ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION FOR LOCATIONS AND DESIGN (U)

JUN 76 I OLKIN, M SOBEL, Y L TONG

N00014-75-C-0561

END

DATE FILMED
2-77
ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION
FOR LOCATION AND SCALE PARAMETER FAMILIES

by

I. OLKIN, M. SOBEL and Y.L. TONG

TECHNICAL REPORT NO. 174
JUNE 28, 1976

SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT N00014-75-C-0561 (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government
Approved for public release; distribution unlimited.

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION
FOR LOCATION AND SCALE PARAMETER FAMILIES

BY

I. Olkin, M. Sobel and Y. L. Tong
Stanford University; University of California, Santa Barbara; University of Nebraska

Abstract

The usual formulation of the problem of selecting the best of k populations has been studied from the point of view of designing an experiment that will guarantee a certain probability of making a correct selection (PCS). However, in retrospective studies the sample has already been taken and we wish to estimate the probability of making a correct selection using this data. The present study deals with obtaining a class of upper and lower bounds for the true PCS and its estimator using the methods of majorization. We also study the large sample behavior of the estimator of the PCS, which turns out to be asymptotically normal.
1. Introduction.

Consider an independent observation, \( X_{ij} \), from \( k \) populations \( F(x, \theta_i) \), \( i = 1, \ldots, k \), \( j = 1, \ldots, n \). The experimenter wishes to select the population associated with the largest parameter \( \theta_{[k]} \). To do this we define an appropriate statistic

\[
Y_i = Y(X_{i1}, \ldots, X_{in}),
\]

and use the decision rule "designate the population corresponding to the largest \( Y_i \) value as the one with the largest \( \theta \)-value." If the c.d.f., \( G(y, \theta_i) \), of \( Y_i \) has the form \( G(y-\theta_i) \), then the probability of a correct selection (PCS) is

\[
\alpha(\theta_1, \ldots, \theta_k) = \int_{-\infty}^{\infty} \left[ \frac{1}{k-1} \sum_{i=1}^{k-1} G(y+\theta_i) \right] dG(y),
\]

where \( \theta_1 \leq \cdots \leq \theta_k \), \( \theta_i = \theta_{[k]} - \theta_{[i]} \).

In the usual methods employing the "indifference zone" approach, as introduced into ranking and selection by Bechhofer (1954), we determine the smallest fixed sample size \( n \) needed in order to satisfy a preassigned probability requirement. This led to a consideration of least favorable configuration, which frequently takes the form \( \theta_1 = \cdots = \theta_{k-1} = \theta_0 \) and suggests an inequality for \( \alpha(\theta_1, \ldots, \theta_{k-1}) \), where we "move downhill" in the sense of majorization to the case of equal \( \theta_i \)-values, regardless of whether the common value \( \theta_0 \) is specified or not. Because this integral is often difficult to evaluate (especially in high dimensions), we are interested in both upper and lower bounds.
that hold for any true configuration, without regard to any least favorable configuration.

In the present paper we consider the problem of estimating the true PCS by using information provided by the sample. This is extremely important for those problems in which we are carrying out a retrospective study -- rather than designing an experiment.

If \( \hat{Y} \) is a consistent estimator of \( \hat{\theta} \), then we may let

\[ Y_{[1]} \leq \cdots \leq Y_{[k]} \]

be the ordered \( Y \)-values, and define

\[ \hat{\theta}_i = Y_{[k]} - Y_{[i]} \quad i = 1, \ldots, k-1. \]

Then \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_{k-1}) \) will be a consistent estimator of \( \theta \), and \( \alpha(\hat{\theta}_1, \ldots, \hat{\theta}_{k-1}) \) will be a consistent estimator of \( \alpha(\theta_1, \ldots, \theta_{k-1}) \).

As a consequence, we wish to further study the behavior of the integral \( \alpha(\hat{\theta}_1, \ldots, \hat{\theta}_{k-1}) \), or \( \alpha(\hat{\theta}_1, \ldots, \hat{\theta}_{k-1}) \), as a function of the configuration of the parameters. In particular, we are interested in the following question: how does the PCS behave if some of the \( \theta \)-values are replaced by their average? This problem is studied in the next section.

Remark: The above development was based on a location parameter family. However, a similar analysis holds for scale parameter families. In this case the true PCS (1.1) is replaced by

\[ \alpha(\theta_1, \ldots, \theta_{k-1}) = \int_{-\infty}^{\infty} \frac{k-1}{i=1} G(\theta_i y) dG(y), \]
where in the case of the largest scale parameter we define $\xi_1 = \theta^{[k]} / \theta^{[1]}$ 
($i = 1, 2, \ldots, k-1$) and hence $\xi_1 \geq \cdots \geq \xi_{k-1} \geq 1$. For the case of the 
smallest scale parameter we define $\xi_{k-1+1} = \theta^{[i]} / \theta^{[1]}$ and then (1.2) 
and the above inequalities on $\xi_1$ still hold.

