Third Revision 10/26/82

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[1]

BU-791-M

[1] Frederick W. Faltin is Staff Statistician in Management Sciences at Air Products and Chemicals, Inc., Trexlertown, Pensylvania, 18105. Charles E. McCulloch was Assistant Professor of Statistics, Florida State University, Tallahassee, Florida, 32306 during the preparation of this article. He is currently Assistant Professor, Biometrics Unit, Cornell University, Ithaca, New York, 14853.

This 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.

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 Panchepakesan [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}, $1 \le i \le k$, $1 \le j \le n$ be independent observations from k populations with c.d.f.'s $F(x;\theta)$. We wish to select the population associated with the largest θ_i. We consider decision procedures as follows: Define an appropriate statistic $Y_i = Y(X_{i1},X_{i2},\cdots,X_{in})$ and select the population giving rise to the largest Y_i as the population associated with the largest θ_i. For example, in the case where $F(\cdot;\theta_i)$ is a normal c.d.f. with mean θ_i and known variance σ^2, considered in Bechhofer [1954], $Y_i = Y(X_{i1},X_{k2},\cdots,X_{in})$ $= \frac{1}{n} \Sigma_{j=1}^{n} X_{ij} = \bar{X}_{i.}$.

Let $\theta_{[i]}$ denote the ranked parameter values, $\theta_{[1]} \le \theta_{[2]} \le \cdots \le \theta_{[k]}$. Let $Y_{[i]}$ denote the ranked statistics, $Y_{[1]} \le Y_{[2]} \le \cdots \le Y_{[k]}$, 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 $\theta_{[j]}$ $(1 \le j \le k)$.

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

$$P\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} G_n(y + \delta_i)dG_n(y) , \qquad (1.1)$$

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

Olkin, Sobel and Tong [1976] and Gibbons, Olkin and Sobel [1977] have presented estimators of P{CS} which consist of replacing the δ_i in (1.1) by estimates, $\hat{\delta}_i$, equal to

$$\hat{\delta}_i = Y_{[k]} - Y_{[i]}$$

which gives

$$\hat{P} = \hat{P}\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} G_n(y+Y_{[k]}-Y_{[i]})dG_n(y) .$$

It is of interest 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 θ_i values were nearly equal, then $\hat{P}$ would tend to over-estimate P{CS} since the $\hat{\delta}_i$ values are always positive. Since the estimation technique described above has been presented in Gibbons, Olkin and Sobel [1977] and Gupta and Panchapakesan [1979] and justified on the basis of large sample properties, a corresponding small sample study is needed (Bechhofer [1980], p. 753). We begin our investigation of $\hat{P}$ with the special case of k = 2 populations.

2. k = 2 Populations

For k = 2 populations we will find it convenient to rewrite formula (1.1) for P{CS}

$$P\{CS\} = P\{Y_{(1)} \le Y_{(2)}\} = P\{Y_{(1)} - Y_{(2)} + \delta_1 \le \delta_1\}$$

$$= H_n(\delta_1) \quad , \qquad (1.2)$$

where $H_n(s) = P\{Y_{(1)} - Y_{(2)} + \delta_1 \le s\}$ is independent of δ_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.

First note that $\hat{\delta}_1 = Y_{[2]} - Y_{[1]} = |Y_{(1)} - Y_{(2)}|$. Therefore, for $s \in (\frac{1}{2},1)$ we have,

$$P\{\hat{P} \le s\} = P\{H_n(|Y_{(1)}-Y_{(2)}|) \le s\} .$$

Taking $H_n^{-1}(\cdot)$ on both sides of the inequality and rearranging gives,

$$\begin{aligned} P\{\hat{P} \le s\} &= P\{|Y_{(1)} - Y_{(2)}| \le H_n^{-1}(s)\} \\ &= P\{-H_n^{-1}(s) + \delta_1 \le Y_{(1)} - Y_{(2)} + \delta_1 \le H_n^{-1}(s) + \delta_1\} \\ &= H_n[H_n^{-1}(s)+\delta_1] - H_n[-H_n^{-1}(s)+\delta_1] \\ &= H_n[H_n^{-1}(s)+\delta_1] + H_n[H_n^{-1}(s)-\delta_1] - 1 . \end{aligned}$$

The last step follows from the symmetry of $H_n(\cdot)$. We have thus proved the following theorem.

