APPROXIMATE CONFIDENCE REGIONS FOR CONSTRAINT PARAMETERS

BY

ALBERT MADANSKY and INGRAM OLKIN

TECHNICAL REPORT NO. 32
FEBRUARY 22, 1969

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT 6681

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
APPROXIMATE CONFIDENCE REGIONS FOR CONSTRAINT PARAMETERS

BY

ALBERT MADANSKY and INGRAM OLKIN

TECHNICAL REPORT NO. 32
FEBRUARY 22, 1969

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT 6681

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Approximate Confidence Regions for Constraint Parameters

By

Albert Madansky and Ingram Olkin

DataPlan Inc., Stanford University

1. Introduction. Suppose we have a probability distribution \( p_\theta(x) \), where the natural parameter \( \theta \) may be vector-valued, and we wish to find a confidence region for some vector-valued function, \( h(\theta) = (h_1(\theta), \ldots, h_q(\theta)) \), of \( \theta \). For reasons which will be clear later, we will call \( h(\theta) \) a "constraint parameter". If \( h(\theta) \) is complicated, it may be difficult to obtain exact confidence regions. However, asymptotic confidence regions frequently are available by using "linearization", i.e., if \( \hat{h}(\theta) \) is an estimate of \( h(\theta) \), and the appropriate regularity conditions are satisfied, \([h(\theta) - h(\theta)] V_n^{-1}(\theta) [h(\theta) - h(\theta)]'\) is approximately \( \chi^2_q \), where \( V_n(\theta) \) is an appropriately normalized estimate of the asymptotic covariance matrix of \( \hat{h}(\theta) \) based on a sample of size \( n \). For \( q = 1 \), we obtain the usual normal approximation, namely, \( [h(\theta) - h(\theta)]/\sqrt{V_n(\theta)} \) is approximately \( N(0,1) \), where \( V_n(\theta) \) is the normalized estimate of the asymptotic variance of \( \hat{h}(\theta) \).

The purpose of this paper is to present an alternative method for obtaining confidence regions. The method is also asymptotic and is based on the asymptotic distribution of the likelihood ratio statistic (LRS). The particular novelty of the procedure rests in the fact that 1/This work was sponsored in part by the Rand Corporation, Center for Advanced Study in the Behavioral Sciences, and the National Science Foundation.
the LRS is not determined explicitly as a function of \( \hat{\theta} \), the maximum
likelihood estimator (MLE) of \( \theta \) under the null hypothesis, but instead,
as a function of certain Lagrangian multipliers. If, then, \( h(\hat{\theta}) \) as a
function of the Lagrangian multipliers possesses appropriate monotonicity
properties, then the LRS also possesses certain monotonicity or convexity
properties, which permits the determination of a confidence region for
\( h(\theta) \) in a simple way.

Lagrangian multipliers arise naturally in the problem of finding
the constrained MLE \( \hat{\theta} \) of \( \theta \) when \( h(\theta) \) is specified to be a particular
vector, \( \rho \). Since the Lagrangian multipliers, as functions of \( \hat{\theta} \), should
equal zero when \( h(\theta) = \rho \), Aitchison and Silvey (1958) suggest that one
test the hypothesis that \( h(\theta) = \rho \) by testing the equivalent hypothesis
that the Lagrangian multipliers equal zero. To this end, Silvey (1959)
obtains the asymptotic distribution of the Lagrangian multipliers as a
function of the constrained MLE \( \hat{\theta} \).

Approximate confidence regions for the Lagrangian multipliers
may be obtained using either Silvey's result or by using the approximate
distribution of the LRS as a function of the multipliers. The essence
of this paper is that it presents a method for converting a confidence
region for the multipliers into a confidence region for \( h(\theta) \).

The general procedure and relevant preliminaries are presented
in Sections 2 and 3. Section 4 is concerned with one sample problems
for the Wishart distribution. There confidence bounds for \( \text{tr } \Sigma A \), \( \text{tr } \Sigma^{-1} A \),
\( |\Sigma| \), and the characteristic roots of \( \Sigma \) are obtained. Two sample problems
are in Section 5 and include confidence bounds for \( \text{tr}(A_1 \Sigma_1 + A_2 \Sigma_2) \),
\( \text{tr}(A_1 \Sigma_1^{-1} + A_2 \Sigma_2^{-1}) \), \( |\Sigma_1| |\Sigma_2|^{a_2}, a_1|\Lambda_1| + a_2|\Lambda_2| \), and \( \text{tr } \Sigma_1^{-1} \Sigma_2 \).
Section 6 deals with confidence bounds for $\mu \Sigma^{-1} \mu'$, and Section 7 with the Behrens-Fisher problem.

2. **A method for obtaining approximate confidence regions.** Let $X$ be a random variable (possibly vector valued) with distribution function $F(x; \theta)$, where $\theta = (\theta_1, ..., \theta_m)$ is a vector of real parameters of the distribution, and assume that $F(x; \theta)$ has a density $f(x; \theta)$. Given $n$ independent observations $X_1, ..., X_n$ on $X$, define

$$L_n(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

Let the parameter space $\Omega$ be a subset of Euclidean $m$-space and $\omega$ be an $(m-q)$-dimensional subset of $\Omega$ defined by the constraints $h_i(\theta) = \rho_i$, $i = 1, ..., q$, where the $\rho_i$ are real numbers. Let $\lambda$ be the likelihood ratio statistic (LRS) for testing the hypothesis $\theta \in \omega = \{ \theta : h_i(\theta) = \rho_i, i = 1, ..., q \}$, and suppose that $-2 \log \lambda$ is asymptotically distributed as a chi-square variable with $q$ degrees of freedom.

An approximate confidence region, with confidence coefficient $\gamma$, for the vector of parameters $\rho = (\rho_1, ..., \rho_q)$ is then given by

$$\{ \rho : -2 \log \lambda \leq \chi^2_\alpha(q) \},$$

where the level of significance is $\alpha = 1 - \gamma$, and $\chi^2_\alpha(q)$ is the upper $100\%$ point of the chi-square distribution with $q$ degrees of freedom. In the expression

$$\lambda = \frac{\sup_{\theta \in \omega} L_n(x; \theta)}{\sup_{\theta \in \Omega} L_n(x; \theta)},$$
it is usually a simple task to determine the denominator of \( \lambda \). To determine the numerator, we must find the maximum likelihood estimator (MLE) of \( \theta \in \omega \), i.e., that value \( \hat{\theta} \in \omega \) which satisfies

\[
\log L_n(x_j \hat{\theta}) = \sup_{\theta \in \omega} \log L_n(x_j \theta).
\]

Consider the Lagrangian expression

\[
M_n(x_j \theta) = \log L_n(x_j \theta) - \sum_{i=1}^{q} \xi_i [h_i(\theta) - \rho_i].
\]

(For simplicity of notation, the subscript \( n \) in \( M_n(x_j \theta) \) and \( L_n(x_j \theta) \) is omitted whenever the context is clear.)

Let \( \hat{\theta}(\xi) = (\hat{\theta}_1(\xi), \ldots, \hat{\theta}_m(\xi)) \), where \( \xi = (\xi_1, \ldots, \xi_q) \), denote a generic member of the family of vectors in \( \omega \) which satisfy the equations

\[
(2.1) \quad \frac{\partial M(x_j \theta)}{\partial \theta_\alpha} = \frac{1}{L(x_j \theta)} \frac{\partial L(x_j \theta)}{\partial \theta_\alpha} - \sum_{i=1}^{q} \xi_i \frac{\partial h_i(\theta)}{\partial \theta_\alpha} = 0, \quad \alpha = 1, \ldots, m.
\]

The usual procedure at this point is to find the vector \( \xi_0 \) which satisfies the constraint equations

\[
h_i(\hat{\theta}(\xi)) = \rho_i, \quad i = 1, \ldots, q.
\]

Then the MLE of \( \theta \in \omega \) is a solution, \( \hat{\theta}(\xi_0) \), of (2.1).

In many cases it is difficult or cumbersome to explicitly express the solution \( \xi_0 \) of the constraint equations as a function of the \( \rho_i \).
However, suppose there exists a transformation taking \( \theta \) into 
\[
h^*(\theta) = (h_1(\theta), \ldots, h_q(\theta), h_{q+1}(\theta), \ldots, h_m(\theta)),
\]
where \( h_{q+1}(\theta), \ldots, h_m(\theta) \) are so defined that the transformation \( \theta \to h^*(\theta) \) is a one-to-one and continuously invertible transformation of \( \Omega \) into itself. Suppose further that there exists a transformation taking \( h^*(\theta) \) into \( \xi^* = (\xi_1, \ldots, \xi_q, h_{q+1}(\theta), \ldots, h_m(\theta)) \), where the transformation \( h^*(\theta) \to \xi^* \) is a one-to-one and continuously invertible transformation of \( \Omega \) into itself. Then we can successively reparameterize our distribution from \( \theta \) through \( h^*(\theta) \) to \( \xi^* \). Given a confidence region for \( \xi = (\xi_1, \ldots, \xi_m) \), the image region for \( h(\hat{\theta}(\xi)) = (h_1(\hat{\theta}(\xi)), \ldots, h_m(\hat{\theta}(\xi))) \) induced by the transformation \( \xi \to h(\hat{\theta}(\xi)) \) is a confidence region for \( \rho = (\rho_1, \ldots, \rho_m) \).

The following conditions are necessary to carry out this procedure:

(a) The MLE of \( \theta \in \Omega \) is obtained as the solution of the equation
\[
\frac{\partial}{\partial \theta} \log L_n(x; \theta) \bigg|_{\theta = \hat{\theta}} = 0,
\]

(b) The MLE of \( \theta \in \omega \) is obtained as the solution of the equations
\[
\frac{\partial M_n(x; \theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} = 0, \quad h(\hat{\theta}(\xi)) = \rho,
\]

(c) \(-2 \log \lambda\) is asymptotically distributed as a chi-square variable with \( q \) degrees of freedom when \( \theta = \theta_0 \).

Various sufficient conditions and specific examples for (a)-(c) are given in the literature (cf., Aitchison and Silvey (1960), Wald (1943)).
However, the hypotheses are sufficiently restrictive that one cannot apply them to the applications we shall deal with, namely the normal and Wishart families. Instead we show that (a) and (b) hold; it is relatively straightforward to show that (c) holds using standard asymptotic results.

2.1. The simple constraint case. In general, it may be hard to characterize this region, even when the region for $\xi$ is simple. However, in the important special case when $q = 1$ and $h(\hat{\theta}(\xi))$ is a continuous monotonic function of $\xi$, the characterization is simple, since intervals are carried into intervals. The following lemma points out that this monotonicity is equivalent to a very convenient property of the likelihood ratio as a function of $\xi$.

**Lemma 2.1:** When there is only one constraint, i.e., when $q = 1$, the derivative of $g(\xi) = -2 \log \lambda$, with respect to $\xi$, changes sign only once, when $\xi = 0$, if and only if $h(\hat{\theta}(\xi))$ is a monotonic function of $\xi$.

