Convergence Results for Sequential Estimation of the Largest Mean

by

Raymond J. Carroll*

University of North Carolina at Chapel Hill

Abstract

We consider the sequential estimation of the largest mean of k populations when the observations are normally distributed with a common unknown variance and the goal is to control the mean square error (MSE) at a prespecified level; this is a generalization of problems considered by Blumenthal (1976) and Carroll (1977). By eliminating from the experiment populations which the data indicate are not associated with the largest mean, it is shown that, compared to existing procedures, significant savings in sample size can be obtained. Weak convergence results are obtained for the stopping times and the estimate of the largest mean as consequences of more general results; these are used to compute the asymptotic MSE.

Key Words and Phrases: Sequential Estimation, Elimination, Largest Normal Mean, Weak Convergence, Random Time Change, Ranking and Selection

*This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

1. Introduction

Let $\theta_1,\ldots,\theta_k$ be the unknown means of k normal populations with common unknown variance σ^2, and let $\overline{X}_{1n},\ldots,\overline{X}_{kn}$ be the sample means for n observations taken from the k populations. Define the ordered population and sample means by $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ and $\overline{X}_{[1]n} \leq \cdots \leq \overline{X}_{[k]n}$. Blumenthal (1973, 1976) constructed sequential procedures for estimating the largest mean $\theta_{[k]}$ with a prespecified bound r on the mean square error (MSE). His procedures were mildly data sensitive in that they depend on estimates $\Delta_{in} = \overline{X}_{[k]n} - \overline{X}_{[i]n}$ of $\Delta_i = \theta_{[k]} - \theta_{[i]}$, and he obtained some partial asymptotic results. Carroll (1978a) extended the asymptotic results when σ^2 is known.

The purpose of this paper is twofold. In Section 2 we consider weak convergence results which greatly extend the asymptotic theory for Blumenthal's stopping time N_B and generalize the results of Carroll (1978a). We define and study the weak convergence of a stopping time process $\{N_r\}$ that includes N_B as a special finite-dimensional case. We then define a new random change of time process for sample means that is based on $\{N_r\}$ and consider its weak convergence. We finally study the limit distribution and MSE for the maximum sample mean upon stopping.

The second purpose of this paper is based upon our belief that N_B can be made more data sensitive and efficient by the simple expedient of grafting onto it the ability to eliminate populations (early in the experiment) which are obviously not associated with and hence give no information about $\theta_{[k]}$. In Sections 3 and 4 we define the elimination procedure, study its asymptotic behavior and give some Monte-Carlo results which show that the savings (over N_B) in sample size can be considerable.

To fix notation, define $H_1 \equiv 1$ and let $(\sigma^2/n)H_k(n^{\frac{1}{2}}\Delta_1/\sigma,\ldots,n^{\frac{1}{2}}\Delta_{k-1}/\sigma)$ be the MSE due to estimating $\theta_{[k]}$ by $\theta_n^* = \overline{X}_{[k]n}$. In order to control MSE at a level r, when σ^2 is known the following procedures have been proposed: take N_B^* observations from each population, where either (Blumenthal (1973))

$$(1.1) \qquad N_B^* = N_B^*(k) = \inf\{n: nr \geq \sigma^2 H_k(n^{\frac{1}{2}}\Delta_{in}/\sigma,\ldots,n^{\frac{1}{2}}\Delta_{k-1,n}/\sigma)\},$$

or (Blumenthal (1976))

$$(1.2) \qquad N_B^* = N_B^*(k) = \inf\{m: \hat{n}(m) \leq m\}, \text{ where}$$

$$\hat{n}(m) = \inf\{t: rt \geq \sigma^2 H_k(t^{\frac{1}{2}}\Delta_{im},\ldots,t^{\frac{1}{2}}\Delta_{k-1,m})\}.$$

Although an analogue of (1.2) (for the case σ^2 unknown) is possible and requires the same basic techniques as used here, for notational reasons we prefer to consider (1.1) and take N_B observations from each population, where

$$(1.3) \quad N_B = N_B(k) = \inf\{n \geq ar^{-1}: nr \geq \sigma_{nk}^2 H_k(n^{\frac{1}{2}}\Delta_{im}/\sigma_{nk},\ldots,n^{\frac{1}{2}}\Delta_{k-1,n}/\sigma_{nk})\}.$$

Here $a = a(r) > 0$ are a set of small bounded constants with finite positive limit a_0 and with ar^{-1} an integer, while σ_{nk}^2 is the usual pooled sample variance with $k(n-1)$ degrees of freedom.

We make the following

Assumptions. The i.i.d. observations $X_{j1}, X_{j2}, \ldots$ from the jth population have finite fourth moment. The functions H_k are continuous and satisfy $0 < H_{min} \leq H_k(x_1,\ldots,x_{k-1}) \leq H_{max} < \infty$. Further, for every k and p,

$$\lim_{x_1,\ldots,x_p \to \infty} H_k(x_1,\ldots,x_{k-1}) = H_{k-p}(x_{p+1},\ldots,x_{k-1}) .$$

Finally, for every k and u, the Lebesgue measure of the set

$$\{(x_1,\ldots,x_{k-1}): H_k(x_1,\ldots,x_{k-1}) = u\}$$

is zero.