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

$$P\{\hat{P} \le s\} = \begin{cases} 1 & s \ge 1 \\ H_n[H_n^{-1}(s) + \delta_1] + H_n[H_n^{-1}(s) - \delta_1] - 1 & \frac{1}{2} < s < 1 . \\ 0 & s \le \frac{1}{2} \end{cases}$$

Certain facts are immediate from the theorem:

1. When $\delta_1 = 0$ ($\theta_1 = \theta_2$), $\hat{P}$ is uniformly distributed on $(\frac{1}{2},1)$.
2. $P_{\delta_1}\{\hat{P} > P\{CS\}\} = \frac{3}{2} - H_n(2\delta_1) > \frac{1}{2}$.

3. The density of $\hat{P}$ is given by

$$\frac{d}{ds} P\{\hat{P} \leq s\} = \begin{cases} [h_n(H_n^{-1}(s) + \delta_1) + h_n(H_n^{-1}(s) - \delta_1)]/[h_n(H_n^{-1}(s))] & \frac{1}{2} < s < 1 \\ 0 & \text{elsewhere} \end{cases}$$

4. If the X_{ij} are i.i.d. $N(\theta_i, \sigma^2)$, σ^2 known, and $Y_i = \bar{X}_{i.}$, then $\hat{P}$ has density $f_{\hat{P}}(s;\Delta)$ given by

$$\frac{d}{ds} P\{\hat{P} \leq s\} = 2\sqrt{2\pi}\, \phi(\Delta) \cosh [\Delta \Phi^{-1}(s)] , \qquad \tfrac{1}{2} \leq s \leq 1$$

where ϕ, Φ are, respectively, the standard normal density and distribution functions and $\Delta = \sqrt{n/2}\, (\theta_{[2]} - \theta_{[1]})/\sigma = \sqrt{n/2}\, \delta_1/\sigma$.

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

5. Using the fact that $E[g(X)] = \int_0^\infty g'(w)[1 - F_X(w)]dw$, when $g(w) = w^r$ and $P\{X > 0\} = 1$, we have

$$E_{\delta_1}[\hat{P}^r] = 2^{-r} + \int_0^\infty rH_n^{r-1}(t)[2 - H_n(t+\delta_1) - H_n(t-\delta_1)]dH_n(t) .$$

In particular, $E_{\delta_1}[\hat{P}] = \frac{1}{2} + \int_0^\infty [2 - H_n(t+\delta_1) - H_n(t-\delta_1)]dH_n(t)$.

In addition, for the normal case, we can derive $E[\hat{P}]$ in closed form.

Theorem 2. *If the* X_{ij} *are* i.i.d. $N(\theta_i, \sigma^2)$, σ^2 *known, and* $Y_i = \bar{X}_{i.}$ *then*

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

where $\Delta = \sqrt{n/2}\, (\theta_{[2]} - \theta_{[1]})/\sigma = \sqrt{n/2}\, \delta_1/\sigma$.

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

$$E_\Delta[\hat{P}] = \int_{-\infty}^{\infty} \Phi(|w + \Delta|)\phi(w)dw$$

$$= \int_{-\infty}^{\infty} \Phi(|u|)\phi(u - \Delta)du$$

$$= \int_{\Delta}^{\infty}\int_{-\infty}^{z-\Delta} \phi(y)\phi(z)\, dydz + \int_{-\infty}^{\Delta}\int_{-\infty}^{\Delta-z} \phi(y)\phi(z)\, dydz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\Delta-z} \phi(y)\phi(z)\, dydz + \int_{\Delta}^{\infty}\int_{\Delta-z}^{z-\Delta} \phi(y)\phi(z)\, dydz .$$

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

$$E[\hat{P}] = \int_{-\infty}^{\Delta/\sqrt{2}}\int_{-\infty}^{\infty} \phi(u)\phi(v)\, dudv + \int_{\Delta/\sqrt{2}}^{\infty}\int_{\Delta/\sqrt{2}}^{\infty} \phi(u)\phi(v)\, dudv$$

$$= \tfrac{3}{4} + [\Phi(\Delta/\sqrt{2}) - \tfrac{1}{2}]^2 .$$

A plot of $E[\hat{P}]$ and $P\{CS\}$ for $k = 2$ normal populations is given in Figure 2. Note that (1.2) implies $P\{CS\} = \Phi(\Delta)$.

