* This research was supported by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2796.

# On Sequential Elimination Procedures

by

Raymond J. Carroll

*Department of Statistics*
*University of North Carolina at Chapel Hill*

Institute of Statistics Mimeo Series No. 1078

July, 1976

# On Sequential Elimination Procedures

by

Raymond J. Carroll*

*Department of Statistics*
*University of North Carolina at Chapel Hill*

## Summary

The asymptotic properties of a general class of nonparametric sequential ranking and selection procedures which possess an elimination feature is studied. If the correct selection probability is to be at least $1 - \alpha$ and the length of the indifference zone is $\Delta$, different results are obtained as $\alpha \to 0$ depending on whether one assumes $\Delta$ fixed or $\Delta \to 0$. A Monte-Carlo study confirms the superiority of elimination procedures.

* This research was supported by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2796.

Key Words and Phrases: Sequential Analysis, Elimination Selection Rules, Asymptotic Distributions.

AMS 1970 Subject Classifications: Primary 62F07; Secondary 62L12.

## Introduction

In the large literature on sequential ranking and selection (Bechhofer, Kiefer and Sobel (1968)), surprisingly little work has been devoted to procedures which are asymptotically nonparametric in nature and which eliminate obviously inferior populations early in the experiment (see Swanepoel (1976) for a recent approach). It is the purpose of this paper to propose and study a general class of sequential eliminating procedures which includes as special cases Swanepoel's rule as well as a competitor to a noneliminating rule mentioned by Wackerly (1975).

We take the indifference zone approach that the best and second best populations are $\mu$ $(\geq \Delta \geq 0)$ units apart and attempt to guarantee a correct selection (CS, the selection of the best population) with probability at least $1 - \alpha$. Suppose $N$ is the number of stages in the experiment. As a sample of our results, we show (Theorem 1) that a normed version of $N$ looks very much like a linear combination of the "mean" and "variance" on which the selection is based. Surprisingly, the conclusions differ as $\alpha \to 0$ depending on the cases $\mu$ fixed or $\mu \to 0$, which eventually leads us to the conclusion that there truly is a cost of ignorance in estimating the variance. The class proposed includes many procedures in the literature, including one which asymptotically takes at most ¼ the observations needed by the Robbins, Sobel and Starr (1968) procedure. The theory is investigated in a convincing Monte-Carlo study.

In the general problem, we have $k$ populations $\pi_1, \ldots, \pi_k$ and a sample $X_{i1}, X_{i2}, \ldots$ from $\pi_i$ with distribution function $F_i$ .

Statistics $T_{in}$ are formed from $X_{i1}, \ldots, X_{in}$, and for some constants $\mu_i$, $\sigma_i^2$, $n^{\frac{1}{2}}(T_{in} - \mu_i)/\sigma_i$ is asymptotically normal. Consistent estimates $\sigma_{in}^2$ of $\sigma_i^2$ are assumed to be available. Assume without loss of generality that $\mu_1 \leq \ldots \leq \mu_k$ and that it is desired to select the population with the largest value of $\mu_i$. The indifference zone approach adopted here assumes $\mu_k - \mu_{k-1} \geq \Delta$. Let $h$ be a given nonnegative function. Then, we propose to eliminate population $\pi_i$ at the $n^{th}$ stage of the experiment if there is a population $\pi_j$ still in contention at the $n^{th}$ stage for which

$$T_{jn} - T_{in} \geq h(\alpha,n)(\sigma_{in}^2 + \sigma_{jn}^2)^{\frac{1}{2}}/n^{\frac{1}{2}} - \Delta.$$

The experiment stops at the $N^{th}$ stage when only one population remains. As we will show, these elimination stopping rules contain those proposed by Swanepoel (1976), and Swanepoel and Geertsema (1976), as well as an eliminating version of Wackerly's (1975) rule.

When $k = 2$, the number of stages in the experiment becomes

$$N_1 = \inf\left\{ n\geq 1: \begin{array}{c} T_{2n}-T_{1n}-(\mu_2-\mu_1) \geq h(\alpha,n)(\sigma_{1n}^2+\sigma_{2n}^2)^{\frac{1}{2}}/n^{\frac{1}{2}} - \Delta - (\mu_2-\mu_1) \\ \text{or} \\ T_{2n}-T_{1n}-(\mu_2-\mu_1) \leq -h(\alpha,n)(\sigma_{1n}^2+\sigma_{2n}^2)^{\frac{1}{2}}/n^{\frac{1}{2}} + \Delta - (\mu_2-\mu_1) \end{array} \right\}.$$

Letting $d = \Delta + \mu_2 - \mu_1$, $T_n = T_{2n} - T_{1n} - (\mu_2-\mu_1)$ and $\sigma_n^2 = \sigma_{1n}^2 + \sigma_{2n}^2$, this is

$$N_1 = \inf\left\{ n\geq 1: \begin{array}{c} T_n \geq h(\alpha,n)\sigma_n/n^{\frac{1}{2}} - d \\ \text{or} \\ T_n \leq -h(\alpha,n)\sigma_n/n^{\frac{1}{2}} + 2\Delta - d \end{array} \right\}.$$