If the observations are normally distributed, the assumptions hold for the MSE function H_k. We take $0 < ka_0 < H_{min} < H_{max}$, and it will simplify computations without affecting results if we take $H_{max}=1$. With k=2, define $\Delta_n=\overline{X}_{[2]n}-\overline{X}_{[1]n}$, $\Delta = \theta_{[2]} - \theta_{[1]}$ and without loss of generality take $\theta_1 \leq \theta_2$ and $\sigma^2 = 1$.

2. Weak Convergence Results for N_B

In this section we prove a number of weak convergence results for a stopping time process that includes N_B as a special (finite dimensional) case. All results are given for k=2 but are easy to generalize to $k \geq 3$. To outline important special cases of the results, in Lemma 1 we give the limit distribution of rN_B, in Lemma 2 we discuss a random change of time for sample means, and in Lemma 3 we establish the limit distribution of the maximum sample mean upon stopping. We assume throughout this section that $\Delta \sim r^{\beta}$ for some $0 \leq \beta < \infty$ and $\Delta^2/r \to \eta_0^2$ $(0 \leq \eta_0 \leq \infty)$. Let W be Brownian motion with mean zero and variance 2t at time t, and define

$$W^*(s,\eta_0) = s^{-\frac{1}{2}} \left|W(s) + s\eta_0\right| .$$

Letting $[\cdot]$ denote the greatest integer function, define a stochastic process

$$G_r(s) = [s/r]^{\frac{1}{2}} \Delta_{[s/r]}/\sigma_{[s/r]} .$$

The proofs of all results are delayed to the end of the section.

<u>Proposition 1</u>. Let $0 < b_1 < b_2 < \infty$ and "$\Rightarrow$" denote weak convergence. On the space $D[b_1,b_2]$ (Billingsley (1968)),

$$G_r \Rightarrow W^*(\cdot,\eta_0) \qquad (\beta \geq \tfrac{1}{2})$$

$$G_r \xrightarrow{p} \infty \qquad (\beta < \tfrac{1}{2}) .$$

In studying the stopping time N_B, we have found a more general approach to be as convenient and to yield much stronger results. Consider processes for $0 \le t \le 1$ given by

$$N_r(t) = \inf\{m \ge ar^{-1}: mr(t+1) \ge \sigma_m^2 H_2(m^{\frac{1}{2}} \Delta_m/\sigma_m)\}$$

$$Q(t) = \inf\{s \ge a: s(t+1) \ge H_2(W^*(s,\eta_0))\} .$$

Note that $N_r(0) = N_B$. Both N_r and Q are monotone non-increasing in t and are easily verified to be members of D[0,1]. Lemma 1 is comparable in spirit to work of Gut (1975).

Lemma 1. (Weak convergence of the stopping time process). For $\beta < \frac{1}{2}$, $rN_B = rN_r(0) \overset{p}{\to} 1$. For $\beta \ge \frac{1}{2}$,

(2.1) $$rN_r \Rightarrow Q \quad \text{on} \quad D[0,1] .$$

Define G_t^* by

(2.2) $$G_t^*(u) = \Pr\{Q(t) > u\} = \Pr\{(H_2(W^*(s,\eta_0)) - s(t+1)) > 0 \text{ for all } a \le s \le u\} .$$

Since $2a < H_{min}$ and H is bounded, it is easy to show that $1-G_t^*$ is a proper distribution function. Of more interest is the following result.

Corollary 1. (Distribution of Blumenthal's stopping time). In Lemma 1 for $\beta \ge \frac{1}{2}$,

$$\Pr\{rN_B > u\} \to G_0^*(u) = \Pr\{Q(0) > u\}.$$

The next result will be useful in discussing the limit distribution of the larger sample mean when sampling is stopped, but it is quite general and may be of some interest in its own right for the following reasons. Typical results in the theory of weak convergence with random indices (Durrett and Resnick (1976) have a nice review) start with processes $\{V_r\}$ in $D[0,\infty)$ and a sequence of integer-valued random variables M_r and consider the process $V_r(trM_r)$ on $D[0,\infty)$, where rM_r has a limit distribution. In other words, V_r is perturbed by a "random time change" proportional to rM_r. In the next result, we allow the random time change rM_r to be itself a stochastic process. Define $m = [t/r]$ and for $j=1,2$ let

$$V_r^{(j)}(t) = r^{\frac{1}{2}} \{ \sum_{i=1}^{m} (X_{ji} - \theta_j) + (t/r - m)(X_{jm+1} - \theta_j)\} .$$

The processes $V_r^{(j)}$ are elements of $C[0,\infty)$ with weak limits $V^{(j)}$.

Lemma 2. (Weak convergence for random change of time processes). Define processes

$$W_r^{(j)}(s,t) = V_r^{(j)}(srN_r(t)) \tag{2.3}$$

on $D_2\{[0,\infty) \times [0,1]\}$ (Bickel and Wichura (1971)). Then for $\beta \geq \frac{1}{2}$,

$$(W_r^{(1)}, W_r^{(2)}, rN_r) \Rightarrow (W^{(1)}, W^{(2)}, Q) ,$$

where $W^{(j)}(s,t) = V^{(j)}(sQ(t))$.