**Proof:** Differentiating $g(\xi)$ with respect to $\xi$, and using (2.1), we obtain

$$
\frac{\partial g(\xi)}{\partial \xi} = -2 \sum_{\alpha=1}^{m} \frac{\partial L(x_j \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi)} \frac{\partial \hat{\theta}_\alpha(\xi)}{\partial \xi} = -2 \xi \frac{\partial h(\hat{\theta}(\xi))}{\partial \xi},
$$

from which the result follows. ||
An approximate confidence region for $\xi$ is the set

$$C(\xi) = \{ \xi : g(\xi) \leq \chi^2_\alpha(q) \}.$$ 

When $h(\hat{\theta}(\xi))$ is a monotonic function of $\xi$, the equation $g(\xi) = \chi^2_\alpha(q)$ will, by Lemma 2.1, typically have two solutions, so that $C(\xi)$ will either be a bounded interval or its complement. By the monotonicity of $g(\xi) = -2 \log \lambda$ for $\xi > 0$ and for $\xi < 0$ separately, one can easily compute the solution of $g(\xi) = \chi^2_\alpha(q)$. Since $h(\hat{\theta}(\xi))$ carries intervals into intervals, this region can be converted directly into a confidence region for $\rho$. (The cases when $g(\xi) = \chi^2_\alpha(q)$ has no or one solution can be handled similarly.)

An alternative procedure, which exploits the monotonicity of $h(\hat{\theta}(\xi))$ as a function of $\xi$ but not the "butterfly" property of $g(\xi)$, is to use Silvey's result to obtain a confidence region for $\xi$ and convert this region into a confidence region for $\rho$. The confidence region for $\xi$ based on Silvey's work is

$$\left\{ \xi : \frac{n(\hat{\xi} - \xi)^2}{\hat{r}(\hat{\xi})} \leq \chi^2_\alpha(1) \right\},$$

where $\hat{\xi}$ satisfies

$$\left. \frac{\partial \log L_n(x; \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}(\xi)} = \hat{\xi} \left. \frac{\partial h(\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}(\xi)},$$

$h(\hat{\theta}(\xi)) = \rho$, and

7
\[
\hat{r}(\hat{\xi}) = \frac{\frac{\partial^2}{\partial \theta^2} \mathbb{E} \log f(x; \theta) + \frac{\partial^2}{\partial \theta^2} h(\theta)}{\left( \frac{\partial h(\theta)}{\partial \theta} \right)^2} \bigg|_{\xi = \hat{\xi}, \theta = \hat{\theta}(\hat{\xi})}.
\]

Our method has the advantage that we need never compute \( \hat{\xi} \) in order to determine a confidence interval for \( \xi \).

Ascertaining that \(-2 \log \lambda\) is a convex function of \( \xi \) may be useful for computational purposes, for then the confidence region \( C(\xi) \) is a convex set. One can then solve the equation \( g(\xi) = \chi^2_{\alpha}(q) \) numerically, e.g., by using the efficient algorithm of Gross and Johnson (1959), and thereby determine the boundary of this set. Similarly, if \(-2 \log \lambda\) is a concave function of \( \xi \), \( C(\xi) \) is the complement in the subspace of \( \xi \)'s of a convex set. When \( q = 1 \), the following lemma gives a necessary and sufficient condition for the concavity or convexity of \( g(\xi) \).

**Lemma 2.2:** For \( q = 1 \), \( g(\xi) \) is concave in \( \xi \) if and only if

\[
\frac{\partial h(\hat{\theta}(\xi))}{\partial \xi} + \xi \frac{\partial^2 h(\hat{\theta}(\xi))}{\partial \xi^2} > 0.
\]

**Proof:** The result follows immediately upon differentiation of (2.2).

**Remark.** A sufficient condition for concavity is that \( h(\hat{\theta}(\xi)) \) be an increasing function of \( \xi \) and \( \xi \frac{\partial^2 h(\hat{\theta}(\xi))}{\partial \xi^2} > 0 \).

When \( h(\theta) \) is linear in \( \theta \), we have the following lemma.
Lemma 2.3: If $L(x; \theta)$ is log concave in $\theta$ and $h(\theta) = \sum_{i=1}^{m} a_{i} \theta_{i}$, then $h(\theta(\xi))$ is monotone decreasing in $\xi$.

Proof: In (2.1) let $S_{i}(\theta) = \partial \log L(x; \theta) / \partial \theta_{i}$, so that the equations $S_{i}(\theta) - a_{i} = 0, i = 1, \ldots, m$, define the $\hat{\theta}(\xi)$. Then

$$\frac{\partial S_{i}(\hat{\theta}(\xi))}{\partial \xi} = \sum_{j=1}^{m} \frac{\partial S_{i}(\hat{\theta}(\xi))}{\partial \hat{\theta}_{j}(\xi)} \frac{\partial \hat{\theta}_{j}(\xi)}{\partial \xi} = a_{i}, \quad i = 1, \ldots, m.$$

But

$$\frac{\partial S_{i}(\hat{\theta}(\xi))}{\partial \hat{\theta}_{j}(\xi)} = \frac{\partial^{2} \log L(x; \theta)}{\partial \theta_{i} \partial \theta_{j}} \bigg|_{\theta = \hat{\theta}(\xi)} = d_{ij}.$$

Let

$$a = (a_{1}, \ldots, a_{m}), \quad \frac{\partial \hat{\theta}(\xi)}{\partial \xi} = \left(\frac{\partial \hat{\theta}_{1}(\xi)}{\partial \xi}, \ldots, \frac{\partial \hat{\theta}_{m}(\xi)}{\partial \xi}\right)$$

and $D = (d_{ij})$, so that $a = \frac{\partial \hat{\theta}(\xi)}{\partial \xi} D$. Then

$$\frac{\partial h(\hat{\theta}(\xi))}{\partial \xi} = \left(\frac{\partial \hat{\theta}(\xi)}{\partial \xi}\right) a' = a D^{-1} a'. $$

By hypothesis $D$ is negative definite so that $a D^{-1} a' < 0.$

2.2. The multiconstraint case Lemmas 2.1 and 2.2 do not generalize readily, as the region $\{\xi : \xi > 0\}$ when $q = 1$ must be translated to an orthant of Euclidean $q$-space, and the appropriate orthant in this case would depend on the magnitudes of the components of the vectors.
of partial derivatives of \( h_1(\hat{\theta}(\xi)) \) with respect to the \( \xi \)'s, and not just their signs. However, we may be able to find conditions for the convexity of

\[
g(\xi_1, \ldots, \xi_q) = -2 \log \sup_\omega L(x; \theta) + 2 \log \sup_\Omega L(x; \theta) .
\]

Since the second term is independent of \( \xi \), we need only show that \(-2 \log \sup_\omega L(x; \theta)\) is convex in the vector \( \xi \). As \(-\log\) is a monotone decreasing and convex function, if we could show that \( \sup_\omega L(x; \theta) \) is concave in \( \xi \), then \( g(\xi) \) will be convex. But \( \sup \) is a convex function, so that we cannot expect simple conditions on \( L(x; \theta) \) to guarantee the convexity of \( g(\xi) \). The following discussion leads to a sufficient condition for the convexity of \( g(\xi) \).

First note that

\[
(2.3) \quad \frac{\partial^2 g(\xi)}{\partial \xi_i \partial \xi_j} = -2 \frac{\partial^2 \log L(x; \hat{\theta}(\xi))}{\partial \xi_i \partial \xi_j} \]

\[
= \frac{2}{[L(x; \hat{\theta}(\xi))]^2} \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \xi_i} \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \xi_j} - \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^m \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi)} \frac{\partial^2 \hat{\theta}_\alpha(\xi)}{\partial \xi_i \partial \xi_m} - \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^m \sum_{\beta=1}^m \frac{\partial^2 L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi) \partial \hat{\theta}_\beta(\xi)} \frac{\partial \hat{\theta}_\alpha(\xi)}{\partial \xi_i} \frac{\partial \hat{\theta}_\beta(\xi)}{\partial \xi_j},
\]

and hence the quadratic form of second derivatives can be written as
\[
\sum_{i=1}^{m} \sum_{j=1}^{m} u_i u_j \frac{\partial^2 g(\xi)}{\partial \xi_i \partial \xi_j} = \frac{2}{[L(x; \hat{\theta}(\xi))]^2} \left[ \sum_{i=1}^{q} u_i \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \xi_i} \right]^2
\]

\[
- \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^{m} \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi)} \sum_{i=1}^{q} u_i \frac{\partial^2 \hat{\theta}_\alpha(\xi)}{\partial \xi_i \partial \xi_j}
\]

\[
- \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} \frac{\partial^2 L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi) \partial \hat{\theta}_\beta(\xi)} \left[ \sum_{i=1}^{q} u_i \frac{\partial \hat{\theta}_\alpha(\xi)}{\partial \xi_i} \right] \left[ \sum_{j=1}^{m} u_m \frac{\partial \hat{\theta}_\beta(\xi)}{\partial \xi_j} \right].
\]

If we let
\[
y_\alpha = \sum_{i=1}^{q} u_i \frac{\partial \hat{\theta}_\alpha(\xi)}{\partial \xi_i},
\]

then the sum of the first and last terms is

\[
\frac{2}{[L(x; \hat{\theta}(\xi))]^2} \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} y_\alpha y_\beta \left[ \frac{\partial^2 L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi) \partial \hat{\theta}_\beta(\xi)} - \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi)} \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\beta(\xi)} \right]
\]

\[
= - \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} y_\alpha y_\beta \frac{\partial^2 \log L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi) \partial \hat{\theta}_\beta(\xi)}
\]

and hence

\[
(2.4) \quad \sum_{i=1}^{q} \sum_{j=1}^{q} u_i u_j \frac{\partial^2 g(\xi)}{\partial \xi_i \partial \xi_j} = - \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} y_\alpha y_\beta \frac{\partial^2 \log L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi) \partial \hat{\theta}_\beta(\xi)}
\]

\[
- \frac{2}{L(x; \hat{\theta}(\xi))} \sum_{\alpha=1}^{m} \frac{\partial L(x; \hat{\theta}(\xi))}{\partial \hat{\theta}_\alpha(\xi)} \sum_{i=1}^{q} u_i \sum_{j=1}^{m} u_j \frac{\partial^2 \hat{\theta}_\alpha(\xi)}{\partial \xi_i \partial \xi_j}.
\]

Thus, conditions for which (2.4) is of one sign will yield convexity or concavity of $g(\xi)$. One such condition is given in the following lemma.
Lemma 2.4: Define

\[ R_1 = \{ \theta_1 : \theta_1 \in \omega, L(x; \theta) \text{ is non-decreasing in } \theta_1 \}, \]

\[ S_1 = \{ \theta_1 : \theta_1 \in \omega, L(x; \theta) \text{ is non-increasing in } \theta_1 \}, \]

\[ i = 1, \ldots, m. \]

If \( \log L(x; \theta) \) is concave in \( \theta \), and for \( i = 1, \ldots, m \), \( \hat{\theta}_i(\xi) \) is concave in \( \xi \) for \( \hat{\theta}_i(\xi) \in R_1 \), and convex for \( \hat{\theta}_i(\xi) \in S_1 \), then \( g(\xi) \) is convex.