2. The Main Bounds.

The key motivating point is that under certain circumstances we 
can approximate $\alpha(\xi)$ by

\[
\alpha(\xi_1, \ldots, \xi_p) \approx \alpha(\xi_1', \ldots, \xi_1', \xi_2', \ldots, \xi_r', \ldots, \xi_r'),
\]

where $\sum p_i = p = k-1$ and the $\xi_i'$s are certain averages of the $\xi_i$'s. 
If the approximation sign is an inequality sign, then we are "moving 
downhill" by averaging and further averaging will move us more downhill. 
This suggests the use of majorization, and we show how this arises.

We now develop the main result. For any fixed $r \leq p$, let 
a_1, \ldots, a_r be distinct numbers satisfying

\[
0 = a_0 < a_1 < a_2 < \cdots < a_r = p,
\]

$p_{\nu} = a_{\nu} - a_{\nu-1}$, $\nu = 1, \ldots, r$, and define the partition

\[
\xi(A) = (\xi_1, \ldots, \xi_a, \xi_{a+1}, \ldots, \xi_{a_2}, \ldots, \xi_{a_{r-1}+1}, \ldots, \xi_p).
\]
We also define the means within each cluster:

$$
\bar{a}_v = \frac{\delta_{v-1} + \cdots + \delta_v}{a_v - a_{v-1}}, \quad v = 1, \ldots, r,
$$

and generate a vector of "averaged" $\delta$-values:

$$
\bar{\delta}(A) = (\bar{\delta}_1(A), \ldots, \bar{\delta}_p(A))
$$

$$
= (\frac{\delta_1}{p_1}, \ldots, \frac{\delta_1}{p_1}, \frac{\delta_2}{p_2}, \ldots, \frac{\delta_2}{p_2}, \ldots, \frac{\delta_r}{p_r}, \ldots, \frac{\delta_r}{p_r}),
$$

where $A$ denotes the set $A = \{a_1, \ldots, a_r\}$.

Let $A \subset B$ mean that $B$ is a refinement of $A$. A question that arises immediately is whether we can compare the $\alpha$-values of two partitions $A$ and $B$ if $A \subset B$?

**Theorem 1.** If $\log G(y)$ is concave, and $A \subset B$, then

$$
(2.1) \quad \alpha(\bar{\delta}_1, \ldots, \bar{\delta}_p) \leq \alpha(\bar{\delta}_1(B), \ldots, \bar{\delta}_p(B)) \leq \alpha(\bar{\delta}_1(A), \ldots, \bar{\delta}_p(A)).
$$

The proof of this result is given in the next section. We note that the $\alpha$'s in (2.1), though written as a single integral, are actually $p$-fold integrals. However, we would also like to have bounds that reduce the dimensionality. To accomplish this we consider the following integral. For any partition $A$, define
(2.2) \[ \mathcal{B}(\varepsilon_1(A), \ldots, \varepsilon_p(A)) \]

\[ = \left[ \int_{-\infty}^{\infty} \frac{a_1}{a_1+1} G(y+\varepsilon_1) \, dG(y) \right] \left[ \int_{-\infty}^{\infty} \frac{a_2}{a_2+1} G(y+\varepsilon_1) \, dG(y) \right] \]

\[ \ldots \left[ \int_{-\infty}^{\infty} \frac{a_p}{a_p+1} G(y+\varepsilon_1) \, dG(y) \right]. \]

Here the reduction in dimensionality is in terms of the size of the intervals \( a_v - a_{v-1} \) (\( v = 1, 2, \ldots, r \)). Note that when \( r = p \), then \( \mathcal{B}(\varepsilon_1, \ldots, \varepsilon_p) = \alpha(\varepsilon_1, \ldots, \varepsilon_p) \).

**Theorem 2.** If \( A \subset B_A \), then

\[ (2.3) \quad \mathcal{B}(\varepsilon_1(B), \ldots, \varepsilon_p(B)) \leq \mathcal{B}(\varepsilon_1(A), \ldots, \varepsilon_p(A)) \leq \alpha(\varepsilon_1, \ldots, \varepsilon_p). \]

Of particular interest are the extremes obtained by the partition \( A = \{1, \ldots, p\} \), in which case we obtain

\[ (2.4) \quad \prod_{i=1}^{p} \int_{-\infty}^{\infty} G(y+\varepsilon_1) \, dG(y) \leq \alpha(\varepsilon_1, \ldots, \varepsilon_p) = \int_{-\infty}^{\infty} \prod_{i=1}^{p} G(y+\varepsilon_1) \, dG(y). \]

**Remarks:**

(i) We should note that Theorem 2 is due to Kimball (1951).