We are going to investigate $N_1$ as $\alpha, d \to 0$, and in Lemma 3 we present conditions under which $\Pr\{CS\} \to 1$ as $\alpha \to 0$ $(\alpha, d \to 0)$. Thus, for purposes of computing the distribution of $N_1$ and finding constants $n$ for which $N_1/n \xrightarrow{p} 1$, it will suffice to consider

$$N = \inf\{n \geq 1 : T_n \geq h(\alpha,n)\sigma_n/n^{\frac{1}{2}} - d\}.$$

Asymptotic Normality of $N$

The standing assumptions of this section are that $\left(n^{\frac{1}{2}}T_n, n^{\frac{1}{2}}(\sigma_n^2 - \sigma^2)\right)$ are jointly asymptotically normal, uniformly continuous in probability (Anscombe (1952)), and $n^{\frac{1}{2}}T_n = O(\log_2 n)$ (a.s.). These conditions hold for the sample mean, M-estimators and linear functions of order statistics (and their variance estimates) under a variety of conditions (Carroll (1975)).

Theorem 1 Suppose as $\alpha \to 0$,

(2.1a) $$h(\alpha,N) \to \infty \quad \text{(a.s.)}$$

(2.1b) $h(\alpha,N) - h(\alpha, N-1) \to 0$ (a.s.)

(2.1c) There exist constants $n_\alpha(d)$ for which $N/n_\alpha(d) \xrightarrow{p} 1$.

Then for some constant $A(F,d)$, as $\alpha \to 0$,

(2.2) $$\left(\sigma^2 h^2(\alpha,N) - d^2 N\right)\left(4d^2 n_\alpha(d)\right)^{-\frac{1}{2}} = N^{\frac{1}{2}}\left(T_N - d(\sigma_N^2 - \sigma^2)/2\sigma^2\right) + o_p(1)$$

(2.3) $$\left(\sigma^2 h^2(\alpha,N) - d^2 N\right)\left(4d^2 n_\alpha(d)\right)^{-\frac{1}{2}} \xrightarrow{L} N\left(0, A(F,d)\right).$$

Proof of Theorem 1 By (2.1a), $N \to \infty$ (a.s.) so that

$$h^2(\alpha,N)/N \to (d/\sigma)^2 \quad \text{(a.s.)}. \tag{2.4}$$

Now, by definition, $\sigma h(\alpha,N) - dN^{\frac{1}{2}} \le N^{\frac{1}{2}}T_N - h(\alpha,N)(\sigma_N - \sigma)$. Also, $h(\alpha, N-1)\sigma_{N-1} \ge (N-1)^{\frac{1}{2}}(T_{N-1} + d)$, so a little algebra together with the fact that uniform continuity and (2.1c) imply $N^{\frac{1}{2}}(T_N - T_{N-1}) \xrightarrow{p} 0$, $N^{\frac{1}{2}}(\sigma_N - \sigma_{N-1}) \xrightarrow{p} 0$ yields

$$\begin{aligned} N^{\frac{1}{2}}T_N - h(\alpha,N)(\sigma_N - \sigma) &\le \sigma h(\alpha,N) - dN^{\frac{1}{2}} + \sigma_{N-1}h(\alpha,N)\left((1 - 1/N)^{-\frac{1}{2}} - 1\right) \\ &\quad + (1 - 1/N)^{-\frac{1}{2}}\sigma_{N-1}\left(h(\alpha, N-1) - h(\alpha,N)\right) + o_p(1) \\ &= \sigma h(\alpha,N) - dN^{\frac{1}{2}} + o_p(1) \quad \text{(by (2.1b) and (2.4))}. \end{aligned}$$

Now, (2.1c), (2.4) and the uniform continuity of $\sigma_n$ show that $h(\alpha,N)(\sigma_N - \sigma) - dN^{\frac{1}{2}}(\sigma_N^2 - \sigma^2)/2\sigma^2 = o_p(1)$, yielding

$$\sigma h(\alpha,N) - dN^{\frac{1}{2}} = N^{\frac{1}{2}}\left(T_N - d(\sigma_N^2 - \sigma^2)/2\sigma^2\right) + o_p(1).$$

Hence $\sigma h(\alpha,N) - dN^{\frac{1}{2}}$ is asymptotically normal and (2.1c) and (2.4) thus show

$$\sigma h(\alpha,N) - dN^{\frac{1}{2}} = \left(\sigma^2 h^2(\alpha,N) - d^2N\right)\left(4d^2 n_\alpha(d)\right)^{-\frac{1}{2}} + o_p(1),$$

completing the proof. □

The choice of $h(\alpha,n)$ below is suggested by Swanepoel and Geertsema for the normal case with $\sigma_1^2 = \ldots = \sigma_k^2 = \sigma_0^2$ (known). They show that, if $\mu_2 - \mu_1 \ge \Delta$, $\Pr\{CS\} \ge 1 - \alpha$.

