (u_1^*, \ldots, u_p^*)' \]

such that the region \( \omega_i(u_i^*) \), a section of the acceptance region \( \omega \) in the space of \( u_i^* \) for a set of fixed values of \( u_j \)'s \((j \neq i)\) and \( V \), is transformed into the region \( \omega_i^*(u_i^*) \) which has the following property: Any section of \( \omega_i^*(u_i^*) \) for fixed values of \( u_k^* \) \((k \neq j)\) is an interval, symmetric about \( u_j^* = 0 \). Then the power function of the test, having acceptance region \( \omega \), monotonically increases in each \( \theta_i \).

We use Theorem 5.1 to prove that the tests for the general linear model based on the \( T_{mk} \) criterion introduced in Section 3.2 satisfy the monotonicity property.

**Theorem 5.2.** In testing the general linear hypothesis, the test with acceptance region,

\[
T_{mk} = E_m [1 + \lambda_1, 1 + \lambda_2, \ldots, 1 + \lambda_k] \leq c
\]

has a power function monotonically increasing in each \( \theta_i \). (The \( \lambda_j \) are the \( k \) largest characteristic roots of \( UU'(VV')^{-1} \).)

**Proof.** The idea in this proof is to show that Theorem 5.1 is satisfied for \( M = I_p \). Let \( V \) and all columns of \( U \) except \( u_1 \) be fixed. (Without loss of generality, we can assume \( i = 1 \).)

Next partition \( U \) into \((u_1, u_2)\), where \( u_1 \) is the first column of \( U \). Then
\[ I_p = UU'(VV')^{-1} = I_p + U_2'U_2(VV') + u_1' u_1(VV')^{-1} = G \text{ (say)} \]  

Let  
\[ E = \{ u_1: E_m(\lambda_1(G), \lambda_2(G), \ldots, \lambda_k(G)) \leq c; \ U_2 \text{ and } V \text{ fixed} \} \]  

The conditions of Theorem 5.1 are satisfied if we can show that for any \( u_1 \in E \), it follows that \( au_1 \in E \) for \(-1 \leq a \leq 1\). Let  
\[ G(\alpha) = I_p + U_2'U_2(VV')^{-1} + \alpha^2 u_1' u_1(VV')^{-1} \]  
\[ = (VV' + U_2'U_2 + \alpha^2 u_1' u_1)(VV')^{-1} \]  

\( G(\alpha) \) is formed from \( G \) by replacing \( u_1 \) with \( au_1 \). Since \( G(\alpha) \) is the product of two positive-definite matrices, the characteristic roots of \( G(\alpha) \) are positive. [See Bellman (1960, p. 134)]. For \( \alpha^2 \leq 1 \),  
\[ G - G(\alpha) = G(1) - G(\alpha) = (1 - \alpha^2) u_1' u_1(VV')^{-1} \]  

Since the characteristic roots of \( G - G(\alpha) \) are nonnegative for \(-1 \leq \alpha \leq 1\), it follows that  
\[ 0 \leq \lambda_j(G(\alpha)) \leq \lambda_j(G) \]  

for \( j = 1, 2, \ldots, k \). Therefore, since \( E_m \) is a nondecreasing function of each of its arguments when they are all nonnegative,  
\[ E_m(\lambda_1(G(\alpha)), \ldots, \lambda_k(G(\alpha))) \leq E_m(\lambda_1(G), \ldots, \lambda_k(G)) \]
Consequently, \( \alpha_1 \in E \) for \(-1 \leq \alpha \leq 1\), and thus the region \( E \) satisfies Theorem 5.1 as was to be shown. \( \|
\)

5.2 **INDEPENDENCE BETWEEN TWO SETS OF VARIATES.**

The canonical form for testing for independence between two sets of variates has been given by Roy (1957). Let \( X: p \times n \) and \( Y: q \times n \) have the joint distribution,

\[
(5.9) \quad f(X, Y) = (2\pi)^{-\frac{n}{2}} (p+q)^{-\frac{p}{2}} \prod_{i=1}^{p} (1-\rho_i^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \left( \sum_{i=1}^{p} (1-\rho_i^2) \sum_{j=1}^{n} (x_{ij}^2 + y_{ij}^2 - 2\rho_i x_{ij} y_{ij}) + \sum_{i=p+1}^{q} \sum_{j=1}^{n} y_{ij}^2 \right) \right\}.
\]

The \( \rho_i \) are the population canonical correlations. The null hypothesis \( H \) holds if and only if \( \rho_1 = \cdots = \rho_p = 0 \). (It is tacitly assumed that \( p \leq q \).) The class of invariant test procedures are based on the characteristic roots of \((XX')^{-1}(XY')(YY')^{-1}(YX')\). Define

\[
(5.10) \quad S_H = (XY')(YY')^{-1}(YX'),
\]

\[
S_E = XX' - S_H.
\]

The following relationship holds between the characteristic roots of \((XX')^{-1}(XY')(YY')^{-1}(YX')\) and \( S_H S_E^{-1} \). If \( \theta_j = \lambda_j [(XX')^{-1}(XY')(YY')^{-1}(YX')] \), then

\[
(5.11) \quad \lambda_j(S_H S_E^{-1}) = \theta_j / (1-\theta_j).
\]
Matrices $U: p \times q$ and $V: p \times (n-q)$ can be found such that the characteristic roots of $UU'(VV')^{-1}$ are the same as the characteristic roots of $S_H S_E^{-1}$. The transformation from $S_H$ and $S_E$ to $U$ and $V$ is not unique. The following theorem of Anderson and Das Gupta (1964a) giving a sufficient condition for monotonicity is valid for any such $U$ and $V$.

**Theorem 5.3.** An invariant test for which the acceptance region is convex in each column vector of $U$ for each set of fixed values $V$ and of the other columns of $U$ has a power function which is monotonically increasing in each $\rho_i$.

We apply Theorem 5.3 to prove that the tests for independence between two sets of variates introduced in Section 3.5, (3.43), satisfy the monotonicity property.

**Theorem 5.4.** In testing for independence between two sets of variates, the test with acceptance region

$$T_{m_{k_1}k_2} = E_m [(1-r_1^2)^{-1}, (1-r_2^2)^{-1}, \ldots, (1-r_k^2)^{-1}, 1, \ldots, 1] \leq c,$$

where there are $k_1$ - $k$ ones in the argument of $E_m$, has a power function monotonically increasing in each $\rho_i$. (The $r_j^2$, $j=1, \ldots, k$, are the $k$ largest characteristic roots of $(XX')^{-1}(XX)(YY')^{-1}(YY')$, where $k = \min(k_1, k_2)$.)

**Proof.** If the matrices $S_H$ and $S_E$ are as defined in (5.11) and $\lambda_j$ are the ordered characteristic roots of $S_H S_E^{-1}$, then

$$\lambda_j = \frac{r_j^2}{1-r_j^2}$$

and

$$r_j^2 = \lambda_j/(1+\lambda_j) \quad (5.13)$$
Let $U: p \times q$ and $V: p \times (n-q)$ be matrices such that the characteristic roots of $UU'(VV')^{-1}$ are the same as the characteristic roots of $S_H S_E^{-1}$. From (5.13), $(1-r_j^2)^{-1} = 1+\lambda_j$. Thus, (5.12) becomes

\begin{equation}
E_m[1+\lambda_1,1+\lambda_2,\ldots,1+\lambda_k,1,\ldots,1] \leq c, \tag{5.14}
\end{equation}

where $\lambda_j$ are the ordered characteristic roots of $UU'(VV')^{-1}$.

