Product Inequalities Involving the Multivariate Normal Distribution

BY
R. L. Dykstra

Technical Report No. 85
Department of Statistics
July 1979

Mathematical Sciences
**Title:** Product Inequalities Involving the Multivariate Normal Distribution

**Authors:** Richard L. Dykstra

**Report Number:** 85

**Type of Report & Period Covered:** Technical Report

**Performing Organization Name & Address:**
Department of Statistics
University of Missouri
Columbia, MO 65211

**Controlling Office Name & Address:**
Office of Naval Research
Department of the Navy
Arlington, VA

**Controlled Code:** Unclassified

**Distribution Statement:** Approved for public release; distribution unlimited.

**Key Words:**
Product inequality, multivariate normal, Wishart matrix, simultaneous confidence intervals in multivariate regression.

**Abstract:**
Suppose \( Y^* = (Y_1, \ldots, Y_k) \) possesses a multivariate normal distribution with mean vector \( \mu \) and positive semidefinite covariance matrix \( \Sigma \). If \( C_i \subset \mathbb{R}^p \) are convex regions symmetric about the origin, then conditions are given such that...
20.

$$P(Y_i \in C_i, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i \in C_i)$$

and/or

$$P(Y_i \in \bar{C}_i, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i \in \bar{C}_i)$$

obtain. These conditions imply that chi-squared random variables defined from a multivariate normal distribution are always positively dependent and non-negatively correlated. Other applications involve conservative simultaneous confidence regions in a multivariate regression setting.
Product Inequalities Involving the Multivariate Normal Distribution

by

R. L. Dykstra

Technical Report No. 85
July 1979

Prepared under contract N00014-78-C-0655
for the Office of Naval Research

Reproduction in whole or in part is permitted
for any purpose of the United States Government

Department of Statistics
University of Missouri
Columbia, Missouri 65211
Product Inequalities Involving the Multivariate Normal Distribution

Abstract

Suppose $Y = (Y_1, \ldots, Y_k)$ possesses a multivariate normal distribution with mean vector $\mu$ and positive semidefinite covariance matrix $\Sigma$. If $C_i \subset \mathbb{R}^p$ denote convex regions symmetric about the origin, then conditions are given such that

$$P(Y_i \in C_i, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i \in C_i)$$

and/or

$$P(Y_i \not\in \overline{C_i}, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i \not\in \overline{C_i})$$

obtain. These conditions imply that chi-squared random variables defined from a multivariate normal distribution are always positively dependent and non-negatively correlated. Other applications involve conservative simultaneous confidence regions in a multivariate regression setting.

AMS Classification numbers: Primary 62H99; Secondary 62E10.

Key words and phrases: product inequality, multivariate normal, Wishart matrix, simultaneous confidence intervals in multivariate regression.
1. INTRODUCTION.

The following question has received the attention of several investigators. Suppose \( Y = (Y_1, Y_2, \ldots, Y_k) \) possesses a multivariate normal distribution with mean vector \( \mu \) and positive semidefinite (p.s.d.) covariance matrix \( \Sigma \), i.e. \( Y \) is \( n(\mu, \Sigma) \). The dimension of \( Y_i \) is \( p_i \) so that the dimension of \( Y \) must be \( \sum_{i=1}^{k} p_i \). We let \( C_i \) denote a convex region in \( \mathbb{R}^{p_i} \) (and \( \overline{C_i} \) its complement) which is symmetric about the origin (\( x \in C_i \) implies \( -x \in C_i \)). Under what conditions will at least one of the two inequalities

\[
P(Y_i \in C_i, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i \in C_i)
\]

and

\[
P(Y_i \in \overline{C_i}, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i \in \overline{C_i})
\]

hold? A primary application of such inequalities has been obtaining conservative simultaneous confidence bounds which are easily computable although the inequality is useful in other areas as well.

Dunn (1958) showed the inequality (1.1) would hold if \( p_i = 1 \) for all \( i \) and the \( C_i \) were equal, providing \( k = 2 \) or \( k = 3 \). (For \( k = 2 \) it is easily shown that (1.1) and (1.2) are equivalent.) Moreover she proved her conjecture for arbitrary \( k \) if the correlation matrix corresponding to \( \Sigma \) was of the form \( \rho_{ij} = b_i b_j \), \( i, j = 1, \ldots, k \), \( i \neq j \) and \( 0 < b_i < 1 \), \( i = 1, \ldots, k \).
Sidak (1967) was able to extend Dunn's results to the case of an arbitrary covariance matrix and different $C_i$'s (although still one dimensional).