(ii) If \( G \) is a normal distribution, then the right-hand side of (2.3) can be regarded as the integral of a multivariate normal distribution with variances 1 and correlations \( \rho = 1/2 \). This integral has been considered by Milton (1970, 1972). If \( G \) is normal then the left-hand side of (2.4) is easily shown to be \( \prod_{i=1}^{p} G(\varepsilon_1/\sqrt{2}) \), so that no integration is required.
(iii) If $G(y, \theta) = G(y/\theta)$, with $G(0) = 0$, is a scale parameter family, then we define geometric means instead of arithmetic means.

For any partition $A$ we have

$$\bar{a}_1 = \left( \prod_{1}^{a_1} \bar{a}_1 \right)^{1/a_1}, \quad \bar{a}_2 = \left( \prod_{a_1+1}^{a_2} \bar{a}_1 \right)^{1/(a_2-a_1)},$$

$$\cdots, \bar{a}_r = \left( \prod_{a_{r-1}+1}^{a_r} \bar{a}_1 \right)^{1/(a_r-a_{r-1})}.$$

The results of Theorems 1 and 2 hold with the obvious changes.

3. **Proof of Inequalities via Majorization.**

We first review several facts concerning the partial ordering majorization. A vector $u_1, \ldots, u_p$ is said to majorize a vector $v_1, \ldots, v_p$, written $u \succ v$, if, upon a rearrangement with $u_1 \geq \cdots \geq u_p$, $v_1 \geq \cdots \geq v_p$, we have $\sum_{1}^{m} u_i \geq \sum_{1}^{m} v_i$, $m = 1, \ldots, p-1$, $\sum_{1}^{p} u_i = \sum_{1}^{p} v_i$. The class of (differentiable) order preserving functions $f(x_1, \ldots, x_p)$ are called Schur functions, or Schur-convex functions, and are defined by the property

$$\left( \frac{\partial f(x)}{\partial x_i} - \frac{\partial f(x)}{\partial x_j} \right) (x_i - x_j) \geq 0 \quad \text{for all } i, j = 1, \ldots, p.$$  

(For the Schur-concave functions we merely turn the inequality around.)

Then one characterization of the Schur functions is through the fact that $u \succ v$ if and only if $f(u) \geq f(v)$ for all Schur-convex functions.
Finally, there is another characterization in terms of doubly-stochastic matrices \( P = (p_{ij}) \), \( p_{ij} \geq 0 \), \( \sum_j p_{ij} = \sum_i p_{ij} = 1 \), namely, 

\( u \succ v \) if and only if \( v = u^P \) for some doubly stochastic matrix \( P \).

This says that one "moves downhill" from \( u \) to \( v \) (in the sense of majorization) by taking a convex combination of elements of \( u \).

We need the following fact.

**Fact.** (Marshall and Olkin (1974)). If \( \log G(y) \) is concave, then

\[
\prod_1^p G(y+\xi_j) \text{ is Schur-concave. Consequently, } \alpha(\delta) = \int_{-\infty}^{\infty} \prod_1^p G(y+\xi_j) \, dG(y)
\]

is Schur-concave.

The proof of Theorem 1 is a direct consequence of the above facts. Recall that in (2.1) we have that if \( A \subset B \) then

\[
\alpha(\delta) \leq \alpha(\delta(B)) \leq \alpha(\delta(A)).
\]

We need only prove the left-hand inequality since the right-hand inequality is obtained by a repeated argument where \( \alpha(\delta(B)) \) is used in place of \( \alpha(\delta) \) and \( \alpha(\delta(A)) \) is used in place of \( \alpha(\delta(B)) \).

It is immediate that \( \delta \succ \delta(B) \) since \( \delta(B) = \delta P \), where

\( P = \text{block diag}(P_{y_1}, \ldots, P_{y_r}) \) is a doubly-stochastic matrices with blocks \( P_{y_j} \) on the diagonal consisting of \( 1/(a_{ij} - a_{i-1,j}) \). This majorization plus the Schur concavity proves Theorem 1.

Theorem 2 follows from an inequality of Kimball (1951) and to apply this it is sufficient that the distribution functions \( G(y+\xi_j) \) all be positive and monotone in the same direction for all real \( y \).
4. **Approximations.**

We now make use of the bounds to approximate \( \alpha(\hat{\sigma}) \). Although \( \alpha(\hat{\sigma}) \) can be computed exactly with the help of quadrature methods, the necessary programming is not always available and the computations may be costly. Thus, it is desirable to be able to approximate the integrals.

As an illustration of the bounds, consider a numerical example for the case of normality with common \( \sigma = 1 \), and the location parameter model with \( \hat{\sigma} = (3.2, 2.7, 1.5, 1.9, 1.7) \); we wish to evaluate \( \alpha(\hat{\sigma}) \).

The exact value is \( \alpha = 0.8016 \). To compute our smallest lower bound note that \( \int g(y; \hat{\sigma}) \, d\Phi(y) = \Phi(5\sqrt{5}) \), so that we obtain \( \int \frac{5}{\sqrt{5}} \, \phi(5\sqrt{5}) = 0.7443 \)

The largest upper bound is \( \int \phi^2(y; \hat{\sigma}) \, d\Phi(y) = 0.8483 \), obtained from Milton (1970).