Remark The facts following Theorem 1 show that $\hat{P} > P\{CS\}$ and $E[\hat{P}] = \frac{3}{4}$ when $\delta_1 = 0$. Also $P_{\delta_1}\{\hat{P} > P\{CS\}\} \to 1$ monotonically as $\delta_1 \to 0$; a similar result holds for $E_\delta[\hat{P}]$ under mild regularity conditions on H_n. In fact, $\hat{P}$ is likely to overestimate the true $P\{CS\}$ regardless of δ_1. This tendency of $\hat{P}$ to be high in its estimate of $P\{CS\}$ would seem to be a substantial flaw, which, as we have seen is especially severe for small δ_1. Interestingly, as Figure 2 shows, $E[\hat{P}]$ can be considerably smaller than $P\{CS\}$. This is due to the extremely skewed distribution of $\hat{P}$ for large δ_1, as evidenced in Figure 1.

3. $k \geq 2$ Populations, Normal Means

For $k \geq 2$ populations we considered the special case

X_{ij} distributed independently $N(\theta_i, \sigma^2)$ $\quad 1 \leq i \leq k \quad 1 \leq j \leq n$

$\sigma^2 > 0$ known

$$Y_i = \bar{X}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} X_{ij} \quad .$$

The performance of $\hat{P}$ was studied via Monte Carlo simulation for the parameter configurations and number of populations given in Table 1. Details of the simulation techniques and computational techniques are given in the Appendix.

For the normal means case, we have

$$P\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + n^{\frac{1}{2}}\delta_i/\sigma) \; d\,\Phi(y) \; .$$

This depends on the parameters θ_i only through the quantities $\delta_i = \theta_{[k]} - \theta_{[i]}$.

The results of the simulation showed that, in the slippage configuration $(\delta_i = \delta,\ i = 1,2,\cdots,k)$, $\hat{P}$ tends to overestimate $P\{CS\}$ for small values of $n^{\frac{1}{2}}\delta/\sigma$. 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, with $k = 4$ populations and $\delta = 0$, the simulated $E[\hat{P}]$ exceeded $P\{CS\}$ by .33. When $n^{\frac{1}{2}}\delta/\sigma = 3$, the simulated $E[\hat{P}]$ was biased downward by .10. A graph of the true

$P\{CS\}$ and the simulated $E[\hat{P}]$ for $k = 4$ populations is given in Figure 3. The results for $k = 3$ and $k = 10$ were very similar.

4. Discussion and Conclusions

Our analysis has shown that the Olkin-Sobel-Tong estimator has some serious deficiencies, tending to overestimate $P\{CS\}$ when the means are close together and tending to be biased downward when $n^{\frac{1}{2}}\delta/\sigma$ is large (see the Remark in Section 2). 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 $P\{CS\}$ 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 of a correct selection is a sound one, but the estimator considered here appears to be seriously flawed. The estimator demonstrates the poor behavior possible when estimating nuisance parameters when general theory such as Randles [1982] is not applicable.

Faltin [1980] has considered an alternative procedure for estimating $P\{CS\}$ for the case $k = 2$ normal populations with common, known variance. That approach estimates $P\{CS\}$ directly instead of estimating the nuisance parameters δ_i and avoids some of the corresponding problems. It was shown that for a broad class of loss functions, one need only consider procedures which (like $\hat{P}$) are non-randomized and monotone in Δ and was able to derive for any $\alpha \in (0,1)$, a unique estimator $\tilde{P}_\alpha$ of this type for which

$$P_\Delta\{\tilde{P}_\alpha \geq P\{CS\}\} = \alpha \qquad \forall\Delta \geq 0 .$$

This work has recently been generalized to certain non-normal location and scale parameter families as well. A manuscript is in preparation.

Table 1: Parameter Configurations Simulated

Number of Populations k	Mean Vectors $(\theta_1, \theta_2, \ldots, \theta_k)$ (in units of $\sqrt{n}/\sigma$)
2	slippage: (0, 0), (0, 0.5), (0, 1), (0, 1.25), (0, 1.5), (0, 2), (0, 3), (0, 4)
3	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)
	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)
4	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)
10	slippage: (0, 0, ..., 0), (0, 0, ..., (0, 0, ..., 2), (0, 0, ..., 3), (0, 0, ..., 4).

