A K-SAMPLE REGRESSION MODEL WITH COVARIANCE

BY

LEON GLESE and INGRAM OLKIN

TECHNICAL REPORT NO. 17
February 26, 1965
Revised July 23, 1965

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-3837

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A K-SAMPLE REGRESSION MODEL
WITH COVARIANCE

By

Leon Gleser*
Columbia University

Ingram Olkin
Stanford University

TECHNICAL REPORT No. 17
February 26, 1965
Revised July 23, 1965

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-3837

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

* The research of Leon Gleser was sponsored in part by the Office of Naval Research at Columbia University; this report is also being issued by Columbia University.
1. **Introduction and Summary**

Let \( y^{(1)}, \ldots, y^{(k)} \) be \( k \) independent random \( p \)-dimensional row vectors, each having a multivariate normal distribution with common covariance matrix \( \Psi \) and mean vector \( \mathbf{\gamma}^{(i)} = \mathbf{\beta}^{(i)} \mathbf{X} \), where \( \mathbf{\beta}^{(i)} \) is a \( q \)-dimensional row vector and \( \mathbf{X} \) is a known \( q \times p \) matrix of rank \( q \leq p \), \( i = 1, \ldots, k \). We observe \( N_i \) independent replications of the vector \( y^{(i)} \), \( i = 1, \ldots, k \).

The problems of concern in the paper are: (i) to test the hypothesis \( H: \mathbf{\beta}^{(1)} = \mathbf{\beta}^{(2)} = \ldots = \mathbf{\beta}^{(k)} = \mathbf{\beta} \), and (ii) to find maximum likelihood estimates (MLE) of the \( \mathbf{\beta}^{(i)} \)'s and of \( \Psi \), and simultaneous confidence bounds for \( \mathbf{\beta}^{(1)}, \ldots, \mathbf{\beta}^{(k)} \). Associated with each of the above problems is a related distributional problem. These distributional problems are extremely untractable but permit representations in integral form.

A chronology of previous papers is as follows. For the case \( k = 1 \), the regression model was considered independently by Rao [10] and by Cochran and Bliss [2]. These authors obtained a conditional likelihood ratio test of \( H: \mathbf{\beta} = \mathbf{\beta}_0 \) and the null distribution of that test. Rao [12] obtained conditional and unconditional distributions under general alternatives to \( H \); the unconditional distribution was also derived independently by Narain [7]. More recently, Kabe [5] has given an alternative derivation of the non-null distribution.

We note that a conditional likelihood ratio test need not be the same as the unconditional likelihood ratio test. In some cases the conditional and unconditional likelihood ratio tests are the same, but a proof of this fact is always necessary. Such a proof can be given for
this case. Alternatively, one could begin \textit{ab initio} and directly obtain the unconditional likelihood ratio test. This was done by Olkin and Shrikhande\cite{8}, who, in addition, obtained the null and non-null distributions of the likelihood ratio test statistic.

Properties of the likelihood ratio test of a closely related problem were considered by Giri\cite{3}. His results show that the likelihood ratio test for $H: \beta = \beta_0$ is UMP similar invariant. Kiefer and Schwartz\cite{6} prove that the likelihood ratio test is admissible.

Rao\cite{15} considered the $k$-sample regression model for $k > 1$ in connection with its application to growth curves. Implicit in his discussion is the conditional likelihood ratio test for $H: \beta^{(1)} = \beta^{(2)} = \ldots = \beta^{(k)} = \beta$. Both he and Potthoff and Roy\cite{9} have discussed a more general regression model and testing problem than the one we consider in this paper, but none of these authors have explicitly derived the unconditional likelihood ratio test. (Actually the testing problem that linear combinations of the $\beta$'s are equal to given constants is equivalent to the Potthoff-Roy model. The inter-relation of various such models will be given by the authors in a later paper.) Potthoff and Roy give certain ad hoc procedures, while Rao has indicated that he uses the conditional likelihood ratio test. As in the case $k = 1$, we treat our special case of the more general Potthoff-Roy model by deriving the unconditional likelihood ratio test.

From the point of view of estimation, the papers by Rao\cite{10},\cite{11},\cite{12},\cite{14} deal with various conditional and unconditional confidence bounds for both $\beta$ and linear combinations of $\beta$ in the case $k = 1$. Gieser and Olkin\cite{4} compare certain confidence procedures,
and further obtain the distributions of \( \hat{\beta} \) and of \( \hat{\psi} \) (the maximum likelihood estimates of \( \beta \) and \( \psi \), respectively). The present paper extends some of these results to the case \( k > 1 \).

In Section 2 we reduce the model and problems (i) and (ii) to a canonical form. Section 3 contains the likelihood ratio test for problem (i), the distribution of this test under \( H \), and the asymptotic distribution of this test under the alternatives to \( H \). Finally, Section 4 deals with MLE of the \( \beta^{(i)} \)'s and \( \psi \), their distributions, and related confidence bounds for \( \beta^{(1)}, \ldots, \beta^{(k)} \).