Scott (1967) purported to prove the same result as Sidak as well as inequality (1.2) with the same conditions holding. However a subtle conditioning error occurs in Scott's arguments and invalidates his proofs. Moreover, Sidak (1971) has constructed a counterexample to Scott's second inequality so that (1.2) does not hold under these conditions.

In light of Sidak's 1967 results it is tempting to conjecture an analogue of a one sided result due to Slepian (1962), namely that decreasing the absolute value of the correlations should decrease $P(c_i \leq Y_i \leq c_i, i = 1, \ldots, k)$. However a counter example of Sidak (1968) showed that this is not true in general. He did show however that the conjecture held if the absolute values of the correlations decreased in a particular manner.

Khatri (1967) was able to prove (1.1) for higher dimensions (although the $p_i$ must be equal) when the $C_i$'s are ellipsoids of the form $\{x'x'Ax \leq c_i\}$, the covariance structure of $Y$ is the Kronecker product $\Psi^{k\times k} \otimes I^{p\times p}$, and $A$ and $\Psi$ are positive semidefinite.

By adding a restriction similar to Dunn's on the matrix $\Psi$, Khatri was able to prove (1.2) in the same setting. He also considered some situations where the ellipsoids are random. Later Khatri (1970) had apparently extended his
1967 results to the point that (1.1) and (1.2) hold without any restrictions on \( \Sigma \) or the \( C_i \)'s (other than those initially assumed). However Sidak (1975) isolated an unobtrusive error in both Scott's 1967 paper and Khatri's 1970 paper. Moreover, Sidak's 1971 counterexample has shown that the general inequality (1.2) without restrictions is incorrect. Das Gupta et. al. (1972) considered inequalities of the type (1.1) and (1.2) in the more general setting of elliptically contoured distributions. They were able to establish inequality (1.1) without restrictions on \( \Sigma \) providing every \( P_i \) except one equals one.

Tong (1970) and Sidak (1973) have established some inequalities related to (1.1) and (1.2) when all the correlations determined by \( \Sigma \) are identical.

Whether inequality (1.1) holds in general without additional assumptions has been an open question discussed in several of the above references. It is the purpose of this paper to give some new conditions under which inequalities (1.1) or (1.2) hold and discuss briefly a few applications which result.

2. THEOREMS AND PROOFS.

We will have need of the following lemmas in proving our theorems.

**LEMMA 1.** If \( Z = (Z_1, \ldots, Z_p) \) is a random vector with density \( f(z) \) symmetric about 0 (\( f(z) = f(-z) \) for all \( z \)), then...
such that \( f(\tilde{z}) \leq c \) is convex for all non-negative \( c \), \( \tilde{z} \) is an arbitrary fixed vector in \( \mathbb{R}^p \), and \( C \) is a convex set symmetric about the origin, then \( P(\tilde{z} \in C + \lambda \tilde{a}) \) is a nonincreasing function of \( \lambda (0 \leq \lambda < \infty) \).

This lemma is due to T. A. Anderson (1955) and is well known.

**Lemma 2.** Suppose the square, symmetric matrix \( A \) is partitioned as \( A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \) where \( A_{11} \) and \( A_{22} \) (nonsingular) are square. Then \( A \) is positive definite (p.d.) iff both \( A_{11} - A_{12} A_{22}^{-1} A_{21} \) and \( A_{22} \) are positive definite.

**Proof.** It is well known that if \( E \) is a p.d. matrix and \( D \) is nonsingular (same dimension), then \( DED^T \) is p.d. If we define the matrix

\[
B = \begin{pmatrix} I & -A_{12} A_{22}^{-1} \\ 0 & I \end{pmatrix},
\]

then \( B \) is clearly nonsingular. However then

\[
BAB^T = \begin{pmatrix} A_{11} - A_{12} A_{22}^{-1} A_{21} & 0 \\ 0 & A_{22} \end{pmatrix},
\]

which implies the desired result.

**Lemma 3.** If \( h_1(x), h_2(x) \) are nonnegative functions either both nonincreasing or both nondecreasing and \( X \) is a random variable, then
\[ E h_1(X) h_2(X) \geq E h_1(X) \cdot E h_2(X). \]

This result is well known and appears in various places, one place being Kimball (1951).