Bibliography

Bechhofer, R.E. (1954). "A single sample multiple decision procedure for ranking means of normal populations with known variances", Annals of Mathematical Statistics, 25, 16-39.

Bechhofer, R.E. (1980). Review of "Selecting and Ordering Populations: A New Statistical Methodology" by Gibbons, J.D., Olkin, I., and Sobel, M., Journal of the American Statistical Association, 75, 751-6.

Faltin, F. (1980). "A quantile unbiased estimator of the probability of correct selection achieved by Bechhofer's single-stage procedure for the two population normal means problem"., Abstract 80t-60, IMS Bulletin, 9, 180-1.

Gibbons, J.D., Olkin, I., and Sobel, M. (1977). Selecting and Ordering Populations: A New Statistical Methodology, John Wiley, New York.

Gupta, S.S., and Panchapakesan (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations, John Wiley, New York.

Olkin, I., Sobel, M., and Tong (1976). "Estimating the true probability of a correct selection for location and scale parameter families" Technical Report 110, Department of Statistics, Stanford University, Stanford, CA.

Randles, R. (1982). "On the asymptotic normality of statistics with estimated parameters." Annals of Statistics, 10, 462-474.

Tong, Y.L. (1980). Probability Inequalities in Multivariate Distributions. John Wiley, New York.

Appendix

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:

$$\Phi(t) = .5* \text{DERFC}(-\sqrt{.5} * t) ,$$

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

Each $E[\hat{P}]$ was determined using 1,000 replications and can be expected to be accurate to ± .01.