Before beginning our discussion, we call the reader's attention to the following notational conventions: Small Latin letters denote vectors; capital Latin letters denote matrices. Vectors in general are row vectors. In order to partition a vector we often use a dot notation, i.e., \( z = (\hat{z}, \tilde{z}) \). The notation \( Z: p \times q \) means that the matrix \( Z \) has \( p \) rows and \( q \) columns; the notation \( Z > 0 \) means that the matrix \( Z \) is positive definite. For a random vector \( z, \mathcal{N}(\mu, \Sigma) \) means that \( z \) has a multivariate normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \). For a random \( p \times p \) matrix \( Z, \mathcal{W}(\Sigma; p, n) \) means that \( Z \) has a Wishart distribution, i.e.,

\[
p(Z) = c(p, n) |Z|^{n-p-1/2} |\Sigma|^{-n/2} e^{-\frac{1}{2} tr \Sigma^{-1} Z}, \quad Z > 0, \Sigma > 0,
\]

where

\[
[c(p, n)]^{-1} = 2^{pn/2} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(\frac{n-i+1}{2}).
\]

By \( \mathcal{B}(z) = \beta(a, b) \) we mean that the random variable \( z \) has a Beta distribution with parameters \( a \) and \( b \).
2. A canonical form.

We are given $N_j$ independent observations on $y^{(j)}$, say
$y_1^{(j)}, y_2^{(j)}, \ldots, y_{N_j}^{(j)}$, where $z(y_i^{(j)}) = \mathcal{N}(\beta^{(j)} X, \psi)$, $i = 1, \ldots, N_j$, $j = 1, \ldots, k$. We need only consider the sufficient statistic ($\bar{y}^{(1)}, \bar{y}^{(2)}, \ldots, \bar{y}^{(k)}, S$) where
$$\bar{y}^{(j)} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_i^{(j)}, \quad i = 1, \ldots, k,$$
and
$$S = \sum_{j=1}^{k} \frac{N_j}{\bar{y}^{(j)}} \left( y^{(j)} - \bar{y}^{(j)} \right), \quad \left( y^{(j)} - \bar{y}^{(j)} \right).$$
The $\bar{y}^{(j)}$ and $S$ are mutually independent with $z(\bar{y}^{(j)}) = \mathcal{N}(\beta^{(j)} X, \psi/N_j)$, $z(S) = \mathcal{W}(\psi; p, n)$, $j = 1, \ldots, k$, where $n = \sum_{j=1}^{k} N_j - k$ and $n \geq p$.

Since $X$ is assumed to be of full rank $q$, there exists a $p \times p$ orthogonal matrix $\Gamma$ and a $q \times q$ non-singular, lower-triangular matrix $T$ such that
$$X = T(I_q, 0) \Gamma.$$
Make the transformations $z^{(j)} = \bar{y}^{(j)} \Gamma'$, $j = 1, \ldots, k$, $V = \Gamma S \Gamma'$. Then the $z^{(j)}$ and $V$ are mutually independent,

$$(2.1) \quad z(z^{(j)}) = \mathcal{N}(\theta^{(j)}, \Sigma/N_j), \quad z(V) = \mathcal{W}(\Sigma; p, n),$$
where $\Sigma = \Gamma \Psi \Gamma'$, $\delta^{(j)} = (\delta^{(j)}(0), (\beta^{(j)} T, 0)$, and each $\delta^{(j)}$, $1 \times q$, $j = 1, \ldots, k$.

Since $\Gamma$ is a known orthogonal matrix independent of the parameters, $(z^{(1)}, \ldots, z^{(k)}, V)$ is a sufficient statistic for problems
(i) and (ii). Further, since \( \Gamma \) and \( T \) are non-singular and known, questions of estimation of \( (\beta(1), \ldots, \beta(k), \psi) \) reduce to questions of estimation of \( (\dot{\theta}(1), \ldots, \dot{\theta}(k), \Xi) \). Finally, testing \( H: \beta(1) = \beta(2) = \ldots = \beta(k) \equiv \beta \) is equivalent to testing \( H^* : \dot{\theta}(1) = \dot{\theta}(2) = \ldots = \dot{\theta}(k) \equiv \theta \). We have thus reduced the model and problems (i) and (ii) to a canonical form.

For notational convenience let

\[
Z = (\dot{Z}, \ddot{Z}) \equiv \begin{pmatrix} Z^{(1)} \\ \vdots \\ Z^{(k)} \end{pmatrix}, \quad \Theta = (\ddot{\theta}, 0) \equiv \begin{pmatrix} \theta^{(1)} \\ \vdots \\ \theta^{(k)} \end{pmatrix},
\]

where \( Z \) and \( \Theta \) are \( k \times p \), \( \dot{Z} \) and \( \dot{\Theta} \) are \( k \times q \). Similarly, let

\[
\bar{\beta} = \begin{pmatrix} \beta^{(1)} \\ \vdots \\ \beta^{(k)} \end{pmatrix}; \quad k \times q, \quad \bar{Y} = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(k)} \end{pmatrix}; \quad k \times p.
\]

Then the sufficient statistic for \( (\bar{\beta}, \bar{\psi}) \) is \( (\bar{Y}, \bar{Z}) \), the sufficient statistic for \( (\dot{\Theta}, \Xi) \) is \( (Z, V) \).