**Lemma 4.** If \( x = (x_1, \ldots, x_p) \) is a \( p \)-variate vector and \( \phi(x) \) is a real-valued function invariant under orthogonal transformation \( \phi(Ax) = \phi(x) \) for every orthogonal matrix \( A \), then \( \phi(x) \) is a function of \( x \) only through \( \sum x_i^2 \).

**Proof.** For a given vector \( x = (x_1, \ldots, x_p) \), let \( O(x) \) denote an orthogonal matrix such that \( O(x) x = ((\sum x_i^2)^{1/2}, 0, \ldots, 0) \). (WLOG we may assume \( O(ax) = O(x) \) for any positive constant \( a \).) The assumption then guarantees that \( \phi(x) = \phi(O(x)x) \) is a function of only \( \sum x_i^2 \).

Theorem 1 is, in one sense, a generalization of a theorem by Das Gupta et al. (1973).

**Theorem 1.** Suppose \( \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \) is \( n(0, \Sigma) \) with covariance matrix \( \Sigma \) expressable as \( \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \) and that \( C \) is an arbitrary convex set symmetric about \( 0 \). Then if \( A \) is idempotent \( (A^2 = A) \),

\[ P(Y_1 \in C, Y_2^*AY_2 \leq c) \geq P(Y_1 \in C)P(Y_2^*AY_2 \leq c) \]

for all \( c \geq 0 \).
PROOF. By making linear transformations on $Y_1$ and $Y_2$ which may reduce the number of random variables, we may assume WLOG that both $\Sigma_{11}$ and $I_{22}$ are of full rank and that the left side of (2.1) is

$$P(Y_1 < c, Y_2^TW Y_2 < c).$$

We may also assume WLOG that $\Sigma$ is full rank, since if it is not, we may define a sequence of positive definite matrices of the same form which converge elementwise to $\Sigma$ and then use a limiting argument to obtain the desired conclusion.

By a continuity argument, we may replace $I_{22}$ by $(1 - \epsilon)I_{22}$ and still have $\Sigma$ be positive definite if $\epsilon$ is sufficiently small. Thus the augmented matrix,

$$\Sigma^* = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{12} \\ \Sigma_{21} & I_{22} & (1-\epsilon)I_{22} \\ \Sigma_{21} & (1-\epsilon)I_{22} & (1-\epsilon)I_{22} \end{pmatrix},$$

is positive definite for $\epsilon$ sufficiently small.

To see this, partition off the last row and column of submatrices and note from Lemma 2 that $\Sigma^*$ is p.d. iff

$$\begin{pmatrix} \Sigma_{11} - \Sigma_{12} \Sigma_{21} / (1 - \epsilon) & 0 \\ 0 & \epsilon I_{22} \end{pmatrix}$$

is positive definite.
and \((1 - \varepsilon)I_{22}\) are p.d. Suppose now that \((Y'_1, Y'_2, Z')\) is \(n(\Omega, \Sigma^*)\). Then we may write

\[
P(Y_1 \in C, Y_2 \leq c) = \mathbb{E}_Z P(Y_1 \in C, Y_2 \leq c | Z)
\]

(2.4) \[= \mathbb{E}_Z \{ P(Y_1 \in C | Y_1) \} \text{ is } n(\Sigma_{12}/(1-\varepsilon), \Sigma_{11} - \Sigma_{12}^2 \Sigma_{21}/(1-\varepsilon)) \cdot P(Y_2 \leq c | Y_2) \text{ is } n(Z, \varepsilon I_{22})
\]

where \(Z\) is \(n(\Omega, (1 - \varepsilon)I_{22})\). If we let \(|Z| = (\Sigma_{22})^{1/2}\) denote the distance from \(Z\) to the origin, we may express \(Z = Q(|Z|, 0, \ldots, 0)'\) where the random orthogonal matrix \(Q\) and \(|Z|\) are independent. Moreover, since the distribution of \(Q\) is invariant if multiplied on the left by a fixed orthogonal matrix, \(Q\) must possess the Haar invariate distribution discussed in Anderson (1958). Since the conditional distribution of \(Y_2 Y_2\) is (except for a constant) a noncentral chi squared distribution with noncentrality parameter \(|Z|^2\), and hence free of \(Q\), we may express the quantity in (2.4) as