We now consider some refinements. Suppose we use \( A = (1, 3, 5) \) so that we cluster into 3 groups \( \{3.2\}, \{2.7, 2.5\}, \{1.9, 1.7\} \), then the lower bound \( B(5(A)) \) in (2.2) is given by

\[
(4.1) \quad \int \phi(y; 3.2) \, d\Phi(y) \cdot \int \phi(y; 2.7) \, \phi(y; 2.5) \, d\Phi(y) \\
\quad \cdot \int \phi(y; 1.9) \, \phi(y; 1.7) \, d\Phi(y),
\]

which may be obtained from the Milton (1970) tables to be 0.7698.

But, if we now approximate (4.1) by

\[
(4.2) \quad \int \phi(y; 3.2) \, d\Phi(y) \int \phi^2(y; 2.6) \, d\Phi(y) \int \phi^2(y; 1.8) \, d\Phi(y),
\]

then we obtain the result 0.7717.
If we use fewer clusters, namely, $A = \{3, 5\}$, yielding $[3.2, 2.7, 2.5] \ [1.9, 1.7]$ then we obtain the approximation

$$(4.3) \quad \int \phi^3(y+2.8) \, d\phi(y) \int \phi^2(y+1.8) \, d\phi(y)$$

to yield a value of $0.7802$.

It is intuitively clear that if we can take advantage of the close proximity of any two $\hat{\sigma}$ values, then the worst case for getting close bounds and approximations may be the equally spaced case. We have, therefore, attached at the end of this paper a table (Table 1) for the normal distribution under equally spaced configurations with a common spacing $d$. The table gives the numerical values of the integral

$$\int \prod_{i} \phi(y+id) \, d\phi(y) = P[Z_i \leq \frac{id}{\sqrt{2}}, \ i = 1, \ldots, p]$$

for $d = 0.00(0.02)0.20(0.10)4.00$ and $k = 2(1)10, 12, 15$ and 20; where $\phi$ is the standard normal cdf.

5. **Limiting Distribution of the Estimator of the PCS for the Normal Location Parameter Model.**

In this section we obtain the asymptotic distribution of the estimator, $\alpha(\hat{\sigma})$, of the true PCS $\alpha(\sigma)$ when the underlying distribution is normal. Although results for the location parameter case are obtained, similar results may be obtained for scale parameter families.

Let $Y = (Y_1, \ldots, Y_k)$ and $\hat{Y} = (Y_{[1]}, \ldots, Y_{[k]})$, where $Y_{[1]} \leq \cdots \leq Y_{[k]}$ are the ordered $Y$-values. Similarly, we define
\[ \theta = (\theta_1, \ldots, \theta_k) \) and \( \tilde{\theta} = (\theta_{[1]}, \ldots, \theta_{[k]}) \). Finally we let
\[ \Omega = \{ \theta : \theta_{[1]} < \cdots < \theta_{[k]} \} \] with strict inequalities between \( \theta \)-values. For any two vectors \( u \) and \( v \) we write \( u \leq v \) to mean \( u_i \leq v_i \) for all \( i \).

Now suppose \( x(\sqrt{n} (Y_i - \theta_i)) \overset{d}{\to} \mathcal{N}(0, I_k) \). We first obtain the asymptotic distribution of \( \tilde{Y} \), then of estimators, \( \hat{\theta}_i \), of \( \theta \), and finally of \( \alpha(\hat{\theta}) \). In the following development \( \sigma \) is assumed to be known, and we take it to be unity.

**Theorem 3.** As \( n \to \infty \), \( x(\sqrt{n} (Y_i - \theta_i)) \overset{d}{\to} \mathcal{N}(0, I_k) \) for every \( \tilde{\theta} \in \Omega \), where \( I_k \) is the identity matrix of order \( k \).

**Proof.** The result follows by bounding the right-hand side of

\[
(5.1) \quad |P (|\sqrt{n} (Y - \tilde{\theta})| \leq z) - \prod_{i=1}^{k} \phi(z_i)|
\]

\[
\leq |P (|\sqrt{n} (Y - \tilde{\theta})| \leq z, Y = \tilde{Y}) - P (|\sqrt{n} (Y - \tilde{\theta})| \leq z)|
\]

\[
+ P (|\sqrt{n} (Y - \tilde{\theta})| \leq z, Y \neq \tilde{Y}) + |P (|\sqrt{n} (Y - \tilde{\theta})| \leq z) - \prod_{i=1}^{k} \phi(z_i)|,
\]

for every \( z = (z_1, \ldots, z_k) \) and every \( \tilde{\theta} \in \Omega \).