We might now ask for a group of transformations which leave the canonical model invariant. As in the case \( k = 1 \), one such group is the group \( A \) of non-singular \( p \times p \) matrices \( A \) of the form

\[
A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \quad A_{11} \equiv q \times q,
\]

with the transformation taking \( Z \) into \( ZA \), \( V \) into \( A'VA \). It is well known that if \( ZV^{-1}Z' = Z^*(V^*)^{-1}(Z^*)' \), then there exists a \( p \times p \) non-singular matrix \( B \) such that \( Z^* = ZB \), \( V^* = B'VB \). Also if \( \bar{Z} V_{22}^{-1} \bar{Z}' = \bar{Z}^* (V_{22}^*)^{-1} \bar{Z}^* \), where
\[ V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad V_{11}: q \times q, \]

and where \( V^* \) is similarly partitioned, then there exists a
\( p-q \times p-q \) non-singular matrix \( H \) such that 
\[
\begin{pmatrix} \bar{Z}^* \\ \bar{Z}^{*}_{22} \end{pmatrix} = H^* V_{22} H, \quad V_{22}^* = H^* V_{22} H.
\]

Combining these two facts, if the ordered pair
\[
(2.2) \quad U = (U_1, U_2) = (Z^{*-1}Z', \bar{Z}^{*}_{22} V_{22}^{-1} \bar{Z}'_2)
\]
equals the ordered pair \( U^* = (Z^*(V^*)^{-1} Z', \bar{Z}^* (V_{22}^*)^{-1} \bar{Z}'_2) \), then
there exists a matrix \( A \) in \( \mathcal{A} \) such that 
\[
Z^* = Z A, \quad V^* = A' V A.
\]
Since \( U \) is invariant under the group \( \mathcal{A} \), we conclude that \( U \) is the
maximal invariant in the sample space under the group \( \mathcal{A} \).

A similar proof shows that the maximal invariant in the parameter
space under \( \mathcal{A} \) is 
\[
\phi = \tilde{\phi} \Lambda_{11} \tilde{\theta}', \quad \text{where}
\]
\[
\Lambda = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}, \quad \Lambda_{11}: q \times q.
\]

In terms of the original (non-canonical) model,
\[
(2.3) \quad U = (\bar{Y} S^{-1} \bar{Y}', \bar{Y} X'_0 (X'_0 S X'_0)^{-1} X_0 \bar{Y}'), \quad \phi = \tilde{\phi} \Sigma^{-1} \tilde{X}' \tilde{\beta}',
\]
where \( X_0 \) is any \( p-q \times p \) matrix whose rows are orthogonal to the
q-dimensional subspace spanned by the rows of \( X \).

The testing problem (1) is also invariant under the transformation
group \( \mathcal{A} \times \mathcal{T} \), where \( \mathcal{A} \times \mathcal{T} = \{(A, \iota): A \in \mathcal{A}, \iota \in T: 1 \times q \} \) and where the
transformation takes $Z$ into $ZA + e'_k(f,0)$, $V$ into $\Lambda'VA$,
\[ e_k = (1,1,\ldots,1) \times k. \]
A modification of the previous proof shows that the maximal invariant in the sample space under $A \times \mathcal{F}$ is the ordered pair
\[ [(Z-e'_k(Z,0)V^{-1}(Z-e'_k(Z,0))', \overline{Z}V_{22}^{-1}\overline{Z})] \]
where $\overline{Z} = (\overline{z}, \overline{\overline{z}}) = k^{-1} e_k Z$.

The maximal invariant in the parameter space under $A \times \mathcal{F}$ is
\[ (I-k^{-1}e'_k e_k)\Phi (I-k^{-1}e'_k e_k). \]

3. The likelihood ratio test for testing the equality of the $\beta^{(i)}$.

We now find the likelihood ratio test for the hypothesis
\[ H^*: \delta^{(1)} = \delta^{(2)} = \ldots = \delta^{(k)} = \delta \]
versus general alternatives $H_A$ in terms of the canonical variables $(Z,V)$. From this result the likelihood ratio test for $H: \beta^{(1)} = \beta^{(2)} = \ldots = \beta^{(k)} = \beta$ in terms of the original variables $(\overline{Y}, S)$ can easily be found.

From (2.1) the joint distribution of $(Z,V)$ is
\[ p(Z,V) \propto \Lambda^{(n+k)/2} |V|^{(n-p-1)/2} \exp\left(-\frac{1}{2} \text{tr} \Lambda[V+(Z-\Theta)'D_N(Z-\Theta)]\right), \]
where $\Lambda = \Sigma^{-1}$ and $D_N = \text{diag}(N_1, \ldots, N_k)$. To find the MLE of $(\hat{\Theta}, \Sigma)$ we maximize $p(Z,V)$ first with respect to $\Sigma$ and then with respect to $\Theta$.

From (3.1) it is immediate that $p(Z,V)$ is maximized at
\[ \Sigma(\Theta) = \frac{V + (Z-\Theta)'D_N(Z-\Theta)}{n + k}, \]
so that the MLE of $\Theta$ is obtained by minimizing
\[ |\hat{\Sigma}(\Theta)| = \frac{|V+(Z-\Theta)'D_N(Z-\Theta)|}{(n+k)^p} = \frac{|V|}{(n+k)^p} |I + \frac{1}{N}(Z-\Theta)V^{-1}(Z-\Theta)'D_N^2| \]

The symbol $\propto$ means "proportional to."
Note that

\[(3.4) \quad (Z-\Theta) V^{-1}(Z-\Theta)^\prime = (\hat{Z} - \bar{Z}) V^{-1}_{22}(\hat{Z} - \bar{Z})^\prime + \bar{Z} V^{-1}_{22} \bar{Z} \, .\]