Lemma 1 Let $h(\alpha,n) = (\beta_\alpha^2 + c \log n)^{\frac{1}{2}}$, where $c \ge 0$ and $\beta_\alpha$ satisfies $1 - \Phi(\beta_\alpha) + \beta_\alpha\phi(\beta_\alpha) + \phi^2(\beta_\alpha)/\Phi(\beta_\alpha) = \alpha$. Define $n_\alpha(d) = (\beta_\alpha\sigma/d)^2$.

Then, as $\alpha \to 0$, $N/n_\alpha(d) \xrightarrow{p} 1$ and

$$(2.5) \quad \{N - n_\alpha(d)\}/\{2n_\alpha(d)^{\frac{1}{2}}/d\} = N^{\frac{1}{2}}\{d(\sigma_N^2 - \sigma^2)/2\sigma^2 - T_N\} + o_p(1).$$

Further, if $\alpha, d \to 0$ in such a way that $(\log d)/\beta_\alpha \to 0$, the result still holds.

Through the case $d$ fixed, Lemma 1 plainly shows the effect of estimating $\sigma^2$, an effect disguised in the case $d \to 0$. For example, if $X_1, X_2, \ldots$ are i.i.d. $N(0,1)$, $T_n = \bar{X}_n$, $\sigma_n^2 = (n-1)^{-1} \sum_1^n (X_i - \bar{X}_n)^2$, then

$$\frac{d\{N-n_\alpha(d)\}}{2n_\alpha(d)^{\frac{1}{2}}} \xrightarrow{L} N(0,1) \qquad (d \to 0 \text{ or setting } \sigma_n \equiv \sigma)$$

$$\xrightarrow{L} N(0,\ 1+d^2/2) \qquad (d \text{ fixed, } \sigma^2 \text{ unknown}).$$

If we fix $d = 2$ and let $\alpha \to 0$, we see that the lack of knowledge of $\sigma^2$ triples the variance of $N$. Hence, there is a "cost of ignorance" due to estimating $\sigma^2$.

<u>Proof of Lemma 1</u> Recall $T_n = O(n^{-\frac{1}{2}} \log_2 n)$ (a.s.). Since

$$N^{\frac{1}{2}} T_N(\beta_\alpha^2 + c \log N)^{-\frac{1}{2}} \geq \sigma_N - dN^{\frac{1}{2}}(\beta_\alpha^2 + c \log N)^{-\frac{1}{2}}$$

and the opposite holds if $N$ is replaced by $N - 1$, we have

$$(2.6) \qquad (\beta_\alpha^2 + c \log N)\sigma^2/d^2N \to 1 \quad \text{(a.s.)}.$$

If $d$ is fixed, this shows $n_\alpha(d) = (\beta_\alpha\sigma/d)^2$ is the right choice in Theorem 1. If $d \to 0$, we must show

$$(2.7) \qquad (\log N)/\beta_\alpha^2 \to 0 \quad \text{(a.s.)}.$$

Now, if for some sequence $\beta_\alpha^2/d^2N \to 0$, then (2.6) shows $(\log N)/\beta_\alpha^2 \to \infty$. But with probability approaching one, $(\beta_\alpha^2 + cN^{\frac{1}{2}})\sigma^2/(d^2N) \geq 1$, which would imply $c\sigma^2/(d^2N^{\frac{1}{2}}) \geq 1$ and hence that

$$(\log N)/\beta_\alpha^2 \leq (\log c\sigma^2 - 2\log d)/\beta_\alpha^2 \to 0.$$

This contradiction shows that there exists $\varepsilon_* > 0$ for which $(\beta_\alpha^2/d^2N) \geq \varepsilon_*$ (a.s.) as $\alpha, d \to 0$. This gives $\log N/\beta_\alpha^2 \leq (\log \beta_\alpha^2 - \log \varepsilon_* d^2)/\beta_\alpha^2 \to 0$, so that $N/n_\alpha(d) \xrightarrow{p} 1$. Now, if $d$ is fixed or $d \to 0$, the proof of Theorem 1 gives

$$\left(N - \sigma^2h^2(\alpha,N)/d^2\right)\left(4n_\alpha(d)/d^2\right)^{-\frac{1}{2}} = N^{\frac{1}{2}}\left(d(\sigma_N^2 - \sigma^2)/2\sigma^2 - T_N\right) + o_p(1).$$

Since $\log N/\beta_\alpha \xrightarrow{p} 0$, this completes the proof. □

The next choice of $h(\alpha,n)$ is motivated by the normal case with unknown variance. Let $b = 1 + (cd/\sigma)^2$ ; Swanepoel and Geertsema choose $c^2 = 2$.

