On the Small-Sample Properties of the Olkin-Sobel-Tong Estimator of the Probability of Correct Selection

by

Frederick W. Faltin$^1$ and Charles E. McCulloch$^2$

Air Products and Chemicals, Inc.$^3$ and Florida State University$^3$

FSU Statistics Report No M581

$^1$Frederick W. Faltin is Staff Statistician in Management Sciences at Air Products and Chemicals, Inc. Trexlertown, PA 18105

$^2$Charles E. McCulloch is Assistant Professor of Statistics, Florida State University, Tallahassee, FL 32306

$^3$Research partially supported by U.S. Army Research Office-Durham Contract DAAG-29-80-C-0036, Office of Naval Research Contract N00014-75-C-0586 while the authors were at Cornell University.
Abstract

In the problem of selecting the best of k populations, Olkin, Sobel and Tong [1976] have introduced the important idea of a posteriori analysis of the data, as opposed to the usual formulation, in which design of the experiment is the major consideration. They considered the large-sample properties of an estimator which has been discussed further by Gibbons, Olkin and Sobel [1977], Gupta and Pancheckesan [1979] and Tong [1980]. In this paper we study the small sample performance of their estimator, analytically for \( k = 2 \) populations and via Monte Carlo simulation for \( k \geq 2 \) populations in the normal means, common, known variance case. This small-sample performance is found to possess some serious shortcomings.

Keywords and Phrases: ranking and selection, a posteriori analysis, estimating the probability of correct selection, small-sample results.
1. Introduction and Notation

Let $X_{ij}$, $1 \leq i \leq k$, $1 \leq j \leq n$ be independent observations from $k$ populations with c.d.f.'s $F(x; \theta_i)$. We wish to select the population associated with the largest $\theta_i$, and we consider decision procedures as follows: Define an appropriate statistic $Y_i = Y(X_{i1}, X_{i2}, \ldots, X_{in}) = \frac{1}{n} \sum_{j=1}^{n} X_{ij} = \bar{X}_i$.

Let $\theta_{[i]}$ denote the ranked parameter values, $\theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k]}$. Let $Y_{[1]}$ denote the ranked statistics, $Y_{[1]} \leq Y_{[2]} \leq \ldots \leq Y_{[k]}$, and let $Y_{(i)}$ denote the statistic associated with $\theta_{[i]}$. It is assumed that there is no a priori knowledge as to the pairings of the $Y_i$ and $\theta_i$.

We will treat the situation where $\theta_i$ is a location or scale parameter for $Y_i$. For the location parameter case, i.e., $P(Y_i \leq t) = G_n(t; \theta_i) = G_n(t - \theta_i)$, the probability of a correct selection ($P(\text{CS})$) is given by

$$ P(\text{CS}) = \int_{\delta_{[1]}}^{\delta_{[k]}} \prod_{i=1}^{k-1} G_n(y + \delta_i) \, dG_n(y), \quad (1.1) $$

where $\delta_i = \theta_{[k]} - \theta_{[i]}$.

For the scale parameter case we have

$$ P(\text{CS}) = \int_{\delta_{[1]}}^{\delta_{[k]}} \prod_{i=1}^{k-1} G_n(y \delta_i) \, dG_n(y), \quad (1.2) $$

where $\delta_i = \theta_{[k]} / \theta_{[i]}$. 
Olkin, Sobel and Tong [1976] and Gibbons, Olkin and Sobel [1977] have presented estimators of $P(\text{CS})$ which consist of replacing the $\delta_i$ in (1.1) and (1.2) by estimates, $\hat{\delta}_i$, equal to
\[
\hat{\delta}_i = Y_{[k]} - Y_{[i]}, \quad \text{location parameter case}
\]
\[
\hat{\delta}_i = \frac{Y_{[k]}}{Y_{[i]}}, \quad \text{scale parameter case.}
\]
For the location parameter case this gives
\[
\hat{P} = \hat{P}(\text{CS}) = \prod_{i=1}^{k-1} F_n(Y + Y_{[k]} - Y_{[i]})dF_n(Y).
\]

It is important to examine the small sample properties of these estimators before recommending that they be used in practice. For example, if, unknown to the experimenter, the $\theta_i$ values were nearly equal, then $\hat{P}$ would tend to overestimate $P(\text{CS})$ since the $\hat{\delta}_i$ values are always positive (the $Y_{[i]}$ are nearly order statistics from $G_n(t; \theta)$).

