THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

ON SEQUENTIAL ESTIMATION OF THE LARGEST NORMAL MEAN
WHEN THE VARIANCE IS KNOWN

Raymond J. Carroll

Institute of Statistics Mimeo Series No. 1133
September, 1977

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

ON SEQUENTIAL ESTIMATION OF THE LARGEST NORMAL MEAN WHEN THE VARIANCE IS KNOWN

by

Raymond J. Carroll*

University of North Carolina at Chapel Hill

Abstract

We define a class of stopping times for estimating the largest of k normal means when the variance is known. The class can achieve significant reduction in sample size compared to a related procedure due to Blumenthal (1976) because it employs an elimination feature which halts sampling on populations furnishing no information about the largest mean. The asymptotic behavior of the stopping times and the mean square consistency of the estimators are studied.

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

Key Words and Phrases: Sequential Estimation, Elimination, Largest Normal Mean, Ranking and Selection

*This research was supported by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2796. The author wishes to thank the Department of Theoretical Statistics, University of Minnesota for providing office space during part of this research.

1. Introduction

Let $\theta_1,\ldots,\theta_k$ be the unknown means of k normal populations with common known variance σ^2 (henceforth taken as unity). Let $\bar{X}_{1n},\ldots,\bar{X}_{kn}$ be the sample means of n observations taken from the k populations, and define the ordered population and sample means by $\theta_{[1]}\leq\ldots\leq\theta_{[k]}$ and $\bar{X}_{[1]n}\leq\ldots\leq\bar{X}_{[k]n}$. The problem is to construct a sequential stopping time for the estimation of $\theta_{[k]}$ (by an estimator θ_n^*, often takes as $\bar{X}_{[k]n}$) with prespecified mean seuqre error (MSE) r. The procedures we investigate depend on the estimates $\Delta_{in} = \bar{X}_{[k]n}-\bar{X}_{[i]n}$ of $\Delta_i = \theta_{[k]}-\theta_{[i]}$ $(i=1,\ldots,k)$.

Blumenthal (1976) constructed a purely sequential stopping time N_B and a related two-stage procedure N_B^*, obtaining results which may be summarized as follows. For $\Delta_1,\ldots,\Delta_{k-1}$ fixed, rN_B and rN_B^* have almost sure limits as $r\to0$, but asymptotic mean square consistency is verified only for the two-stage procedure for k=2. If each Δ_i is proportional to $r^{1/2}$ ($\Delta_i\sim r^{1/2}$), neither N_B^* nor N_B^* has an almost sure limit, but the limit distribution is computed only for the two-stage procedure N_B^* when k=2; asymptotic mean square consistency is not checked in this case for the sequential procedure N_B. Blumenthal indicates that for k=2, his procedures will achieve approximately 10% savings in sample size when compared to a conservative, fixed-sample procedure.

In this note we generalize Blumenthal's results in two ways: first, we introduce a class of stopping times N for which N_B is (in a sense to be made precise) a "least favorable" member and second, we answer a number of open questions in his paper by obtaining limit distributions and asymptotic mean square errors. In particular, when the parameters

Δ_i are proportional to r^{β_i} $(0 \le \beta_i < \infty)$, we compute the limit distributions of this class, finding that if $\beta_{k-1} \ge 1/2$, the limit distribution is a stopping time for a function of Brownian motion, and that in general, the mean square error is proportional to r.

Define $H_1 \equiv 1$ and let $n^{-1} H_k(n^{1/2}\Delta_1, \ldots, n^{1/2}\Delta_{k-1})$ be the MSE due to estimating $\theta_{[k]}$ by $\theta_n^* = \bar{X}_{[k]n}$. Define

$$n(\boldsymbol{\Delta}) = n_k(\Delta) = \inf\{n: H_k(n^{1/2}\Delta_1, \ldots, n^{1/2}\Delta_{k-1}) \le nr\} \ .$$