<u>Lemma 2</u> Define $h(\alpha,n) = n^{\frac{1}{2}}\{(t_\alpha n)^{1/n} - 1\}^{\frac{1}{2}}/c$, where $c > 0$, $t_\alpha = (1 + a^2)^2/2$ and $\frac{1}{2} - (\arctan a)/\pi + a/(1+a^2)\pi = \alpha/2$. Then $\pi a\alpha \to 4$ as $\alpha \to 0$. Let $n_\alpha(d) = \log t_\alpha/\log b$. Then for $d$ fixed

(2.8) $$\sigma^2 n_\alpha(d)^{-\frac{1}{2}}(\log b)\left(N - n_\alpha(d)\right)/2dc^2 \xrightarrow{L} N\left(0, A(F,d)/b^2\right).$$

If as $\alpha, d \to 0$ there exists $\varepsilon > 0$ for which $d(-\log \alpha)^{\frac{1}{2}-\varepsilon} \to \infty$, then (2.8) still holds with the convergence to $N\left(0, A(F,0)\right)$.

<u>Proof of Lemma 2</u> First consider $d$ fixed. Then, clearly $N \to \infty$, $h(\alpha,N) \to \infty$ and $h(\alpha,N)/N^{\frac{1}{2}} \to d/\sigma$ (all a.s.). Thus, $t_\alpha^{1/N} \to b$ (a.s.). Multiplying and

dividing by $h(\alpha, N-1) + h(\alpha,N)$, (2.1b) will follow if

$$N^{\frac{1}{2}}\{(t_\alpha(N-1))^{1/N-1} - (t_\alpha N)^{1/N}\} \to 0 \quad \text{(a.s.)}.$$

This last term is given by

$$N^{\frac{1}{2}}\left\{\begin{array}{l}(t_\alpha N)^{1/N}((1 - 1/N)^{1/N} - 1) \\ + (t_\alpha(N-1))^{1/N}(1 - t_\alpha^{1/N(N-1)}) \\ + (t_\alpha(N-1))^{1/N-1}((N-1)^{-1/N(N-1)} - 1)\end{array}\right\}.$$

Now, $t_\alpha^{1/N} \to b$ (a.s.), $N^{1/N} \to 1$ (a.s.) and one shows by L'Hospital's rule that for $\mu > 0$, $n^{\frac{1}{2}}(\mu^{1/n} - 1) \to 0$, giving (2.1b). Since $(\log b)N/\log t_\alpha \xrightarrow{p} 1$, Theorem 1 tells us that with $n_\alpha(d) = (\log t_\alpha)/\log b$ ,

$$(2.9)\quad n_\alpha(d)^{\frac{1}{2}}\left[c^{-2}\sigma^2((t_\alpha N)^{1/N} - 1) - d^2\right]/2d$$

$$= N^{\frac{1}{2}}(T_N - d(\sigma_N^2-\sigma^2)/2\sigma^2) + o_p(1), \text{ i.e.,}$$

$$\sigma^2 n_\alpha(d)^{\frac{1}{2}}((t_\alpha N)^{1/N} - b)/2dc^2 \xrightarrow{L} N(0, A(F,d)).$$

Since $N(N^{1/N} - 1)/\log N \to 1$ (a.s.) and $\alpha^4 t_\alpha = O(1)$ we see that as long as $(-\log d)^2/-\log \alpha \to 0$,

$$n_\alpha(d)^{\frac{1}{2}}((t_\alpha N)^{1/N} - t_\alpha^{1/N})/d \xrightarrow{p} 0,$$

so that

$$(2.10)\quad n_\alpha(d)^{\frac{1}{2}}(c^{-2}\sigma^2(t_\alpha^{1/N} - 1) - d^2)/2d = N^{\frac{1}{2}}(T_N - d(\sigma_N^2 - \sigma^2)/2\sigma^2) + o_p(1).$$

By Lehmann (1959, page 274) and a little algebra,

$$\sigma^2 n_\alpha(d)^{-\frac{1}{2}}(\log t_\alpha - N \log b)/2dc^2 \xrightarrow{L} N(0, A(F,d)/b^2),$$

which gives the result. If $d \to 0$, we need only verify (2.9) above. First, since $T_N \to 0$, $t_\alpha^{1/N} \to 1$ so that $(\log t_\alpha)/N \to 0$. Since $\alpha^4 t_\alpha = O(1)$, this gives $(-\log \alpha)/N \to 0$. Since $d(-\log \alpha)^{\frac{1}{2}-\varepsilon} \to \infty$, we get $dN^{\frac{1}{2}}/(\log N)^2 \to \infty$ so by the Law of the Iterated Logarithm, $T_N/d \to 0$ (all statements above being a.s.). Thus, $d^{-2}\{(t_\alpha N)^{1/N} - 1\} \to c^2/\sigma^2$ , i.e.,

$$d^{-2}\{t_\alpha^{1/N} - 1\} + d^{-2}t_\alpha^{1/N}\{N^{1/N} - 1\} \to c^2/d^2 .$$

Since the second term of this last equation is of the order $(\log N)/Nd^2 \to 0$ (a.s.), we have $d^{-2}\{t_\alpha^{1/N} - 1\} \to c^2/\sigma^2$ . From here, the steps of Theorem 1 go through, although the algebra is a bit more complicated. □

Denoting the centering constants in Lemma 1 by $n_\alpha^{(1)}(d)$ and those in Lemma 2 by $n_\alpha^{(2)}(d)$, we see that for $c^2 = 2$,

$$\lim_{\alpha \to 0} n_\alpha^{(2)}(d)/n_\alpha^{(1)}(d) = 2(d/\sigma)^2\{\log(1 + 2(d/\sigma)^2)\}^{-1} .$$

Thus, the two choices of $h(\alpha,n)$ are not equivalent.