3. Preliminaries and Notation. As certain distributions appear throughout this paper, we establish some notation and properties needed later.

By \( \mathcal{L}(X) \) we mean the law of the random variable (vector) \( X \). In particular, \( \mathcal{L}(X) = N(\mu, \Sigma) \) means that the random vector \( X \) has a normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \), with density function

\[
(3.1) \quad p(x; \mu, \Sigma) = c_0 |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu) \right],
\]

\[-\infty < x_i < \infty, \quad i = 1, \ldots, k, \quad c_0 = (2\pi)^{-k/2} N^{-kN/2}. \]

By \( \mathcal{L}(S) = W(\Sigma; k, n) \) we mean that the random \( k \times k \) matrix \( S \) has a Wishart distribution with parameter \( \Sigma \) and \( n \) degrees of freedom, with density function

\[
(3.2) \quad p(S; \Sigma) = c(k, n) |\Sigma|^{-n/2} |S|^{(n-k-1)/2} \exp \left[ -\frac{1}{2} \text{tr } \Sigma^{-1} S \right],
\]

\( S > 0, \Sigma > 0, \frac{1}{2} \) where \( c(k, n) = \left[ 2^{nk/2} \pi^{k(k-1)/2} \prod_{i=1}^{k} \Gamma \left( \frac{n-i+1}{2} \right) \right]^{-1}. \)

\footnote{If \( A \) and \( B \) are symmetric matrices, the inequality \( A > B \) means that \( A - B \) is positive definite.}
We note immediately that \( p(x; \mu, \Sigma) \) is log concave in \( \mu \) (but not concave). Neither \( p(x; \mu, \Sigma) \) nor \( p(S; \Sigma) \) is convex, concave, log convex, or log concave in \( \Sigma \). However, both \( p(x; \mu, \Sigma) \) and \( p(S; \Sigma) \) are log concave in \( \Lambda = \Sigma^{-1} \). This follows from the fact that \( |\Lambda| \) is log concave in \( \Lambda \), (e.g., see Beckenbach and Bellman (1961)).

Other common distributions whose densities are log concave are the multinomial and Poisson with the usual parameterization. A general class of (univariate) densities with this property is the class of Polya frequency functions of order 2, i.e., densities of the form \( p(x; \theta) \) which are totally positive of order 2. One can then use the above method to construct approximate confidence regions for appropriate parameters from these distributions. An application to the binomial distribution is given by Madansky (1965).

If we observe \( N \) independent observations \( x_\alpha = (x_{1\alpha}, \ldots, x_{k\alpha}), \alpha = 1, \ldots, N, \) from a \( k \)-variate normal population, we need only consider the joint distribution of the sample mean vector \( \bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha/N, \) and the \( k \times k \) sample covariance matrix \( V = (v_{ij}), \quad v_{ij} = \frac{1}{N} \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)/n, \) \( n = N-1. \) Since \( \bar{x} \) and \( S = nV \) are independently distributed, with \( \mathcal{L}(\bar{x}) = N(\mu, \Sigma/N), \mathcal{L}(S) = W(\Sigma; k, n) \), only the marginal densities are needed, and this will generally be our starting point.

The extremal problems encountered in verifying that the maximizers under \( \Omega \) and \( \omega \) satisfy the standard first order conditions given earlier generally fall into two categories. The simplest is where we require the maximum of a concave function, \( \mathcal{L}(\theta) \), (the logarithm of the density function) subject to a concave constraint \( h(\theta) = 0. \) In this instance, the maximizer of \( \mathcal{L}(\theta) \) in \( \omega \) is an interior point; it is obtained from \( \partial M(\theta)/\partial \theta = 0, h(\theta) = 0, \) (e.g., Kuhn and Tucker (1951)).
When $l(\theta)$ is concave, but $h(\theta)$ is not, the maximum occurs either at a unique interior point or on the boundary. Thus we need only check whether it occurs on the boundary. Usually $l(\theta) \to -\infty$ as $\theta$ approaches a boundary point.

When neither $l(\theta)$ nor $h(\theta)$ is concave, there may be many local maxima or there may be a maximum at a boundary point. Then all the local maxima must be compared to determine the global maximum, and must be compared to boundary points.

The two results needed later are the following. If

$$M(\Lambda) = n \log |\Lambda| - \xi \text{ tr } AS,$$

$\Lambda > 0$, $S > 0$ is a fixed matrix, then $M(\Lambda) \to -\infty$ as $\Lambda$ approaches a singular matrix (or equivalently, the smallest characteristic root approaches 0), or as the largest characteristic root approaches $\infty$.

If

$$M(\Lambda, \mu) = n \log |\Lambda| - \text{ tr } AS - \xi (t-\mu) \Lambda(t-\mu)' ,$$

$-\infty < \mu_1 < \infty$, $\Lambda > 0$, $S$ and $t$ are fixed with $S > 0$, $-\infty < t_1 < \infty$, then $M(\Lambda, \mu) \to -\infty$ as $\Lambda$ approaches a singular matrix or as the largest characteristic root of $\Lambda$ approaches $\infty$. 

14
4. One sample problems for the Wishart distribution. Let $S = nV$ be a $k \times k$ random symmetric matrix having a Wishart distribution, $W(\Sigma, k, n)$ with density function given by (3.2), let $\Lambda = \Sigma^{-1}$ and let $A$ be an arbitrary symmetric matrix. We obtain approximate confidence intervals for $\text{tr} \; \Lambda \; A$ (Section 4.2), $\text{tr} \; \Sigma \; A$ (Section 4.3), $|\Sigma|$ (Section 4.3), and all the characteristic roots of $\Sigma$ (Section 4.4), using the method proposed in Section 2, as well as by the method of linearization (Section 4.1). Finally, we present an example of each of these intervals.

4.1. Linearization. The method of linearization for functions of $\Sigma$ is based on the approximation

$$h(V) \approx h(\Sigma) + \sum_{i,j=1}^{k} \frac{\partial h(\Sigma)}{\partial \sigma_{ij}} (v_{ij} - \sigma_{ij}),$$

(4.1)

$$\text{Var} \; h(V) \approx \frac{2}{n} \text{tr} \; H(\Sigma) \Sigma \; H(\Sigma),$$

(4.2)

where $H(\Sigma) = (\partial h(\Sigma)/\partial \sigma_{ij})$. Since $h(V)$ is a function of moments, and

$$\lim_{n \to \infty} v_{ij} = \sigma_{ij},$$

it follows that

$$\mathcal{L} \left( \frac{h(V) - h(\Sigma)}{\sqrt{\frac{2}{n} \text{tr} \; [VH(V)]^2}} \right) \to N(0,1).$$

Thus, an equal tailed approximate 100 $\gamma\%$ confidence interval for $h(\Sigma)$ is given by

$$h(V) \pm z_{\gamma} \sqrt{\frac{2}{n} \text{tr} \; [VH(V)]^2},$$

15
where \( z_\gamma \) is the \( 100(1+\gamma)/2 \) % point at the standard normal distribution.

To apply this, we need the derivatives \( \partial H(\Sigma)/\partial \sigma_{ij} \) for the functions \( \text{tr} \Sigma^{-1} A \), \( \text{tr} \Sigma A \), and \( |\Sigma| \) playing the role of \( h(\Sigma) \). These derivatives and the corresponding expressions for \( \text{tr} (\Sigma H(\Sigma))^2 \) are given by

\[
(4.3) \quad \frac{\partial \text{tr} \Lambda A}{\partial \sigma_{ij}} = - (\Lambda A \Lambda)_{ij}, \quad \text{tr} (\Sigma H(\Sigma))^2 = \text{tr} (\Lambda A)^2,
\]

\[
(4.4) \quad \frac{\partial \text{tr} \Sigma A}{\partial \sigma_{ij}} = a_{ij}, \quad \text{tr} (\Sigma H(\Sigma))^2 = \text{tr} (\Sigma A)^2,
\]

\[
(4.5) \quad \frac{\partial |\Sigma|}{\partial \sigma_{ij}} = |\Sigma| \lambda_{ij}, \quad \text{tr} (\Sigma H(\Sigma))^2 = k|\Sigma|^2.
\]

Consequently,

\[
(4.6) \quad \mathcal{L} \left( \frac{\sqrt{n} (\text{tr} V^{-1} A - \text{tr} \Sigma^{-1} A)}{\sqrt{2} \text{tr} (V^{-1} A)^2} \right) \rightarrow \mathcal{N}(0,1),
\]

\[
(4.7) \quad \mathcal{L} \left( \frac{\sqrt{n} (\text{tr} VA - \text{tr} \Sigma A)}{\sqrt{2} \text{tr} (VA)^2} \right) \rightarrow \mathcal{N}(0,1),
\]

\[
(4.8) \quad \mathcal{L} \left( \frac{\sqrt{n}}{\sqrt{2k}} \left[ \frac{|V|}{|\Sigma|} - 1 \right] \right) \rightarrow \mathcal{N}(0,1),
\]

from which we can easily obtain confidence intervals for the constraint parameters.
4.2. **Confidence interval for tr Λ A.** From (3.2) with \( \Omega = \{\Sigma : \Sigma > 0\} \), we have \( \hat{\Sigma} = V = S/n \), and

\[
(4.9) \quad \sup_{\Omega} p(S;\Sigma) = c(k,n) \left| S \right|^{-(k+1)/2} n^{kn/2} e^{-kn/2} .
\]

To find \( \sup_{\omega} p(S;\Sigma) \), where \( \omega = \{\Lambda : \Lambda > 0, \text{tr} \, \Lambda A = \rho\} \), we form the Lagrangian

\[
M(\Lambda) = n \log |\Lambda| - \text{tr} \, \Lambda S - \xi(\text{tr} \, \Lambda A - \rho) ,
\]

and note that \( p(S;\Sigma) \) and \( \text{tr} \, \Lambda A \) are concave functions of \( \Lambda \). Setting the derivatives of \( M(\Lambda) \) with respect to \( \lambda_{ij} \) equal to zero yields the matrix equation

\[
n \Sigma - S - A = 0 ,
\]

so that

\[
\hat{\Lambda}(\xi) = n(S + \xi A)^{-1} ,
\]

where \( \xi \) is such that \( \hat{\Lambda}(\xi) > 0 \). Consequently,

\[
\sup_{\omega} p(S;\Sigma) = c(k,n) \left| S \right|^{(n-k-1)/2} n^{kn/2} \left| S + \xi A \right|^{-n/2} \times \exp \left[ - \frac{n}{2} \text{tr} \, S(S + \xi A)^{-1} \right] ,
\]

and

\[
g(\xi) = n \text{tr} \, S(S + \xi A)^{-1} - n \log |S(S + \xi A)^{-1}| - kn .
\]
Since \( p(S; \Sigma) \) is log concave in \( \Lambda \) and \( h(\Lambda) = \text{tr} \Lambda A \) is linear in \( \Lambda \), by Lemma 2.3, \( h(\hat{\Lambda}(\xi)) \) is a monotonically decreasing function of \( \xi \).

Hence, from Lemma 2.1, \( g(\xi) \) is monotone increasing for \( \xi > 0 \) and decreasing for \( \xi < 0 \).