The stopping time $N_B = N_B(k)$ is defined as follows: after obtaining t observations from each population, compute the estimates $\Delta_{1t}, \ldots, \Delta_{k-1,t}$ and compute $\hat{n}(t) = n(\hat{\Delta})$ by using the estimates $\Delta_{1t}, \ldots, \Delta_{k-1,t}$ in the definition of $n(\Delta)$; then

$$N_B = N_B(k) = \inf\{m: \hat{n}(m) \le m\} \ .$$

The difficulty with N_B is that it continues to sample populations which obviously are not associated with the largest population mean, i.e., it fails to eliminate inferior populations. We correct this difficulty by defining a "selection sequence" (Swanepoel and Geertsema (1976)); define b to be the solution to $1-\Phi(b)+b\phi(b)+\phi^2(b)/\Phi(b) = \alpha/k-1$ (here Φ and ϕ are the standard normal distribution function and density), so that as $\alpha \to 0$, $b^2 \sim 2\log((k-1)/\alpha)$, and define

$$M_i = \inf\{n: \bar{X}_{[k]n} - \bar{X}_{in} \ge 2^{1/2}((b^2 + \log n)/n)^{1/2} = 2^{1/2} g(n,b)\} \ .$$

Assuming $\Delta_{k-1} > 0$ and $\theta_j = \theta_{[k]}$, it follows (Robbins (1970)) that

$$\Pr\{M_j > M_i \text{ for all } i \ne j\} \ge 1-\alpha.$$

Thus, our plan will be to continue to sample from population i as long as $N_B \le M_i$; once $N_B > M_i$, we will discontinue sampling from that population. Formally, we make the following definition.

Definition: Reorder the populations so that $M_1 \le M_2 \le \ldots \le M_k$. If $N_B(k) \le M_1$, take $N_B(k)$ observations from each population. Otherwise, completely eliminate the first population from further consideration, and continue as if there were k-1 populations in the experiment (although b^2 is unchanged). Then, if $N_B(k-1) \le M_2$, take $N_B(k-1)$ observations from each population; otherwise, eliminate population two. Continue in this manner until stopping, denoting the number of observations on each population by $(N_1 \le N_2 \le \ldots \le N_k) = \underline{N}$. The total sample size is $T = N_1 + \ldots + N_k$.

Note that by choosing $b^2 = \infty$ $(\alpha=0)$, we obtain $N_1 = \ldots = N_k = N_B(k)$, so that Blumenthal's stopping time is a special case of ours. The benefits of this class of stopping times are discussed in the next section.

For notational convenience, we limit ourselves to the special case k = 2, but the proofs are structured so as to extend immediately. In order to indicate the precise nature of the extension, we make no use of the following facts which hold only for k = 2:

$$\max(\overline{X}_{1n}, \overline{X}_{2n}) = (\overline{X}_{1n} + \overline{X}_{2n})/2 + |\overline{X}_{1n} - \overline{X}_{2n}|/2 ,$$

$$N_B = N_B(2) = \inf\{n: H_2(n^{\frac{1}{2}}\Delta_1) \le nr\} .$$

In what follows, we define $\Delta_n = \overline{X}_{[2]n} - \overline{X}_{[1]n}$, $\Delta = \theta_{[2]} - \theta_{[1]}$, and $M = \min(M_1, M_2)$.

2. Asymptotic Distributions

For k = 2, the limit distributions of $\underline{N}$ and $\underline{T}$ are basic functions of N_B and M. We assume throughout this section that $\Delta \sim r^{\beta}$ for some $\beta \ge 0$ and $\Delta r^{\frac{1}{2}} \to c_0$ $(0 \le c_0 \le \infty)$. Let W be Brownian motion with mean zero and variance 2t at time t, and define

$$(1) \quad W^*(s,t,c_0) = (s^{\frac{1}{2}}/t)\,|W(t)+c_0| = (s^{\frac{1}{2}}t)\{\max(W(t)+c_0,0) - \min(W(t)+c_0,0)\}$$

if $c_0 < \infty$, while $W^*(s,t,\infty) = \infty$. Let $[\cdot]$ denote the greatest integer function. Consider $0 < a < b < \infty$ and define $G_r(s,t) = [s/r]^{\frac{1}{2}}\Delta_{[t/r]}$, which is a stochastic process on the multidimensional time parameter space $D_2[a,b]$ (Bickel and Wichura (1971), Billingsley (1968)). Assuming that $\theta_1<\theta_2$, we see that

(2) $$\Delta_n = \max((\overline{X}_{2n}-\overline{X}_{1n}-\Delta)+\Delta,0) - \min((\overline{X}_{2n}-\overline{X}_{1n}-\Delta)+\Delta,0)$$

and, since $\overline{X}_{2n} - \overline{X}_{1n}-\Delta$ is an average of mean zero normal random variables, (2) tells us that on $D_2[a,b]$,