(2.5) \[\mathbb{E}_Z \left[ \mathbb{E} [P(Y_1 \in C | Q, |Z|)] \cdot P(Y_2 \leq c | |Z|) \right].
\]

By Lemma 1, the second factor of the integrand is a decreasing function of \(|Z|\). Similarly, for a fixed value of \(Q\),
the integrand of $E$ is nonincreasing in $|z|$, and hence $E \mathbb{P}(Y_1 \in C|Q, |Z|)$ is nonincreasing in $|Z|$. Then by applying Lemma 3, the expression in (2.5) must be

$$
\leq E \mathbb{P}(Y_1 \in C|Q, |Z|) \cdot E \mathbb{P}(Y_2 \leq c| |Z|)
$$

$$
= \mathbb{P}(Y_1 \in C) \cdot \mathbb{P}(Y_2 \leq c)
$$

which was the desired result.

By replacing the words "nonincreasing in $|Z|$" by "nondecreasing in $|Z|$" the following corollary is immediate.

**COROLLARY 1.** Under the same assumptions as in Theorem 1,

$$
\mathbb{P}(Y_1 \in C, Y_2 \leq c) \geq \mathbb{P}(Y_1 \in C) \mathbb{P}(Y_2 \leq c).
$$

If we assume that $C$ is of the form

$$
C = \bigcap_i \{Y, Y_1 \Lambda_i Y_1 \leq c_i\}
$$

where the $\Lambda_i$ are idempotent matrices, then repeated application of Theorem 1 gives the following corollary.

**COROLLARY 2.** If $Y = (Y_1, \ldots, Y_k)$ is $n(\emptyset, \Sigma)$ where the corresponding partition of $\Sigma$ is
then

\[(2.0) \quad P(Y_i^* \sim \Lambda_1, \Lambda_i \leq c_1, i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_i^* \sim \Lambda_1, \Lambda_i \leq c_1)\]

for all positive constants \(c_1, \ldots, c_k\).

This is a generalization of Sidak's (1967) inequality in the sense that a symmetric interval generalizes to a symmetric spheroid in higher dimensions and a \(n(0,1)\) random variable generalizes to a \(n(0,1)\) random vector. Corollary 2 states that the chi squared random variables \(Y_i^* \sim \Lambda_1, \Lambda_i \) are positively orthant dependent. Stronger forms of dependency such as associativity, positive regression dependence and monotone likelihood ratio dependence as discussed for example in Dykstra, Hewett, and Thompson (1973) do not hold in general for these chi squared variables.

However, when \(k = 2\), more can be said about the nature of the inequality in Corollary 2 as indicated in the following theorem.

**Theorem 2.** If \((Y_1, Y_2) \sim n(0, \Sigma)\), where \(\Sigma\) may be partitioned as

\[\Sigma = \begin{pmatrix}
I_{11} & \Sigma_{12} \\
\Sigma_{21} & I_{22}
\end{pmatrix},\]
then for all \( c_1 \) and \( c_2 \)

\[
P(Y_{i1}^2 \leq c_1, Y_{i2}^2 \leq c_2)
\]

is a nondecreasing function of the characteristic roots of \( \Sigma_{12}^{\Sigma_{21}} \).

**PROOF.** There exist orthogonal matrices \( Q_1 \) and \( Q_2 \) such that

\[
Q_1 \Sigma_{12}^{\Sigma_{21}} Q_2^T = \text{diag}(\rho_1^2, \ldots, \rho_m^2)
\]

where \( \rho_1, \ldots, \rho_m \) are the characteristic roots of \( \Sigma_{12}^{\Sigma_{21}} \). Since we may express (2.7) as

\[
\prod_{Y_{11}^*, Y_{12}^*} \sum_{Y_{11}^*, Y_{12}^*} \frac{P(Y_{11}^2 \leq c_1}{Y_{11}^*, Y_{12}^*} \sum_{i=2}^{P_1} \frac{P(Y_{i1}^2 \leq c_1 - \Sigma Y_{i1}^2}{Y_{i1}^*, Y_{11}^*} \sum_{i=2}^{P_2} \frac{P(Y_{i2}^2 \leq c_2 - \Sigma Y_{i2}^2}{Y_{i2}^*, Y_{12}^*}
\]

where \((Y_{11}^*, Y_{12}^*)\) denote the random vectors \((Y_1, Y_2)\) with the first components removed, if we increase only \( \rho_1 \), it is well known that the integrand, and hence the whole expression must increase. Since \( \rho_1 \) is not special, the theorem easily follows.