Using this expansion we see that \((3.3)\) is minimized by

\[(3.5) \quad \hat{\Theta} = \bar{Z} - \bar{Z} V^{-1}_{22} \hat{V}_{21} \, .\]

Substituting in \((3.2)\) yields

\[(3.6) \quad \hat{\Sigma} = \frac{1}{(n+k)^p} \left[ V + (\bar{Z} V^{-1}_{22} \hat{V}_{21})^\prime D_N^0 (\bar{Z} V^{-1}_{22} \hat{V}_{21}) \right] \, .\]

We remark that

\[|\hat{\Sigma}| = \frac{|V|}{(n+k)^p} \left| I + \frac{1}{n} \bar{Z} V^{-1}_{22} \bar{Z}^\prime \frac{1}{n} \bar{Z} \right| \, .\]

For the hypothesis \(H^\ast\) the sufficient statistic becomes \((\bar{Z}, W)\)

where \(\bar{Z} = (n+k)^{-1} e_k D_N Z\), and

\[W = V + Z^\prime \frac{1}{n} D_N^0 C_k \frac{1}{n} \bar{Z} \, ,\]

\[(3.7) \quad C_k = I - (n-k)^{-1} \frac{1}{n} e_k e_k^\prime \frac{1}{n} \bar{Z} \, .\]

Here \(\bar{Z}\) and \(W\) are independent, \(z(\bar{Z}) = \chi (\hat{\Theta}, 0), \Sigma/(n+k))\), and

\(z(W) = W(\Sigma; p, n+k-1)\). Thus the form of this model is similar to that treated in this section with \(\bar{Z}\) in place of \(Z\), \(W\) in place of \(V\), \((n+k)\) in place of \(D_N\), and \(1\) in place of \(k\). It follows that

\[\hat{\Theta} = \bar{Z} - \bar{Z} W^{-1}_{22} W_{21}, \quad \hat{\Sigma}_{H^\ast} = \frac{1}{n+k} \left[ W + (n+k) (\bar{Z} W^{-1}_{22} W_{21})^\prime (\bar{Z} W^{-1}_{22} W_{21}) \right] \, ,\]

so that

\[|\hat{\Sigma}_{H^\ast}| = \frac{|W|}{(n+k)^p} \left| I + (n+k) \bar{Z} W^{-1}_{22} \bar{Z}^\prime \right| \, .\]
We conclude that the likelihood ratio statistic $\lambda_{H^*}$ for testing $H^*$ versus general alternatives is determined by

$$(3.8) \quad \frac{\chi^2}{(n+k)} = \frac{|\hat{\Sigma}|}{|\Sigma_{H^*}|} = \frac{|I + C_k \frac{1}{N} V_{12} \Sigma V_{22}^{-1} Z N_k C_k N^{-\frac{1}{2}}|}{|V_{22}|} \frac{|V_{22}^{-1} Z D_k C_k N^{-\frac{1}{2}} \Sigma Z|}{|V + Z D_k C_k N^{-\frac{1}{2}} D_k Z|},$$

where we recall from (3.7) that $C_k = (I-(n+k)^{-1} D_k N^{-\frac{1}{2}} e_k e_k^T D_k N^{-\frac{1}{2}})$.

In terms of the original observations, the likelihood ratio statistic $\lambda_H$ for $H: \beta^{(1)} = \beta^{(2)} = \ldots = \beta^{(k)} \equiv \beta$ is determined by

$$(3.9) \quad \frac{\chi^2}{(n+k)} = \frac{\chi^2_{H^*}}{(n+k)} = \frac{|I + C_k \frac{1}{N} U_2 D_k N^{-\frac{1}{2}} C_k N^{-\frac{1}{2}}|}{|I + C_k \frac{1}{N} U_1 D_k N^{-\frac{1}{2}} C_k N^{-\frac{1}{2}}|},$$

where $U = (U_1, U_2)$ is given by (2.3).

3.1 The distribution of the likelihood ratio statistic in the central case.

Since $\lambda_{H^*} = \lambda_H$ and since the hypothesis $H$ and $H^*$ are equivalent, we need only consider the distribution of $\lambda_{H^*}$ when $H^*$ is true. Further, from (3.8) and from (2.2) we see that $\lambda_{H^*}$ is a function of the maximal invariant in the sample space under $a$ and is thus invariant under $a$. There exists $A_0 \in a$, $A_0$ $p \times p$ lower triangular, such that $A_0^T \Sigma A_0 = I$, and thus it follows that we can assume $\Sigma = I$ in deriving the distribution of $\lambda_{H^*}$. Finally, when $H^*$ holds, we may assume that $\theta = 0$, for $\lambda_{H^*}$ is also a function of the maximal invariant in the sample space under the group $a \times \mathbb{F}$.  