Since the estimation technique described above has been presented in Gibbons, Olkin and Sobel [1977] and Gupta and Panchepekesan [1979] on the basis of large sample properties, a corresponding small sample study is needed. We begin our investigation of $\hat{P}$ with the special case of $k = 2$ populations.

2. $K = 2$ Populations

We consider first the case of a location parameter family and remark later on the extension to scale parameter families. For $k = 2$ populations we will find it convenient to rewrite formula (1.1) for $P(\text{CS})$:
\[
P(\text{CS}) = P(Y_{(1)} \leq Y_{(2)}) = P(Y_{(1)} - Y_{(2)} + \delta_1 \leq \delta_1)
\]
\[
= H_n(\delta_1), \quad (1.3)
\]
where $H_n(s) = P(\hat{Y}_1 - \hat{Y}_2 = t)$ is independent of $\delta_1$. Therefore, $\hat{P} = H_n(\hat{\delta}_1)$, where $\hat{\delta}_1 = Y_2 - Y_1$. Note also that, since $(Y_1 + \delta_1) - Y_2$ is a difference of i.i.d. random variables, symmetry yields $H_n(s) + H_n(-s) = 1$. We next derive the distribution and density functions of $\hat{P}$, subject to certain assumptions on $H_n$.

Theorem 1: If $H_n^{-1}(\cdot)$ exists, then the distribution function of $\hat{P}$ is given by,

$$
P(\hat{P} \leq s) = \begin{cases} 1 & s \geq 1 \\ H_n(H_n^{-1}(s) + \delta_1) + H_n(H_n^{-1}(s) - \delta_1) - 1 & \frac{1}{2} < s < 1 \\ 0 & s \leq \frac{1}{2} \end{cases}
$$

Proof: First note that $\hat{\delta}_1 = Y_2 - Y_1 = |Y_1 - Y_2|$.

For $s \in (\frac{1}{2}, 1)$ we have,

$$
P(\hat{P} \leq s) = P(H_n(Y_1 - Y_2) \leq s)
\leq P(-H_n^{-1}(s) \leq Y_1 - Y_2 \leq H_n^{-1}(s))
\leq P(-H_n^{-1}(s) + \delta_1 \leq Y_1 - Y_2 + \delta_1 \leq H_n^{-1}(s) + \delta_1)
= H_n(H_n^{-1}(s) + \delta_1) - H_n(-H_n^{-1}(s) + \delta_1)
= H_n(H_n^{-1}(s) + \delta_1) + H_n(H_n^{-1}(s) - \delta_1) - 1.
$$

The last step follows from the symmetry of $H_n(\cdot)$.

Corollary 1: When $\delta_1 = 0$ ($\delta_1 = \delta_2$), $\hat{P}$ is uniformly distributed on $(\frac{1}{2}, 1)$.

Corollary 2: If $H_n^{-1}(\cdot)$ exists and $H_n(\cdot)$ is absolutely continuous w.r.t. Lebesque measure on $(-\infty, \infty)$ with density $h_n(\cdot)$, then the density of $\hat{P}$ is given by,

$$
d\hat{P}(\hat{P} \leq s) = \begin{cases} h_n(H_n^{-1}(s) + \delta_1) + h_n(H_n^{-1}(s) - \delta_1)/[h_n(H_n^{-1}(s))] & \frac{1}{2} < s < 1 \\ 0 & \text{elsewhere} \end{cases}
$$
Proof: Using the fact that
\[
\frac{d}{ds} H_n^{-1}(s) = [h_n(H_n^{-1}(s))]^{-1},
\]
straightforward differentiation gives the result.

Remark 1: Under slight regularity conditions on $h_n(\cdot)$, we have,
\[
E_{\delta_1} [\hat{P}] \rightarrow \frac{3}{4} \text{ as } \delta_1 \rightarrow 0.
\]

This implies $\hat{P}$ is biased upwards by a serious amount for small $\delta_1$, since $P(\text{CS}) \rightarrow \frac{1}{2}$ as $\delta_1 \rightarrow 0$.

Corollary 3: If the $X_{ij}$ are i.i.d. $N(\theta_1, \sigma^2)$, $\sigma^2$ known, and $Y_i = \overline{X}_i$, then $\hat{P}$ has density $f_{\hat{P}}(s; \Delta)$ given by,
\[
\frac{d}{ds} P(\hat{P} \leq s) = 2\sqrt{2\pi} \phi(\Delta) \cosh [\Delta\delta_1^{-1}(s)], \quad 1/2 \leq s \leq 1
\]
where $\phi, \phi$ are, respectively, the standard normal density and distribution functions and $\Delta = \sqrt{\frac{\theta_1}{2}}(\theta_2 - \theta_1)/\sigma = \sqrt{\frac{\delta_1}{2}}/\sigma$.