Figure 1: $f_{\hat{p}}(s;\,\Delta)$ vs s

$\Delta = 0, 1, 2, 3$

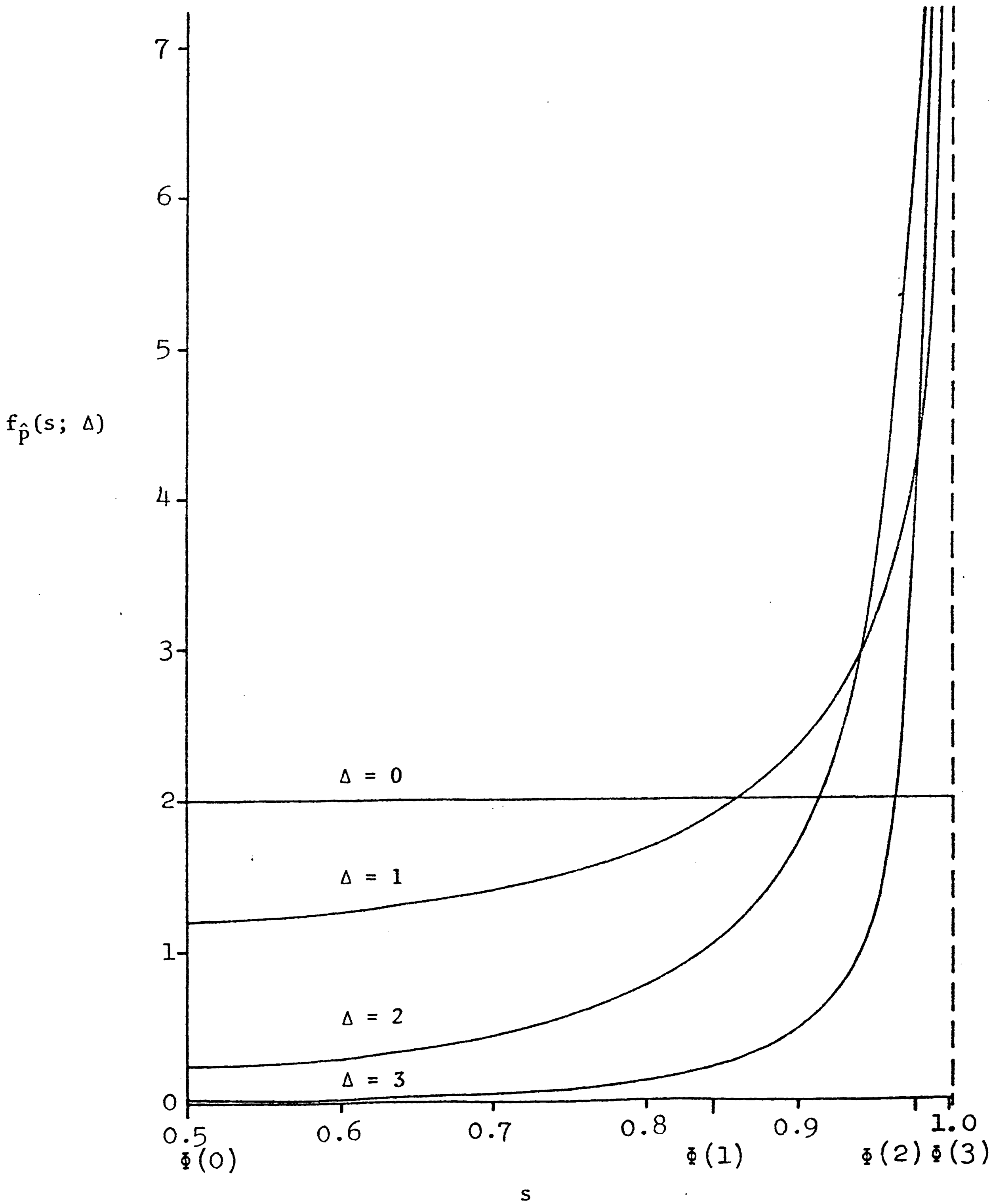


Figure 2: $E_\Delta[\hat{P}]$ and $P_\Delta\{CS\}$ versus Δ

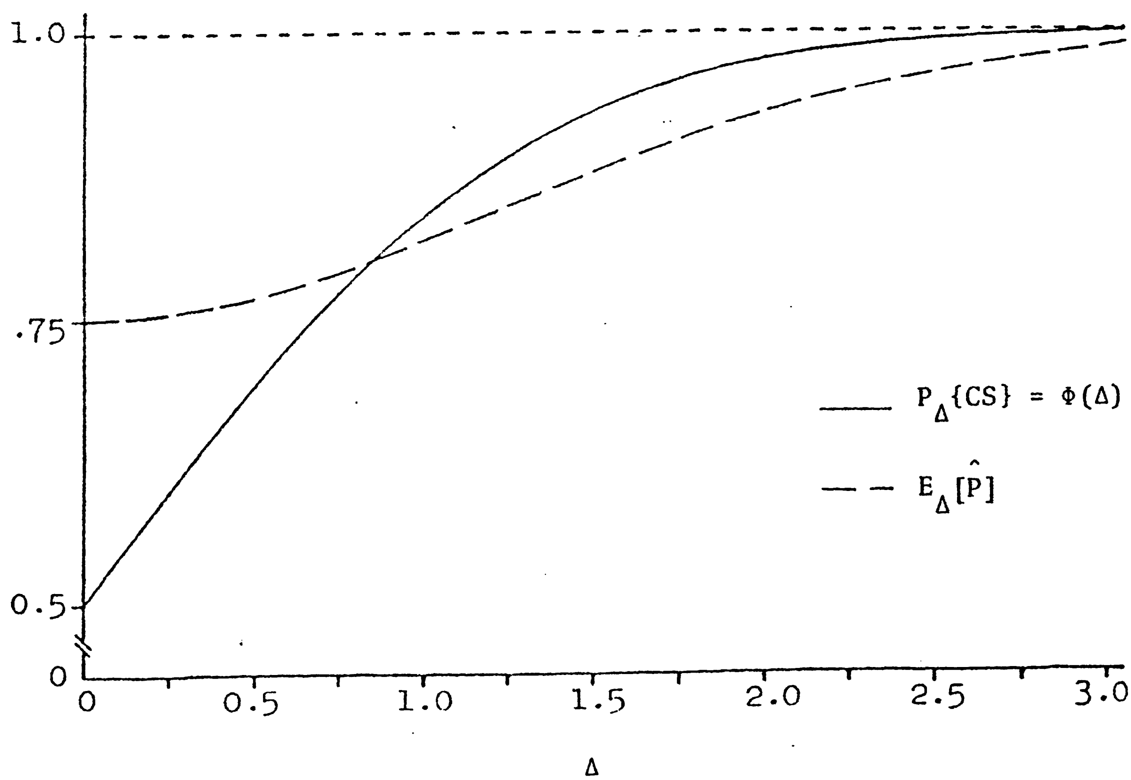