In attempting to extend Corollary 2 to ellipsoidal regions rather than spheroids, difficulties are encountered. However, by putting a rather stringent condition on the covariance structure, a result rather similar to Khatri's (1967) result, though stated differently, is possible.

**THEOREM 3.** Let \( Y' = (Y_1', \ldots, Y_k') \) denote a kp \( \times 1 \) random vector possessing an \( n(0, B \otimes \Sigma) \) distribution where
is the Kronecker product of $B$ and $\Sigma$. Then if $A$ is a symmetric p.d. matrix,

$$P(Y_1^\top A^{-1} Y_i \leq c_i, \ i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_1^\top A^{-1} Y_i \leq c_i)$$

for all positive constants $c_1, \ldots, c_k$.

**PROOF.** We may assume WLOG that $B$ and $\Sigma$ are full rank. Then, by using the nonsingular matrix $Q$ such that

$$Q \Sigma Q^\top = I \quad \text{and} \quad Q A Q^\top = \text{diag} \Lambda = (\lambda_i)$$

where $\lambda_i$ are the roots of $|A - \lambda \Sigma| = 0$, we may express (2.10) as

$$P(Y_1^\top A^{-1} Y_i \leq c_i, \ i = 1, \ldots, k) \geq \prod_{i=1}^{k} P(Y_1^\top A^{-1} Y_i \leq c_i)$$

where $Y$ has covariance matrix $B \otimes I$. The augmented matrix

$$B \otimes I = \begin{pmatrix} b_{11}^I & \cdots & b_{1k}^I & b_{1k}^I \\ \vdots & \ddots & \vdots & \vdots \\ b_{21}^I & \cdots & b_{2k}^I & b_{2k}^I \\ \vdots & \ddots & \vdots & \vdots \\ b_{k1}^I & \cdots & b_{kk}^I & b_{kk}^I \\ \vdots & \ddots & \vdots & \vdots \\ (1-\varepsilon)b_{kk}^I & \cdots & (1-\varepsilon)b_{kk}^I & (1-\varepsilon)b_{kk}^I \end{pmatrix}$$
will be p.d. if $\epsilon > 0$ is sufficiently close to 0. Thus if 
$(Y'_1, \ldots, Y'_k, Z')'$ is $n(0, B^* \sigma I)$, the left side of (2.11) 
is expressable as 
$$E \sum_{Z' \in \mathcal{Z}} P(Y_k' A_{k-1} Y_1' \leq c_i, i=1, \ldots, k-1, \|Z\|) \cdot P(Y_k' A_{k-1} Y_k' \leq c_k | Z) .$$

However, by the diagonal structure of $A^{-1}$ and the conditional covariance structure of $Y$ given $Z$, it follows from Lemma 1 that each factor of the integrand is a non-
increasing function of $Z_1$ when $Z_2, \ldots, Z_p$ are held fixed. Repeated applications of Lemma 3 imply the desired result.

One might hope that Theorem 3 would extend in certain situations to the case when $A$ is random. Under the right conditions, this does indeed happen; the right conditions being that $A$ possess a central Wishart distribution with covariance matrix $\Sigma$ and that $A$ be independent of $Y$.

**COROLLARY 3.** If $Y' = (Y'_1, \ldots, Y'_k)'$ is $n(0, B\sigma I)$, $S$ 
has a central Wishart distribution with covariance matrix $\Sigma$, 
($S$ is $W(v, \Sigma)$), and $Y$ and $S$ are independent, then 

$$(2.12) \ P(Y_i S^{-1} Y_i' \leq c_i, i=1, \ldots, k) \geq \pi P(Y_i S^{-1} Y_i' \leq c_i) \ |_{i=1}^{k}$$

for all positive constants $c_1, c_2, \ldots, c_k$.

**PROOF.** Clearly we may assume $\Sigma = I$ WLOG. Then the conditional distribution of $(Y_1' S^{-1} Y_1, \ldots, Y_k' S^{-1} Y_k)$ given
S depends upon $S$ only through its characteristic roots 
$
\Psi = (\psi_1, \ldots, \psi_p).
$
By Theorem 3

$$
P(Y_i S^{-1} Y_i \leq c_i, i = 1, \ldots, k) 
= \exp \psi \sum_{i=1}^{k} P(Y_i S^{-1} Y_i \leq c_i | \psi) 
(2.13)
$$

Since the characteristic roots of a Wishart matrix with identity covariance matrix are stochastically increasing in sequence (Dykstra and Newett (1978)), and since each factor of the integrand in (2.14) is nondecreasing in $\psi_i$, Theorem 1 of Dykstra, Newett and Thompson (1973) preserves the desired inequality when the product sign is brought outside the expectation sign.