Define the matrix $G$ as in (5.3) and the matrix $G(\alpha)$ as in (5.5). Let

\begin{equation}
E = \{ u_\perp: E_m(\lambda_1(G),\ldots,\lambda_k(G),1,\ldots,1) \leq c; \ U_2 \text{ and } V \text{ fixed} \}. \tag{5.15}
\end{equation}

From (5.7), $0 \leq \lambda_j(G(\alpha)) \leq \lambda_j(G)$, for $j = 1,2,\ldots,k$. Therefore, since $E_m$ is a nondecreasing function of each of its arguments when they are all nonnegative,

\begin{equation}
E_m[\lambda_1(G(\alpha)),\ldots,\lambda_k(G(\alpha)),1,\ldots,1] \leq \tag{5.16}
E_m[\lambda_1(G),\ldots,\lambda_k(G),1,\ldots,1].
\end{equation}

Hence, $\alpha u_\perp \in E$ for $-1 \leq \alpha \leq 1$, when $u_\perp \in E$, and thus, $E$ is a convex set in $u_\perp$. \|
6. COMPARISON OF TESTS BY BAHAUDUR EFFICIENCY.

6.1 BAHAUDUR EFFICIENCY.

In this chapter, we seek to compare the various tests introduced in Chapter 3 with respect to the indices \( m \) and \( k \). In general, it can not be expected that there will be a single pair of indices which will be uniformly best with respect to all parameter points. However, by using Bahadur efficiency as a means of comparison, we are able to compare the tests with respect to the indices without requiring the explicit non-null distribution of the test statistics.

The non-null distributions of the test statistics proposed in this study are so complicated that direct comparison of power functions is impractical. A method of comparison of tests which avoids these difficulties has been proposed by Bahadur (1960) and generalized by Gleser (1963, 1964). Bahadur efficiency requires only knowledge of the limiting distribution of the test statistic under the null hypothesis and of the probability limit of the statistic for every alternative parameter point.

The basic idea of Bahadur efficiency in comparing two tests is to consider a fixed point \( \theta \) in the alternative parameter space. Fix the power at \( \theta \) for each \( n \) at a constant \( \beta \) for the two tests and compare the rates at which the type one errors \( \alpha_n \) converge to zero. The ratio of rates is called Bahadur efficiency. In computing Bahadur efficiency, we require the following definitions.

**DEFINITION 6.1.** Let \( D(a,t) \) denote the set of cdfs \( F \) such that

\[
2 \log(1-F(x)) = -ax^t(1+o(1)), \quad t > 0, \ x \to \infty.
\]
We note that the normal distribution, $N(0,1) \in D(1,2)$, and the chi-square distribution $\chi_k^2 \in D(l, 1)$ for all $k$.

**Definition 6.2.** If the distribution of the statistic $T_N(x) \in D(a, t)$, then the "slope" of the test based on $T_N$ is defined by

\[ C(\theta) = a \lim_{N \to \infty} N^{-t} [T_N(x)]^t. \]  

**Definition 6.3.** Let $C_1(\theta)$ and $C_2(\theta)$ be the slopes for $T_N^{(1)}(x)$ and $T_N^{(2)}(x)$, respectively. The Bahadur efficiency of $T_N^{(1)}(x)$ to $T_N^{(2)}(x)$ is defined by

\[ \varphi_{12}(\theta) = C_1(\theta)/C_2(\theta). \]  

The greater the value of $\varphi_{12}$, the more efficient is the test based on $T_N^{(1)}(x)$ to $T_N^{(2)}(x)$.

In section 6.2 we obtain some results which, along with those given by Gleser (1963), form a calculus for computing $\varphi(\theta)$. In sections 6.3-6.6, these results are applied to the specific tests which have been proposed in this study.

6.2 SOME THEOREMS REGARDING THE CLASS $D(a, t)$.

In computing slopes and consequently Bahadur efficiencies, it is necessary to know the class $D(a, t)$ in which the distribution of the statistic $T_N$ belongs. Gleser has given some useful results in this area.
THEOREM 6.1 (Gleser (1963)). If $X_i$ has c.d.f. $F_i \in D(a,t)$, $i = 1, \ldots, k$, then

(i) $\max_{i} X_i$ has c.d.f. $F \in D(a,t)$;

(ii) $\min_{i} X_i$ has c.d.f. $F \in D(a,t)$;

(iii) $bX_i + c$ has c.d.f. $G_i \in D(ab^{-t}, t)$, provided $b > 0$ and $t \geq 1$;

(iv) if $X_i \geq 0$, it follows that $X_i^d$, $d > 0$ has a c.d.f. $H_i \in D(a,t/d)$;

(v) if $X_i \leq Y \leq X_i$, the c.d.f. of $Y \in D(a,t)$;

(vi) if $X_i \geq 0$ and $\sum_{i=1}^{k} x_i$ has a c.d.f. $\in D(a,t)$; then for any set of coefficients $a_1, \ldots, a_k$, $0 \leq a_i \leq \beta$, $\beta > 0$, $\sum_{i=1}^{k} a_i x_i$ has a c.d.f. $\in D(ab^{-t}, t)$.

The following additional theorems are required in our applications.

Theorem 6.2 is a generalization of Theorem 6.1(iii).

THEOREM 6.2. Suppose that $X_n \overset{D}{=} x_0$ (constant) and $n^{1/t}(X_n - x_0)$ converges to a distribution in $D(a,t)$. Let $f(x)$ be an analytic function. Then $n^{1/t}(f(X_n) - f(x_0))$ converges to a distribution in $D(a(f'(x_0))^{-t}, t)$ if $f'(x_0) \neq 0$ and in $D(a(f''(x_0)/2)^{-t/2}, t/2)$ if $f'(x) = 0$ and $f''(x_0) \neq 0$.

Proof. Since $f(x)$ is an analytic function, $f(X_n)$ can be expanded about $X_n = x_0$ as

$$f(X_n) = f(x_0) + f'(x_0)(X_n - x_0) + \frac{f''(x_0)}{2} (X_n - x_0)^2 + \frac{f'''(Z_n)}{6} (X_n - x_0)^3,$$

where $Z_n$ is between $X_n$ and $x_0$. 

60
Now, $n^{1/t}(f(X_n) - f(x_0))$ has asymptotically the same distribution as $n^{1/t}f'(x_0)(X_n - x_0)$ if $f'(x_0) \neq 0$. By Theorem 6.1(iii), this distribution belongs to $D(a(f'(x_0))^{-t}, t)$.

If $f'(x_0) = 0$ and $f''(x_0) \neq 0$, then $n^{1/t}(f(X_n) - f(x_0))$ has asymptotically the same distribution as $n^{1/t}f''(x_0)(X_n - x_0)^2/2$. By Theorem 6.1(iv), the limiting distribution of $n^{1/t}(X_n - x_0)^2$ belongs to $D(a,t/2)$. Applying Theorem 6.1(iii), the limiting distribution of $n^{1/t}(f(X_n) - f(x_0))$ belongs to $D(a(f''(x_0)/2)^{-t/2}, t/2)$.

**THEOREM 6.2.** If the c.d.f. $F_i$ of $f(X_i)$ belongs to $D(a,t)$, $i = 1, \ldots, k$, then the c.d.f. $G$ of $\sum_{i=1}^{k} f(X_i)$ also belongs to $D(a,t)$.