9
From (3.8) and our assumptions above we have that

\[ \lambda_{h^*} = \frac{|V|}{V_{22}} \left| \frac{|V_{22} + Z^1}{|V + Z^1|} \right| \left( C_k^2 \right) \left( \frac{1}{2} \right)^{\frac{1}{2}} \left| \frac{Z}{N} \right| \] ,

where \( Z \) and \( V \) are independently distributed,

\[ p(Z, V) \propto |V|^{(n-p+1)/2} \exp - \frac{1}{2} \left( \text{tr} V + \text{tr} D_{N} ZZ' \right) . \]  (3.10)

Now \( C_k^2 = C_k \) (i.e., \( C_k \) is idempotent), \( C_k \) is of rank \( k - 1 \), and thus there exists a \( k \times k \) orthogonal matrix \( \Delta \) such that \( \Delta C_k \Delta' = (0, I_{k-1})' (0, I_{k-1}) \). Letting

\[ F = (0, I_{k-1}) \Delta \frac{1}{N} Z = (\hat{F}, \hat{F}) \],

where \( \hat{F}: k - 1 \times q \), \( F: k - 1 \times p \), we therefore have

\[ t = \frac{2}{(n-k)} = \frac{|V|}{V_{22}} \left| \frac{|V_{22} + \hat{F}' \hat{F}|}{|V + \hat{F}' \hat{F}|} \right| \frac{|I + \hat{F}' \hat{F}^{-1} \hat{F}'|}{|I + \hat{F}' \hat{F}^{-1} \hat{F}'|} \] .

We can rewrite \( t \) as

\[ t = \frac{|I + \hat{F}' \hat{F}^{-1} \hat{F}'|}{|I + \hat{F}' \hat{F}^{-1} \hat{F}' + (\hat{F}' \hat{F}^{-1} \hat{F}' - \hat{F}' \hat{F}^{-1} \hat{F}') |} . \]  (3.11)

Make the transformations

\[ L = V_{22}^{-\frac{1}{2}} V_{21}, \ M = V_{11} - L'L, \ V_{22} = V_{22}, \ Q = \hat{F}' V_{22}^{-\frac{1}{2}}, \ R = \hat{F}' V_{22}^{-1} V_{21}, \ P = \hat{F}' - Q \left( L'L \right)^{-1} Q' \] ,

so that

\[ t = \frac{|I + QQ'|}{|I + QQ' + RM^{-1}R'|} \]

From (3.10) we obtain
\[ p(R, Q, M, L, V_{22}) \propto |V_{22}|^{(n-(p-q)+k-2)/2} |M|^{(n-p-1)/2} \times \exp - \frac{1}{2} \left[ \text{tr}(R+QL)(R+QL)' + \text{tr} Q V_{22} Q' + \text{tr} M + \text{tr} LL' + \text{tr} V_{22} \right], \]

for \( V_{22} > 0, M > 0, \) and \( R, L, Q \) unrestricted. Integrating over \( V_{22} > 0 \) we have

\[ p(R, Q, M, L) \propto \frac{|M|^{\frac{n-p-1}{2}} \exp - \frac{1}{2} \left[ \text{tr}(R+QL)(R+QL)' + \text{tr} LL' + \text{tr} M \right]}{\left| I + QQ' \right|^{\frac{n+k-1}{2}}} . \]

The exponent may be written as

\[ \text{tr}[L' + R'Q(I+Q'Q)^{-1}] (I+Q'Q) [L + (I+Q'Q)^{-1}Q'R] \]

\[ + \text{tr} M + \text{tr} R'R' - \text{tr} R'Q(I+Q'Q)^{-1} Q'R , \]

and hence, for \( R \) and \( Q \) fixed, the matrix \( L \) is normally distributed. We thus may integrate over \( L \) yielding

\[ p(R, Q, M) \propto \frac{|M|^{\frac{n-p-1}{2}}}{\left| I + QQ' \right|} \exp - \frac{1}{2} \left[ \text{tr} M + \text{tr} R'[I - Q(I+Q'Q)^{-1} Q']R \right] . \]

Note that \( (I+QQ')^{-1} = I - Q(I+Q'Q)^{-1} Q' \), and

\[ t^{-1} = \left| I + (I+QQ')^{-\frac{1}{2}} R M^{-\frac{1}{2}} R'(I+QQ')^{-\frac{1}{2}} \right| . \]

Thus, by letting \( B = (I+QQ')^{-\frac{1}{2}} R M^{-\frac{1}{2}} \), we find that \( t^{-1} = |I + BB'| \)

and

\[ p(B, Q, M) \propto \frac{|M|^{\frac{n-p+k}{2} - 1}}{\left| I + QQ' \right|^{\frac{n+k-1}{2}}} \exp - \frac{1}{2} \text{tr} M(I+BB') . \]
A final integration over \( M \) and \( Q \) yields

\[
p(B) \propto \left| I + BB' \right|^{-\frac{n+k+q-p-1}{2}},
\]

for \( B \) unrestricted. Since \( B \) is a \( k \times q \) matrix, we must consider the two cases \( k \geq q + 1 \) and \( k < q + 1 \).

If \( k \geq q + 1 \), then from Hsu's Theorem, \( [1, p. 319] \), the distribution of \( G = B'B \) is

\[
p(G) \propto \frac{|G|^{(k-q-2)/2}}{|I + G|^{(n+k+q-p-1)/2}}, \quad G > 0.
\]

If \( k < q + 1 \), then the distribution of \( G = BB' \) is

\[
p(G) \propto \frac{|G|^{(q-k)/2}}{|I + G|^{(n+k+q-p-1)/2}}, \quad G > 0.
\]

In either case we ask for the distribution of \( t = |I + G|^{-1} \).