The last Lemma will be shown useful when we discuss the specific proposal for eliminating the inferior population early in the experiment. It is a simple Corollary of Lemma 2 which delineates the asymptotic behavior of the larger sample mean when the number of observations are approximately N_B.

Lemma 3. (Limit distribution of the larger sample mean). Let $M_r^{(j)}$ and $\overline{X}_M^{(j)}$ be the number of observations and the sample mean after $M_r^{(j)}$ observations on the jth population, where

$$1 \geq M_r^{(j)}/N_B = M_r^{(j)}/N_r(0) \overset{p}{\to} 1 \qquad (j = 1,2) .$$

Let $\theta_r^* = \max\{\overline{X}_M^{(1)}, \overline{X}_M^{(2)}\}$. Then, for $\beta \geq \frac{1}{2}$,

$$r^{-\frac{1}{2}}(\theta_r^* - \theta_2) \Rightarrow Q(0)^{-1} \max\{V^{(1)}(Q(0)) - \eta_0, V^{(2)}(Q(0))\}.$$

Proof of Proposition 1: The process $G_r(s)$ can be written as

$$(2.4)\qquad G_r(s) = |[s/r]^{\frac{1}{2}} (\bar{X}_{2[s/r]} - \bar{X}_{1[s/r]} - \theta_2 + \theta_1) + [s/r]^{\frac{1}{2}} \Delta|/\sigma_{[sk]} .$$

The denominator of (2.4) converges almost surely to $\sigma = 1$ while the numerator converges weakly to $W^*(\cdot, \eta_0)$, completing the proof. □

Proof of Lemma 1: We first prove (2.1) by verifying tightness and the convergence of the finite dimensional distributions. Note first that

$$(2.5)\qquad \Pr\{rN_r(t_i) > u_i \quad (i = 1,\ldots,p)\}$$
$$= \Pr\{mr(t_i+1) < \sigma_m^2 H_2(m^{\frac{1}{2}}\Delta_m/\sigma_m) \text{ for all } a \le mr \le u_i \quad (i = 1,\ldots,p)\}$$
$$= \Pr\{[s/r]r(t_i+1) < \sigma_{[s/r]}^2 H_2(G_r(s)) \text{ for all } a \le s \le u_i \quad (i = 1,\ldots,p)\} ,$$

the last equation using the facts that ar^{-1} is an integer and $2a < H_{min}$. Rewrite (2.5) as

$$(2.6)\qquad \Pr\{rN_r(t_i) > u_i \quad (i = 1,\ldots,p)\}$$
$$= \Pr\{ \inf_{a\le s\le u_i} (\sigma_{[s/r]}^2 H_2(G_r(s)) - [s/r]r(t_i+1)) > 0 \quad (i = 1,\ldots,p) .$$

From Proposition 1, the continuous mapping theorem (since inf is continuous in this context) and Theorem 2.1 of Billingsley (1968), (2.6) shows that as $r \to 0$,

$$\liminf \Pr\{rN_r(t_i) > u_i \quad (i = 1,\ldots,p)\}$$
$$\ge \Pr\{Q(t_i) > u_i \quad (i = 1,\ldots,p)\} ,$$

thus verifying the convergence of the finite dimensional distributions. To prove tightness, we appeal to Theorem 15.2 of Billingsley (1968). The first condition of his Theorem is satisfied in our case because the process rN_r is non-increasing and $rN_r(0)$ has been shown to have a limit distribution. To check the second condition of Billingsley's Theorem 15.2, we must show (in his notation) that for all $\varepsilon > 0$,

$$\lim_{\delta\to 0}\ \lim_{r\to 0} \Pr\{\omega'_{rN_r}(\delta/2) \geq \varepsilon\} = 0 \tag{2.7}$$

Now, since rN_r and Q are non-increasing,

$$\begin{aligned}(2.8)\quad & \lim_{r\to 0} \Pr\{\omega'_{rN_r}(\delta/2) \geq \varepsilon\} \\ & \leq \lim_{r\to 0} \Pr\{rN_r(i\delta) - rN_r((i+1)\delta) \geq \varepsilon \text{ for some } i = 0,1,\ldots,[1/\delta]\} \\ & \leq \Pr\{Q(i\delta) - Q((i+1)\delta) \geq \varepsilon \text{ for some } i = 0,1,\ldots[1/\delta]\} \\ & \leq \Pr\{\omega'_Q(4\delta) \geq \varepsilon/4\}\ ,\end{aligned}$$

the next to last inequality following by the weak convergence of the finite dimensional distributions, while the last follows because Q is non-increasing. Then (2.7) follows from (2.8) because Q is an element of D[0,1].