**Proof.** First, $\min f(X_i) \leq \sum_{i=1}^{k} f(X_i) \leq \max f(X_i)$. By Theorem 6.1(i) and (ii), the c.d.f.'s of $\min f(X_i)$ and $\max f(X_i)$ each belong to $D(a,t)$. From Theorem 6.1(v), the c.d.f. of $\sum_{i=1}^{k} f(X_i)$ belongs to $D(a,t)$.

**THEOREM 6.4.** Let $S_1$ and $S_2$ be independently distributed, each according to $W(\Sigma, p, n+o(n))$ and let $\lambda_1(S_1S_2^{-1}), \lambda_2(S_1S_2^{-1}), \ldots, \lambda_p(S_1S_2^{-1})$ be unordered characteristic roots of $S_1S_2^{-1}$. Then, if $f(x)$ is an analytic function and $k$ is any integer between $m$ and $p$, the limiting distribution of

\[
(6.4) \quad \sqrt{n} E_m[e^{f(\lambda_1)}, e^{f(\lambda_2)}, \ldots, e^{f(\lambda_k)}]
\]

belongs to

\[
D([2m^{(k)}e^{mf(1)f'(1)}]^{-2}, 2)
\]

if $f'(1) \neq 0$, and
\[ D(\frac{1}{\sqrt{2m}}e^{m^2(1)x^2(1)}x^2(1))^{-1}, 1) \]

if \( f'(1) = 0 \) and \( f''(1) \neq 0 \).

**Proof.** Let \( z_1 = \sqrt{\frac{2}{p}} (\lambda_1 - 1) \). The \( z_1 \) can be considered as the characteristic roots of a random \( p \times p \) symmetric matrix \( A \) whose elements \( a_{ij}, i \leq j \) are independently normally distributed with zero means and variances equal to one for the diagonal elements and \( 1/2 \) for the off-diagonal elements (Anderson (1958, p. 323)). The characteristic roots of \( A \) are real but may be negative. The matrix \( B = A^2 \) is positive definite with positive characteristic roots. The characteristic roots of \( B \) are the squares of the characteristic roots of \( A \).

Let \( y_k \) equal the sum of the \( k \) largest characteristic roots of \( B \).

Then

\[
y_k = \sum_{i=1}^{k} z[i] ,
\]

where \( z[i] \) are the characteristic roots of \( A \) arranged in decreasing order according to their absolute value. Consider the distribution of \( y_k \). We have the following inequality on \( y_k \).

\[
\sum_{i=1}^{k} b_{ii} \leq y_k \leq \text{tr } B .
\]

The left side of (6.6) equals \( \sum_{i=1}^{k} b_{ii} = \sum_{i=1}^{k} \sum_{j=1}^{p} a_{ij}^2 = \sum_{i=1}^{k} a_{ii}^2 + 2 \sum_{i < j} a_{ij}^2 \)

which is distributed as \( \chi^2 + \frac{1}{2} \chi^2_{p(k-1)/2} \), which belongs to \( D(1,1) \) by Theorem 1(vi). Similarly, \( \text{tr } B \) is distributed as \( \chi^2_p + \frac{1}{2} \chi^2_{p(p-1)/2} \) which also belongs to \( D(1,1) \). Thus, by Theorem 6.1(v), the distribution of \( y_k \)
belongs to $D(1,1)$. Recalling that $y_k = \sum_{i=1}^{k} z_{[i]}^2$ and each $z_{[i]}^2 \geq 0$, the application of Theorem 6.1(vi) with $\alpha_i = 1$, $\alpha_j = 0$ for $j \neq i$, $\beta = 1$, $a = 1$, and $t = 1$ yields the result that the distribution of each $z_{[i]}^2$ belongs to $D(1,1)$.

From Theorem 6.1(iv), it follows (with $d = \frac{1}{2}$) that the distribution of each $z_{[1]}$ belongs to $D(1,2)$. Since this is true for each $k = 1, 2, \ldots, p$ and the set of $z_i$ is the same as the set of $z_{[1]}$, it follows that the distribution of each $z_i$ belongs to $D(1,2)$.

Now we note that

$$E_m[e^{f(\lambda_1)}, e^{f(\lambda_2)}, \ldots, e^{f(\lambda_k)}] = \sum_{i_1 < \cdots < i_m} \exp(\sum_{j=1}^{m} f(\lambda_{i_j})) .$$

(6.7)

Since $z_i = \sqrt{n} (\lambda_i - 1)$ converges to a distribution in $D(1,2)$, or, equivalently, $\sqrt{n} (\lambda_i - 1)$ converges to a distribution in $D(\frac{1}{4}, 2)$, we can apply Theorem 6.2 with $x_n = \lambda_i$ and $x_0 = 1$. First assume that $f'(1) \neq 0$. Then by Theorem 6.2, $\sqrt{n} (f(\lambda_{i_j}) - f(1))$ converges to a distribution in

$$D(\frac{1}{4f'(1)^2}, 2) ,$$

(6.8)

for each $j = 1, 2, \ldots, m$. Next, note that $\sqrt{n} \sum_{j=1}^{m} [f(\lambda_{i_j}) - f(1)]$ converges to a distribution in $D(\frac{1}{4m^2[f'(1)]^2}, 2)$.

Let $g(x) = \exp(x)$. Reapplication of Theorem 6.2 using $g(x)$ instead of $f(x)$ yields the result that
\begin{equation}
\sqrt{n} \exp \left[ \sum_{j=1}^{m} \left[ f(\lambda_{i_j}) - f(1) \right] \right] = \sqrt{n} e^{-mf(1)} \sum_{j=1}^{m} f(\lambda_{i_j})
\end{equation}

converges to a distribution in

\[ D\left( \frac{1}{4m^2 [f'(1)]^2}, 2 \right) \]

Summing (6.9) over all permutations \( 1 \leq i_1 < \cdots < i_m \leq k \) of the integers 1 through k, and averaging them yields

\begin{equation}
\sqrt{n} e^{-mf(1)} \prod_{i=1}^{k} \left[ e^{f(\lambda_{i})}, e^{f(\lambda_{2})}, \ldots, e^{f(\lambda_{k})} \right]
\end{equation}

which has its distribution in the class

\[ D\left( \frac{1}{4m^2 [f'(1)]^2}, 2 \right) \]

This is equivalent to the distribution of

\[ \sqrt{n} \prod_{i=1}^{k} \left[ e^{f(\lambda_{i})}, e^{f(\lambda_{2})}, \ldots, e^{f(\lambda_{k})} \right] \]

being in the class

\begin{equation}
D\left( \left( \sum_{i=1}^{k} m e^{f(1)} f'(1) \right)^{-2}, 2 \right)
\end{equation}

If \( f'(1) = 0 \) and \( f'(1) \neq 0 \), the limiting distribution of

\[ \sqrt{n} \left( f(\lambda_{i_j}) - f(1) \right) \]

now becomes, from Theorem 6.2,
(6.12) \[ D\left( \frac{1}{2m^m(1)}, 1 \right), \]

for each \( j = 1, 2, \ldots, m \). Next, \( \sqrt{n} \sum_{j=1}^{m} [f(\lambda_j) - f(1)] \) converges to a distribution in \[ D\left( \frac{1}{2m^m(1)}, 1 \right). \]

Continuing the derivation analogous to the case when \( f'(1) \neq 0 \), we finally obtain the class of the limiting distribution of

\[ \sqrt{n} \mathbb{E}_m \left[ e^{f(\lambda_1)}, e^{f(\lambda_2)}, \ldots, e^{f(\lambda_k)} \right] \]

as

(6.13) \[ D\left( \frac{k}{2m^m(1)} e^{\frac{m^m(1)}{f''(1)}} - 1, 1 \right). || \]

Note that if we are dealing with a single covariance matrix \( S \sim W(\Sigma, p, n) \) with characteristic roots \( \lambda_1(S), \lambda_2(S), \ldots, \lambda_p(S) \), then the distribution of each \( z_i = \sqrt{n/2} (\lambda_i - 1) \) belongs to \( D(1,2) \). With this modification in the derivation of Theorem 6.4, we arrive at a single sample version in Theorem 6.5.