4.3. **Confidence interval for** \( \text{tr} \Sigma A \). The region \( \Omega = \{ \Lambda : \Lambda > 0 \} \),

so that \( \sup_{\Omega} p(S; \Sigma) \) is given by (4.9). Now \( \omega = \{ \Lambda : \Lambda > 0, \text{tr} \Lambda^{-1} A = \rho, \ A \geq 0 \} \) and we form the Lagrangian

\[
M(\Lambda) = n \log |\Lambda| - \text{tr} \Lambda S - \xi (\text{tr} \Lambda^{-1} A - \rho).
\]

Recalling that at the boundary of \( \Omega \), \( n \log |\Lambda| - \text{tr} \Lambda S \) approaches infinity, we see that the interior saddlepoint is a maximizer.

Setting the derivatives of \( M(\Lambda) \), with respect to the \( \lambda_{ij} \), equal to zero yields the matrix equation

\[
(4.10) \quad n S - S - \xi \Lambda^{-1} A \Lambda^{-1} = 0,
\]

or equivalently, after pre- and post-multiplying (4.10) by \( S^{1/2} A \),

\[
(4.11) \quad (S^{1/2} A S^{1/2})^2 - n(S^{1/2} A S^{1/2}) - \xi S^{1/2} A S^{1/2} = 0.
\]

Letting \( \psi = S^{1/2} A S^{1/2} \) and completing the square in (4.11) yields

\[
(\psi - \frac{n}{2} I)^2 = \frac{n^2}{4} I + \xi S^{1/2} A S^{1/2} = B,
\]

18
where $\xi$ is such that $B > 0$. Hence

$$\psi = \frac{n}{2} I + B^{1/2},$$

where $B^{1/2}$ is any symmetric square root of $B$.

From

$$n \log |\Lambda| - \text{tr} \Lambda S = n \log |\psi| - \text{tr} \psi - n \log |S|,$$

we require the square root $B^{1/2}$ which maximizes

$$s(B^{1/2}) = n \log \left| \frac{n}{2} I + B^{1/2} \right| - \text{tr} B^{1/2}.$$

But $s(B^{1/2})$ is a function of the eigenvalues of $B^{1/2}$, and the eigenvalues of any square root $B^{1/2}$ differ only in sign, i.e.,

$$\lambda(B^{1/2}) = \pm [\lambda(B^2)]^{1/2}.$$

Consequently the problem reduces to the following. If $\theta_1, \ldots, \theta_k$ are the positive eigenvalues of $B$, we must determine the set of signs $\epsilon_i = \pm 1, i = 1, \ldots, k$ which maximizes

$$n \log \Pi \left( \frac{n}{2} + \epsilon_i \theta_i \right) - \sum \epsilon_i \theta_i.$$

We assert that $\epsilon_1 = \cdots = \epsilon_k = 1$ is the maximizer, i.e.,

$$n \sum \log \left( \frac{n}{2} + \epsilon_i \theta_i \right) - \sum \epsilon_i \theta_i \leq n \sum \log \left( \frac{n}{2} + \theta_i \right) - \sum \theta_i.$$
For any single term for which \( \varepsilon_i = -1 \), we have

\[
 n \log\left(\frac{n}{2} - \theta\right) + \theta \leq \log\left(\frac{n}{2} + \theta\right) - \theta ,
\]

which holds if and only if

\[
x \leq \frac{\log(1+x)}{1 - x} , \quad 0 < x < 1 .
\]

But \( t(x) = e^{-x}(1+x)/(1-x) \) is a monotone increasing function with \( t(0) = 1 \), from which the result follows.

Hence we obtain

\[
\sup_{\omega} p(S; \Sigma) = c(k, n) \left|S\right|^{-\frac{k+1}{2}} \left(\frac{n}{2}\right)^{kn/2} \left|I + (I+\xi C)^{1/2}\right|^{n/2} \times \exp \left(-\frac{n}{2} \left[k + \text{tr}(t + \xi C)^{1/2}\right]\right) ,
\]

where \( C = 4S^{1/2} A S^{1/2}/n^2 \), so that

\[
(4.12) \quad g(\xi) = kn \log 2 - n \log\left|I + (I+\xi C)^{1/2}\right| + \frac{n}{2} \left[-k + \text{tr}(I+\xi C)^{1/2}\right] .
\]

Also,

\[
h(\hat{\Lambda}(\xi)) = \text{tr} \hat{\Sigma} A = \frac{2}{n} \text{tr} C[I + (I+\xi C)^{1/2}]^{-1} = \frac{2}{n} \sum_{i=1}^{k} \frac{c_j}{1 + \sqrt{1 + \xi c_j^2}}
\]

where \( c_1, \ldots, c_k \) are the characteristic roots of \( C \). Thus \( h(\hat{\Lambda}(\xi)) \) is a monotone function of \( \xi \).
4.4. Confidence interval for $|\Sigma|$. The region $\Omega$ is as in Section 4.2, so that $\sup_{\Omega} p(S; \Sigma)$ is given by (4.9). Now $\omega = \{\Lambda : \Lambda > 0 > \log|\Lambda| = \rho\}$, and we form the Lagrangian

$$M(\Lambda) = n \log|\Lambda| - \text{tr} \Lambda S - \xi(\log|\Lambda| - \rho),$$

and note that $\log|\Lambda|$ is a concave function of $\Lambda$. Equating the derivatives of $M(\Lambda)$ with respect to the $\lambda_{ij}$ to zero yields the matrix equation

$$n\Sigma - S - \xi\Sigma = 0,$$

so that $\hat{\Sigma} = S/(n-\xi)$, $\xi < n$. However, in this case we can easily solve for $\xi$ in terms of $\rho$, namely,

$$n - \xi(|S| e^{\rho})^{1/k} = [\eta(S, \rho)]^{1/k}.$$

Hence

$$g(\eta) = -n \log \eta(S, \rho) + k [\eta(S, \rho)]^{1/k} + kn \log n - kn.$$

Since $g(\eta)$ is convex in $\eta$, the interval $[\eta_1, \eta_2]$ is a confidence interval for $\eta$, with confidence coefficient $1 - \alpha$, where $\eta_1$ and $\eta_2$ are the roots of $g(\eta) = \chi^2_\alpha(1)$. Clearly, for any $S$, we can convert to a confidence interval $[\exp(-\rho_1), \exp(-\rho_2)]$ for $|\Sigma|$, where $\exp(-\rho_j) = |S|/\eta_j^j$, $j = 1, 2$.

To make a comparison between linearization and likelihood ratio inversion confidence limits for $|\Sigma|$, we recall from (4.8) that the linearization confidence limits are given by
\[ P(a_1 | V | \leq | \Sigma | \leq a_2 | V |) = \gamma, \]

where

\[ a_1 = \left( 1 + \frac{\sqrt{2k}}{n} \right)^{-1}, \quad a_2 = \left( \max \left( 0, 1 - \frac{\sqrt{2k}}{n} \right) \right)^{-1}. \]

The limits based on inverting the LRT are given by

\[ P(b_1 | V | \leq | \Sigma | \leq b_2 | V |) = \gamma, \]

where \( b_i = n^{k/n_i}, \ i = 1, 2, \) and \( (\eta_1, \eta_2) \) are the roots of

\[ g(\eta) = \chi^2_{1-\gamma}(l). \]

Values of \( b_1 \) to \( 5 \sigma \) are given in Table I for \( n = 10(10)40, \ k = 2(2)n, \ \gamma = .95. \) We have computed, though not tabled, \( a_1 \) as well, and note that \( a_1 \leq b_1, \ a_2 \leq b_2, \) and \( a_2 \) is typically infinite for the tabled values. Though as \( n \) gets large, \( a_1 \) and \( b_1 \) should converge to the same value, the convergence is slow even for small values of \( k, \) for lower one-sided intervals, linearization yields a larger lower limit, but for two-sided intervals, likelihood ratio inversion yields a finite band which is quite usable.
<table>
<thead>
<tr>
<th>k</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 30</th>
<th>n = 40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lower</td>
<td>upper</td>
<td>lower</td>
<td>upper</td>
</tr>
<tr>
<td>2</td>
<td>.251</td>
<td>3.076</td>
<td>0.389</td>
<td>2.263</td>
</tr>
<tr>
<td>4</td>
<td>.151</td>
<td>5.123</td>
<td>0.271</td>
<td>3.250</td>
</tr>
<tr>
<td>6</td>
<td>.102</td>
<td>7.585</td>
<td>0.205</td>
<td>4.292</td>
</tr>
<tr>
<td>8</td>
<td>.073</td>
<td>10.563</td>
<td>0.162</td>
<td>5.426</td>
</tr>
<tr>
<td>10</td>
<td>.055</td>
<td>14.144</td>
<td>0.132</td>
<td>6.672</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>0.109</td>
<td>8.043</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>0.092</td>
<td>9.551</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>0.078</td>
<td>11.208</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>0.068</td>
<td>13.026</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>0.059</td>
<td>15.015</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.5. Confidence intervals for the characteristic roots of $\Sigma$. Consider in detail the development of Section 4.2 when $A = u'u$, $u = (u_1, \ldots, u_k)$, $uu' = 1$, then using

$$|S + \xi u'u| = |S| \left(1 + \xi u S^{-1} u'\right),$$

and

$$(S + \xi u'u)^{-1} = S^{-1} - \xi(S^{-1} u'u S^{-1}) \left(1 + \xi u S^{-1} u'\right)^{-1},$$

we see that $g(\xi)$ can be expressed as a function $t(x)$ of $x = \xi u S^{-1} u'$, namely

$$t(x) = n \log (1+x) - n x/(1+x).$$

Solving $t(x) = \chi^2_{1-\gamma}(1)$ for the two roots $x_1$ and $x_2$, $x_1 > x_2$, we have

$$P\left\{\frac{n u S^{-1} u'}{1 + x_1} \leq u \wedge u' \leq \frac{n u S^{-1} u'}{1 + x_2}\right\} = \gamma.$$

Hence

$$(4.15) \quad P\left\{\frac{\lambda_1(V^{-1})}{1 + x_1} \leq u \wedge u' \leq \frac{\lambda_k(V^{-1})}{1 + x_2}\right\} \geq \gamma,$$

where $\lambda_1(V^{-1})$ and $\lambda_k(V^{-1})$ are the smallest and largest characteristic roots of $V^{-1} = nS^{-1}$, respectively. To obtain confidence bounds for the characteristic roots of $\Sigma$, we take reciprocals in (4.15).
When \( A = u'u \), the asymptotic variance of \( u\nu^{-1}u' \) is
\[
\frac{2}{n} (u \Lambda u')^2,
\]
and need not be estimated, so that (4.6) becomes
\[
\mathcal{L} \left( \frac{\sqrt{n} (u\nu^{-1}u' - u \Lambda u')}{\sqrt{2} u \Lambda u'} \right) \longrightarrow N(0,1),
\]
from which
\[
(4.16) \quad P(a_1 u\nu^{-1}u' < u \Lambda u' < a_2 u\nu^{-1}u') = \gamma,
\]
where \( a_1 = (1 + z_\gamma \sqrt{2/n})^{-1} \) and \( a_2 = [\max(0, 1 - z_\gamma \sqrt{2/n})]^{-1}. \)