Let \( \epsilon_1 > 0, \epsilon_2 > 0 \) be arbitrary fixed real numbers such that
\[ 2\epsilon_1 \leq \min_{i=1}^{k} (\theta_{[i]} - \theta_{[i-1]}) \]. Since \( \tilde{Y} \) is a consistent estimator of \( \tilde{\theta} \) and the \( Y_i \)'s are independent, it follows that for every \( \tilde{\theta} \in \Omega \)
and every \( z = (z_1, \ldots, z_k) \) there exist numbers \( N_1 \equiv (\epsilon_1, \epsilon_2) \) and
\[ N_2 \equiv (\epsilon_2, z) \) such that for \( n > N_1 \), \( n \geq 1, \quad P (Y_i = \tilde{Y}) \geq P \left(|Y_i - \tilde{\theta}_i| \leq \epsilon_1 \right) \]
\[ i = 1, \ldots, k \) \( \geq 1 - \epsilon_2 \), and for \( n > N_2 \), \( |P (\sqrt{n} (Y - \tilde{\theta})| \leq z) - \prod_{i=1}^{k} \phi(z_i)| \leq \epsilon_2 \).
Here $\varepsilon_1$ is small enough so that $|v_1 - \bar{v}_1| \leq \varepsilon_1$ implies that $Y = \bar{Y}$, so that the first difference on the right side of (5.1) is zero. Then for $n > \max(N_1, N_2)$, (5.1) is bounded by $2\varepsilon_2$, which completes the proof.

As a consequence of Theorem 3, we have the

Corollary 4. For every $\tilde{\varepsilon} \in \mathbb{R}$, $\mathcal{L}(\sqrt{n} (\hat{\varepsilon} - \varepsilon))$ is asymptotically normal with mean vector zero, variances $2$ and covariances $1$.

Now consider the PCS for the normal case

$$\alpha(\varepsilon) = a(\varepsilon_1, \ldots, \varepsilon_p) = \int_{-\infty}^{\infty} \prod_{i=1}^{p} \phi(t + \sqrt{n} \varepsilon_i) \, d\phi(t),$$

and its estimator,

$$\alpha(\hat{\varepsilon}) = a(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_p) = \int_{-\infty}^{\infty} \prod_{i=1}^{p} \phi(t + \sqrt{n} \hat{\varepsilon}_i) \, d\phi(t).$$

Theorem 4. $\mathcal{L}(\sqrt{n} (a(\hat{\varepsilon}) - a(\varepsilon))) \xrightarrow{d} \eta(0, \nu_n)$, where $\nu_n(\varepsilon) = \sum_1^p d_1^{\varepsilon_1} + \sum_1^{p-1} d_1^{\varepsilon_1} d_2^{\varepsilon_2}$, and

$$d_1 = \frac{\sqrt{n}}{\sqrt{2}} \phi(\sqrt{n} \varepsilon_1) \int_{-\infty}^{\infty} \prod_{j \neq 1} \phi \left( t + \sqrt{n} \hat{\varepsilon}_j - \frac{\sqrt{n} \varepsilon_1}{2} \right) \, d\phi(t).$$

Proof. A Taylor expansion about $\varepsilon = \varepsilon_0 = (\varepsilon_{10}, \ldots, \varepsilon_{p0})$ yields

$$\sqrt{n} (a(\hat{\varepsilon}) - a(\varepsilon_0)) = \sqrt{n} \sum_1^p d_1 (\hat{\varepsilon}_1 - \varepsilon_{10}) + \left( \sqrt{n} \sum_1^p (\hat{\varepsilon}_1 - \varepsilon_{10})^2 \frac{\partial^2 a(\varepsilon)}{\partial \varepsilon_1^2} \right) + \left. \frac{\partial^2 a(\varepsilon)}{\partial \varepsilon_1 \partial \varepsilon_j} \right|_{\varepsilon = \varepsilon_0} \left( \hat{\varepsilon}_1 - \varepsilon_{10} \right) \left( \hat{\varepsilon}_j - \varepsilon_{j0} \right)$$

$$= w_1(n) + w_2(n),$$
where \( \hat{\varepsilon}_1 \leq \varepsilon_{10} \leq \varepsilon_1 \). Corollary 1 and the boundedness condition

\[
\left| \frac{d a(\varepsilon)}{d \varepsilon_1} \right| \leq c/\varepsilon_1 \implies \mathcal{L}(w_1(n)) \xrightarrow{d} \mathcal{N}(0, v_n). \]

Consequently, the proof will be complete if we show that \( \text{plim } w_2(n) = 0 \). To see this, we use the fact that \( x^n(x) \) and \( x^2 \phi(x) \) are bounded, so that there exist positive constants \( c_1, c_2, c_3 \) such that

\[
\left| \frac{d^2 a(\varepsilon)}{d \varepsilon_1^2} \right| \leq \frac{c_1}{\varepsilon_1}, \quad \left| \frac{d^2 a(\varepsilon)}{d \varepsilon_1 d \varepsilon_j} \right| \leq \frac{c_2}{\varepsilon_1} + \frac{c_3}{\varepsilon_1^2}. \]