The rest of Lemma 1 ($\beta < \frac{1}{2}$) follows in a similar but easier fashion. □

<u>Proof of Corollary 1</u>: By Lemma 1, we need only show that G_1^* is continuous. Letting I(A) be the indicator of the event A,

$$\lim_{\varepsilon\downarrow 0} I\{Q(0) > u + \varepsilon\} = I\{Q(0) > u\}$$

$$\lim_{\varepsilon\uparrow 0} I\{Q(0) > u + \varepsilon\} = I\{Q(0) > u\} + I\{H_2(W^*(u,\eta_0)) - u = 0\}\ .$$

Now, by assumption, $\Pr\{H_2(W^*(u,\eta_0)) = u\} = 0$, so that G_1^* is continuous. □

In order to prove Lemma 2, we need the following supplementary results. For intervals T_1, T_2 of the real line, we define $D_2\{T_1\times T_2\}$ to be the space of functions $x(s,t)$ ($s\in T_1$, $t\in T_2$) which are continuous from above with limits from below (Bickel and Wichura (1971)).

Proposition 2. Let a,b be arbitrary positive numbers and define $D_0(0,a]$ to be the set of functions ϕ in $D[0,a]$ which are nonincreasing and satisfy $0 \leq \phi(t) \leq b$ for $0 \leq t \leq a$. Let $\{V_n\}$ be elements of $C[0,\infty)$, $\{\Phi_n\}$ be elements of $D_0[0,a]$ and suppose there exists an element (V,Φ_0) in $C[0,\infty] \times D_0(0,a]$ satisfying

$$(V_n, \Phi_n) \Rightarrow (V,\Phi_0) .$$

Define random elements $\{V_n^*\}$, V^* of $D_2\{[0,\infty) \times [0,a]\}$ by $V_n^*(s,t) = V_n(s\ \Phi_n(t))$. Then on $D_2\{[0,\infty) \times [0,a]\}$,

$$V_n^* \Rightarrow V^* .$$

Proof: Define $V_n^{(1)}(s,t) = V_n(s,t)$, $V^{(1)}(s,t) = V(s,t)$, so that $\{V_n^{(1)}\}$, $V^{(1)}$ are elements of $C_2[0,\infty]$. Denote by A the space $C_2\{[0,\infty) \times [0,a]\} \times D_0[0,a]$ and define a function $h: A \to D_2\{[0,\infty) \times [0,b]\}$ by

$$h(x,v)(s,t) = x(s,v(t)) .$$

This is shown to be a measurable mapping following Billingsley (1968, page 232). Since $V_n^* = h(V_n^{(1)}, \Phi_n)$ and

$$(V_n^{(1)}, \Phi_n) \Rightarrow (V^{(1)}, \Phi_0) ,$$

the continuous mapping theorem completes the proof once we show that h is continuous at elements of A. Let $(x_n, \phi_n) \to (x,\phi) \in A$. Then there exists functions λ_n mapping $[0,a]$ into $[0,a]$ such that for every $c > 0$,

$$\sup\{|x_n(s,t) - x(s,t)|: 0 \leq s \leq c,\ 0 \leq t \leq a\} \to 0$$

$$\sup\{\max[|\phi_n(\lambda_n(t)) - \phi(t)|,\ |\lambda_n(t) - t|]: 0 \leq t \leq a\} \to 0 .$$

These two facts imply that for every $c_0 > 0$,

$$\sup\{|x_n(s, \phi_n(\lambda_n(t))) - x(s,\phi(t))| \; 0 \le s \le c_0, \; 0 \le t \le a\} \to 0 \; .$$

Since $c_0 > 0$ is arbitrary, following Lindvall (1973) and Whitt (1970), this shows that h is continuous at (x,ϕ) and completes the proof. □

Proposition 3. On $C[0,\infty) \times C[0,\infty) \times D[0,1]$, if $\beta \ge \frac{1}{2}$,

$$(V_r^{(1)}, V_r^{(2)}, rN_r) \Rightarrow (V^{(1)}, V^{(2)}, Q) \; . \tag{2.9}$$

Proof: By Lemma 1, each of the elements of (2.9) are individually and hence jointly tight, so it suffices to prove convergence of the finite dimensional distributions. We show this only for a special case, noting that

$$\Pr\{V_r^{(1)}(t_1) > u_1, V_r^{(2)}(t_2) > u_2, rN_r(t_3) > u_3\} \tag{2.10}$$

$$= \Pr\{V_r^{(1)}(t_1) > u_1, V_r^{(2)}(t_2) > u_2, \inf_{a \le s \le u_3} (\sigma^2_{[s/r]} H_2(G_r(s)) - [s/r]r(t_3+1))>0\}.$$

We assume with no loss of generality that $0 \le t_1, t_2, u_3 \le 1$. Since fourth moments are finity, for any u

$$\sup\{r^{\frac{1}{2}} |X_{jm} - \theta_j|: 1 \le m \le ur^{-1}\} \overset{p}{\to} 0. \tag{2.11}$$

This shows that the second term in the definition of $V_r^{(j)}$ is negligible, so that (since G_r is a continuous function of the first terms in $V_r^{(j)}$), on $C[0,1]\times C[0,1]\times D[a,1]$,

$$(V_r^{(1)}, V_r^{(2)}, G_r) \Rightarrow (V^{(1)}, V^{(2)}, W^*(\cdot,\eta_0)) \; .$$

Thus, as $r \to 0$, the continuous mapping theorem and Theorem 2.1 of Billingsley (1968) show that

$$\liminf \Pr\{V_r^{(1)}(t_1) > u_1, V^{(2)}(t_2) > u_2, rN_r(t_3) > u_3\}$$

$$\ge \Pr\{V^{(1)}(t_1) > u_1, V^{(2)}(t_2) > u_2, Q(t_3) > u_3\} \; ,$$

which proves convergence of the finite dimensional distributions and completes the proof. □