3. APPLICATIONS.

Since the right side of the (2.6) is just the product of central chi-squared probabilities, Corollary 2 essentially states that chi-squared random variables which are quadratic forms of a multivariate normal vector are always positively orthant dependent as defined in Dykstra et al. (1973).

However an example of Sidak's (1971) shows that stronger forms of positive dependence like "association", stochastically increasing in sequence" and "positively likelihood ratio dependence" do not hold in general. However, positive orthant
dependence does imply non-negative correlations by the expression for the covariance given in Lehmann (1966).

(c) Corollary 2 seems somewhat related to the bivariate chi-squared inequality given by Jensen (1969). However, Jensen's inequality, while two-sided, only hold for \( k = 2 \), equal degrees of freedom, and identical intervals. Moreover, since bivariate chi-squared random variables defined in this manner are conditionally independent and identically distributed as shown by Shaked (1977), Jensen's inequality would also hold for any Borel set.

(b) Corollary 2 implies that the product of the marginal c.d.f.'s of multivariate chi-squared random variables serves as a lower bound for the joint c.d.f. of the random variables. A similar statement applies for multivariate F random variables such as discussed by Schuurmann, Krishnaiah and Chattopadhyay (1975). This follow by conditioning on the independent denominator, applying Corollary 2, and then using Lemma 3.

(c) If goodness of fit statistics are defined on different co-ordinates of multivariate data, then the asymptotic chi-squared distributions, assuming the null hypotheses to be true, will satisfy the inequality in Corollary 2. Thus if one rejects the hypothesis that all univariate hypotheses are true whenever a univariate hypothesis is rejected, he will, asymptotically, have a conservative estimate of the significance level if he treats the individual tests as being
(d) An application of Corollary 3 involves simultaneous inference in multivariate linear regression. That is, let \( Y \) be a \( N \times p \) data matrix of \( N \) independent observations on \( p \) responses, \( X \) be a \( N \times q \) design matrix of fixed known independent variables, \( B \) be a \( q \times p \) matrix of parameters and \( E \) be a \( N \times p \) matrix of random errors whose rows are distributed independently as \( n(Q, \Sigma) \) random vectors. It is well known from least squares theory that the estimator of \( B \) which minimizes \( \text{Tr}[(Y - XB)'(Y - XB)] \) is given by

\[
\hat{B} = (X'X)^{-1}X'Y.
\]

assuming \( X'X \) is of full rank. The distribution of the row-wise rolled out version of \( \hat{B} \) is then multivariate normal with covariance matrix \( (X'X)^{-1}\Sigma \) and mean equal to the row-wise rolled out \( B \).

Moreover, \( Q_E = Y'Y - \hat{B}'X'XB \) is independent of \( \hat{B} \) and possesses a central Wishart distribution with \( N - q \) degrees of freedom and covariance matrix \( \Sigma \). Thus if we construct ellipsoidal confidence regions for the \( i^{th} \) row of \( B \) based on the \( i^{th} \) row of \( \hat{B} \) and \( Q_E \) from the expression

\[
P(\hat{b}_i - b_i)'Q_E^{-1}(\hat{b}_i - b_i) \leq c_i) = 1 - \alpha_i,
\]

then Corollary 3 guarantees that the confidence coefficient for all the ellipsoidal regions to contain the respective para-
meters must be at least \( \Pi(1 - \alpha_1) \). This generalizes the known comparable result for univariate linear regression.

(c) Siotani (1959) is concerned with

\[
\hat{T}_{\text{MAX}} \leq \max_h \left( z_h^2 S^{-1} z_h \right)
\]

where \( z_1, \ldots, z_N \) is a random sample from a \( n(0, \Sigma) \) distribution and \( S \) possesses an independent \( \mathcal{W}(v, \Sigma) \) distribution. Siotani approximates the distribution of \( \hat{T}_{\text{MAX}}^2 \) by using Bonferroni inequalities. However the product inequality given in Theorem 3 will be closer to the true probabilities and hence could be used to improve Siotani's approximations.
REFERENCES