If we define

\[
U(n) = \sqrt{n} \left( \frac{1}{2} \sum_{i=1}^p \left( \hat{\varepsilon}_i - \varepsilon_{10} \right)^2 + \frac{1}{2} \sum_{i,j} \left| (\hat{\varepsilon}_i - \varepsilon_{10})(\hat{\varepsilon}_j - \varepsilon_{10}) \right| \left( \frac{c_2}{\varepsilon_1} + \frac{c_3}{\varepsilon_1^2} \right) \right),
\]

then \( \text{plim } U(n) = 0 \). Consequently, if \( \mathcal{Q} = \{ \hat{\varepsilon}_i > \varepsilon^*, i = 1, ..., p \} \), for any \( \varepsilon^* < \varepsilon_f \), it follows that

\[
(5.3) \quad P(|w_2(n)| > \varepsilon) = P(|w_2(n)| > \varepsilon, \mathcal{Q}) + P(|w_2(n)| > \varepsilon, \overline{\mathcal{Q}}) \leq P(U_n > \varepsilon) + P(\overline{\mathcal{Q}}) = \varepsilon.
\]

But for fixed \( \varepsilon > 0 \) and sufficiently large \( n \), each term on the right-hand side of (5.3) is \( \leq \varepsilon \). ||

The asymptotic variance \( v_n \) defined in Theorem 1 is quite complicated. We may use \( \hat{\varepsilon}_1 \) in \( v_n(\hat{e}) \) to yield a confidence interval for the PCS, \( a(\hat{e}) \), namely,
\[
\frac{\sqrt{n} \left( a(\varepsilon) - a(\check{\varepsilon}) \right)}{\sqrt{\nu_n(\check{\varepsilon})}} \sim n(0, 1).
\]

However, we may also wish to approximate the asymptotic variance \( \nu_n(\check{\varepsilon}) \) or its estimate \( \nu_n(\hat{\varepsilon}) \). We again use majorization to obtain some bounds.

To do this we require the following facts.

**Fact 1.** For fixed \( i \), and any \( z \), the function

\[
\prod_{j \neq i} \left( \frac{t}{\sqrt{2}} + \sqrt{n} \varepsilon_j - \frac{\sqrt{n} \varepsilon_i}{2} \right)
\]

is a Schur-concave function in \( \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_{i+1}, \ldots, \varepsilon_p \).

This follows from the result of Section 3.

**Fact 2.** For fixed \( i \) and \( \varepsilon_1 \geq \cdots \geq \varepsilon_p \),

\[
\bar{\varepsilon} = \left( \varepsilon_1 - \frac{\varepsilon_i}{2}, \ldots, \varepsilon_{i-1} - \frac{\varepsilon_i}{2}, \varepsilon_{i+1} - \frac{\varepsilon_i}{2}, \ldots, \varepsilon_p - \frac{\varepsilon_i}{2} \right)
\]

\[
\leq \left( p \bar{\varepsilon} - \frac{(p+1)}{2(p-1)} \varepsilon_i, \ldots, \frac{1}{p-1} \varepsilon - \frac{(p+1)}{2(p-1)} \varepsilon_i \right) = \bar{\varepsilon}_L,
\]

where \( \bar{\varepsilon} = \frac{\sum_{l=1}^{p} \varepsilon_l}{p} \).

To prove this result we note that the elements of \( \bar{\varepsilon}_L \) are means of the elements in \( \bar{\varepsilon} \).

**Fact 3.** For fixed \( i \) and \( \varepsilon_1 \geq \cdots \geq \varepsilon_p \),
\( \tilde{e} = \left( e_1 - \frac{3}{2}, \ldots, e_{i-1} - \frac{3}{2}, e_{i+1} - \frac{1}{2}, \ldots, e_p - \frac{1}{2} \right) \)

\( \prec \left( p\tilde{e} - (p-2)e_p - \frac{3}{2} e_1, e_p - \frac{1}{2}, \ldots, e_p - \frac{1}{2} \right) \equiv \tilde{e}_U. \)

To prove this majorization we first note that

\[ p\tilde{e} - (p-2)e_p - \frac{3}{2} e_1 \geq e_p - e_1/2. \]

Further, the sum of all the elements in each vector is equal to \( p\tilde{e} - \frac{1}{2}(p+1)e_1. \) Consequently, we need show that for \( m = 1, \ldots, p-1, \)

\[ m(e_p - \frac{1}{2}) \leq \sum_{j \neq i}^{p-m+1} e_j - \frac{m e_i}{2}. \]

But this inequality is equivalent to

\[ \frac{p}{p-m+j} \frac{e_j}{(p-1)} \geq \frac{e_p - (e_1 - e_i)/2}{e_j}, \]

which clearly holds.