(3) $$G_r \Rightarrow W^* (\cdot,\cdot,c_0) \qquad (\beta \geq 1/2),$$
$$G_r \overset{p}{\to} \infty \qquad \beta < 1/2) ,$$

where "$\Rightarrow$" denotes weak convergence. Thus we obtain

Lemma 1 Let $H_{min}=\min\{H_2(x)\}$. For $\beta<1/2$, $rN_B \overset{p}{\to} 1$. For $\beta\geq 1/2$ and $u \geq H_{min}$,

$$\Pr\{rN_B>u\} \to \Pr\{\min_{H_{min}\leq s\leq t\leq u} (H_2(W^*(s,t,c_0))-s) > 0\} = G^*(u) .$$

Further, G^* is a proper distribution function.

Proof of Lemma 1 By definition,

(4) $$\Pr\{rN_B>u\} = \Pr\{\hat{n}(m)>m \text{ for all } H_{min}\leq rm\leq u\}$$
$$= \Pr\{\min_{H_{min}\leq rk\leq rm\leq u}(H_2(k^{1/2}\Delta_m)-rk)>0\} .$$

For $\beta<{}^1/_2$, $\min\{k^{1/2}\Delta_m:H_{min}\leq rk\leq rm\leq u\} \overset{p}{\to} \infty$, so that $rN_B \overset{p}{\to} 1$ since $H(x)\to 1$ as $x \to \infty$. For $\beta\geq {}^1/_2$, (3) and (4) show that

$$\Pr\{rN_B>u\}=\Pr\{\min_{H_{min}\leq s\leq t\leq u} \left(H_2(G_r(s,t)\right)-s\big)> 0\}$$
$$\to G^*(u).$$

The following result (Carroll (1976) delineates the behavior of M for a particular choice of b^2.

Lemma 2 If $b^{-1}\log\Delta \to 0$, then $\Delta^2M/b^2 \overset{p}{\to} 1$. Thus, if $r^{1-2\beta_0}b^2 \to 1$ for some $0 < \beta_0 < 1/2$, then

$$M/N_B \overset{p}{\to} 0 \quad (\text{if } \beta_0 > \beta)$$

$$\overset{p}{\to} 1 \quad (\text{if } \beta_0 = \beta)$$

$$\overset{p}{\to} \infty \quad (\text{if } \beta_0 < \beta).$$

Now, letting T be the total sample size using $b^2 = r^{2\beta_0 - 1}$ and $T_B = 2N_B$ the total sample size of the Blumenthal procedure, we see that

<u>Lemma 3</u> $T/T_B \overset{p}{\to} 1/2$ if $\beta<\beta_0$, while $T/T_B \overset{p}{\to} 1$ if $\beta\geq\beta_0$.

<u>Remark</u> Lemma 3 is easily extended to the case of general k as follows. Let $T_B = kN_B$ be the total sample size of the Blumenthal procedure, set $b^2 = r^{2\beta_0 - 1}$, let $\Delta_i \sim r^{\beta_i}$ (i=1,...,k-1). Let p be the number of $\beta_i<\beta_0$, i.e., p is the number of populations furnishing little information about $\theta_{[k]}$. Then

$$T/T_B \overset{p}{\to} 1-p/k.$$

In other words, the elimination can result in a significant savings in sample size.

3. <u>Asymptotic MSE</u>

Our goal is to find an estimate θ_N^* of $\theta_{[2]}$ for which the following mean square consistency result holds:

(5) $r^{-1}E(\theta_N^* - \theta_{[2]})^2 \to 1$ as $r \to 0$.

In the proof given below, we are forced to make the convention that for a small constant a>0, at least ar^{-1} observations are taken from each population. Suppose that upon stopping, N_i observations have been taken on the i^{th} population (i=1,2). Our estimate of $\theta_{[2]}$ is taken to be $\theta_N^* = \max(\bar{X}_{1N_1}, \bar{X}_{2N_2})$. This estimate does allow the possibility of estimating $\theta_{[2]}$ by the mean of an eliminated population, but the nature of the elimination shows this possibility to be quite unlikely.

Lemma 4 Let $b^2 = r^{2\beta_0 - 1}$ and $\Delta \sim r^\beta$, with $0 \le \beta$, $\beta_0 < 1/2$. Then (5) holds.