Proof of Lemma 2: The boundedness of H_2 means that with probability one there exist positive numbers a_1, a_2 such that

$$a_1 < \inf\{Q(t): 0 \leq t \leq 1\} < \sup\{Q(t): 0 \leq t \leq 1\} < a_2 .$$

Define a process $M_r(t)$ by

$$rM_r(t) = rN_r(t)\, I\{a_1 \leq rN_r(t) \leq a_2\} + a_2\, I\{rN_r(t) > a_2\} + a_1\, I\{rN_r(t) < a_1\} ,$$

and define $Z_r^{(j)}(s,t) = V_r^{(j)}(srM_r(t))$. By an extension of Proposition 2 and by Proposition 3,

$$(Z_r^{(1)}, Z_r^{(2)}, rN_r) \Rightarrow (W^{(1)}, W^{(2)}, Q) .$$

Now, since N_r is non-increasing,

$$\Pr\{M_r(t) \neq N_r(t)\} \leq \Pr\{N_r(0) > a_2\} + \Pr\{N_r(1) < a_1\} \to 0 ,$$

so that $Z_r^{(j)} - W_r^{(j)} \overset{p}{\to} 0$. An application of the continuous mapping theorem and Theorem 4.4 of Billingsley completes the proof. □

Proof of Lemma 3. Calculations and (2.11) show that

$$(2.12)\quad r^{-\frac{1}{2}}(\theta_r^* - \theta_2)$$
$$= r^{\frac{1}{2}} \max\{(rM_r^{(1)})^{-1} \sum_{i=1}^{M_r^{(1)}} (X_{1i} - \theta_1) + (\theta_1 - \theta_2),\ (rM_r^{(2)})^{-1} \sum_{i=1}^{M_r^{(2)}} (X_{2i} - \theta_2)\}$$
$$= \max\{(rM_r^{(1)})^{-1} V_r^{(1)}(rM_r^{(1)}) + r^{\frac{1}{2}}(\theta_1 - \theta_2),\ (rM_r^{(2)})^{-1} V_r^{(2)}(rM_r^{(2)})\} + o_p(1)$$
$$= \max\{(\frac{M_r^{(1)}}{N_B})(rN_B)^{-1} V_r^{(1)}(\frac{M_r^{(1)}}{N_B} rN_B) - r^{\frac{1}{2}}\Delta,\ (\frac{M_r^{(2)}}{N_B})(rN_B)^{-1} V_r^{(2)}(\frac{M_r^{(2)}}{N_B} rN_B)\} + o_p(1) .$$

By Lemma 2, the processes $V_r^{(j)}(srN_B)$ are elements of $C[0,1]$ which are tight with weak limits, so that since $M_r^{(j)}/N_B \overset{p}{\to} 1$,

$$V_r^{(j)}\left(\frac{M_r^{(j)}}{N_B} rN_B\right) - V_r^{(j)}(rN_B) \overset{P}{\to} 0 .$$

This means by Lemma 2 that the elements of the last equation in (2.11) are jointly weakly convergent, so that an application of the continuous mapping theorem completes the proof. □

3. Eliminating Populations

The difficulty with the stopping time N_B is that it is only mildly data sensitive in that it estimates $\Delta_1,\ldots,\Delta_{k-1}$ but continues to sample from populations which the data indicate are not associated with the largest population mean, i.e., it fails to eliminate inferior populations. A basic method for correcting this deficiency is to use the technology due to Robbins (1970) and Swanepoel and Geertsema (1976). Suppose then an initial sample of size m is taken. Define $t(\alpha) = m^{-1}(1 + b^2/(m-1))^m$, and let $b = b(\alpha)$ satisfy the equation $1 - F_{m-1}(b) + bf_{m-1}(b) = \alpha/(k-1)$, where $F_{m-1}(f_{m-1})$ is the distribution (density) function of a t-distribution with m-1 degrees of freedom. Define $h(t(\alpha),n) = \{(t(\alpha)n)^{1/n} - 1\}^{\frac{1}{2}}$ and let $s^2(i,j,n) = (n-1)^{-1} \sum_{p=1}^{n} (X_{jp}-X_{ip}-\overline{X}_{jn}+\overline{X}_{in})^2$. We say that the ith population is eliminated at stage M_i if it has not been eliminated at any stage $n < M_i$ and if, when populations $j_1,\ldots,j_p$ also have not been eliminated before stage M_i, we have for some $j \in \{j_1,\ldots,j_p\}$ that

$$\overline{X}_{jM_i} - \overline{X}_{iM_i} > h(t(\alpha), M_i)S(i,j,M_i) . \tag{3.1}$$

Assuming $\Delta_{k-1} > 0$, $\theta_j = \theta_{[k]}$ and an initial sample size m, the previously cited works show that

$$\Pr\{M_j > M_i \text{ for all } i \neq j\} \geq 1 - \alpha. \tag{3.2}$$

In other words, the probability is at most α of eliminating the population with the largest mean. We believe the choice m=5 initial observations will work quite well. The stopping times we consider are then defined formally as follows: choose α (see below) and take an initial sample of size $\max(5, ar^{-1})$ from each population.