**Theorem 6.5.** If \( S \) is distributed according to \( W(\Sigma, p, n) \) and \( \lambda_1, \lambda_2, \ldots, \lambda_p \) are the (unordered) characteristic roots of \( S \), then if \( f(x) \) is an analytic function and \( k \) is any integer between \( m \) and \( p \), the limiting distribution of

(6.14) \[ \sqrt{n} \mathbb{E}_m \left[ e^{f(\lambda_1)}, e^{f(\lambda_2)}, \ldots, e^{f(\lambda_k)} \right] \]
belongs to
\[ D[(\sqrt{E} m \binom{k}{m} e^{mf(1)} f'(1))^{-2}, 2) \]
if \( f'(1) \neq 0 \) and
\[ D[(\sqrt{E} m \binom{k}{m} e^{mf(1)} f''(1))^{-1}, 1) \]
provided \( f'(1) = 0 \) and \( f''(1) \neq 0 \).

6.3 COMPARISON OF TESTS ON THE EQUALITY OF THE COVARIANCE MATRIX TO THE IDENTITY.

Suppose \( S \sim W(\Sigma, p, n) \) and we wish to compare tests of the hypothesis \( H: \Sigma = I_p \). The class of tests introduced in Section 3.2 is equivalent to rejecting \( H \) when

\[ T_{mk} = E_m \left[ e^{f(\lambda[1])}, e^{f(\lambda[2])}, \ldots, e^{f(\lambda[k])} \right] > c, \]

where \( f(\lambda) = \lambda \log \lambda \) and \( \lambda_{[i]} \) are the characteristic roots of \( S \) arranged according to decreasing values of \( f(\lambda) \).

Since \( f'(1) = 1-\lambda^{-1} \) and \( f''(1) = \lambda^{-2} \), we have \( f(1) = 1 \), \( f'(1) = 0 \) and \( f''(1) = 1 \). Applying Theorem 6.5 with \( f(\lambda) = \lambda \log \lambda \) leads to the result that the limiting distribution of \( \sqrt{n} T_{mk} \) belongs to the class

\[ E[(\sqrt{E} m \binom{k}{m} e^{m})^{-1}, 1) \].

Thus, the slope of a test based on \( T_{mk} \) is
\[(6.17) \quad C_{mk}(\lambda) = \frac{E_m[e^{f(\lambda_{[1]}^{(\Sigma)})}, \ldots, e^{f(\lambda_{[k]}^{(\Sigma)})}]}{\sqrt{2} m (\frac{k}{m}) e^m} = \]
\[
\frac{E_m[\lambda_{[1]}^{(\Sigma)} e^{\lambda_{[1]}}, \ldots, \lambda_{[k]}^{(\Sigma)} e^{\lambda_{[k]}}]}{\sqrt{2} m (\frac{k}{m}) e^m}.
\]

The Bahadur efficiency of a test based on $T_{m_1k_1}$ to a test based on $T_{m_2k_2}$ at the parameter point $\lambda$ is

\[(6.18) \quad \varphi_{m_1k_1:m_2k_2}(\lambda) = \frac{C_{m_1k_1}(\lambda)}{C_{m_2k_2}(\lambda)},
\]

where $C_{mk}$ is defined in (6.17).

Consider now a special subset of the parameter space in which

\[(6.19) \quad \Omega_j(\alpha) = \{\lambda_1 = \ldots = \lambda_j = \alpha > 1 \text{ and } \lambda_{j+1} = \ldots = \lambda_p = 1\},
\]

where $\lambda_j = \lambda_j^{(\Sigma)}$. There are two cases to consider: according as $k \leq j$ or $k > j$.

**CASE I.** $1 \leq m \leq k \leq j$.

When $\lambda \in \Omega_j(\alpha)$, (6.17) becomes

\[(6.20) \quad C_{mk}(\lambda) = \frac{E_m[\alpha^{-1} e^\alpha, \alpha^{-1} e^\alpha, \ldots, \alpha^{-1} e^\alpha]}{\sqrt{2} m (\frac{k}{m}) e^m} = \]
\[
\frac{(\frac{k}{m}) \alpha^{-m} e^{m\alpha}}{\sqrt{2} m (\frac{k}{m}) e^m} = \alpha^{-m (\alpha - 1)}.
\]
Thus, when \( \lambda \in \Omega_j(\alpha) \) and \( m \) is fixed, \( C_{mk}(\lambda) \) is the same for all \( k = 1, 2, \ldots, j \), and depends only on \( m \) and \( \alpha \). As a means of analyzing the best values of \( m \), temporarily consider \( m \) as a continuous variable. Then set

\[
\frac{\partial \log C_{mk}}{\partial m} = \alpha - 1 - \frac{1}{m} - \log \alpha = 0.
\]

This leads to

\[
(6.21) \quad \hat{m} = \frac{1}{\alpha - 1 - \log \alpha}
\]

as the only stationary value of \( C_{mk} \). When \( \alpha - 1 \) is small, then

\[
\log \alpha \approx (\alpha - 1) - (\alpha - 1)^2/2.
\]

Substitution into (6.21) yields

\[
(6.22) \quad \hat{m} \approx \frac{2}{(\alpha - 1)^2}.
\]

Thus, when \( \alpha - 1 \) is small, \( \hat{m} \) is large; that is, \( \hat{m} \approx k \). When \( \alpha \) is large, we find that

\[
(6.23) \quad \hat{m} \approx \frac{1}{\alpha}.
\]

In this situation \( \hat{m} \) is small, meaning \( \hat{m} \approx 1 \).

More generally, we find that from (6.21),

\[
(6.24) \quad \frac{\partial (1/\hat{m})}{\partial \alpha} = 1 - \frac{1}{\alpha} > 0
\]
for $\alpha > 1$. Thus, $m$ is a decreasing function of $\alpha$.

**CASE II.** $j < k$.

When $\lambda \in \Omega_j(\alpha)$, (6.17) becomes

$$
(6.25) \quad c_{mk}(\lambda) = \frac{E_m^{-1}(\alpha \cdot \alpha, \ldots, \alpha \cdot \alpha, 1, \ldots, 1)}{\sqrt{2 \cdot m} (\frac{k}{m})^m},
$$

where the number of ones in the argument of $E_m$ is $k-j$. Expanding $E_m^2$, we obtain

$$
(6.25) \quad c_{mk}(\lambda) = \frac{\min(j,m) \sum (j, m-k-j) (k-j) j^m \cdot e^{-i \alpha}}{\sqrt{2 \cdot m} (\frac{k}{m})^m} = \frac{(\frac{k-j}{m}) F \left( \begin{array}{c} k-j, -j, k-j+m+1, \alpha^{-1} \end{array} \right)}{(\frac{k}{m}) \sqrt{2 \cdot m} e^m}
$$

where $F$ is the hypergeometric function defined by the series,

$$
(6.26) \quad F \left( \begin{array}{c} a, b; r, t \end{array} \right) = 1 + \frac{ab \cdot t}{r \cdot 1!} + \frac{a(a+1)b(b+1)}{r(r+1)} \cdot \frac{t^2}{2!} + \ldots
$$

[See Kendall and Stuart (1943, p. 134)].