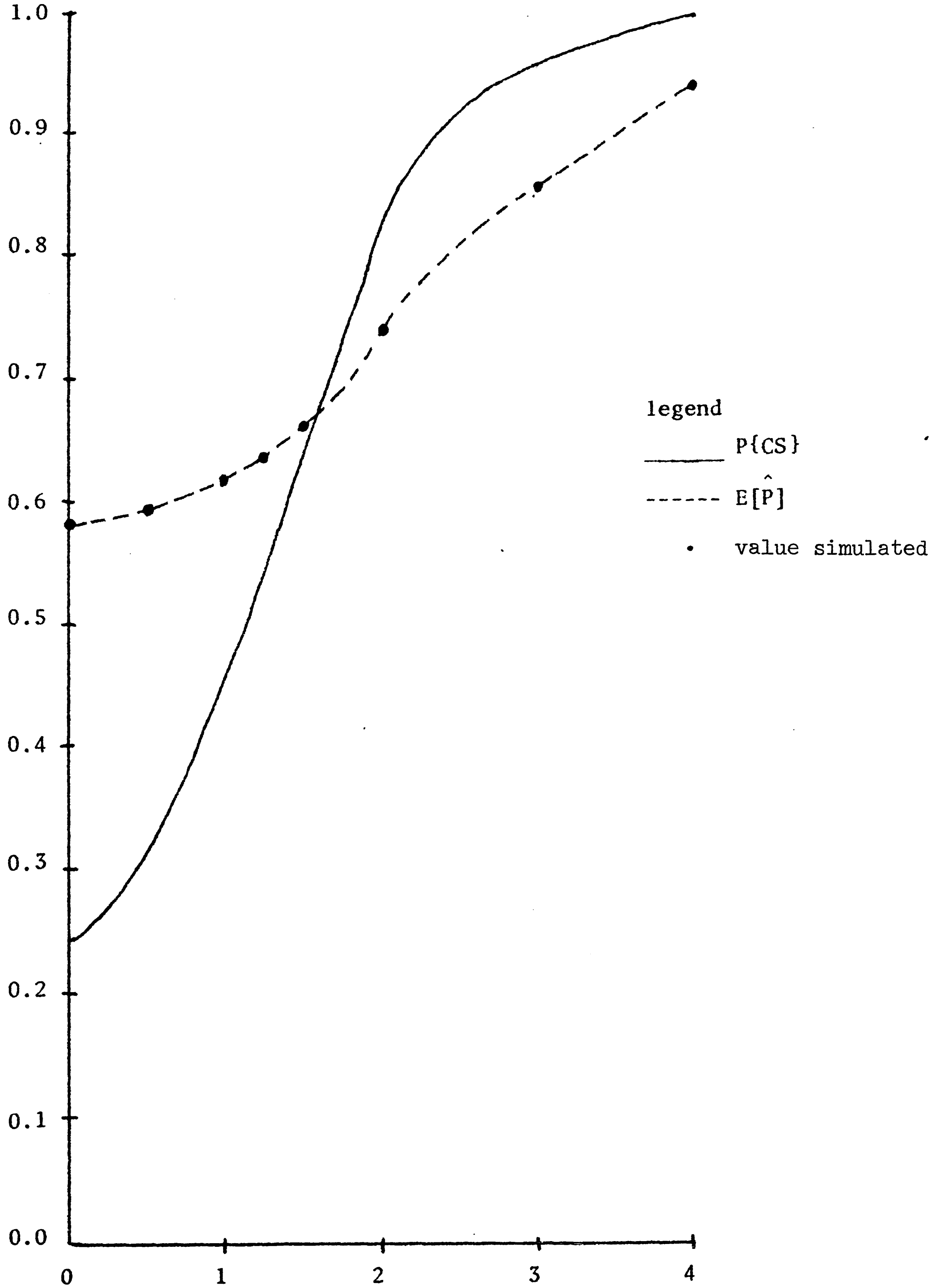


Figure 3: $E[\hat{P}]$ and $P\{CS\}$ versus Δ for the Case of $k = 4$ Normal Populations



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