Comparing moments, we see that \( x(t) = x \prod_{i=1}^{s_1} t_i \), where the \( t_i \) are mutually independent with \( x(t_1) = \beta(\frac{1}{2}(n+a_1-p-1), \frac{1}{2}a_2) \), \( a_1 = \min(q,k-1) \), \( a_2 = \max(q,k-1) \). The method of Box (viz., Anderson [1], pp. 203-209) gives an excellent approximation to the distribution of \( t \) for \( (n+k) \) large, namely,

\[
P\left\{-2 \frac{m}{n+k} \log t \leq x\right\} = \left[ \left( 1 - \frac{v_1^2}{m^2} \right) + \frac{v_1}{m} - \frac{v_2}{m^4} \right] \ P\left\{ \chi^2_{q(k-1)} \leq x \right\} + \frac{v_1}{m^2} \left( 1 - \frac{v_1^2}{m^2} \right) \ P\left\{ \chi^2_{q(k-1)+4} \leq x \right\} + \frac{v_2}{m^4} \ P\left\{ \chi^2_{q(k-1)+8} \leq x \right\} + O((n+k)^{-6}),
\]

12
where

\[ m = n - p - 1 + \frac{q+k}{2}, \quad \nu_1 = \frac{q(k-1)}{48} \left( q^2 + (k-1)^2 - 5 \right), \]

and

\[ \nu_2 = \frac{1}{2} \nu_1^2 + \frac{q(k-1)}{1920} \left[ 3q^4 + 3(k-1)^4 + 10q^2 (k-1)^2 - 50q^2 - 50(k-1)^2 + 159 \right]. \]

3.2 The limiting non-null distribution of the likelihood ratio statistic.

Let us assume that as \( n \to \infty \), \( (n+k)^{-1} D_N \) converges to \( \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) = D_\lambda \). (Recall that \( \Sigma N_\lambda = n + k \), so that \( \Sigma \lambda_1 = 1 \).)

Then from (3.7) \( \lim_{n \to \infty} C_k = I - \begin{bmatrix} 2 \lambda \end{bmatrix} C_\lambda \begin{bmatrix} 2 \lambda \end{bmatrix} = C_\lambda \). Further

\[ \text{plim} \left[ V/(n+k) \right] = \Sigma \quad \text{and} \quad \text{plim} Z = \Theta = (\Theta, 0). \]

Letting \( B = I + C_k \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} C_k \), we can write \( t = \lambda_2^2/(n+k) \) as:

\[ t = \left| I + B^{-\frac{1}{2}} C_k \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} C_k \right|^2 \]

Since \( \text{plim} B = I \) and from the above, the limiting distribution of \( t \) as \( n \to \infty \) is the same as that of

\[ r = \left| I + C_\lambda GG' C_\lambda \right|, \]

where \( G = (n+k)^{-\frac{1}{2}} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \end{bmat
distribution of \( \log r = \log |I + C_\lambda GG' C_\lambda| \), and show that

\[
\lim_{n \to \infty} \mathcal{N}(\log |I + C_\lambda GG' C_\lambda| - \log |I + C_\lambda \Omega C_\lambda| = \mathcal{N}(0, \delta),
\]

where

\[
\Omega = (G G')(G G)' = D_{\lambda}^{-\frac{1}{2}} \Lambda_\lambda \Lambda_\lambda' \Lambda_\lambda \Lambda_\lambda' \Lambda_\lambda D_{\lambda}^{-\frac{1}{2}}
\]

and

\[
\delta = \frac{1}{4} \text{tr}[(I + C_\lambda \Omega C_\lambda)^{-2} C_\lambda \Omega C_\lambda].
\]

From this it follows that

\[
\lim_{n \to \infty} \mathcal{N}(\log t - \log |I + C_\lambda \Omega C_\lambda| = \mathcal{N}(0, \delta),
\]

since \( \log t \) and \( \log r \) have the same limiting distribution.

We prove (3.12) by expanding \( \log r = \log |I + C_\lambda GG' C_\lambda| \) in a Taylor series in \( G \) about the mean of \( G \):

\[
\sqrt{n+k}(\log r - \log |I + C_\lambda \Omega C_\lambda|) = \sqrt{n+k} \left[ \sum_{i=1}^{k-1} \sum_{j=1}^{\sigma} \left( \frac{\partial \log r}{\partial g_{ij}} \right) \right]_{g_{ij}=\sigma g_{ij}} + o_p(1),
\]

where \( G = (g_{ij}), \ i = 1, \ldots, k - 1, \ j = 1, \ldots, q \). But we know that \( \mathcal{N}(\sqrt{n+k}(g_{ij}-\sigma g_{ij})) = \mathcal{N}(0,1) \) and that the \( g_{ij} \) are independently distributed. Thus

\[
\mathcal{N}(0, \sum_{i=1}^{k-1} \sum_{j=1}^{\sigma} \left( \frac{\partial \log r}{\partial g_{ij}} \right)^2)_{g_{ij}=\sigma g_{ij}}.
\]
Lemma. If $L = (L_{ij})_a \times b$ and $C: c \times a,$

$$\Sigma_{i=1}^{a} \Sigma_{j=1}^{b} \left( \frac{\partial \log |I + C L L' C'|}{\partial \ell_{ij}} \right)^2 = 4 \text{ tr } L L' (C' W^{-1} C)^2$$

where $W = I + C L L' C' = (W_{ij}).$

Proof: First let $f_{ij} = \frac{\partial \log |I + C L L' C'|}{\partial \ell_{ij}}$ and let $F = (f_{ij}).$