6.4 **COMPARISON OF TESTS ON THE EQUALITY OF TWO COVARIANCE MATRICES.**

Suppose $S_1$ and $S_2$ are independently distributed according to $W(\Sigma_1, p, n+\sigma(n))$ and $W(\Sigma_2, p, n+\sigma(n))$, respectively. We wish to compare tests of the hypothesis $H: \Sigma_1 = \Sigma_2$. The class of tests introduced in Section 3.3 is equivalent to rejecting $H$ when

$$
T_{mk} = E_m[(\theta_1 - \theta^2_1)^{-1}, \ldots, (\theta_k - \theta^2_k)^{-1}] > c,
$$

69
where \( \theta_{[j]} \) are the characteristic roots of \( S_1(S_1+S_2)^{-1} \) arranged in decreasing order according to \( g(\theta) = (\theta-\sigma^2)^{-1} \).

Denoting the characteristic roots of \( S_1S_2^{-1} \) by \( \lambda \), the relationship between \( \theta \) and \( \lambda \) is \( \theta = \lambda/(1+\lambda) \). Since \( (\theta-\sigma^2)^{-1} = (1+\lambda)^2/\lambda \), (6.26) becomes, in terms of the \( \lambda \)'s

\[
T_{mk} = E_m[e^{f(\lambda_1)}, e^{f(\lambda_2)}, \ldots, e^{f(\lambda_k)}] > c,
\]

where

\[
f(\lambda) = 2 \log(1+\lambda) - \log \lambda
\]

and \( \lambda_{[j]} \) are the characteristic roots of \( S_1S_2^{-1} \) arranged in decreasing order according to \( f(\lambda) \).

Since \( f'(\lambda) = 2(1+\lambda)^{-1}-\lambda^{-1} \) and \( f''(\lambda) = \lambda^{-2}-2(1+\lambda)^{-2} \), we have \( f(1) = 2 \log 2, f'(1) = 0, \) and \( f''(1) = 1/2 \). Applying Theorem 6.4 with \( f(\lambda) = 2 \log(1+\lambda)-\log \lambda \) leads to the result that the limiting distribution of \( \sqrt{n} T_{mk} \) belongs to the class

\[
D([2^{2m}m^{-m}]^{1/2},1) .
\]

Thus, the slope of a test based on \( T_{mk} \) is

\[
C_{mk}(\lambda) = \frac{E_m[e^{f(\lambda_1)(\Sigma_1\Sigma_2^{-1})}, \ldots, e^{f(\lambda_k)(\Sigma_1\Sigma_2^{-1})}]}{2^{2m}m^{-m/(m_k)}}
\]

\[
= \frac{E_m[(1+\lambda_1)(\Sigma_1\Sigma_2^{-1})^2/\lambda_1, \ldots, (1+\lambda_k)(\Sigma_1\Sigma_2^{-1})^2/\lambda_k]}{2^{2m}m^{-m/(m_k)}} .
\]

70
The Bahadur efficiency of one test to another at a parameter point \( \lambda \) is the ratio of their respective slopes \( C_{mk}(\lambda) \). Consider now a special subset of the parameter space in which

\[
\Omega_j(\alpha) = \{ \lambda_1 = \cdots = \lambda_j = \alpha > 1 \text{ and } \lambda_{j+1} = \cdots = \lambda_p = 1 \},
\]

where \( \lambda_j = \sum_1^j \frac{1}{E_2}. \) There are two cases to consider: \( k \leq j \) and \( k > j \).

**CASE I.** \( 1 \leq m \leq k \leq j \).

When \( \lambda \in \Omega_j(\alpha) \), (6.30) becomes

\[
C_{mk}(\lambda) = \frac{E_m[(1+\alpha)^2/\alpha, \ldots, (1+\alpha)^2/\alpha]}{2^m \cdot \binom{k}{m}} = \frac{(\binom{k}{m})(1+\alpha)^2/\alpha^m}{2^m \cdot \binom{k}{m}} = \frac{(\alpha^{1/2} + \alpha^{-1/2})^{2m}}{2^m \cdot \binom{k}{m}}.
\]

Thus, when \( \lambda \in \Omega_j(\alpha) \) and \( m \) is fixed, \( C_{mk}(\lambda) \) is the same for all \( k = 1, 2, \ldots, j \) and depends only on \( m \) and \( \alpha \). Let us temporarily consider \( m \) as a continuous variable. Let

\[
\frac{\partial \log C_{mk}}{\partial m} = 2 \log(\alpha^{1/2} + \alpha^{-1/2}) - 2 \log 2 - \frac{1}{m} = 0.
\]

This leads to

\[
(6.33) \quad \hat{m} = \frac{1}{2[\log(\alpha^{1/2} + \alpha^{-1/2}) - \log 2]}.
\]
When $\alpha$ is large, $\hat{m} \approx 2/\log \alpha$ is small; that is $\hat{m} \approx 1$. When $
abla \alpha-1$ is small, $\log(\alpha^{1/2} + \alpha^{-1/2}) - \log 2$ is small, and in this case $\hat{m}$ is large. In general, we find that from (6.33)

\[
\frac{d(1/\hat{m})}{d \alpha^{1/2}} = \frac{2(1-\alpha^{-1})}{\alpha^{1/2} + \alpha^{-1/2}} > 0
\]

when $\alpha > 1$. Thus, $\hat{m}$ is a decreasing function of $\alpha$.

CASE II. $j < k$.

When $\lambda \in \Omega_j(\alpha)$, (6.30) becomes

\[
C_{mk}(\lambda) = \frac{E_m[(1+\alpha)^2/\alpha, \ldots, (1+\alpha)^2/\alpha, 4, \ldots, 4]}{2^{2m} m^{k \choose m}},
\]

where the number of fours in the argument of $E_m$ is $k-j$. Expanding $E_m$, we obtain

\[
C_{mk}(\lambda) = \frac{\sum_{i=\max(0, m-k+1)}^{\min(j, m)} \binom{i}{j} \binom{k-j}{m-i} (1+\alpha)^{2i} m^{-i}/\alpha^i}{2^{2m} m^{k \choose m}}
\]

\[
= \frac{m^k}{m^{k \choose m}} 2F_1(-m, -j; k-j+m+1; (1+\alpha)^2/4\alpha).
\]

6.5 COMPARISON OF TESTS OF THE GENERAL LINEAR HYPOTHESIS.

In the canonical form for this problem, let $X$ be a $p \times t$ matrix and $Y$ be a $p \times (n-r)$ matrix. Here, $t$ and $n-r$ are the degrees of freedom for hypothesis and error, respectively. The joint density of $X$ and $Y$ is
\begin{equation}
(6.37) \quad f(x, y) = (2\pi)^{-p(t+n-r)/2} \exp\left\{ -\frac{1}{2} \left[ \sum_{j=1}^{n-r} \sum_{i=1}^{p} y_{ij}^2 
+ \sum_{i=1}^{t} (x_{ii} - \theta_1)^2 
+ \sum_{i=t+1}^{p} x_{ii}^2 
+ \sum_{i=1}^{p} \sum_{j=1}^{p} x_{ij}^2 \right] \right\}.
\end{equation}

The $\theta_i$'s are arranged in decreasing order according to $|\theta|$. The hypothesis to be tested is

$$H: \theta_1 = \cdots = \theta_t = 0.$$  

The class of tests introduced in Section 3.2 is equivalent to rejecting $H$ when

\begin{equation}
(6.38) \quad T_{mk} = E_m[l + \lambda_1, l + \lambda_2, \ldots, l + \lambda_k] > c
\end{equation}

where $\lambda_j = \lambda_j[(XX')(YY')^{-1}]$.