A plot of the density of $\hat{P}$ for various values of $\Delta$ is given in figure 1.

Remark 2: If $Y_i$ has scale parameter $\delta_1$ and $P(Y_i > 0) = 1$ then it is easy to see that the same results hold, with $\log \delta_1$ replacing $\delta_1$ and $H_n(s) = P(\log Y_1 < s)$ replacing $H_n(s)$.

Theorem 1 also allows us to calculate exact operating characteristic curves of $\hat{P}$ for $k = 2$ populations with the aid of a well known lemma for the expected values of positive random variables.

Lemma 1: Suppose $P(X > 0) = 1$ and let $F_X(t) = P(X \leq t)$, then
\[
E[g(X)] = \int_0^\infty g'(w)[1 - F_X(w)]dw
\]
Figure 1: $f_{\delta}(s; \Delta)$ vs $s$

$\Delta = 0, 1, 2, 3$
for \( g(\ast) \) such that \( g(0) = 0 \).

**Corollary 4:** Choosing \( g(w) = w^r \) in the above lemma allows us to calculate the \( r \)th moment of \( P \).

\[
E_{\delta_1} [\hat{P}] = 2^{-r} + \int_0^\infty \frac{H_{n-1}^r(t)}{H_n(t)} \left[ 2 - H_n(t + \delta) - H_n(t - \delta) \right] dt \text{d}H_n(t)
\]

In particular, \( E_{\delta_1} [\hat{P}] = \frac{1}{2} + \int_0^\infty \left[ 2 - H_n(t + \delta) - H_n(t - \delta) \right] dt \text{d}H_n(t) \).

**Corollary 5:** If \( X_{ij} \) are i.i.d. \( N(\theta_1, \sigma^2) \), \( \sigma^2 \) known, and

\[ Y_i = \frac{X_i}{\bar{X}_i}, \] then

\[ E_{\Delta} [\hat{P}] = \frac{3}{4} + \frac{[\phi(\Delta/\sqrt{2}) - 1/2]^2}{2}, \]

where \( \Delta = \sqrt{\frac{n}{2}} (\theta[2] - \theta[1])^{1/\sigma} = \frac{n}{\sqrt{2}} \delta_1^{1/\sigma} \).

**Proof:** \( E_{\Delta} [\hat{P}] = E[\phi(|Y + \Delta|)] \), where \( Y \sim N(0, 1) \).

Hence,

\[
E_{\Delta} [\hat{P}] = \int_{-\infty}^{\infty} \phi(|u + \Delta|) \phi(u) \, du
\]

\[
= \int_{-\infty}^{\infty} \phi(|u|) \phi(u - \Delta) \, du
\]

\[
= \int_{-\infty}^{\infty} \int_{-\Delta}^{\infty} \phi(y) \phi(z) \, dy \, dz + \int_{-\Delta}^{\infty} \int_{-\infty}^{-\Delta} \phi(y) \phi(z) \, dy \, dz
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\Delta} \phi(y) \phi(z) \, dy \, dz + \int_{\Delta}^{\infty} \int_{-\infty}^{-\Delta} \phi(y) \phi(z) \, dy \, dz
\]

Transforming with \( u = \frac{1}{\sqrt{2}} (z - y) \)

\[ v = \frac{1}{\sqrt{2}} (z + y), \]

and using the spherical symmetry of the bivariate normal yields the following expression.
\[ E \hat{\Delta} = \int_{-\infty}^{\infty} \phi(u) \phi(v) \, dv \] 

\[ + \int_{-\infty}^{\Delta} \int_{-\infty}^{\Delta} \phi(u) \phi(v) \, dudv \]

\[ = \frac{3}{4} \left[ \phi(\Delta / \sqrt{2}) - \frac{1}{2} \right]^2 . \]

A plot of \( E \hat{\Delta} \) and \( P(\text{CS}) \) for \( k = 2 \) normal populations is given in figure 2. Note that (1.3) implies \( P(\text{CS}) = \phi(\Delta) \).

3. \( K \geq 2 \) Populations, Normal Means

For \( K \geq 2 \) populations we consider the special case

\[ X_{ij} \text{ distributed independently } N(\theta_i, \sigma^2) \]

\[ 1 \leq i \leq k \quad 1 \leq j \leq n, \quad \sigma^2 \text{ known} \]

\[ Y_i = \overline{X}_{.i} = \frac{1}{n} \sum_{j=1}^{n} X_{ij} \]

The performance of \( \hat{\Delta} \) was studied via Monte Carlo simulation for \( k = 2, 3, 4 \) and 10 populations and for various parameter configurations. Histograms of the actual Monte Carlo distributions obtained may be found in appendix 1. Details of the simulation techniques and computational techniques are given in appendix 2.