Note that $\Sigma_{i=1}^{a} \Sigma_{j=1}^{b} f_{ij}^2 = \text{tr } FF'$ and, by the chain rule, that

$$\frac{\partial \log |I + C L L' C'|}{\partial \ell_{ij}} = \Sigma_{\alpha, \beta} \frac{\partial \log |W|}{\partial w_{\alpha \beta}} \frac{\partial w_{\alpha \beta}}{\partial \ell_{ij}} = \text{tr} \left( \frac{\partial \log |W|}{\partial w_{\alpha \beta}} \right) \left( \frac{\partial w_{\alpha \beta}}{\partial \ell_{ij}} \right),$$

where $\left( \frac{\partial \log |W|}{\partial w_{\alpha \beta}} \right)$ is the $c \times c$ matrix of partial derivatives of $\log |W|$ with respect to the $w_{\alpha \beta},$ and where for each $\ell_{ij},$ $\left( \frac{\partial w_{\alpha \beta}}{\partial \ell_{ij}} \right)$ is the $c \times c$ matrix of partial derivatives of $w_{\alpha \beta}$ with respect to $\ell_{ij}.$

But $\frac{\partial \log |W|}{\partial w_{\alpha \beta}} = w_{\alpha \beta},$ where $W^{-1} = (w_{\alpha \beta}),$ and $\frac{\partial w_{\alpha \beta}}{\partial \ell_{ij}}$ is obtained from the equation

$$dW = C (dL L' + L dL') C',$$

namely,

$$\left( \frac{\partial w_{\alpha \beta}}{\partial \ell_{ij}} \right) = C \left[ \prod_{i,j} L' + L \prod_{i,j} \right] C',$n

where $\prod_{i,j}$ is the $a \times b$ matrix having a 1 in the $(i,j)$-th place and zeroes elsewhere. Thus

$$f_{ij} = \text{tr} \left( \frac{\partial \log |W|}{\partial w_{\alpha \beta}} \right) \left( \frac{\partial w_{\alpha \beta}}{\partial \ell_{ij}} \right)$$

$$= \text{tr} W^{-1} C \left[ \prod_{i,j} L' + L \prod_{i,j} \right] C'$$

$$= 2 \text{ tr } L' C' W^{-1} C \prod_{i,j} = 2 (L' C' W^{-1} C)_{ij}.$$
Finally,

$$\text{tr } FF' = 4 \text{ tr } L'C'W^{-1}CC'W^{-1}CL = 4 \text{ tr } LL'(C'W^{-1}C)^2.$$ 

The result (3.12) is now obtained as a direct consequence of the Lemma and the fact that $\frac{C^2}{\lambda} = \frac{C}{\lambda}$.

4. The distribution of the maximum likelihood estimates.

From (3.5) and (3.6) the MLE of $\hat{\phi}$ and $\Sigma$ are

$$\hat{\phi} = \hat{z} - \hat{z}V^{-1}_{22}V_{21},$$

$$\hat{\Sigma} = \frac{1}{n+k} \left[ V + (\hat{z}V^{-1}_{22}V_{21}z')^{'}D_N(zV^{-1}_{22}V_{21}z) \right],$$

which in terms of the original model becomes

$$(4.1) \quad \hat{\beta} = \bar{Y} S^{-1} X^{'} (X S^{-1} X)^{-1},$$

$$\hat{\psi} = \frac{1}{n+k} \left[ S + (\bar{Y} \hat{\beta} X)'^{(\bar{Y} \hat{\beta} X)} \right].$$

4.1 The distribution of $\hat{\phi}$.

To find the distribution of $\hat{\phi}$ we make use of invariance. Let $A_0 \in \mathcal{Q}$ be the $p \times p$ lower triangular matrix such that

$A_0^'* \Sigma A_0 = I$. Partition $A_0$ as

$$A_0 = \begin{pmatrix} A_0^{(1)} & 0 \\ A_0^{(2)} & A_0^{(3)} \end{pmatrix}, \quad A_0^{(1)}: q \times q,$$

and let $F = (\hat{F}, \hat{F}') = D_N^{1/2}(\hat{Z} - \hat{\phi}, \hat{\Sigma})A_0, W = A_0'^*VA_0$. Then $F$ and $W$ are independent, $\mathcal{N}(0, I)$, the rows of $F$ are mutually independent with each row having the distribution $\mathcal{N}(0, I)$, and

$$(4.2) \quad D_N^{1/2}(\hat{\phi} - \hat{\phi}) A_0^{(1)} = \hat{F} - \hat{F}W^{-1}_{22}W_{21}.$$
\[(3.8) \quad I_{\frac{1-p_m}{n-s_m+1, s_m}} J \left( \frac{s_1, \ldots, s_{m-1}}{p_1, \ldots, p_{n-1}, l, m} \right) \]

\[\leq J \left( \frac{s_1, \ldots, s_m; n}{p_1, \ldots, p_m; m+1} \right)\]

\[\leq I_{\frac{1-p_m}{n-s_{m+1}, s_m}} J \left( \frac{s_1, \ldots, s_{m-1}}{p_1, \ldots, p_{m-1}, l, m} \right),\]

from which by iteration, we obtain

\[(3.9) \quad I_{\frac{1-p_m}{n-s_{m+1}, s_m}} I_{\frac{1-p_{m-1}}{p_m}} \cdots I_{\frac{1-p_1}{p_m}} \left( \frac{s_1, \ldots, s_m; n}{p_1, \ldots, p_m; m+1} \right) \]

\[\leq J \left( \frac{s_1, \ldots, s_m; n}{p_1, \ldots, p_m; m+1} \right)\]

\[\leq I_{\frac{1-p_m}{n-s_{m+1}, s_m}} I_{\frac{1-p_{m-1}}{p_m}} \cdots I_{\frac{1-p_1}{p_m}} \left( \frac{s_1, \ldots, s_m; n}{p_1, \ldots, p_m; m+1} \right).\]