Table II permits comparison of the two procedures by presenting values of the \( a_i \) and \( b_i = 1/(1+x_i) \), \( i = 1, 2, \) for \( n = 2(2)50, \gamma = .95. \) Another procedure for obtaining confidence bounds for the characteristic roots is given by Roy and Bose (1953),

Table II

4.6. \textbf{An example.} As an example of the use of our method of obtaining confidence limits for the various functions of the covariance matrix of a multivariate normal distribution treated in Sections 4.2 - 4.5, we consider the data from Barnard's study of Egyptian skulls (see Anderson (1958), p. 214) in which there are four populations: Late Predynastic, Sixth to Twelfth, Twelfth to Thirteenth, and Ptolemaic Dynasties, each being a 4-variate (maximum breadth, basialreolar length,
### Table II

<table>
<thead>
<tr>
<th>$n$</th>
<th>lower</th>
<th>upper</th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.0570</td>
<td>4.4027</td>
<td>.3378</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.1663</td>
<td>3.0876</td>
<td>.4191</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.2487</td>
<td>2.5929</td>
<td>.4691</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.3104</td>
<td>2.322</td>
<td>.5050</td>
<td>50.0000</td>
</tr>
<tr>
<td>10</td>
<td>.3586</td>
<td>2.149</td>
<td>.5328</td>
<td>8.0997</td>
</tr>
<tr>
<td>12</td>
<td>.3974</td>
<td>2.026</td>
<td>.5555</td>
<td>5.0041</td>
</tr>
<tr>
<td>14</td>
<td>.4297</td>
<td>1.933</td>
<td>.5744</td>
<td>3.8581</td>
</tr>
<tr>
<td>16</td>
<td>.4570</td>
<td>1.861</td>
<td>.5906</td>
<td>3.2569</td>
</tr>
<tr>
<td>18</td>
<td>.4805</td>
<td>1.802</td>
<td>.6048</td>
<td>2.8846</td>
</tr>
<tr>
<td>20</td>
<td>.5011</td>
<td>1.753</td>
<td>.6173</td>
<td>2.6302</td>
</tr>
<tr>
<td>22</td>
<td>.5192</td>
<td>1.712</td>
<td>.6285</td>
<td>2.4447</td>
</tr>
<tr>
<td>24</td>
<td>.5355</td>
<td>1.677</td>
<td>.6386</td>
<td>2.3031</td>
</tr>
<tr>
<td>26</td>
<td>.5501</td>
<td>1.646</td>
<td>.6478</td>
<td>2.1910</td>
</tr>
<tr>
<td>28</td>
<td>.5633</td>
<td>1.618</td>
<td>.6562</td>
<td>2.1000</td>
</tr>
<tr>
<td>30</td>
<td>.5754</td>
<td>1.594</td>
<td>.6659</td>
<td>2.0245</td>
</tr>
<tr>
<td>32</td>
<td>.5865</td>
<td>1.573</td>
<td>.6711</td>
<td>1.9607</td>
</tr>
<tr>
<td>34</td>
<td>.5968</td>
<td>1.553</td>
<td>.6777</td>
<td>1.9061</td>
</tr>
<tr>
<td>36</td>
<td>.6062</td>
<td>1.535</td>
<td>.6840</td>
<td>1.8586</td>
</tr>
<tr>
<td>38</td>
<td>.6150</td>
<td>1.519</td>
<td>.6898</td>
<td>1.8170</td>
</tr>
<tr>
<td>40</td>
<td>.6233</td>
<td>1.504</td>
<td>.6952</td>
<td>1.7802</td>
</tr>
<tr>
<td>42</td>
<td>.6309</td>
<td>1.490</td>
<td>.7004</td>
<td>1.7473</td>
</tr>
<tr>
<td>44</td>
<td>.6382</td>
<td>1.477</td>
<td>.7052</td>
<td>1.7178</td>
</tr>
<tr>
<td>46</td>
<td>.6450</td>
<td>1.466</td>
<td>.7098</td>
<td>1.6911</td>
</tr>
<tr>
<td>48</td>
<td>.6514</td>
<td>1.455</td>
<td>.7142</td>
<td>1.6668</td>
</tr>
<tr>
<td>50</td>
<td>.6574</td>
<td>1.444</td>
<td>.7183</td>
<td>1.6447</td>
</tr>
</tbody>
</table>
nasal height, basibregmatic height) normal population with common covariance matrix. The sample matrix of cross products, $S$, distributed as $W(\Sigma, 4, 394)$, is

$$S = \begin{pmatrix} 9661.997 & 445.573 & 1130.624 & 2148.584 \\ 445.573 & 9073.115 & 1239.212 & 2255.813 \\ 1130.624 & 1239.212 & 3958.320 & 1271.055 \\ 2148.584 & 2255.813 & 1271.055 & 8741.509 \end{pmatrix}.$$

We record that the characteristic roots of $V = S/394$ are $32.828$, $22.689$, $15.483$, $8.732$, so that

$$\text{tr } V = 79.732, \quad \text{tr } V^{-1} = .2536, \quad |V| = 1.007 \times 10^5.$$

In determining 95% confidence limits for $\text{tr } \Lambda$, we first find the two solutions of the equation $g(\xi) = \chi^2_{.95}(1) = 3.841$, where $g(\xi)$ is given by (4.10). These are $\xi_1 = 412.264$ and $\xi_2 = -364.683$, so that

$$(\text{tr } \hat{\Lambda}(\xi_1), \text{tr } \hat{\Lambda}(\xi_2)) = (.2344, .2741)$$

is a 95% confidence interval for $\text{tr } \Lambda$.

In determining 95% confidence limits for $\text{tr } \Sigma$, we find that the two solutions of $g(\xi) = 3.841$, where $g(\xi)$ is given by (4.12), are $\xi_1 = -1.118$ and $\xi_2 = 1.408$, so that a 95% confidence interval for $\text{tr } \Sigma$ is
\( (\text{tr} \, \hat{\Sigma}(\xi_1), \text{tr} \, \hat{\Sigma}(\xi_2)) = (73.964, 86.206) \).

To determine a confidence interval for \(|\Sigma|\), we need multiplicative factors akin to those of Table I for \(n = 394, k = 4\), which we apply to \(|V|\) or \(|S|\) as in \((4.14)\). These factors are \(0.75386\) and \(1.31792\), so that a 95\% confidence interval for \(|\Sigma|\) is

\[ 1.007 \times 10^{-5} \times (0.75386, 1.31792) = (0.759, 1.327) \times 10^{-5}. \]

To determine a confidence interval for all the characteristic roots of \(\Sigma\), we need multiplicative factors akin to those of Table II for \(n = 394\), which we apply to the minimum and maximum characteristic roots of \(V^{-1}\) as per \((4.15)\). These factors are \(0.86679\) and \(1.14621\), so that the 95\% confidence interval for all the characteristic roots of \(\Sigma\) are

\[ (0.732/1.14621, 32.828/0.86679) = (7.618, 37.873). \]

5. Two sample problems for the Wishart distribution. Let \(S_1\) and \(S_2\) be independent \(k \times k\) random symmetric matrices having Wishart distributions \(W(\Sigma_1; k, n_1)\) and \(W(\Sigma_2; k, n_2)\), respectively, and let \(\Lambda_1 = \Sigma_1^{-1}\), \(\Lambda_2 = \Sigma_2^{-1}\). Using the method of Section 2, we obtain approximate confidence regions for various functions of \(\Sigma_1\) and \(\Sigma_2\): \(\text{tr}(\Lambda_1 \Lambda_1^T + \Lambda_2 \Lambda_2^T)\) and
\( \text{tr}(A_1 \Sigma_1 + A_2 \Sigma_2) \), where \( A_1 \) and \( A_2 \) are \( k \times k \) symmetric matrices (Section 5.2), \( \Sigma_1^{a_1} \Sigma_2^{a_2} \) (Section 5.3), \( a_1 |\Lambda_1| + a_2 |\Lambda_2| \) (Section 5.4), and \( \text{tr} \Sigma_1^{-1} \Sigma_2 \) (Section 5.5). For example, \( \text{tr}(\Sigma_1 - \Sigma_2) \), \( |\Sigma_1|/|\Sigma_2| \) and \( |\Lambda_1| - |\Lambda_2| \), which are of particular interest, are included as special cases. In addition, approximate confidence regions for these functions by the method of linearization are also obtained in Section 5.1.

The results of Section 5.2 reduce in the case where \( k = 1 \) (the univariate normal case) to confidence limits for a linear combination of variances or reciprocals of variances of two (or more) normal distributions. The results of the other sections specialize similarly when \( k = 1 \).

We note that as in Section 4, for each problem the logarithm of the joint density function is a concave function of \((\Lambda_1, \Lambda_2)\).

5.1. Linearization. The development is similar to that of Section 4.1, and we have

\[
(5.1) \quad h(V_1, V_2) \approx h(\Sigma_1, \Sigma_2) + \sum_{\alpha=1}^{2} \sum_{i,j=1}^{k} \frac{\partial h(\Sigma_1, \Sigma_2)}{\partial \sigma_{ij}^{(\alpha)}} (v_{ij}^{(\alpha)} - \sigma_{ij}),
\]

where \( V_\alpha = (v_{ij}^{(\alpha)}) \), \( \Sigma_\alpha = (\sigma_{ij}^{(\alpha)}) \), and \( V_\alpha = S_\alpha / n_\alpha \), \( \alpha = 1, 2 \). Then

\[
(5.2) \quad \text{Var} h(V_1, V_2) \approx 2 \left( \frac{1}{n_1} \text{tr} [\Sigma_1 H_1 (\Sigma_1, \Sigma_2)]^2 + \frac{1}{n_2} \text{tr} [\Sigma_2 H_2 (\Sigma_1, \Sigma_2)]^2 \right),
\]

29
where $H_{\alpha} = H_{\alpha}(\Sigma_1, \Sigma_2) = [\partial h(\Sigma_1, \Sigma_2)/\partial \sigma_{ij}^{(\alpha)}]$, $\alpha = 1, 2$. The additional necessary derivatives for the functions $h(\Sigma_1, \Sigma_2)$ to be considered in this section are:

**Combinations of traces:**

\[ \frac{\partial \text{ tr}(A_1 A_1 + A_2 A_2)}{\partial \sigma_{ij}^{(\alpha)}} = -(A_\alpha A_\alpha A_\alpha)_{ij}, \quad \text{tr}(\Sigma_{\alpha} H_{\alpha})^2 = \text{tr}(A_{\alpha} A_{\alpha})^2, \]

\[ \frac{\partial \text{ tr}(A_1 \Sigma_1 + A_2 \Sigma_2)}{\partial \sigma_{ij}^{(\alpha)}} = a_{ij}^{(\alpha)}, \quad \text{tr}(\Sigma_{\alpha} H_{\alpha})^2 = \text{tr}(A_{\alpha} A_{\alpha})^2, \]