The results of the simulation agreed very well with the analytic results for \( k = 2 \) populations discussed in Section 2.

For the normal means case, variance equal to 1, we have

\[ P(\text{CS}) = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \phi(y + n^{1/2} \delta_i / \sigma) \]

which depends on the parameters \( \mu_i \) only through the quantities

\[ \delta_i = \theta_i[k] - \theta_i[i], \quad i = 1, 2, \ldots, k - 1. \]

The parameter configuration and number of populations simulated are given in figure 3.
Figure 2: $E_\Delta[P]$ and $P_\Delta[CS]$ versus $\Delta$
Figure 3: Parameter Configurations Simulated

<table>
<thead>
<tr>
<th>Number of Populations</th>
<th>Mean Vectors $(\theta_1, \theta_2, \ldots, \theta_k)$ (in units of $\sqrt{n}/\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>slippage: $(0, 0), (0, 0.5)$, $(0, 1), (0, 1.25), (0, 1.5)$, $(0, 2), (0, 3), (0, 4)$</td>
</tr>
<tr>
<td>3</td>
<td>slippage: $(0, 0, 0), (0, 0, 0.5)$, $(0, 0, 1), (0, 0, 1.25), (0, 0, 1.5)$, $(0, 0, 2), (0, 0, 3), (0, 0, 4)$</td>
</tr>
<tr>
<td></td>
<td>other: $(0, 1, 1), (0, 2, 2), (0, 1, 2)$, $(0, 3, 3), (0, 2, 3), (0, 1, 3)$, $(0, 4, 4), (0, 3, 4), (0, 2, 4)$, $(0, 1, 4)$</td>
</tr>
<tr>
<td>4</td>
<td>slippage: $(0, 0, 0, 0), (0, 0, 0, 0.5)$, $(0, 0, 0, 1), (0, 0, 0, 1.25)$, $(0, 0, 0, 1.5), (0, 0, 0, 2)$, $(0, 0, 0, 3), (0, 0, 0, 4)$</td>
</tr>
<tr>
<td>10</td>
<td>slippage: $(0, 0, \ldots, 0), (0, 0, \ldots, 2), (0, 0, \ldots, 3)$, $(0, 0, \ldots, 4)$</td>
</tr>
</tbody>
</table>
The results of the simulation showed that, in the slippage configuration ($\delta_1 = \delta$, say), $\hat{P}$ tends to overestimate $P(\mathcal{C} \mathcal{S})$ for small values of $n^{1/2} \delta / \sigma$. This is the same pattern as in the analytic results for $k = 2$ populations, but more pronounced for $k > 2$ populations. This bias can be large at times. For example, when $\delta = 0$, the simulated $E[\hat{P}]$ exceeded $P(\mathcal{C} \mathcal{S})$, $1/k$, by $.25$, $.32$, $.33$ and $.33$ for $k = 2$, $3$, $4$ and $10$ populations respectively. When $n^{1/2} \delta / \sigma = 1$, the simulated $E[\hat{P}]$ was biased downwards by $.04$, $.08$, $.10$ and $.17$ for $k = 2$, $3$, $4$ and $10$ populations, respectively. Graphs of the true $P(\mathcal{C} \mathcal{S})$ and the simulated $E[\hat{P}]$ in the slippage configuration follow.

Additional simulations for the case $k = 3$ populations were performed to discover the performance of $\hat{P}$ when the parameters are not in the slippage configuration. The results are summarized in figure 8.

4. Discussion and Conclusions

Our analysis has shown that the Olkin-Sobel-Tong estimator has some serious deficiencies, especially tending to overestimate $P(\mathcal{C} \mathcal{S})$ when the means are close together. This deficiency may be accentuated by the majorization techniques advocated by Gibbons, Olkin and Sobel [1977], Olkin, Sobel and Tong [1976], and Tong [1980]. In the case where the means are nearly equal their upper bound will overestimate $P(\mathcal{C} \mathcal{S})$ more than $\hat{P}$. These upper and lower bounds are presumably advocated for computational simplicity, though numerical evaluation of the integral involved is not difficult with the aid of a computer.