4. Other Tail Probabilities.

In the basic urn model we continue taking observations until cell $C_{k+1}$ contains $s$ observations. Our concern then was with the events that each cell $C_j$ contains at least $s_j$ or at most $s_j$ observations. There are, of course, other models and events which may be of interest.
\[ p(Q) \propto \int \frac{e^{-\frac{1}{2} \text{tr}[Q'Q(I+LL')^{-1}]} dL}{|I + LL'|^{n/2}} \quad \text{L unrestricted} \]

4.2 The distribution of \( \hat{\Sigma} \).

In terms of the random variables \( \hat{F} = \frac{1}{N} \hat{Z} A_0 \) and \( \hat{W} = A_0^\top V A_0 \) we have that

\[
(n+k) \hat{\Sigma} = \Sigma \frac{1}{2} [\hat{W} + (\hat{F} \Sigma^{-1} \hat{W} \Sigma^{-1} \hat{F})^\top (\hat{F} \Sigma^{-1} \hat{W} \Sigma^{-1} \hat{F})]^{\frac{1}{2}},
\]

where the rows of \( \hat{F} \): \( k \times q \) are independent with the common distribution \( \mathcal{N}(0, I_q) \) and independent of \( \hat{W} \), \( \mathcal{W}(I; p, n) \). By an argument similar to that of [4, Section 5], we obtain

\[
p(\hat{\Sigma}) = [c(p, n) \Lambda]^{n/2} |\hat{\Sigma}|^{(n-p-1)/2} \exp -\frac{1}{2} \text{tr} \Lambda \hat{\Sigma}\]

\[\hat{\Sigma} > 0, Y: k \times p - q, D_Y = \text{diag}(\nu_1', \ldots, \nu_{p-q}'), \text{ the } \nu's \text{ are the characteristic roots of } R(\Sigma^\frac{1}{2} \hat{\Sigma}^\frac{1}{2} \Sigma^\frac{1}{2}), \text{ where for a positive definite matrix } H: p \times p \text{ with } H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}, H_{11}: q \times q, \text{ we have } R(H) = H^\frac{1}{2} H_{22} H_{21} H_{12} H^\frac{1}{2}.
\]

The distribution of \( \hat{\psi} \) has the form (4.6) with \( \hat{\Sigma} \) replaced by \( \hat{\psi} \) and \( \Sigma \) replaced by \( \psi \).

4.3. Confidence bounds for \( \hat{\theta} \).

Recall that
After differentiating \( \varphi(p_k) \), collapsing sums, and simplifying,

\[
(4.2) \quad \frac{dQ(p_k)}{dp_k} = \frac{-n! \prod_{l=1}^{k-1} (1-p_l)}{(\sum_{l=1}^{k-1} s_l - 1)! (n - \sum_{l=1}^{k-1} s_l)!} c \left( s_1, \ldots, s_{k-1}; \frac{p_1}{1-p_k}, \ldots, \frac{p_{k-1}}{1-p_k}, \sum_{l=1}^{k-1} s_l - 1 \right). \]

Using induction for \( k-1 \), (4.2) can be expressed as a \((k-2)\)-fold integral. Integration with respect to \( p_k \) (using the fact that \( Q(0) = 0 \)) leads to (4.1). The proof is completed by noting that for \( k = 2 \),

\( c(s_1, s_2; p_1, p_2; n) \) reduces to (2.5) which completes the induction argument.

Remark. There are various ways to prove (4.1). The following method is an alternative which has intrinsic interest and we sketch the underlying idea. Consider \( n \) independent observations on a uniform distribution on \([0,1]\). The points \( \delta_1 = p_1, \delta_2 = p_1 + p_2, \ldots, \delta_{k-1} = p_1 + \cdots + p_{k-1} \) divide the unit interval into \( k \) subintervals (cells) of length \( p_1, \ldots, p_k \). If \( X_j \) denotes the number of observations that falls in the \( j \)-th cell, then the probability that \( X_1 \geq s_1, X_1 + X_2 \geq s_1 + s_2, \ldots, X_1 + \cdots + X_k \geq s_1 + \cdots + s_k \) is given by the LHS of (4.1). On the other hand, if we let \( T_j \) denote the length of the interval from the left endpoint of the \( j \)-th cell to the \( s_j \)-th order statistic in the \( j \)-th cell, then the random variables \( T_1, \ldots, T_{k-1} \) have the multivariate Beta distribution as given by (2.14). The condition \( \sum_{l=1}^{r} X_l \geq \sum_{l=1}^{r} s_l \), \( r = 1, \ldots, k-1 \), is now equivalent to the condition
References