**Combinations of determinants:**

\[ \frac{\partial |\Sigma_1|^{a_1} |\Sigma_2|^{a_2}}{\partial \sigma_{ij}^{(\alpha)}} = a_\alpha |\Sigma_1|^{a_1} |\Sigma_2|^{a_2} \lambda_{ij}^{(\alpha)}, \quad \text{tr}(\Sigma_{\alpha} H_{\alpha})^2 = k a_\alpha^2 |\Sigma_1|^{2a_1} |\Sigma_2|^{2a_2} \]

\[ \frac{\partial (a_1 |\Sigma_1|^{-1} + a_2 |\Sigma_2|^{-1})}{\partial \sigma_{ij}^{(\alpha)}} = a_\alpha |\Sigma_\alpha|^{-1} \lambda_{ij}^{(\alpha)}, \quad \text{tr}(\Sigma_{\alpha} H_{\alpha})^2 = k a_\alpha^2 |\Sigma_\alpha|^2, \]

**A mixed case:**

\[ \frac{\partial \text{ tr}(A_1 \Sigma_2)}{\partial \sigma_{ij}^{(1)}} = -(A_1 \Sigma_2 A_1)_{ij}, \quad \text{tr}(\Sigma_{\alpha} H_{\alpha})^2 = \text{tr}(A_{\alpha} A_{\alpha})^2, \]

\[ \frac{\partial \text{ tr}(A_1 \Sigma_2)}{\partial \sigma_{ij}^{(2)}} = (A_1)_{ij}, \quad \text{tr}(\Sigma_{\alpha} H_{\alpha})^2 = \text{tr}(A_{\alpha} A_{\alpha})^2. \]

It then follows that
\( (5.9) \quad \mathcal{L} \left( \frac{\text{tr}(A_1 V_1^{-1} + A_2 V_2^{-1}) - \text{tr}(A_1 \Sigma_1^{-1} + A_2 \Sigma_2^{-1})}{\sqrt{2\left[\frac{1}{n_1} \text{tr}(A_1 V_1^{-1})^2 + \frac{1}{n_2} \text{tr}(A_2 V_2^{-1})^2\right]}} \right) \to N(0,1), \)

\( (5.10) \quad \mathcal{L} \left( \frac{\text{tr}(A_1 S_1 + A_2 S_2) - \text{tr}(A_1 \Sigma_1 + A_2 \Sigma_2)}{\sqrt{2\left[\frac{1}{n_1} \text{tr}(A_1 V_1)^2 + \frac{1}{n_2} \text{tr}(A_2 V_2)^2\right]}} \right) \to N(0,1), \)

\( (5.11) \quad \mathcal{L} \left( \frac{|V_1|^a_1 |V_2|^a_2 - |\Sigma_1|^a_1 |\Sigma_2|^a_2}{|\Sigma_1|^a_1 |\Sigma_2|^a_2 \sqrt{2k a_1^2 + a_2^2 \left(\frac{a_1}{n_1} + \frac{a_2}{n_2}\right)}} \right) \to N(0,1), \)

\( (5.12) \quad \mathcal{L} \left( \frac{(a_1 |S_1^{-1}| + a_2 |S_2^{-1}|) - (a_1 |\Sigma_1^{-1}| + a_2 |\Sigma_2^{-1}|)}{\sqrt{2k \left(\frac{a_1^2}{n_1 |V_1|^2} + \frac{a_2^2}{n_2 |V_2|^2}\right)}} \right) \to N(0,1), \)

\( (5.13) \quad \mathcal{L} \left( \frac{\text{tr} V_1^{-1} V_2 - \text{tr} \Sigma_1^{-1} \Sigma_2}{\sqrt{2\left[\frac{1}{n_1} + \frac{1}{n_2}\right] \text{tr}(V_1^{-1} V_2)^2}} \right) \to N(0,1). \)

Approximate confidence regions for the various functions

\( h(\Sigma_1, \Sigma_2) \)

may be obtained by using these results as outlined in Section 4.1.

5.2. **Confidence intervals for** \( \text{tr}(A_1 \Lambda_1 + A_2 \Lambda_2) \) and \( \text{tr}(A_1 \Sigma_1 + A_2 \Sigma_2). \)

By a straightforward analog of Sections 4.2, 4.3, we obtain

the following results.

The interval
\[
\left[ \frac{2}{\sum_j n_j \text{tr} A_j (S_j + \xi_1 A_j)^{-1}}, \frac{2}{\sum_j n_j \text{tr} A_j (S_j + \xi_2 A_j)^{-1}} \right]
\]

is a confidence interval for \(\text{tr}(A_1 \Lambda_1 + A_2 \Lambda_2)\) with confidence coefficient 1-\(\alpha\), where \(\xi_1\) and \(\xi_2\) are the two roots of

\[
g(\xi) = \sum_1^2 \left( n_j \log |I + \xi S_j^{-1} A_j| + n_j \text{tr}(I + \xi S_j^{-1} A_j)^{-1} - k_n \right) = \chi^2_\alpha(1).
\]

The interval

\[
\frac{2}{\sum_j n_j / 2} \text{tr} C_j \left[ I + (I - \xi_1 C_j)^{1/2} \right], \frac{2}{\sum_j n_j / 2} \text{tr} C_j \left[ I + (I - \xi_2 C_j)^{1/2} \right],
\]

where \(C_j = 4 S_j^{1/2} A_j S_j^{-1} A_j / n_j^2\), is a confidence interval for \(\text{tr}(A_1 \Sigma_1 + A_2 \Sigma_2)\) with confidence coefficient 1-\(\alpha\), where \(\xi_1\) and \(\xi_2\) are the two roots of

\[
g(\xi) = \sum_1^2 \left( k_n \left( \log 2 - \frac{1}{2} \right) - n_j \log |I + (1 - \xi C_j)^{1/2}| + n_j \frac{1}{2} \text{tr}(I - \xi C_j)^{1/2} \right) = \chi^2_\alpha(1).
\]

5.3. Confidence interval for \(|\Sigma_1|^{a_1} |\Sigma_2|^{a_2}\). Since we can obtain a confidence interval for \(|\Sigma_1|^{a_1} |\Sigma_2|^{a_2}\) from a confidence interval for \(a_1 \log |\Lambda_1| + a_2 \log |\Lambda_2|\) by exponentiation, we consider as in Section 3.2, \(\omega = \{\Lambda_1, \Lambda_2 : \Lambda_1 > 0, \Lambda_2 > 0, a_1 \log |\Lambda_1| + a_2 \log |\Lambda_2| = \rho\}\) and obtain

\[
\sup_{\omega} p(S_1, S_2, \Sigma_1, \Sigma_2).
\]

The appropriate Lagrangian is
\[ M(\Lambda_1, \Lambda_2) = \sum_{\alpha=1}^2 (n_\alpha \log |\Lambda_\alpha| - \text{tr} \, \Lambda_\alpha S_\alpha) - \xi (a_1 \log |\Lambda_1| + a_2 \log |\Lambda_2| - \rho). \]

As in Section 4.4, the MLE for \( \Lambda_\alpha \) in \( \omega \) is

\[ \hat{\Lambda}_\alpha = (n_\alpha - a_\alpha \xi) S_\alpha^{-1}, \]

where \( \xi \) is such that \( n_\alpha - a_\alpha \xi > 0 \) for \( \alpha = 1, 2 \). In contrast to the case of one covariance matrix, we cannot in general solve the constraint equation for \( \xi \) in terms of \( \rho \). It is easily verified that

\[ g(\xi) = -\sum_{\alpha=1}^2 k a_\alpha \xi + kn_\alpha \log \frac{n_\alpha - a_\alpha \xi}{n_\alpha}, \]

and

\[ h(\hat{\Sigma}_1(\xi), \hat{\Sigma}_2(\xi)) = \sum_{\alpha=1}^2 a_\alpha \log |S_\alpha^{-1}(n_\alpha - a_\alpha \xi)|. \]

Since \( h(\hat{\Sigma}_1(\xi), \hat{\Sigma}_2(\xi)) \) is a monotonically decreasing function of \( \xi \), it follows from Lemma 2.1, that \( g(\xi) \) is monotone increasing for \( \xi > 0 \) and monotone decreasing for \( \xi < 0 \).

5.4. Confidence Interval for \( a_1 |\Lambda_1| + a_2 |\Lambda_2| \). To find

\[ \sup_{\omega} \mathcal{P}(S_1, S_2; \Lambda_1, \Lambda_2), \]

where \( \omega = \{ \Lambda_1, \Lambda_2; a_1 |\Lambda_1| + a_2 |\Lambda_2| = \rho \} \), we use the Lagrangian

\[ M(\Lambda_1, \Lambda_2) = \frac{2}{\sum_{j=1}^2 (n_j \log |\Lambda_j| - \text{tr} \, S_j \Lambda_j) - \xi (a_1 |\Lambda_1| + a_2 |\Lambda_2| - \rho). \]
Setting the derivatives of \( M(\lambda_1, \lambda_2) \) with respect to the elements of \( \lambda_1 \) and \( \lambda_2 \) equal to zero yields the matrix equations

\[
n_j \Sigma_j - S_j - \xi a_j |\Lambda_j| \Sigma_j = 0, \quad j = 1, 2.
\]

Thus

\( (n_j - \xi a_j |\Lambda_j|)I = \Lambda_j S_j, \)

so that

\( (\text{tr} \ \Lambda_j S_j)/k = n_j - \xi a_j |\Lambda_j|, \)

and

\( |\Lambda_j S_j|^{1/k} = n_j - \xi a_j |\Lambda_j|, \quad j = 1, 2. \)

But \( \text{tr} \ \Lambda_j S_j/k \) is the arithmetic mean and \( |\Lambda_j S_j|^{1/k} \) is the geometric mean of the characteristic roots of \( \Lambda_j^{1/2} S_j \Lambda_j^{1/2} \). These two means are equal if and only if all the characteristic roots are equal to a constant, say \( c_j \). Thus

\[
\Lambda_j^{1/2} S_j \Lambda_j^{1/2} = c_j I,
\]

so that \( \hat{\Lambda}_j = c_j S_j^{-1}, \ j = 1, 2. \) Since

\[
|\hat{\Lambda}_j S_j| = c_j^k = (n_j - \xi a_j |\hat{\Lambda}_j|)^k = (n_j - \xi a_j c_j |S_j^{-1}|)^k,
\]

we see that \( c_j \) must be a positive root of the equation

\( \xi a_j |S_j^{-1}| x^k + x - n_j = 0, \quad j = 1, 2, \)

or alternatively, \( \hat{a}_j = 1/c_j \) must be a positive root of the equation
$$\varphi_j(y) = n_j y^k - y^{k-1} - \xi b_j = 0, \quad j = 1, 2,$$

where $b_j = a_j |s_j^{-1}|$.

We next show that when there are multiple positive roots, the smallest root $(c_j)$ is the maximizer of $M(\hat{\lambda}_1, \hat{\lambda}_2)$.

When $\xi b_j > 0$, there is exactly one positive root of $\varphi_j(y) = 0$. To see this note that $\varphi_j(0) = -\xi b_j < 0$ and $\lim_{y \to \infty} \varphi_j(y) = \infty$ guarantees the existence of at least one positive root, and by Descartes rule of signs there is at most one positive root. Also, from $\varphi_j(1/x) = 0$, we see that $c_j < n_j$, a fact which will be needed later.