To find the limiting distribution of $T_{mk}$ first consider the statistic $z = n \sum_{j=1}^{k} \lambda_j$. The limiting distribution of $z$ is chi-square with $pt$ degrees of freedom. (See Anderson (1958), p. 224). This is in the class $D(1,1)$. The application of Theorem 6.1(vi) with $\alpha_1 = 1, \alpha_i = 0$ for $j \neq i, \beta = 1, a = 1,$ and $t = 1$ yields the result that the limiting distribution of each $n\lambda_j$ also belongs to $D(1,1)$.

Next, to each $\lambda_1$ we apply Theorem 6.2 with $f(\lambda) = \log(1+\lambda), \lambda_0 = 0$. Noting that $f'(1) = 1$, we arrive at the class of the limiting distribution of $n \log(1+\lambda_j)$ as $D(1,1)$ for each $j$.

Thus, the limiting distribution of $nz = n \sum_{j=1}^{m} \log(1+\lambda_{ij})$ belongs to $D(\frac{1}{m}, 1)$ for any subset $\{\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{im}\}$ of $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$. Reapply Theorem 6.2 to $z$ with $f(z) = \exp(z), z_0 = 0, a = 1/m$ and $t = 1$.  

73
This results in the limiting distribution of \( n(e^z - 1) \) belonging to \( D(\frac{1}{m}, 1) \). Summing over all combinations of \( i_1 < i_2 < \cdots < i_m \) and averaging, the distribution of

\[
(6.39) \quad \frac{n}{\binom{k}{m}} \sum_{i_1 < \cdots < i_m} \left[ \exp\left( \sum_{j=1}^{m} \log(1 + \lambda_{i_j}) \right) - 1 \right] = n\left[ \frac{E_m(1 + \lambda_1, 1 + \lambda_2, \ldots, 1 + \lambda_k)}{\binom{k}{m}} \right] - 1
\]

belongs to \( D(\frac{1}{m}, 1) \). By Theorem 6.1(iii), \( n T_{mk} = n E_m(1 + \lambda_1, 1 + \lambda_2, \ldots, 1 + \lambda_k) \) has a limiting distribution in

\[
(6.40) \quad D([m(\frac{k}{m})]^{-1}, 1).
\]

The slope of a test based on \( T_{mk} \) is

\[
(6.41) \quad C_{mk}(\theta) = \frac{E_m(1 + \theta_1, 1 + \theta_2, \ldots, 1 + \theta_k)}{m(\frac{k}{m})}.
\]

Note that \( \theta_j = 0 \) for \( j > t \).

The Bahadur efficiency of one test to another at a parameter point \( \theta \) is the ratio of their respective slopes \( C_{mk}(\theta) \). Consider now a special subset of the parameter space in which

\[
(6.42) \quad \Omega_j(\alpha) = \{ \theta_1 = \cdots = \theta_j = \alpha > 0, \theta_{j+1} = \cdots = \theta_t = 0 \}.
\]

There are again two cases to consider: \( k \leq j \) and \( k > j \).
CASE I. $1 \leq m \leq k \leq j$.

When $\theta \in \Omega_j(\alpha)$, (6.41) becomes

\[
C_{mk}(\theta) = \frac{E_m[l+\alpha, l+\alpha, \ldots, l+\alpha]}{m^k} = \frac{(l+\alpha)^m}{m^k},
\]

which is the same for all $k = 1, 2, \ldots, j$. Temporarily, consider $m$ as a continuous variable. Set

\[
\frac{\partial \log C_{mk}}{\partial m} = \log(l+\alpha) - \frac{1}{m} = 0.
\]

This leads to

\[
(6.44) \quad \hat{m} = \frac{1}{\log(l+\alpha)}.
\]

Thus $\hat{m}$ is a decreasing function of $\alpha$.

CASE II. $j < k$

When $\theta \in \Omega_j(\alpha)$, (6.41) becomes

\[
C_{mk}(\theta) = \sum_{i=\max(0, m-k+1)}^{\min(j, m)} \binom{j}{i} \binom{k-j}{i} \frac{(l+\alpha)^i}{m^k} = \frac{(k-j)}{m^k} \binom{k}{m} \frac{2^F_1(-m; -j; k-j+m+1, l+\alpha)}{m^k}.
\]
6.6 COMPARISON OF TESTS ON THE INDEPENDENCE BETWEEN TWO SETS OF VARIATES.

In the canonical form of this problem, let

\[
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma'_{12} & \Sigma_{22}
\end{pmatrix},
\]

where \( \Sigma_{11} \) is \( p \times p \) and \( \Sigma_{22} \) is \( q \times q \). Without loss of generality we assume \( p \leq q \). Let \( \rho_1 \geq \rho_2 \geq \cdots \geq \rho_p \geq 0 \) be the characteristic roots of \( \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}' \). The \( \rho \)'s are known as the squares of the canonical correlations. The hypothesis \( H: \Sigma_{12} = 0 \) is equivalent to the hypothesis \( H: \rho_1 = \cdots = \rho_p = 0 \).

Let \( S \) be a sample covariance matrix having a Wishart distribution and partitioned in conformity with the partition of \( \Sigma \). The characteristic roots of \( S_{11}^{-1} S_{12} S_{22}^{-1} S_{12}' \) have the same distribution as the characteristic roots of \( (XX')^{-1}XY'(YY')^{-1}YX' \), where \( X: p \times n \) and \( Y: q \times n \) have the joint density,

\[
f(X,Y) = (2\pi)^{-n(p+q)/2} \prod_{i=1}^{p} (1-\rho_i^2)^{-n/2} \exp\left(-\frac{1}{2} \left[ \sum_{i=1}^{p} (1-\rho_i^2)^{-1} \sum_{j=1}^{n} \left( x_{i,j}^2 + y_{i,j}^2 - 2\rho_i x_{i,j} y_{i,j} \right) + \sum_{i=p+1}^{q} \sum_{j=1}^{n} y_{i,j}^2 \right] \right).\]

Note that if \( r_i^2 \) are the characteristic roots of \( (XX')^{-1}XY'(YY')^{-1}YX' \) and \( \lambda_i \) are the characteristic roots of \( S_H S_E^{-1} \), where \( S_H = XY'(YY')^{-1}YX' \) and \( S_E = XX' - S_H \), the \( r_i^2 \) and \( \lambda_i \) are related by \( \lambda_i = \frac{r_i^2}{1-r_i^2} \) and \( r_i^2 = \frac{\lambda_i}{1+\lambda_i} \).

The class of tests introduced in Section 3.5 is to reject \( H \) when
(6.48) \[ T_{mk_1k_2} = E_m[(1-r_1^2)^{-1}, \ldots, (1-r_n^2)^{-1}, 1, \ldots, 1] = c \]

where \( k = \min(k_1, k_2) \) and the number of ones in the argument of \( E_m \) is \( k_1 - k \). It is seen that (6.48) can also be written as

(6.49) \[ T_{mk_1k_2} = E_m[1+\lambda_1, \ldots, 1+\lambda_k, 1, \ldots, 1] > c \]

where \( \lambda_j \) are the ordered characteristic roots of \( S_H S_E^{-1} \).