Lemma 3 Define $N^* = \inf\{n: T_n \le -h(\alpha,n)/n^{\frac{1}{2}} - (\mu-\Delta)\}$. Then, using either of the $h(\alpha,n)$ in Lemmas 1 and 2, as $d = \mu + \Delta \to 0$,

$$\Pr\{N^* > N\} \to 1.$$

Proof of Lemma 3 Define $M^* = \inf\{n: T_n \le -h(\alpha,n)/n^{\frac{1}{2}} + d/2\}$. Then $N^* \ge M^*$. Under Lemma 1, $M^*/N \xrightarrow{P} 4$, while under Lemma 2, $M^*/N \xrightarrow{P} 4$. □

The upshot of Lemma 3 is that $\Pr\{CS\} \to 1$ when $N \to \infty$ (a.s.) and $d = \mu + \Delta \ge d_0$ , while if $d \to 0$, the same holds for the choices $h(\alpha,n)$ in Lemmas 1 and 2.

## The Ranking Problem

Returning to the ranking problem, the results of the previous section will be illustrated for arbitrary $k$. We assume throughout the rest of this paper that $\mu_1 \le \dots \le \mu_k$, $\mu_k - \mu_{k-1} \ge \Delta$, and define $d_i = \Delta + \mu_k - \mu_i$. Suppose $d_{k-1}/d_i \to \xi_i$ $(0 \le \xi_i \le 1)$ and let $s = s(\Delta)$ be the smallest integer $i \le k-1$ such that $\xi_i = 1$. Assume

$$(3.1) \quad \max_{s \le i \le k-1} n^{\frac{1}{2}}\left(d(\sigma_{ni}^2 + \sigma_{nk}^2 - \sigma_i^2 - \sigma_k^2)/2(\sigma_i^2 + \sigma_k^2) - (T_{nk} - T_{ni} - \mu_k + \mu_i)\right) \xrightarrow{L} F.$$

The distribution of $N$, the number of observations taken on the selected population, will be computed in the following cases.

Lemma 4 Under the conditions of Lemma 1,

$$d\left(N - n_\alpha(d)\right)/2n_\alpha(d)^{\frac{1}{2}} \xrightarrow{L} F.$$

Proof of Lemma 4 If $N_i$ is the time it takes for population $k$ to eliminate population $i$, Lemmas 1 and 3 show $\Pr\{N = \max_{s \le i \le k-1} N_i\} \to 1$. But for $s \le i \le k-1$, $N_i/N \xrightarrow{p} 1$. A multivariate extension of Anscombe's Theorem 1 and (2.5) complete the proof. □

Lemma 5 Define $G(x) = F(bx)$, where $b = 1 + c^2d^2/\sigma^2$. Under the conditions of Lemma 2,

$$\sigma^2 n_\alpha(d)^{-\frac{1}{2}}(\log b)\left(N - n_\alpha(d)\right)/2dc^2 \xrightarrow{L} G.$$

Proof of Lemma 5 By (2.10) and the argument of Lemma 4 we obtain

$$n_\alpha(d)^{\frac{1}{2}}\sigma^2(t_\alpha^{1/N} - b)/2dc^2 \xrightarrow{L} F.$$

By a Taylor expansion,

$$n_\alpha(d)^{\frac{1}{2}}\sigma^2(\log t_\alpha^{1/N} - \log b)/2dc^2 \xrightarrow{L} G,$$

which with a little algebra completes the proof. □

## Other Choices of $h(\alpha,n)$

The selection in Lemmas 1 and 2 by no means exhaust the possibilities for the function $h(\alpha,n)$. For example, one could define $h(\alpha,n) = (r/n)^{\frac{1}{2}} g_\alpha(n/r)$, where

$$g_\alpha(t) = (t + \tfrac{1}{4})^{\frac{1}{2}}\{\log(t + \tfrac{1}{4}) + \beta_\alpha^2\}^{\frac{1}{2}},$$

and where $r = 1/\Delta^a$ $(1 \le a \le 2)$. This is the choice suggested by Swanepoel and Carroll for the case $\alpha$ fixed, $d \to 0$. The analysis is essentially as in Lemma 1.