When $\xi b_j < 0$, by Descartes rule of signs there exist at most two positive roots of $\varphi_j(y) = 0$. When $k$ is even, there must be at least one positive root, for if $y < 0$, $n_j y^k - y^{k-1} - \xi b_j > 0$.

However, when $k$ is odd, it may happen that there are no positive roots of $\varphi_j(y) = 0$ (e.g., if $k = 1$, $\varphi_j(y) = n_j y - 1 - \xi b_j = 0$, implies that $y = (1 + \xi b_j)/n$, which may be negative for arbitrarily $\xi$).

This imposes a constraint on $\xi$, namely that in case $k$ is odd and $\xi b_j < 0$, $\xi$ must be such that there exists at least one positive root of $\varphi_j(y) = 0$. This constraint is equivalent to

$$\xi b_j \geq -\frac{1}{k} \left( \frac{k-1}{nk} \right)^{k-1}.$$

To see this note that $\varphi_j'(y) = 0$ at $y = (k-1)/nk$, and we require that $\varphi_j((k-1)/nk) \leq 0$. Thus when $\xi b_j < 0$, there is at least one, and at most two, positive roots of $\varphi_j(y) = 0$.

In the event there are two positive roots, we show that the larger root is the maximizer of
\[ \psi_j(y) = n_j \log |\Lambda_j| - \text{tr} \; S_j \Lambda_j \]

\[ = -n_j \log |S_j| - kn_j \log y - k/y. \]

Then

\[ \psi_j(y) = \frac{-n_j k}{y} + \frac{k}{y^2}, \]

which is positive if and only if \( y < 1/n_j \). But \( \varphi_j(1/n_j) = -\xi b_j > 0 \), \( \varphi_j'(1/n_j) = 1/n_j^{k-1} > 0 \), so that all positive roots of \( \varphi_j(y) \) are smaller than \( 1/n_j \). Consequently, \( \psi_j(y) > 0 \) and the larger of the two positive roots of \( \varphi_j(y) \) is to be used as \( d_j \), or equivalently, \( c_j \) is the smallest positive root of \( \varphi_j(1/x) = 0 \). Further, note that since \( d_j \) is the largest root of \( \varphi_j(y) \), and \( \lim_{y \to \infty} \varphi_j(y) = \infty \),

\[ \varphi_j'(d_j) = [n_j k d_j - (k-1)] d_j^{k-2} > 0, \]

so that \( c_j = 1/d_j < n_j k/(k-1) \).

We now return to a consideration of the function

\[ g(\xi) = k \sum_{l=1}^{2} (n_j \log n_j - n_j) - k \sum_{l=1}^{2} (n_j \log c_j - c_j). \]

Then

\[ \frac{dg}{d\xi} = k \sum_{l=1}^{2} \frac{(n_j - c_j)}{c_j} \frac{d c_j}{d\xi} = k \sum_{l=1}^{2} (\xi b_j c^{k-1} \frac{b_j c^{k-1}}{c_j} \frac{b_j c^{k-1}}{1 + \xi b_j c^{k-1}}) \]

\[ = k \xi \sum_{l=1}^{2} \frac{b_j^2 c^{2k-1}}{1 + \xi b_j c^{k-1}} = k \xi \sum_{l=1}^{2} \frac{b_j^2 c^{2k}}{(kn - c(k-1))}. \]

Recalling that \( c_j < n_j \) when \( \xi b_j > 0 \) and \( c_j < kn_j/(k-1) \) when \( \xi b_j < 0 \), it is immediate that \( g(\xi) \) is increasing when \( \xi > 0 \) and decreasing when \( \xi < 0 \).
5.5 Confidence Region for \( \text{tr} \, \Sigma_1^{-1} \Sigma_2 \).

For this problem, \( \omega = \{ \Lambda_1, \Lambda_2: \Lambda_1 > 0, \Lambda_2 > 0, \text{tr} \, \Lambda_1 \Lambda_2^{-1} = \rho \} \), and we use the Lagrangian

\[
M(\Lambda_1, \Lambda_2) = n \log |\Lambda_1| + m \log |\Lambda_2| - \text{tr} \, \Lambda_1 S_1 - \text{tr} \, \Lambda_2 S_2 \\
- \xi (\text{tr} \, \Lambda_1 \Lambda_2^{-1} - \rho).
\]

Differentiating with respect to the elements of \( \Lambda_1 \) and \( \Lambda_2 \), and setting the derivatives equal to zero yields the equations

\[
(5.14) \quad n \Sigma_1 - S_1 - \xi \Sigma_2 = 0,
\]

\[
(5.15) \quad m \Sigma_2 - S_2 + \xi \Sigma_2 \Lambda_1 \Sigma_2 = 0.
\]

Pre- and post-multiplying (5.14) by \( \Lambda_1^{1/2} \), (5.15) by \( \Lambda_2^{1/2} \), taking traces and adding yields,

\[
(n + m) k = \text{tr} \, (\Lambda_1 S_1 + \Lambda_2 S_2).
\]

From (5.14), we have \( \hat{\Sigma}_2(\xi) = \left( n \hat{\Sigma}_1(\xi) - S_1 \right) / \xi \), which together with (5.15) yields

\[
S_1 \Lambda_1 S_1 + n(m + n) \Sigma_1 - (m + 2n) S_1 - \xi S_2 = 0,
\]

37
or equivalently,

\[(5.16) \quad \frac{s_1^{1/2} \Lambda_1 s_1^{1/2}}{\sqrt{n(m+n)}} + \frac{\sqrt{n(m+n)}}{s_1^{-1/2} \Sigma_1 s_1^{-1/2}} - \frac{(m+2n)I + \xi s_1^{-1/2} s_2 s_1^{-1/2}}{\sqrt{n(m+n)}} = 0.\]

Letting

\[(5.17) \quad \psi = \frac{s_1^{1/2} \Lambda_1 s_1^{1/2}}{\sqrt{n(m+n)}}, \quad T = \frac{(m+2n)I + \xi s_1^{-1/2} s_2 s_1^{-1/2}}{2\sqrt{n(m+n)}},\]

\[(5.16)\] becomes

\[(5.18) \quad \psi + \psi^{-1} = 2T,\]

where \(\xi\) is such that \(T > I\). Because \(\psi + \psi^{-1}\) commutes with \(\psi\), \(T\) must commute with \(\psi\), so that \(\psi^2 - 2T\psi + I = 0\). Hence

\[(\psi - T)^2 = T^2 - I > 0,\]

and

\[(5.19) \quad \psi = T + (T^2 - I)^{1/2},\]
where \((T^2 - I)^{1/2}\) is any symmetric square root of \(T^2 - I\).

We now determine which square root is the maximizer of

\[
M^*(\hat{A}_1, \hat{A}_2) = n \log |\hat{A}_1| + m \log |\hat{A}_2| - \text{tr} \hat{A}_1 S_1 - \text{tr} \hat{A}_2 S_2.
\]

By virtue of (5.16), and the fact that

\[
|\hat{\Sigma}_2| = |S_1| |\psi^{-1}| |(v\sqrt{n/(n+m)} I - \psi)/\xi|,
\]

\[
M^*(\hat{A}_1, \hat{A}_2) = (n + m) \log |\psi| - m \log |(v\sqrt{n/(n+m)} I - \psi)/\xi|
\]

\[
- k (n + m) + \frac{km}{2} \log n (n + m) - (n + m) \log |S_1|.
\]

Consequently, \(M^*(\hat{A}_1, \hat{A}_2)\) is a function of the characteristic roots of \(\psi\). If \(t_1, \ldots, t_k\) denote the roots of \(T\), then \(\epsilon_j(t_j^2 - 1)\) are the roots of any symmetric square root, \((T^2 - I)^{1/2}\), where \(\epsilon_j = \pm 1, j = 1, \ldots, k\). Hence the roots \(\psi = T + (T^2 - I)^{1/2}\) are

\[
t_j + \epsilon_j \sqrt{t_j^2 - 1}.
\]

The determination of which of the \(2^k\) square roots is the maximizer of \(M^*(\hat{A}_1, \hat{A}_2)\) depends on the maximization of \(M^*(\hat{A}_1, \hat{A}_2)\) with respect to \(\epsilon_j\); i.e., of

\[
(m + n) \sum_{1}^{k} \log (t_j + \epsilon_j \sqrt{t_j^2 - 1})
\]

\[
- m \sum_{1}^{k} \log \left(\frac{\sqrt{n}}{n + m} - t_j - \epsilon_j \sqrt{t_j^2 - 1}\right)
\]
We now show that the maximizer is achieved for \( \epsilon_1 = \ldots = \epsilon_k = 1 \).

Let \( \alpha = n/(n+m) \), \( \bar{\alpha} = 1 - \alpha \), and note that \( \sqrt{\alpha} > t + \epsilon \sqrt{t^2 - 1} > 0 \).

Consider the function

\[
\begin{align*}
    s(t) &= s_+(t) - s_-(t) \\
    &= \left\{ \log [t + \sqrt{t^2 - 1}] - \bar{\alpha} \log [\sqrt{\alpha} - t - \sqrt{t^2 - 1}] \right\} \\
    &\quad - \left\{ \log [t - \sqrt{t^2 - 1}] - \bar{\alpha} \log [\sqrt{\alpha} - t + \sqrt{t^2 - 1}] \right\},
\end{align*}
\]

then \( s(1) = 0 \). A direct computation shows that \( ds/dt > 0 \) for

\[0 < t + \sqrt{t^2 - 1} < \sqrt{\alpha},\]

from which the result follows.

From (5.14), (5.15) and (5.17), note that \( |\Lambda_2 S_2| \)

\[= |(n + m) I - \Lambda_1 S_1| = |(n + m) I - \sqrt{n(m + n)} \psi|,
\]

so that

\[
g(\xi) = -n \log |\psi| - m \log |I - \psi \sqrt{n/(n + m)}|
\]

\[= -mk \log (n + m) + nk \log n + mk \log m.
\]

Recall that the roots of \( \psi \) are \( \psi_j = t_j + \sqrt{t_j^2 - 1} \), \( j = 1, \ldots, p \).

Then

\[
\frac{\partial g}{\partial \xi} = -\sum \left( \frac{n}{\psi_j} + \frac{m}{1 - \psi_j \sqrt{n/m}} \right) \frac{\partial \psi_j}{\partial \xi}.
\]

But \( \frac{\partial \psi_j}{\partial \xi} = \frac{\psi_j}{\sqrt{t_j^2 - 1}} > 0 \), and hence \( g(\xi) \) is a decreasing function.
6. Confidence interval for $\mu \Sigma^{-1} \mu'$. Let $x$ be a $k$-dimensional random vector having a multivariate normal distribution, $N(\mu, \Sigma/N)$, and let $S = \mathbf{V}_n$ be an independent $k \times k$ random symmetric matrix having a Wishart distribution $w(\Sigma, k, N-1)$. Although $x S^{-1} x'$ (except for a constant) has a non-central $F$-distribution with non-centrality parameter $(\mu \Sigma^{-1} \mu')^{1/2}$, it is not clear how to use this exact result to obtain a confidence interval for $\mu \Sigma^{-1} \mu'$.