Definition. Reorder the populations so that $M_1 \leq M_2 \leq \ldots \leq M_k$, the ordering in case of ties being by sample means. If $N_B(k) \leq M_1$, take $N_B(k)$ observations from each population. Otherwise, completely eliminate the first population from further study and continue as if there were k-1 populations in the experiment (this includes changing the values of H_k to H_{k-1} and σ_{nk} to σ_{nk-1}, but the value of $t(\alpha)$ in (3.1) remains unchanged). Then, if $N_B(k-1) \leq M_2$, take $N_B(k-1)$ observations from each population; otherwise eliminate the second population. Continue in this manner until stopping, denoting the number of observations on each population by $(N_1 \leq N_2 \leq \ldots \leq N_k) = \underline{N}$, with total sample size $T = N_1 + \ldots + N_k$.

Note that Blumenthal's $N_B = N_B(k)$ is obtained as a special case by choosing $\alpha = 0$. We again consider only the case k=2 and define $M = \min(M_1, M_2)$. Recall that $\Delta \sim r$. The next result shows how letting $\alpha \to 0$ as $r \to 0$ influences M. The proof is at the end of the section.

Lemma 4. (size of M) Choose $\alpha \to 0$ as $r \to 0$ so that $b^2 = 2 \log t(\alpha) = r^{2\beta_0 - 1}$ $(0 < \beta_0 < \frac{1}{2})$. Then, as $r \to 0$,

$$
\begin{aligned}
rM &\overset{p}{\to} a_0 \quad \text{if} \quad \beta < \beta_0 \\
&\overset{p}{\to} 1 \quad \text{if} \quad \beta = \beta_0 \\
&\overset{p}{\to} \infty \quad \text{if} \quad \beta > \beta_0 .
\end{aligned}
$$

Lemma 4 is rather confusing at first sight. Note that $\Delta = |\theta_2 - \theta_1| \sim r^\beta$, so the smaller the value of β the farther apart the means are and the quicker one should eliminate. This means that for smaller β, rM should be small, as Lemma 4 shows. The constant β_0 (a monotone increasing function of α) merely serves as a cut off point; for small α and hence small β_0, it becomes harder to eliminate because we are insisting on more protection (see (3.2)).

Lemma 5. (Comparison of sample sizes). For general k, let $T_b = kN_B$ be the total sample size of the Blumenthal procedure. Let $\Delta_i \sim r^{\beta_i}$ $(i = 1,\ldots,k-1)$ and let p be the number of $\beta_i < \beta_0$, i.e., p is the number of populations whose means are far from $\theta_{[k]}$ relative to α. Then, if T_E is the total sample size taken by the elimination procedure,

$$T_E/T_B \xrightarrow{p} 1 - (1-a_0)p/k ,$$

where a_0 is defined immediately following (1.3).

Lemma 5 shows that considerable savings in sample size are possible. In the next section we show that this is accomplished without a corresponding increase in MSE.

Proof of Lemma 4. First consider $\beta < \frac{1}{2}$. Since $rM = (\Delta^2 M/2 \log t(\alpha))(2r \log t(\alpha))/\Delta^2 \sim (\Delta^2 M/2 \log t(\alpha))r^{2\beta_0/\beta}$, it suffices to prove that $\Delta^2 M/\log t(\alpha) \to 2$. Recalling that $\theta_1 < \theta_2$ and defining $T_n = \overline{X}_{2n} - \overline{X}_{1n} - \Delta$, equation (3.2) shows that with probability approaching one, $M = M_2$ so that with probability approaching one,

$$\text{(3.3)} \qquad \begin{aligned} T_M + \Delta &> h(t(\alpha),M)\ S(1,2,M) \\ T_{M-1} + \Delta &\le h(t(\alpha),M)\ S(1,2,M) \ . \end{aligned}$$

Using the facts that $\beta < \frac{1}{2}$ and $M \ge ar^{-1}$, the law of the iterated logarithm shows that $T_M/\Delta \xrightarrow{p} 0$, $T_{M-1}/\Delta \xrightarrow{p} 0$. Dividing through by Δ in (3.3) and noting that $S^2(1,2,n) \to 2$ almost surely, a few manipulations show that

$$\text{(3.4)} \qquad (\log t(\alpha) + \log M)/\Delta^2 M \xrightarrow{p} \tfrac{1}{2} \ .$$

Now, since $M \geq ar^{-1}$, $\Delta^2 M \geq M^\gamma$ for some small positive γ, so that (3.4) becomes $(\log t(\alpha))/\Delta^2 M \overset{p}{\to} \frac{1}{2}$. Next we consider the case $\beta \geq \frac{1}{2}$. In (3.3), divide through by $r^{\frac{1}{2}-\varepsilon}$, where $\varepsilon > 0$ is sufficiently small. Since $\Delta/r^{\frac{1}{2}-\varepsilon} \to 0$, some manipulations yield

$$\{\log t(\alpha) + \log M\}/Mr^{1-2\varepsilon} \overset{p}{\to} 0 .$$

Since $Mr^{1-2\varepsilon} \geq M^\gamma$ for some $\gamma > 0$, this gives

$$Mr^{1-2\varepsilon}/\log t(\alpha) \overset{p}{\to} \infty . \tag{3.5}$$

Since $rM = r^{2\varepsilon} \log t(\alpha)(r^{1-2\varepsilon} M/\log t(\alpha))$, we can choose $\varepsilon > 0$ sufficiently small so that $r^{2\varepsilon} \log t(\alpha) \to \infty$ and hence by (3.5), $rM \overset{p}{\to} \infty$, which completes the proof.