Another possibility arises from the recent work of Wackerly (1975) on noneliminating sequential procedures. His idea is that there is a vector $\underset{\sim}{n}(\alpha,\Delta,\underset{\sim}{\mu})$ of fixed sample sizes needed to guarantee a probability requirement under the condition that the distributions are known up to location parameters. The goal is to define a vector $\underset{\sim}{N}$ of observations which satisfy $e'\underset{\sim}{N}/e'\underset{\sim}{n}(\alpha,\Delta,\underset{\sim}{\mu}) \xrightarrow{p} 1$ as $\alpha \to 0$, where $e' = (1,1,\ldots,1)$. He is only partially successful in that the convergence is to $c(\Delta,\underset{\sim}{\mu}) \ge 1$ with equality if and only if $\mu_1 = \ldots = \mu_{k-1} = \mu_k - \Delta$. The following remarks sketch very briefly an elimination rule which satisfies the original goal.

We assume $F_i(x)$ is symmetric about $\mu_i$ and define $Y_{ijn} = X_{in} - X_{jn}$. Then large deviation theory shows that for $i > j$,

$$\lim_{n\to\infty} n^{-1} \log \Pr\{n^{-1} \sum_{p=1}^{n} Y_{ijp} < 0\} = A(\underset{\sim}{F}, \mu_i - \mu_j),$$

$$A(\underset{\sim}{F}, \delta) = \log \inf_t \left\{ \exp(t\delta/2) E \exp\{t(Y_{ij1} - \mu_i + \mu_j)\}\right\}.$$

Thus, if the maximal error probability desired is $\alpha$ and $\beta_\alpha^2/-\log \alpha \to 1$, the correct sample size for fixed $\underset{\sim}{\mu}$ to eliminate $\pi_i$ is

$$n(\alpha, \Delta, \mu_k - \mu_i) \sim \beta_\alpha^2/-A(F, \mu_k - \mu_i).$$

Define $H_{ijn} = \max(\Delta, |n^{-1} \sum_{p=1}^{n} Y_{ijp}|)$ and $H_{ij}^* = \max(\Delta, |\mu_i - \mu_k|)$. Wackerly defined the natural estimate of $A(F,\delta)$ which, for some function $\psi$, satisfies (if $i \geq j$)

$$A_n(\underset{\sim}{F}, H_{ijn}) - A(\underset{\sim}{F}, H_{ij}^*) = n^{-1} \sum_{p=1}^{n} \{\psi(Y_{ijp} - \mu_i + \mu_j) - E\psi(Y_{ij1} - \mu_i + \mu_j\}$$
$$+ o(n^{-\frac{1}{2}}) \quad \text{(a.s.)}.$$

The proof of this fact follows along the lines of Carroll (1976) and the details are omitted for the sake of brevity. The elimination stopping rule eliminates $\pi_i$ at stage $n$ if for some $\pi_j$ still in contention, $n^{-1} \sum_{p=1}^{n} Y_{ijp} < 0$ and $n \geq \beta_\alpha^2/-A_n(\underset{\sim}{F}, H_{ijn})$, i.e., if

$$-\{A_n(\underset{\sim}{F}, H_{ijn}) - A(\underset{\sim}{F}, H_{ij}^*)\} \geq h(\alpha,n)/n^{\frac{1}{2}} - d,$$

$$h(\alpha,n) = \beta_\alpha^2/n^{\frac{1}{2}}, \quad d = A(\underset{\sim}{F}, H_{ij}^*).$$

If one wishes to analyze the number of stage $N$ as $\alpha \to 0$, since $\Pr\{CS\} \to 1$, merely replace $i$ and $j$ in the above by $k$ and $k-1$ and then Theorem 1 applies.

## Moments of N

We present a simple proof that $EN/n_\alpha(d) \to 1$ (confer Swanepoel and Geertsema (1976)). Consider the ranking problem with $k = 2$ and define $M = \inf\{n: h^2(\alpha,n)(\sigma_{1n}^2 + \sigma_{2n}^2) < n\Delta^2\}$. Then $N \le M$ since if $N > M$ were true, then

$$T_{2M} - T_{1M} < h(\alpha,M)(\sigma_{1M}^2 + \sigma_{2M}^2)^{\frac{1}{2}} + \Delta < 0$$

$$T_{2M} - T_{1M} > -h(\alpha,M)(\sigma_{1M}^2 + \sigma_{2M}^2)^{\frac{1}{2}} - \Delta < 0,$$

the final inequalities following from the definition of $M$. Thus $EN/n_\alpha(d) \to 1$ if $M/n_\alpha(d)$ is uniformly integrable. By Bickel and Yahav (1968), it suffices to show that for some $\alpha_0 > 0$,

$$\sum_{p=1}^{\infty} \sup_{0<\alpha<\alpha_0} \Pr\{M/n_\alpha(d) > p\} < \infty. \tag{5.1}$$

Assume $(\sigma_{1n}^2 + \sigma_{2n}^2)$ has bounded $r^{th}$ moments for $n$ large (call this bound $c_0$); then (5.1) is bounded by (setting $m = pn_\alpha(d)$)

$$\sum_{p=1}^{\infty} \sup_{0<\alpha<\alpha_0} \Pr\{h^2(\alpha,m)(\sigma_{1m}^2 + \sigma_{2m}^2) > m\Delta^2\}$$
$$\leq \sum_{p=1}^{\infty} \sup_{0<\alpha<\alpha_0} \left\{\frac{c_0 h^2(\alpha, pn_\alpha(d))}{pn_\alpha(d)^{\ 2}}\right\}^r .$$

<u>**PROPOSITION 1**</u> If $r > 1$ and $h^2(\alpha,n) = (\beta_\alpha^2 + c \log n)$, then (5.1) holds.