It is assumed that \( m \leq k \). By expanding \( E_m \) and taking note of the number of ones in the argument of \( E_m \), the statistic \( n T_{mk_1k_2} \) can be expressed as

(6.50) \[ n T_{mk_1k_2} = \frac{\min(m,k_1-k)}{\sum_{v=0}^{\min(m,k_1-k)}} z_v, \]

where each \( z_v \) is the sum of \( \binom{k}{m-v}\binom{k_1-k}{v} \) terms of the form \( n \exp(\sum_{j=1}^{m-v} \log(1+\lambda_{ij})) \). In Section 6.5 it is shown that the limiting distribution of each of these terms belongs to \( D((m-v)^{-1}, 1) \). Hence, if

(6.51) \[ a_v = [(m-v)\binom{k}{m-v}\binom{k_1-k}{v}]^{-1}, \]

the limiting distribution of \( z_v \) belongs to \( D(a_v, 1) \). Define

(6.52) \[ w_v = a_v z_v. \]
The limiting distribution of each $w_\nu$ belongs to $D(1,1)$. In terms of $w_\nu$, (6.50) becomes

$$n^{n_{MK_1K_2}} = \min_{\nu=0}^{\min(m,k_1-k)} \left( \frac{1}{a_\nu} \right)^{w_\nu}. \tag{6.53}$$

Since $\sum_{\nu=0}^{\min(m,k_1-k)} w_\nu$ has a limiting distribution in

$$D\left(\frac{1}{1+\min(m,k_1-k)}, 1\right), \tag{6.54}$$

The application of Theorem 6.1(vi) with $\alpha_1 = a_\nu^{-1}$ and $\beta = \max a_\nu^{-1} = 1/\min a_\nu$ implies that the limiting distribution of $n^{n_{MK_1K_2}}$ belongs to

$$D\left(\frac{\min a_\nu}{1+\min(m,k_1-k)}, 1\right). \tag{6.55}$$

If $k_1 \leq k_2$, then $k_1 - k = 0$ and $\nu = 0$. In this case, (6.55) simplifies to

$$D\left([m(k_1)]^{-1}, 1\right). \tag{6.56}$$

The slope of a test based on $T_{MK_1K_2}$ then becomes

$$C_{MK_1K_2}(\lambda) = \frac{E_m[1+\theta_1, 1+\theta_2, \ldots, 1+\theta_k]}{m(k_1^m)}, \tag{6.57}$$

which coincides with (6.41) for the general linear hypothesis. Therefore, the analysis over parameter subsets $\Omega_j(\alpha)$ defined by (6.42) remains identical.
7. SIMULTANEOUS CONFIDENCE BOUNDS ON POPULATION ELEMENTARY SYMMETRIC FUNCTIONS.

This chapter contains simultaneous confidence bounds on the elementary symmetric functions of population characteristic roots. The simultaneity is respect to variations in the index matrix $A$. These bounds are not direct inversions of the component tests; a construction which would achieve bounds of simultaneous level exactly $1-\alpha$. The difficulty arises in that the population values can not be separated from the sample values; that is, the confidence bounds can not be expressed in closed form.

The approach which is used in this chapter is based on a lemma of Roy (1954). The implication of this lemma is that if one can obtain confidence bounds on the characteristic roots of a matrix of the form $[T_A][\Theta_A]^{-1}$, simultaneous in $A$, where $T_A$ is a sample matrix and $\Theta_A$ is a parameter matrix, each depending on the index $A$ and each positive definite, then one can obtain simultaneous confidence bounds on $\text{tr}_m(\Theta_A)$ with confidence level at least $1-\alpha$. These bounds may turn out to be rather conservative.

**Lemma 7.1** [Roy (1954)]. If $A \geq 0$ and $B > 0$ are each $p \times p$ matrices and

\begin{equation}
(7.1) \quad \lambda_1 \leq \text{all } \lambda(AB^{-1}) \leq \lambda_2,
\end{equation}

then for $m = 1,2,\ldots,p$,

\begin{equation}
(7.2) \quad \lambda_1^m \text{tr}_m(B) \leq \text{tr}_m(A) \leq \lambda_2^m \text{tr}_m(B).
\end{equation}
Note that (7.1) implies (7.2), but (7.2) does not imply (7.1). Thus if (7.1) is true with probability equal to $1 - \alpha$, then the probability of (7.2) being true is at least $1 - \alpha$.

A more convenient form of Lemma 7.1 is obtained by interchanging $A$ and $B$. Since (7.1) implies that

$$(7.3) \quad d_2^{-1} \leq \text{all } \lambda(AB^{-1}) \leq d_1^{-1},$$

it follows from Lemma 7.1 that

$$(7.4) \quad d_2^{-m} \text{ tr}_m(A) \leq \text{tr}_m(B) \leq d_1^{-m} \text{ tr}_m(A).$$

This version is stated as a corollary.

**COROLLARY 7.2.** If $A \geq 0$ and $B \geq 0$ and

$$d_1 \leq \text{all } \lambda(AB^{-1}) \leq d_2,$$

then for each $m = 1, 2, \ldots, p$,

$$d_2^{-m} \text{ tr}_m(A) \leq \text{tr}_m(B) \leq d_1^{-m} \text{ tr}_m(A).$$

By setting $B = I_p$ in Lemma 7.1, we arrive at a special case, which is also obvious from the definition of $\text{tr}_m(A)$.

**COROLLARY 7.3.** If $A \geq 0$ and $d_1 \leq \text{all } \lambda(A) \leq d_2$, then for each $m = 1, 2, \ldots, p$,

$$(7.5) \quad d_1^{-m} \text{ tr}_m(A) \leq \text{tr}_m(A) \leq d_2^{-m} \text{ tr}_m(A).$$
7.1 SIMULTANEOUS CONFIDENCE BOUNDS ON A SINGLE COVARIANCE MATRIX.

Let

\[ A_{kp} = \{A: k \times p, \text{rank}(A) = k\} , \]

and let \( S \) have a Wishart distribution, \( W(\Sigma, p, n) \). For any \( A \in A_{kp} \), consider the matrix, \((ASA')(A\Sigma A')^{-1}\). By Theorem 2.6, the characteristic roots of \((ASA')(A\Sigma A')^{-1}\) are bounded by

\[
(7.6) \quad \lambda_p(S^{-1}) \leq \text{all } \lambda(ASA')(A\Sigma A')^{-1} \leq \lambda_1(S^{-1})
\]

for every \( A \in A_{kp} \).

Now let \( c_{1\alpha} \) and \( c_{2\alpha} \) be constants such that

\[
(7.7) \quad P[c_{1\alpha} \leq \lambda_p(S^{-1}) \leq \lambda_1(S^{-1}) \leq c_{2\alpha}] = 1-\alpha .
\]

Combining (7.6) and (7.7),

\[
(7.8) \quad P[c_{1\alpha} \leq \text{all } \lambda(ASA')(A\Sigma A')^{-1} \leq c_{2\alpha} \text{ for all } A \in A_{kp}] = 1-\alpha .
\]

Applying Corollary 7.2,

\[
(7.9) \quad c_{2\alpha}^{-m} \text{tr}_m(ASA') \leq \text{tr}_m(A\Sigma A') \leq c_{1\alpha}^{-m} \text{tr}_m(ASA')
\]

is a simultaneous set of confidence bounds on \( \text{tr}_m(A\Sigma A') \) for all \( A \in A_{kp} \) with confidence level \( \geq 1-\alpha \).
7.2 SIMULTANEOUS CONFIDENCE BOUNDS ON TWO COVARIANCES MATRICES.