As a consequence of the majorization

\[ \tilde{e}_L \prec \tilde{e} \prec \tilde{e}_U, \]

we have that for every \( z \) and every \( i \)

\[ \sum_{j \neq i} \phi \left( \frac{t}{\sqrt{2}} + \sqrt{n} e_j - \sqrt{n} \frac{e_i}{2} \right) \]

\[ \left\{ \begin{array}{l} \leq \phi^{p-1} \left( \frac{t}{\sqrt{n}} - \sqrt{\frac{p}{p-1}} e_1 \right) = D_2(t), \\ \geq \phi \left( \frac{t}{\sqrt{2}} + \sqrt{n} e_p - (p-2) \sqrt{\frac{p}{2}} e_1 \right) \phi^{p-2} \left( \frac{t}{\sqrt{n}} + \sqrt{n} \frac{e_p - \frac{e_1}{2}}{2} \right) = D_1(t). \end{array} \right. \]
Remark. The upper bound is a monotone decreasing function of $\delta_1$, and the lower bound is a monotone increasing function of $\delta_p$. Consequently, if $\delta_1 \geq \delta^*$, we may obtain the bounds where $\delta^*$ is inserted in place of $\delta_1$ in $D_2(t)$ and in place of $\delta_p$ in $D_1(t)$. Call these functions $\tilde{D}_j(t)$, $j = 1, 2$. From the definition of $v_n$, we see that

$$\frac{n}{4\pi} H(\delta) \left[ \int \tilde{D}_1(t) \, d\omega(t) \right]^2 \leq v_n \leq \frac{n}{4\pi} H(\delta) \left[ \int \tilde{D}_2(t) \, d\omega(t) \right]^2$$

where

$$H(\delta) = \sum_{i=1}^{p} \exp\left(-n\delta_i^2/2\right) + \left( \sum_{i=1}^{p} \exp\left(-n\delta_i^2/4\right) \right)^2.$$

The main point is that under certain circumstances we may approximate $v_n$ by $n H(\delta)/4\pi$, which is small for large $n$. This holds when the dispersion of the $\delta$-values is not too large, so that $\delta_1 < 2\delta^*$. In this case both integrals above are close to 1 as $n \to \infty$.

Finally, we note that $H(\delta)$ is Schur-convex in $(\delta_1^2, \ldots, \delta_p^2)$, so that further bounds on $H$ may readily be obtained. In particular, it can be shown via such bounds that $v_n$ decreases exponentially in $n$, so that we can expect reasonably good approximations for moderate values of $n$. 