4. Mean Square Error

In this section we consider the MSE $r^{-1} E(\theta_N^* - \theta_2)^2$ both asymptotically and in a Monte-Carlo study for small sample sizes. Suppose that upon stopping, N_i observations have been taken from the ith population ($i = 1,2$). Recall that from previous considerations, we are taking $\theta_1 < \theta_2$, $N_i > ar^{-1}$, and $\sigma=1$. Blumenthal and Cohen (1968) indicate that even for N_B, there are many ways of estimating $\theta_{[2]}$, but that $\theta_n^* = \max(\overline{X}_{1n}, \overline{X}_{2n})$ is a reasonably effective choice. Our stopping time employs an elimination feature, so we must take into account the possibility that an eliminated population has a sample mean (upon stopping) larger than any other sample mean (upon stopping). The estimate we choose then is given by $\,^*_N = \max(\overline{X}_{1N_1}, \overline{X}_{2N_2})$. An alternative estimator is the maximum sample mean over all populations which have not been eliminated, but we have been unable to verify the uniform integrability needed in the proofs to follow. The cases $\Delta \sim r^\beta (\beta < \frac{1}{2})$ and $\Delta \sim r^\beta (\beta \geq \frac{1}{2})$ are different and are treated separately.

Lemma 6. (Asymptotics of θ_N^* when elimination may occur). Consider the conditions of Lemma 4 with $0 < \beta_0 < \frac{1}{2}$, $0 \leq \beta < \frac{1}{2}$. Then the limit distribution of $r^{-\frac{1}{2}}(\theta_N^* - \theta_2)$ is the standard normal and $r^{-1} E(\theta_N^* - \theta_2)^2 \to 1$.

Lemma 6 says that for k=2 if the population means are sufficiently separated, even if elimination does not occur, our general procedure has precisely the same asymptotic behavior (in θ_N^*) as does N_B. The next result shows this to be true when the means are not separated, i.e., $\theta_2 - \theta_1 = \Delta \sim r^{\beta}$, $\beta \geq \frac{1}{2}$. Note in this case that Lemma 4 says that elimination will probability not happen.

Lemma 7. (Asymptotics of θ_N^* when elimination is unlikely to occur). Consider the conditions of Lemma 4 with $\beta \geq \frac{1}{2}$. Let ξ be the limit distribution in the conclusion to Lemma 3. Then $r^{-\frac{1}{2}}(\theta_N^* - \theta_2) \Rightarrow \xi$, $E\xi^2$ exists, and

$$r^{-1} E(\theta_N^* - \theta_2)^2 \to E\xi^2 .$$

The same results hold if N_B is used without elimination.

The results of Lemma 6 and Lemma 7 are rather unusual, in that they say that for k=2 the simple elimination idea employed here can save the user in terms of sample size with no (asymptotic) change in MSE. In order to see how this works with small samples, we ran a Monte-Carlo experiment with 500 simulations. The complete results are reported in Carroll (1978b), but here we consider $\alpha = .01$, $r = .10$, $.01$ and $\Delta = 2.00$, 1.00 and $.20$. An initial sample of size m=5 was chosen as suggested in Section 3. The results are given below, with T_E/T_B being the ratio of sample size needed for the elimination procedure relative to Blumenthal's procedure.

		$\Delta=2.00$	$\Delta=1.00$	$\Delta=.20$
T_E/T_B	$r = .10$	1.00	1.00	1.00
	$r = .01$	.82	.64	1.00
r^{-1} MSE for elimination	$r = .10$	.86	.89	.72
	$r = .01$	.93	.90	.63
r^{-1} MSE for Blumenthal	$r = .10$	.86	.89	.73
	$r = .01$	.93	.77	.63

Apparently, both procedures achieve their goal of controlling MSE. The elimination stopping time can lead to substantial savings in sample size while achieving its bound on MSE. The Blumenthal procedure appears to have slightly lower MSE overall, but this is achieved at the cost of increased sample size.