<u>**PROPOSITION 2**</u> If $r \geq 2$ and $h(\alpha,n)$ is given in Lemma 2, then (5.1) holds.

<u>Proof of Proposition 2</u> $t_\alpha^{1/n_\alpha(d)} \to b$ so an application of L'Hospital's rule shows that if $\eta < 1$,

$$p^\eta\left\{\frac{h^2(\alpha, pn_\alpha(d))}{pn_\alpha(d)\Delta^2}\right\} \to 0,$$

uniformly in $\alpha$.

<u>Estimation by Sample Means</u>

In the one-sided rule of Section 2, we have implicitly shown that if $T_n - \mu + T_n^* - \mu + o(n^{-\frac{1}{2}}) = n^{-1} \sum_1^n \psi(X_i) - E\psi(X_1) + o(n^{-\frac{1}{2}})$ and $M$ is the stopping rule for $T_n^*$, then $M$ and $N$ will typically have the same asymptotic distributions. One might conjecture than that $M - N \xrightarrow{p} 0$. If $M - N \xrightarrow{p} 0$, $h(\alpha,n) = \beta_\alpha \sim (-2 \log \alpha)^{\frac{1}{2}}$, $\sigma_n^2 = 1$, and we neglect overshoot, we obtain

$$\beta_\alpha^{-1} \sqrt{M}\,\sqrt{N}\,(\sqrt{N} + \sqrt{M})(T_N - T_M^*) \sim \sigma(M - N) \xrightarrow{p} 0, \text{ i.e.,}$$

$$N(T_N - T_N^*) \xrightarrow{p} 0,$$

the last following from the fact that $\sqrt{N}/\beta_\alpha$ has a limit. If one defines $T_n$ to be the Huber M-estimate (the solution to $\sum_1^n \psi^*(X_i - T_n) = 0$ ), one can show by Taylor expansions that $n(T_n - T_n^*) \xrightarrow{L} F$, where $F$ is non-degenerate. This shows that $M - N \xrightarrow{p} 0$ is probably not true, although $E(M - N) \to 0$ may hold.

Monte-Carlo

A simulation experiment was performed ( $k = 2$, $\alpha = .10$, 200 iterations) when the sampling distribution for $\pi_i$ was $F^{(j)}(x - \mu_i)$, where

$$F^{(1)}(x) = \Phi(x) \quad \text{(the standard normal)}$$
$$F^{(2)}(x) = .95\Phi(x) + .05\Phi(x/3)$$
$$F^{(3)}(x) = .85\Phi(x) + .15\Phi(x/3).$$

In all cases, $\mu_3 - \mu_2 = \Delta$ (= .25, .375, or .50) and the two cases $\mu_2 - \mu_1 = 0$ (slippage configuration) and $\mu_2 - \mu_1 = \Delta$ (equally spaced means) were considered. The statistics $T_{ni}$, $\sigma_{ni}^2$ were either the sample mean and variance or a 10% trimmed mean and its winsorized variance estimate. $h(\alpha,n) = (\beta_\alpha^2 + (1.05)(\log n)/m)^{\frac{1}{2}}$, where $m = 10$ was the initial number of observations $M$ taken being tabulated. "Ratio to RSS" denotes the ratio of $M$ to the expected total number of observations taken by the Robbins, Sobel and Starr procedure.

The results are clear. First, the elimination rule attains (and usually exceeds) $.90 = 1 - \alpha$, its predetermined Pr{CS}. Second, the rule is much superior to the nonelimination rule, the superiority tending to be more pronounced as $\Delta$ decreases. Third, the trimmed mean is more robust than the sample mean to heavy tails (this is the situation $F^{(3)}$ ); the rules using the trimmed mean tending to take less than 75% of the number of observations needed by the sample mean.