The method of linearization is based on

$$L \left( \frac{\mathbb{E}(x \Sigma^{-1} x' - \mu \Sigma^{-1} \mu')}{\sqrt{\frac{2}{n} [(x \Sigma^{-1} x')^2 + (x \Sigma^{-1} x')]}}, \mu \Sigma^{-1} \mu' \right) \rightarrow N(0,1),$$

and yields an approximate confidence interval for $\mu \Sigma^{-1} \mu'$.

We now obtain an approximate confidence interval for $\mu \Sigma^{-1} \mu'$ using the method of Section 2. From the joint distribution of $x$ and $S$ given by (2.11) and (2.12) and $\Omega = \{\mu, \Sigma : \Sigma > 0, -\infty < \mu_i < \infty, i = 1, \ldots, k\}$, we obtain $\hat{\Sigma} = S/N, \hat{\mu} = x$, so that

$$\sup_{\Omega} p(x, S; \mu, \Sigma) = C \frac{c(k, N-1) |S|^{-(k+2)/2} N^{kN/2} e^{-kN/2}}{N}.$$

To find $\sup_{\omega} p(x, S; \mu, \Sigma)$, where $\omega = (\mu, \Sigma; \Sigma > 0, \mu \Sigma^{-1} \mu' = \rho)$, we use the Lagrangian

$$M(\mu, \Lambda) = N \log |\Lambda| - \text{tr} \Lambda S - N(x - \mu) \Lambda (x - \mu)' - \xi(\mu \Lambda \mu' - \rho),$$

where $\Lambda = \Sigma^{-1}$. Differentiating $M(\mu, \Lambda)$ with respect to the $\mu_i$ and $\lambda_{ij}$ and setting the derivatives equal to zero yields the equations
\[ N(x-\mu)^2 - \xi \mu \lambda = 0 , \]

\[ N \Sigma - S - N(x-\mu)'(x-\mu) - \xi \mu'\mu = 0 , \]

so that

\[(6.1) \quad \hat{\mu}(\xi) = \frac{N}{N+\xi} x , \]

\[(6.2) \quad \hat{\Sigma}(\xi) = \frac{1}{N} (S + \frac{N\xi}{N+\xi} x'x) . \]

Consequently,

\[
\sup_{\omega} p(x,S; \mu,\Sigma) = C \ c(k,n) |S|^{-(k+2)/2} N^{kN/2} e^{-kN/2}
\times (1 + \frac{N\xi}{N+\xi} x \ S^{-1} x')^{-N/2} \exp \left[ \frac{-N^2 \xi \ xS^{-1} x'}{2(N+\xi) (N+\xi+N\xi \ xS^{-1} x')} \right].
\]

Thus

\[
g(\xi) = -2 \log \frac{p_{\omega}}{p_\Omega} = N \log (1 + \frac{N\xi}{N+\xi} xS^{-1} x') - \frac{N^2 \xi \ xS^{-1} x'}{(N+\xi)(N+\xi+N\xi \ xS^{-1} x')} .
\]

The function \( g(\xi) \) is monotone increasing in \((-\infty , -1) \) with \( g(-\infty) = N \log (1 + N xS^{-1} x') \) and \( g(-1) = \infty ; \) \( g(\xi) \) is undefined on \([1, -1/1+q] \); and \( g(\xi) \) is bowl-shaped in \((-1/1+q, \infty) \), monotone decreasing in \((-1/1+q, 0) \) and increasing in \((0,\infty) \) with \( g(-1/1+q) = \infty , \)

\( g(0) = 0, \ g(\infty) = N \log(1 + N xS^{-1} x') . \)
Consequently, the equation \( g(\xi) = \text{constant} \) has two solutions. If \( \chi^2(1) < N \log(1 + N xS^{-1} x') \) we obtain an interval about zero, and if \( \chi^2(1) > N \log(1 + N xS^{-1} x') \) we obtain two semi-infinite intervals.

7. Behrens-Fisher Model. Here we study the multivariate Behrens-Fisher problem, which in its canonical form may be expressed as follows. Let \( x \) and \( y \) be random vectors having \( k \)-variate normal distributions with \( \mathcal{L}(x) = N(\mu_1, \Sigma_1) \), \( \mathcal{L}(y) = N(\mu_2, \Sigma_2) \); let \( S_1 \) and \( S_2 \) be random \( k \times k \) matrices having Wishart distributions \( \mathcal{L}(S_1) = W(\Sigma_1; k, n) \), \( \mathcal{L}(S_2) = W(\Sigma_2; k, m) \), where \( x, y, S_1, S_2 \) are independent. The joint density is given by

\[
p(x, y, S_1, S_2; \mu, \nu, \Lambda_1, \Lambda_2) = C |\Lambda_1|^{N/2} |\Lambda_2|^{M/2}
\exp\left(-\frac{1}{2} \left[ (x-\mu)^\top \Lambda_1 (x-\mu)^\top + (y-\nu)^\top \Lambda_2 (y-\nu)^\top + \text{tr} \Lambda_1 S_1 + \text{tr} \Lambda_2 S_2 \right]\right),
\]

where \( \Lambda_1 = \Sigma_1^{-1} \), \( \Lambda_2 = \Sigma_2^{-1} \), \( N = n+1 \), \( M = m+1 \). The problem is to find a confidence region for \( \mu-\nu = (\mu_1-\nu_1, \ldots, \mu_k-\nu_k) \).

For \( \Omega = \{\mu, \nu, \Lambda_1, \Lambda_2 : -\infty < \mu_i < \infty, -\infty < \nu_i < \infty, i = 1, \ldots, k, \Lambda_1 > 0, \Lambda_2 > 0\} \), a straightforward calculation yields

\[
\sup_{\Omega} p(x, y, S_1, S_2; \mu, \nu, \Lambda_1, \Lambda_2) = C \frac{|S_1|^{-N/2} |S_2|^{-M/2}}{|N|^{N/2} |M|^{M/2}} \exp\left[-\frac{1}{2} k(N+M)\right].
\]

In \( \omega = \{\mu, \nu, \Lambda_1, \Lambda_2 : \mu-\nu = \rho, \Lambda_1 > 0, \Lambda_2 > 0\} \), we consider the Lagrangian
\begin{align}
N(\mu, \nu, \Lambda_1, \Lambda_2) &= N \log|\Lambda_1| + M \log|\Lambda_2| - (x-\mu)' \Lambda_1 (x-\mu)', \\
- (y-\mu)' \Lambda_2 (y-\nu)' - \text{tr} \Lambda_1 S_1 - \text{tr} \Lambda_2 S_2 - 2\xi[(\mu-\nu) - \rho]', \\
\end{align}

where now $\xi = (\xi_1, \ldots, \xi_k)$. Equating to zero the derivatives of (7.1) with respect to the elements of $\mu, \nu, \Lambda_1, \Lambda_2$ yields the equations

\begin{align}
(7.2) & \quad (x-\mu)' \Lambda_1 = \xi, \\
(7.3) & \quad (y-\nu)' \Lambda_2 = -\xi, \\
(7.4) & \quad NE_1 = S_1 + (x-\mu)' (x-\mu), \\
(7.5) & \quad ME_2 = S_2 + (y-\nu)' (y-\nu).
\end{align}

These equations need not be solved explicitly since we only need the quantities $|\Lambda_1(\xi)|, |\Lambda_2(\xi)|, \text{tr} \Lambda_1(\xi) [S_1 + (x-\mu(\xi))' (x-\mu(\xi))], \text{and tr} \Lambda_2(\xi) [S_2 + (y-\nu(\xi))' (y-\nu(\xi))], \text{in order to compute sup}_{\omega} \log p(x, y, S_1, S_2, \mu, \nu, \Lambda_1, \Lambda_2) as a function of $\xi$.}

Premultiplication of (7.4) and (7.5) by $\Lambda_1$ and $\Lambda_2$, respectively, and taking the trace of both sides yields

\[ k(N+M) = \text{tr}[\Lambda_1[S_1 + (x-\mu)' (x-\mu)] + \Lambda_2[S_2 + (y-\nu)' (y-\nu)]], \]

where the dependence on $\xi$ is omitted.
From (7.4),

$$|\hat{S}_1| N^k = |S_1| \left[ 1 + (x-\hat{\mu}) S_1^{-1}(x-\hat{\mu})' \right],$$

$$\frac{1}{N} (x-\hat{\mu}) \Lambda_1(x-\hat{\mu})' = \frac{(x-\hat{\mu}) S_1^{-1}(x-\hat{\mu})'}{1 + (x-\hat{\mu}) S_1^{-1}(x-\hat{\mu})'},$$

and from (7.2)

$$(x-\hat{\mu}) \Lambda_1(x-\hat{\mu})' = \xi(x-\mu)' .$$

Multiplication of the left-hand and right hand sides of (7.2) and (7.4) respectively yields

$$N(x-\hat{\mu}) = \xi S_1 + \xi(x-\hat{\mu})' (x-\hat{\mu}) .$$

Post-multiplication by $\xi$ yields a quadratic in $\xi(x-\hat{\mu})'$ the solution of which is

$$\xi(x-\hat{\mu})' = \frac{1}{2} \left[ N \pm (N^2 - 4 \xi S_1 \xi')^{1/2} \right] ,$$

so that

$$\left[ 1 + (x-\hat{\mu}) S_1^{-1}(x-\hat{\mu})' \right]^{-1} = \left[ N \pm (N^2 - 4 \xi S_1 \xi')^{1/2} \right] / 2N .$$

A simple direct calculation shows that the positive sign is the maximizer, so that
$$\sup_{\omega} p(x, y, S_1, S_2; \mu, \nu, \Lambda_1, \Lambda_2) = C \frac{|S_1|^{-N/2}}{|M|^{-M/2}} e^{-1/2} k(N+M)$$

$$\times \frac{[N + (N^2 - 4\xi S_1 S_1')^{1/2}]^{N/2}}{(2N)^{N/2}} \frac{[M + (M^2 - 4\xi S_2 S_2')^{1/2}]^{M/2}}{(2M)^{M/2}}.$$ 

Consequently,

$$g(\xi_1, \ldots, \xi_k) = -2 \log \frac{P_0}{P_{\Omega}} = N \log 2N + M \log 2M$$

$$- N \log [N + (N^2 - 4\xi S_1 S_1')^{1/2}] - M \log [M + (M^2 - 4\xi S_2 S_2')^{1/2}].$$

Finally, we note that $\xi S_j \xi'$ is a convex function of $\xi$ for $j = 1, 2$, and as both the square-root and logarithm are monotone increasing and concave, it follows that $-g(\xi)$ is concave, so that $g(\xi)$ is a convex function of $\xi$. 
REFERENCES


Gross, O. A. and Johnson, J. M. (1959), Sequential minimax search for a zero of a convex function, M.T.A.C.


Madansky, A. (1965), Approximate confidence limits for the reliability of series and parallel systems, Technometrics, 7, 495-503.


Wald, A. (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, Trans. Amer. Math. Soc., 54, 426-482.