Proof of Lemma 6. By Lemmas 1 and 4, rN_i converges in probability to a constant (either a_0 or 1 depending on β_0,β). Thus, by Anscombe (1952, Theorem 1) the vector

$$(4.1)\qquad (r^{-\frac{1}{2}}(\overline{X}_{1N_1} - \theta_1),\ r^{-\frac{1}{2}}(\overline{X}_{2N_2} - \theta_2))$$

converges in distribution to a normal random vector. This gives $\Pr\{\overline{X}_{2N_2} \geq \overline{X}_{1N_1}\} \to 1$ since $r^{-\frac{1}{2}}(\theta_2 - \theta_1) \to \infty$. Hence

$$\Pr\{r^{-\frac{1}{2}}(\theta_N^* - \theta) \leq z\} = \Pr\{r^{-\frac{1}{2}}(\overline{X}_{1N_2} - \theta_2) \leq z\} + o(1)\ ,$$

so that $r^{-\frac{1}{2}}(\theta_N^* - \theta)$ has the required limit distribution. By Bickel and Yahav (1968), to complete the proof it suffices to show that for some $r_0 > 0$,

$$(4.2)\qquad \sum_{m=1}^{\infty} \sup_{0<r<r_0} \Pr\{r^{-1}(\theta_N^* - \theta_2)^2 > m\} < \infty\ .$$

Since $N_i > ar^{-1}$ $(i = 1,2)$, our definition of θ_N^* shows that

$$\Pr\{r^{-1}(\theta_N^* - \theta_2)^2 > m\}$$
$$\leq \Pr\{|\overline{X}_{1n} - \theta_1| > (mr)^{\frac{1}{2}} \text{ for some } n \geq ar^{-1}\}$$
$$+ \Pr\{|\overline{X}_{2n} - \theta_2| > (mr)^{\frac{1}{2}} \text{ for some } n \geq ar^{-1}\}$$
$$\leq c_0\, m^{-2} \quad (\text{for some } c_0 > 0) \ ,$$

this last following by the maximal inequality for reverse martingales (Doob (1953)). This verifies (4.2). □

Proof of Lemma 7. Under our conditions, $N_i/N_B \to 1$ $(i = 1,2)$. Thus, by Lemma 3, $r^{-\frac{1}{2}}(\theta_N^* - \theta_2) \Rightarrow \xi$. The rest of the proof now follows from Bickel and Yahav (1968) and (4.2).

5. The Case of More Than Two Populations

The results of the previous sections can be generalized for $k \geq 3$ by basic notational changes. Lemma 5 already discusses the case $\Delta_i \sim r^{\beta_i}$, with $0 \leq \beta_i < \frac{1}{2}$ for $i = 1,\ldots,p$ and $\beta_i \geq \frac{1}{2}$ for $i = p+1,\ldots,k-1$. Lemma 6 does not change if $p = k-1$, so that all populations but one may be eliminated. Lemmas 1-3 and 7 can handle the mixture situation $p < k - 1$ with some simple changes in definition. If one makes the further reasonable assumption that $H_k(x_1,\ldots,x_{k-1}) = H_{k-p}(x_{p+1},\ldots,x_{k-1})$ when $x_1 = \ldots = x_p = \infty$, then, as in Section 4, it is possible to show that N_B and the elimination procedure lead to the same asymptotic MSE.

References

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

Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist. 12* 1656-1670.

Bickel, P.J. and Yahav, J.A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statisti. 39* 442-456.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley & Sons, Inc.

Blumenthal, S. (1973). Sequential estimation of the largest normal mean when the variance is unknown. Department of Operations Research Technical Report No. 164, Cornell University.

Blumenthal, S. (1976). Sequential estimation of the largest normal mean when the variance is known. *Ann. Statist. 4* 1077-1087.

Blumenthal, S. and Cohen, A. (1968). Estimation of the larger of two normal means. *J. Amer. Statist. Assoc. 63* 861-876.

Carroll, R.J. (1978a). On sequential estimation of the largest normal mean when the variance is known. To appear in Sankhyā, Series A.

Carroll, R.J. (1978b). A study of sequential procedures for estimating the larger mean . University of North Carolina at Chapel Hill Mimeo Series # 1189 .

Doob, J.L. (1953). *Stochastic Processes.* Wiley, New York.

Durrett, R.J. and Resnick, S.I. (1976). Weak convergence with random indices. Technical Report No. 83, Department of Statistics, Stanford University.

Gut, A. (1975). Weak convergence and first passage times. *J. Appl. Prob. 12* 324-332.

Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space D[0,∞), *J. Appl. Prob.* *10* 109-121.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* *41* 1397-1409.

Swanepoel, J.W.H. and Geertsema, J.C. (1976). Sequential procedures with elimination for selecting the best of k normal populations. *S. Aft. Statist. J.* *10* 9-36.

Whitt, W. (1970). Weak convergence of probability measures on the function sapce C[0,∞). *Ann. Math. Statist.* *41* 939-944.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Convergence Results for Sequential Estimation of the Largest Mean		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1192
7. AUTHOR(s) Raymond J. Carroll		8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research, Bolling AFB, Washington, D.C.		12. REPORT DATE August, 1978
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

BIOMATHEMATICS TRAINING PROGRAM

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sequential Estimation, Elimination, Largest Normal Mean, Weak Convergence, Random Time Change, Ranking and Selection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider the sequential estimation of the largest mean of k populations when the observations are normally distributed with a common unknown variance and the goal is to control the mean square error (MSE) at a prespecified level; this is a generalization of probalems considered by Blumenthal (1976) and Carroll (1977) By eliminating from the experiment populations which the data indicate are not associated with the largest mean, it is shown that, compared to existing procedures, significant savings in sample size can be obtained. Weak convergence results are obtained for the stopping times and the estimate of the largest mean as consequences of more general results;these are used to compute the asymptotic MSE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)