TABLE 1

The Monte-Carlo experiment for $\phi(x)$.

| | Pr(CS) | # of Observations on $\pi_3$ | Total # of Observations = M | Ratio to RSS |
|---|---|---|---|---|
| Sample Mean | | | | |
| $\mu_2-\mu_1=0$, $\Delta=.50$ | .95 | 20 | 53 | .89 |
| $=\Delta$ | .96 | 17 | 46 | .77 |
| $=0$, $\Delta=.375$ | .95 | 33 | 84 | .79 |
| $=\Delta$ | .96 | 27 | 67 | .63 |
| $=0$, $\Delta=.25$ | .92 | 61 | 156 | .63 |
| $=\Delta$ | .98 | 54 | 130 | .54 |
| 10% Trimmed Mean | | | | |
| $\mu_2-\mu_1=0$, $\Delta=.50$ | .96 | 21 | 57 | .85 |
| $=\Delta$ | .97 | 18 | 47 | .71 |
| $=0$, $\Delta=.375$ | .93 | 31 | 82 | .69 |
| $=\Delta$ | .95 | 28 | 70 | .59 |
| $=0$, $\Delta=.25$ | .94 | 63 | 164 | .62 |
| $=\Delta$ | .96 | 57 | 136 | .51 |

TABLE 2

The Monte-Carlo experiment for $.95\Phi(x) + .05\Phi(x/3)$.

| | Pr(CS) | # of Observations on $\pi_3$ | Total # of Observations = M | Ratio to RSS |
|---|---|---|---|---|
| Sample Mean | | | | |
| $\mu_2-\mu_1=0$, $\Delta=.50$ | .98 | 25 | 64 | .77 |
| $=\Delta$ | .95 | 23 | 58 | .69 |
| $=0$, $\Delta=.375$ | .92 | 39 | 102 | .69 |
| $=\Delta$ | .95 | 36 | 88 | .59 |
| $=0$, $\Delta=.25$ | .92 | 75 | 193 | .58 |
| $=\Delta$ | .94 | 71 | 171 | .51 |
| 10% Trimmed Mean | | | | |
| $\mu_2-\mu_1=0$, $=.50$ | .92 | 24 | 64 | .77 |
| $=\Delta$ | .99 | 19 | 50 | .59 |
| $=0$, $\Delta=.375$ | .95 | 34 | 89 | .68 |
| $=\Delta$ | .96 | 30 | 76 | .57 |
| $=0$, $\Delta=.25$ | .94 | 76 | 189 | .64 |
| $=\Delta$ | .94 | 56 | 138 | .47 |

TABLE 3

The Monte-Carlo experiment for $.85\Phi(x) + .15\Phi(x/3)$.

| | Pr(CS) | # of Observations on $\pi_3$ | Total # of Observations = M | Ratio to RSS |
|---|---|---|---|---|
| Sample Mean | | | | |
| $\mu_2-\mu_1=0,\ \Delta=.50$ | .94 | 33 | 87 | .66 |
| $=\Delta$ | .96 | 31 | 77 | .59 |
| $=0,\ \Delta=.375$ | .91 | 60 | 155 | .66 |
| $=\Delta$ | .94 | 50 | 123 | .53 |
| $=0,\ \Delta=.25$ | .91 | 115 | 297 | .57 |
| $=\Delta$ | .98 | 105 | 252 | .48 |
| 10% Trimmed Mean | | | | |
| $\mu_2-\mu_1=0,\ \Delta=.50$ | .97 | 27 | 70 | .78 |
| $=\Delta$ | .97 | 25 | 63 | .70 |
| $=0,\ \Delta=.375$ | .94 | 46 | 118 | .74 |
| $=\Delta$ | .95 | 36 | 89 | .56 |
| $=0,\ \Delta=.25$ | .88 | 89 | 233 | .65 |
| $=\Delta$ | .95 | 70 | 171 | .47 |

## ACKNOWLEDGEMENT

Mr. Vernon Chinchilli programmed the Monte-Carlo simulation and his help is gratefully acknowledged.

## REFERENCES

ANSCOMBE, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc. 48*, 600-617.

BECHHOFER, R. E., KIEFER, J. and SOBEL, M. (1968). *Sequential Identification and Ranking Procedures*. University of Chicago Press.

CARROLL, R. J. (1975). An asymptotic representation for M-estimators and linear functions of order statistics. Institute of Statistics Mimeo Series #1006, University of North Carolina at Chapel Hill.

__________, (1976). On the Serfling-Wackerly alternate approach. Institute of Statistics Mimeo Series #1052, University of North Carolina at Chapel Hill.

__________, (1976). On Swanepoel's elimination procedures. Institute of Statistics Mimeo Series #1074, University of North Carolina at Chapel Hill.

GEERTSEMA, J. C. (1972). Nonparametric sequential procedures for selecting the best of k populations. *J. Am. Statist. Assoc. 67*, 614-616.

ROBBINS, H., SOBEL, M. and Starr, N. (1968). A sequential procedure for selecting the largest of k means. *Ann. Math. Statist. 39*, 88-92.

SWANEPOEL, J. W. H. (1976). Nonparametric elimination selection procedures based on robust estimators. Unpublished manuscript.

__________, and GEERTSEMA, J. C. (1976). Sequential procedures with elimination for selecting the best of k normal populations. To appear in *S. Afr. Statist. J.*

WACKERLY, D. D. (1975). An alternative sequential approach to the problem of selecting the best of k populations. University of Florida Technical Report #91.