Let $S_1 \sim W(\Sigma_1, p, n_1)$ and $S_2 \sim W(\Sigma_2, p, n_2)$. As the starting point, we consider the confidence limits on $\lambda(\Sigma_1 \Sigma_2^{-1})$ given by Roy (1954).

\[
(7.10) \quad c_{2\alpha}(p, n_1, n_2) \lambda_p(S_1 S_2^{-1}) \leq \lambda(\Sigma_1 \Sigma_2^{-1}) \leq c_{1\alpha}(p, n_1, n_2) \lambda_1(S_1 S_2^{-1}),
\]

where $c_{1\alpha}(p, n_1, n_2)$ and $c_{2\alpha}(p, n_1, n_2)$ are constants such that

\[
P[c_{1\alpha}(p, n_1, n_2) \leq \lambda_p(S_1 S_2^{-1}) \leq \lambda_1(S_1 S_2^{-1}) \leq c_{2\alpha}(p, n_1, n_2)] = 1-\alpha,
\]

when $\Sigma_1 = \Sigma_2$.

For any fixed $A \in A_{kp}$, note that $AS_1 A' \sim W(A\Sigma_1 A', k, n_1)$ and $AS_2 A' \sim W(A\Sigma_2 A', k, n_2)$. Thus, by applying (7.10) to $(A\Sigma_1 A')(A\Sigma_2 A')^{-1}$, we obtain

\[
(7.11) \quad c_{2\alpha}(k, n_1, n_2) \lambda_k[(A\Sigma_1 A')(A\Sigma_2 A')^{-1}] \leq \lambda[(A\Sigma_1 A')(A\Sigma_2 A')^{-1}] \leq c_{1\alpha}(k, n_1, n_2) \lambda_1[(A\Sigma_1 A')(A\Sigma_2 A')^{-1}],
\]

The confidence interval in (7.11) holds with probability $1-\alpha$ for each fixed $A \in A_{kp}$. To hold with probability $1-\alpha$ simultaneously for all $A \in A_{kp}$, we must maximize the upper confidence limit and minimize the lower limit with respect to $A$. But by Theorem 2.6,

\[
\max_{A_{kp}} \lambda_1[(A\Sigma_1 A')(A\Sigma_2 A')^{-1}] = \lambda_1(S_1 S_2^{-1}),
\]

82
and
\[
\min_{\kappa \in \mathcal{A}_{kp}} \lambda_k \left( (A^T \Sigma_1 A') (A^T \Sigma_2 A')^{-1} \right) = \lambda_p \left( S_1 S_2^{-1} \right).
\]

Thus, a simultaneous set of confidence bounds of level 1-\( \alpha \) on the characteristic roots of \( (A^T \Sigma_1 A') (A^T \Sigma_2 A')^{-1} \), for all \( A \in \mathcal{A}_{kp} \), is given by
\[
(7.12) \quad c_{2\alpha}^{-1}(k, n_1, n_2) \lambda_p (S_1 S_2^{-1}) \leq \alpha \left( (A^T \Sigma_1 A') (A^T \Sigma_2 A')^{-1} \right) \leq c_{1\alpha}^{-1}(k, n_1, n_2) \lambda_1 (S_1 S_2^{-1}).
\]

We next apply Corollary 7.2 to (7.12). The following set of bounds hold simultaneously with probability at least 1-\( \alpha \) as \( A \) varies throughout \( \mathcal{A}_{kp} \).
\[
(7.13) \quad c_{2\alpha}^{-1}(k, n_1, n_2) \lambda_p (S_1 S_2^{-1}) \leq \text{tr}_m \left( (A^T \Sigma_1 A') (A^T \Sigma_2 A')^{-1} \right) \leq c_{1\alpha}^{-1}(k, n_1, n_2) \lambda_1 (S_1 S_2^{-1}).
\]

7.3 THE GENERAL LINEAR MODEL.

Let \( X: p \times n \) be an observable random matrix with
\[
(7.14) \quad E X' = B \theta,
\]
where \( B: n \times m \) is the design matrix of rank \( r \leq m < n \) and \( \theta: m \times p \) is a set of unknown parameters. It is assumed that \( n > p \) and that the columns of \( X \) have a multivariate normal distribution with unknown covariance matrix \( \Sigma \).
Partition $B$ and $\theta$ into

$$B = (B_I, B_D),$$
$$\theta = (\theta_I, \theta_D),$$

where $B_I$ is an $n \times r$ and $\theta_I$ is $r \times p$. Let $C: s \times m$ be a "hypothesis" matrix, where $m \leq r$. Define the matrix due to hypothesis $S_H$ and the matrix due to error $S_E$ by

$$(7.15) \quad S_H = X(B_I'B_I)^{-1} C' [C(B_I'B_I)^{-1} C']^{-1} C(B_I'B_I)^{-1} B_I X'/s,$$
$$S_E = X[I_n - B_I(B_I'B_I)^{-1} B_I] X'/(n-r).$$

Denoting $C\theta$ by $\eta: s \times p$, Roy and Gnanadesikan (1959) give confidence bounds for $\eta$ as follows.

$$(7.16) \quad \lambda_1^{1/2}(\eta'(C(B_I'B_I)^{-1} C')^{-1}\eta) \leq s^{1/2} \lambda_1^{1/2}(S_H) + [sc^2]^{1/2} \lambda_1^{1/2}(S_E).$$

This confidence bound holds with probability $1-\alpha$. Now replacing $\eta$ by $\eta A'$, $S_H$ by $A S_H A'$ and $S_E$ by $A S_E A'$ where $A: k \times p$ and $AA' = I_k$, we obtain the following simultaneous confidence bounds at level $1-\alpha$,

$$(7.17) \quad \lambda_1^{1/2} [A\eta'(C(B_I'B_I)^{-1} C')^{-1} A\eta'] \leq s^{1/2} \lambda_1^{1/2}(A S_H A') + [sc^2]^{1/2} \lambda_1^{1/2}(A S_E A').$$

But $\lambda_1(A S_H A') \leq \lambda_1(S_H)$ and $\lambda_1(A S_E A') \leq S_E$. Thus,
(7.18) \[ \lambda_1^{1/2} [A \eta' (C(B'B_L)^{-1}C')^{-1} \eta A'] \leq s^{1/2} \lambda_1^{1/2} (S_H) + (sc_\alpha)^{1/2} \lambda_1^{1/2} (S_E) \]

holds simultaneously in \( A \) with probability \( 1 - \alpha \). Squaring both sides and applying Corollary 7.3 we obtain

(7.19) \[ \text{tr}_m [A \eta' (C(B'B_L)^{-1}C')^{-1} \eta A'] \leq \]

\[ s^m \left( \sum_{k=1}^{K} \lambda_k (S_H) + c_\alpha \lambda_1 (S_E) + 2 \sqrt{c_\alpha \lambda_1 (S_H) \lambda_1 (S_E)} \right)^m \]

holds simultaneous for \( A \in A_{kp} \) with probability \( \geq 1 - \alpha \).
BIBLIOGRAPHY


10. Eckhart, C. and Young, G. (1936). The approximation of one matrix by
another of lower rank. Psychometrika, 1, 211-218.

11. Fan, K. (1949). On a theorem of Weyl concerning the eigenvalues of

12. Fan, K. (1951). Maximum properties and inequalities for the eigen-
Sci. 37, 760-766.


15. Gleser, L.J. (1964). On a measure of test efficiency proposed by


the air testing of sample bombsights. Techniques of Statistical
Analysis, 11-184.


30, 180-187.