In conclusion, the idea of an a posteriori analysis of the probability
Figure 4: $E[P]$ and $P(CS)$ versus $\Delta$ for the case of $K = 2$ Normal Populations

Legend:
- $P(CS)$
- $E[P]$
- * simulated value of $\Delta$

$$\Delta = \frac{n^{1/2}(\theta[2] - \theta[1])}{\sigma}$$
Figure 5: $E[P]$ and $P(CS)$ versus $\Delta$ for the Case of $k = 3$ Normal Population

\[ \Delta = \frac{n^{1/2}(\theta [3] - \theta [2])}{\sigma} \]

$s[2] = \theta [1]$
Figure 6: $E[P]$ and $P(CS)$ versus $\Delta$ for the Case of $K = 4$ Normal Populations

$\Delta = \frac{n^{1/2}(\theta[4] - \theta[3])}{\sigma}$

Figure 7: $E[P]$ and $P(CS)$ versus for the Case of $K = 10$ Normal Populations

\[ \Delta = \frac{n^{1/2} \theta[10] - \theta[9]}{\sigma} \quad \theta[9] = \ldots \theta[1] \]
Figure 8: Mean Observed $\hat{P}(CS)$ in 1000 replications
(True $P(CS)$)

$K = 3$ populations

\[
\frac{\mu[3] - \mu[2]}{(\sigma^2/n)^{1/2}} = \delta_2
\]

\[
\frac{\mu[3] - \mu[1]}{(\sigma^2/n)^{1/2}} = \delta_1
\]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.650</td>
<td>(\ .333)</td>
<td>(\ .333)</td>
<td>NOT</td>
<td>POSSIBLE</td>
</tr>
<tr>
<td>1</td>
<td>.681</td>
<td>.681</td>
<td>(\ .443)</td>
<td>(\ .634)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.708</td>
<td>.730</td>
<td>.786</td>
<td>(\ .489)</td>
<td>(\ .729)</td>
</tr>
<tr>
<td>3</td>
<td>.737</td>
<td>.758</td>
<td>.832</td>
<td>.893</td>
<td>(\ .499)</td>
</tr>
<tr>
<td>4</td>
<td>.749</td>
<td>.783</td>
<td>.856</td>
<td>.927</td>
<td>.961</td>
</tr>
</tbody>
</table>
of a correct selection is a sound one, but the estimator considered here appears to be seriously flawed. Alternative procedures for estimating the \( P(\text{CS}) \) have been considered. Faltin (1980) showed that for \( k = 2 \) normal populations with common known variance,

\[
P_\Delta \{ \hat{P} \geq P(\text{CS}) \} = \frac{3}{2} - 2\Phi(\Delta) \geq 1/2.
\]

Thus the Olkin-Sobel-Tong estimator overestimates the \( P(\text{CS}) \) with probability always exceeding 1/2. For this case he was able to derive an estimator, \( \tilde{P} \), such that

\[
P_\Delta \{ \tilde{P} \geq P(\text{CS}) \} = 1/2.
\]

Clearly, further research needs to be done in this area.
Bibliography


Appendix 1

Monte Carlo sampling frequency tabulations
AI: Histograms of Monte Carlo Sampling
Distribution of P for K = 2 Normal Populations and various values of

\[ \Delta = n^{1/2} \left( \frac{\sigma_k}{\sigma} \right) \]

(slipage configuration)
A2: Histograms of Monte Carlo Sampling Distribution of $P$ for $K = 3$ Normal Populations and various values of

$$\delta = \frac{1}{2} \left( \frac{S^2[k] - S^2[k-1]}{\sigma} \right)$$

(slipage configuration)
A3: Histograms of Monte Carlo Sampling Distribution of $P$ for $K = 4$ Normal Populations and various values of

$$
\Delta = \frac{n^2}{2} \left( \frac{\theta_k - \theta_{k-1}}{\sigma} \right)
$$

(slippage configuration)
Histograms of Monte Carlo Sampling Distribution of $P$ for $k = 10$ Normal Populations and various values of $\Lambda$:

$$\Lambda = n^{1/2} \left( \frac{\mu [k] - \mu [k-1]}{\sigma} \right)$$

(slippage configuration)
Appendix 2

All random number generation and computation was done on Cornell's IBM 370 computer. The pseudo random numbers were generated by the IMSL subroutine GGNOF. Numerical evaluation of the P(CS) integral was done via IMSL's numerical integration technique DCADRE, which uses cautious Romberg extrapolation methods. The normal cdf was evaluated using the identity:

\[ \phi(t) = 0.5 \times \text{DERFC}(-\sqrt{2} \times t) \]

Where DERFC is the double precision version of the complemented error function (a built-in FORTRAN function). All computations were performed in double precision.