Proof of Lemma 4 By Bickel and Yahav (1968), it suffices to show that $r^{-1}(\theta_N^* - \theta_{[2]})^2$ has a limit distribution and that for some $r_0 > 0$,

$$(6) \quad \sum_{m=1}^{\infty} \sup_{0<r<r_0} \Pr\{r^{-1}(\theta_N^* - \theta_{[2]})^2 > m\} < \infty.$$

Now, set $\theta_1 < \theta_2$ without loss of generality. Then the event $\{|\theta_N^* - \theta_2| > (mr)^{1/2}\}$ is contained in the union of the events

$$\{|\bar{X}_{1N_1} - \theta_1| > (mr)^{1/2}\}, \quad \{|\bar{X}_{2N_2} - \theta_2| > (mr)^{1/2}\}.$$

Now, by the maximal inequality of reverse martingales (Doob (1953), pp. 317-318) and the fact that $\min(N_1, N_2) \ge ar^{-1}$,

$$\Pr\{|\bar{X}_{2N_2} - \theta_2| > (mr)^{1/2}\}$$
$$\le \Pr\{\sup_{k \ge ar^{-1}} |\bar{X}_{2k} - \theta_2| > (mr)^{1/2}\} \le c_0 m^{-2}$$

for some $c_0 > 0$. This verifies (6). To show that $r^{-1}(\theta_N^* - \theta_{[2]})^2$ has a limit distribution, first note that $r^{-1}N_i$ converges in probability to a constant $(i=1,2)$, so by an extension of Anscombe's (1952) Theorem 1, the vector

$$(r^{-\frac{1}{2}}(\bar{X}_{1N_1} - \theta_1), \quad r^{-\frac{1}{2}}(\bar{X}_{2N_2} - \theta_2))$$

converges in law to a jointly normal random vector. Since $r^{-1/2}(\theta_2 - \theta_1) \to \infty$, we complete the proof by noting that $\Pr\{\bar{X}_{2N_2} \ge \bar{X}_{1N_1}\} \to 1$ and

$$\Pr\{r^{-1/2}(\theta_N^* - \theta_2) \le z\}$$
$$= \Pr\{r^{-1/2}(\bar{X}_{1N_1} - \theta_1) \le z + r^{1/2}(\theta_2 - \theta_1) \text{ and } \bar{X}_{1N_1} > \bar{X}_{2N_2}\}$$
$$+ \Pr\{r^{-1/2}(\bar{X}_{1N_2} - \theta_2) \le z \text{ and } \bar{X}_{1N_1} \le \bar{X}_{2N_2}\}. \quad \square$$

An alternate definition of θ_N^* takes the maximum of the sample means only if elimination has not occured; we have not been able to verify (6) in this case. Note that by choosing $b^2 = \infty$, the proof of Lemma 4 shows that the Blumenthal procedure satisfies (5).

The situation for $\Delta \sim r^{1/2}$ is considerably more complicated. Define $m = [t/r]$, assume that $\theta_1 < \theta_2$, and let

$$(7) \quad V_r(t) = m^{1/2}(\max(\bar{X}_{1m}, \bar{X}_{2m}) - \theta_2)$$
$$= m^{1/2}(\max(\bar{X}_{1m} - \theta_1 - (\theta_2 - \theta_1) \ , \ \bar{X}_{2m} - \theta_2)).$$

This is a stochastic process on $D[1/2,2]$ which, for $\Delta \sim r^{\beta}$ $(\beta \geq 1/2)$, converges weakly to a process V in $C[1/2,2]$. We also know that rN_B converges in law to a random variable (call it W_1) with distribution function as in Lemma 1. Define a process

$$V^*(t) = W_1^{-1/2} V(tW_1).$$

Lemma 5 Let $b^2 = r^{2\beta_0 - 1}$ and $\Delta \sim r^{\beta}$ with $0 < \beta_0 < \frac{1}{2} \leq \beta < \infty$. Then $E(V^*(1))^2$ exists and

$$r^{-1} E(\theta_N^* - \theta_{[2]})^2 \to E(V^*(1))^2.$$

Proof of Lemma 5 By Lemma 4, since (6) holds, it suffices to show that $r^{-1/2}(\theta_N^* - \theta)$ converges in distribution to $V^*(1)$. By Lemma 3, $N_i/N_B \overset{p}{\to} 1$, so we may take $\theta_N^* = \max(\bar{X}_{1N_B}, \bar{X}_{2N_B})$. Note that $1/2 \leq rN_B \leq 2$; we first show that on $D_2[1/2,2] \times D_2[1/2,2] \times R^1$,