15
<table>
<thead>
<tr>
<th>(d)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>12</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>.3333</td>
<td>.2500</td>
<td>.2000</td>
<td>.1667</td>
<td>.1429</td>
<td>.1250</td>
<td>.1111</td>
<td>.1000</td>
<td>.0909</td>
<td>.0769</td>
<td>.0625</td>
<td>.0476</td>
</tr>
<tr>
<td>0.02</td>
<td>.3418</td>
<td>.2694</td>
<td>.2118</td>
<td>.1795</td>
<td>.1567</td>
<td>.1396</td>
<td>.1264</td>
<td>.1159</td>
<td>.1074</td>
<td>.0945</td>
<td>.0814</td>
<td>.0683</td>
</tr>
<tr>
<td>0.04</td>
<td>.3504</td>
<td>.2709</td>
<td>.2238</td>
<td>.1928</td>
<td>.1710</td>
<td>.1549</td>
<td>.1425</td>
<td>.1329</td>
<td>.1251</td>
<td>.1135</td>
<td>.1023</td>
<td>.0920</td>
</tr>
<tr>
<td>0.06</td>
<td>.3589</td>
<td>.2815</td>
<td>.2360</td>
<td>.2063</td>
<td>.1857</td>
<td>.1707</td>
<td>.1593</td>
<td>.1506</td>
<td>.1437</td>
<td>.1337</td>
<td>.1246</td>
<td>.1172</td>
</tr>
<tr>
<td>0.08</td>
<td>.3676</td>
<td>.2922</td>
<td>.2484</td>
<td>.2202</td>
<td>.1988</td>
<td>.1862</td>
<td>.1767</td>
<td>.1689</td>
<td>.1630</td>
<td>.1547</td>
<td>.1477</td>
<td>.1429</td>
</tr>
<tr>
<td>0.10</td>
<td>.3762</td>
<td>.3030</td>
<td>.2610</td>
<td>.2343</td>
<td>.2163</td>
<td>.2036</td>
<td>.1945</td>
<td>.1877</td>
<td>.1827</td>
<td>.1761</td>
<td>.1710</td>
<td>.1681</td>
</tr>
<tr>
<td>0.12</td>
<td>.3849</td>
<td>.3139</td>
<td>.2737</td>
<td>.2486</td>
<td>.2319</td>
<td>.2205</td>
<td>.2125</td>
<td>.2068</td>
<td>.2027</td>
<td>.1975</td>
<td>.1940</td>
<td>.1924</td>
</tr>
<tr>
<td>0.14</td>
<td>.3936</td>
<td>.3249</td>
<td>.2865</td>
<td>.2630</td>
<td>.2478</td>
<td>.2376</td>
<td>.2279</td>
<td>.2204</td>
<td>.2124</td>
<td>.2168</td>
<td>.2165</td>
<td>.2157</td>
</tr>
<tr>
<td>0.16</td>
<td>.4024</td>
<td>.3360</td>
<td>.2994</td>
<td>.2775</td>
<td>.2547</td>
<td>.2449</td>
<td>.2367</td>
<td>.2279</td>
<td>.2197</td>
<td>.2192</td>
<td>.2192</td>
<td>.2191</td>
</tr>
<tr>
<td>0.18</td>
<td>.4111</td>
<td>.3470</td>
<td>.3124</td>
<td>.2922</td>
<td>.2737</td>
<td>.2639</td>
<td>.2559</td>
<td>.2479</td>
<td>.2397</td>
<td>.2392</td>
<td>.2391</td>
<td>.2389</td>
</tr>
<tr>
<td>0.20</td>
<td>.4209</td>
<td>.3581</td>
<td>.3254</td>
<td>.3067</td>
<td>.2899</td>
<td>.2828</td>
<td>.2755</td>
<td>.2679</td>
<td>.2597</td>
<td>.2592</td>
<td>.2591</td>
<td>.2589</td>
</tr>
<tr>
<td>0.30</td>
<td>.4636</td>
<td>.4136</td>
<td>.3901</td>
<td>.3786</td>
<td>.3730</td>
<td>.3704</td>
<td>.3692</td>
<td>.3687</td>
<td>.3684</td>
<td>.3683</td>
<td>.3683</td>
<td>.3693</td>
</tr>
<tr>
<td>0.40</td>
<td>.5067</td>
<td>.4679</td>
<td>.4524</td>
<td>.4462</td>
<td>.4439</td>
<td>.4432</td>
<td>.4429</td>
<td>.4428</td>
<td>.4428</td>
<td>.4428</td>
<td>.4428</td>
<td>.4428</td>
</tr>
<tr>
<td>0.50</td>
<td>.5487</td>
<td>.5199</td>
<td>.5014</td>
<td>.4907</td>
<td>.4902</td>
<td>.4907</td>
<td>.4907</td>
<td>.4907</td>
<td>.4907</td>
<td>.4907</td>
<td>.4907</td>
<td>.4907</td>
</tr>
<tr>
<td>0.60</td>
<td>.5932</td>
<td>.5686</td>
<td>.5633</td>
<td>.5621</td>
<td>.5619</td>
<td>.5619</td>
<td>.5619</td>
<td>.5619</td>
<td>.5619</td>
<td>.5619</td>
<td>.5619</td>
<td>.5619</td>
</tr>
<tr>
<td>0.70</td>
<td>.6301</td>
<td>.6164</td>
<td>.6138</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
<td>.6134</td>
</tr>
<tr>
<td>0.80</td>
<td>.6638</td>
<td>.6546</td>
<td>.6533</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
<td>.6532</td>
</tr>
<tr>
<td>0.90</td>
<td>.6976</td>
<td>.6916</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
<td>.6892</td>
</tr>
<tr>
<td>1.00</td>
<td>.7288</td>
<td>.7253</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
<td>.7251</td>
</tr>
<tr>
<td>1.10</td>
<td>.7574</td>
<td>.7574</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
<td>.7573</td>
</tr>
<tr>
<td>1.20</td>
<td>.7836</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
<td>.7825</td>
</tr>
<tr>
<td>1.30</td>
<td>.8074</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
<td>.8068</td>
</tr>
<tr>
<td>1.40</td>
<td>.8293</td>
<td>.8297</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
<td>.8287</td>
</tr>
<tr>
<td>1.50</td>
<td>.8485</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
<td>.8483</td>
</tr>
<tr>
<td>1.60</td>
<td>.8661</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
<td>.8660</td>
</tr>
</tbody>
</table>

Table 1. Numerical Values of \(\int_0^\infty \prod_{i=1}^p d(y) dy\)
for \(d=0.00(0.02)0.20(0.1)4.00\), \(p=2(1)10,12,15,20\)
References


ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION FOR LOCATION AND SCALE PARAMETER FAMILIES

I. Olkin, M. Sobel and Y.L. Tong

DEPT. OF OPERATIONS AND DEPT. OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

OCEANS & ATMOSPHERIC PROGR. CODE 436
OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA 22217

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

This report was also supported by National Science Foundation
Grant MPS 7509450-000

MULTIPLE COMPARISONS
RANKING & SELECTION
MAJORIZATION
PROBABILITY OF CORRECT SELECTION

The usual formulation of the problem of selecting the best of k populations has been studied from the point of view of designing an experiment that will guarantee a certain probability of making a correct selection (PCS). However, in retrospective studies the sample has already been taken and we wish to estimate the probability of making a correct selection using this data. The present study deals with obtaining a class of upper and lower bounds for the true PCS and its estimator using the methods of majorization. We also study the large sample behavior of the estimator of the PCS, which turns out to be asymptotically normal.