$$(8) \quad (V_r^{(1)}, V_r^{(2)}, rN_B) \Rightarrow (V^{(1)}, V^{(2)}, W_1).$$

Here, for $m_1 = [s/r]$, $m_2 = [t/r]$,

$$V_r^{(j)}(s,t) = m_1^{1/2} \ (\bar{X}_{jm_2} - \theta_j) \qquad (j=1,2).$$

and V_j is the limit distribution of $V_r^{(j)}$. Because $V_r^{(1)}$ and $V_r^{(2)}$ are tight, (8) will follow if we can prove the convergence of the finite dimensional distributions. To do this, define two processes:

$$V_r^{(3)}(s,t) = H_2(|V_r^{(2)}(s,t) - V_r^{(1)}(s,t) + [s/r]^{\frac{1}{2}}(\theta_2-\theta_1)|) - [s/r]r \ ,$$

$$Z_r(u) = Z_r(u,H_{min}) = \inf\{V_r^{(3)}(s,t) : H_{min} \le s \le t \le u\} \ .$$

If $u<H_{min}$, define $Z_r(u) = Z_r(H_{min})$. By the continuous mapping theorem, since H_2 is continuous and both $V_r^{(1)}$ and $V_r^{(2)}$ have weak limits in $C_2[1/2,2]$, it follows that $V_r^{(3)}$ and Z_r have weak limits (call them $V^{(3)}, Z$) in $C_2[1/2,2]$ and $C[1/2,2]$ and on $D_2[1/2,2] \times D_2[1/2,2] \times D[1/2,2]$,

$$(9) \qquad (V_r^{(1)}, V_r^{(2)}, Z_r) \Rightarrow (V^{(1)}, V^{(2)}, Z) \ .$$

To check the convergence of the finite dimensional distributions in (8), we consider only a special case; note that

$$(10) \qquad \begin{aligned} &\Pr\{V_r^{(1)}(s,t) \le u_1,\ V_r^{(2)}(s,t) \le u_2,\ rN_B \le u_3\} \\ &= \Pr\{V_r^{(1)}(s,t) \le u_1,\ V_r^{(2)}(s,t) \le u_2\} \\ &- \Pr\{V_r^{(1)}(s,t) \le u_1,\ V_r^{(2)}(s,t) < u_2,\ Z_r(u_3) > 0\} \ . \end{aligned}$$

Since the first term on the right hand side of (10) has a limit, equation (9) and Theorem 2.1 of Billingsley (1968) prove the weak convergence of the finite dimensional distributions. We next apply a modification of Theorem 17.2 of Billingsley, replacing his equation (17.18) by our (8) and remembering in the proof that $\frac{1}{2} \le rN_B,\ W_1 \le 2$ with probability one. To see this, define $\Phi_r(t) = trN_B$, $\Phi_0(t) = tW_1$, and note that (8) implies that on $D[1/2,2] \times D[1/2,2] \times R$,

$$(V_r, \Phi_r, rN_B) \Rightarrow (V, \Phi_0, W_1) \ .$$

Then by the continuous mapping theorem,

$$(rN_B)^{-\frac{1}{2}}(V_r o \Phi_r) \to W_1^{-\frac{1}{2}}(V o \Phi_0) ,$$

where "o" denotes composition. Since

$$(rN_B)^{-\frac{1}{2}}(V_r o \Phi_r)(1) = r^{-\frac{1}{2}}(\theta^*_{N_B} - \theta_{[2]}) ,$$

the proof is complete. □

4. Conclusions

In this note we have defined a class of stopping times for estimating the largest of k means. This class includes a procedure due to Blumenthal but, by building in an elimination feature, allows the possibility for significant savings in sample size. We have obtained the asymptotic behavior of the stopping times, showing that they are related to stopping times for a function of Brownian motion. Finally, we define an estimator with asymptotic MSE proportional to r.

References

Anscombe, F. J. (1952). Large sample theory of sequential estimation. Proc. Camb. Phil. Soc. 48, 600-607.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. 42, 1656-1670.

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

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

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

Carroll, R. J. (1976). On sequential elimination procedures. Institute of Statistics Mimeo Series No. 1078, University of North Carolina at Chapel Hill.

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

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. (1967). Sequential procedures with elimination for selecting the best of k normal populations. S. Afr. Statist. J., 10, 9-36.