A STUDY OF SEQUENTIAL PROCEDURES FOR ESTIMATING THE LARGEST OF THREE NORMAL MEANS

by

Raymond J. Carroll*
and
Robert Smith

University of North Carolina at Chapel Hill

Abstract

We study the problem of estimating the largest mean from three populations when data are normally distributed. A Monte-Carlo study is performed to compare two sequential procedures, one of which eliminates populations during the experiment, while the other does not. The elimination procedure is shown to be preferable.

Key Words and Phrases: Monte-Carlo, Sequential Analysis, Ranking and Selection, Estimating the Larger Mean, Elimination, Mean Square Error

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

1. Introduction

Blumenthal (1976) considered the problem of sequential estimation of the largest of k normal means when a bound is set on the acceptable mean square error. He showed that his procedure results in only a small savings in sample size when compared to a conservative fixed sample procedure for the case of known variance. Carroll (1978a) noted that this procedure does not give the user the flexibility of sampling selectively from the k populations. Carroll (1978a) defined a procedure which early in the experiment eliminates from further consideration those populations which are obviously not associated with the largest mean and hence provide little relevant information; his theoretical large-sample calculations indicate possible large savings in sample size with no corresponding increase in mean square error. Carroll (1978b) verified the theory for small sample size by a Monte-Carlo experiment, but only for $k = 2$. In this paper, the simulation study of Carroll (1978b) is extended to the case $k = 3$.

2. The Procedures

We are dealing with independent identically distributed observations $X_{i1}, X_{i2}, \ldots$ from the i-th population, $i = 1,2,3$. These are assumed to be normally distributed with means μ_1, μ_2 and μ_3 and common variance σ^2. The goal is to estimate the largest mean μ_* with a prespecified bound on the mean square error (MSE) r. The asymptotic theorems in Blumenthal (1976) and Carroll (1977) take place as $r \to 0$. If $\mu_{[1]} \le \mu_{[2]} \le \mu_{[3]}$ is the true (unknown) ordering, the mean square error for estimating μ_* by the largest sample mean based on n observation can be written as

$$\text{MSE} = (\sigma^2/n)H_3\left(\frac{n^{\frac{1}{2}}\mu_{[1]}}{\sigma}, \frac{n^{\frac{1}{2}}\mu_{[2]}}{\sigma}, \frac{n^{\frac{1}{2}}\mu_{[3]}}{\sigma}\right) .$$

When k = 2, a similar formula can be written as

$$\text{MSE} = (\sigma^2/n)H_2\left(\frac{n^{\frac{1}{2}}\mu_{[1]}}{\sigma}, \frac{n^{\frac{1}{2}}\mu_{[2]}}{\sigma}\right) .$$

In order to control the MSE at a prespecified level r when σ is known, Blumenthal (1976) defined the following stopping times.

<u>Definition 2.1</u>. Suppose there are k populations. After obtaining m observations from each population, define the sample means by $(\overline{X}_{1m}, \overline{X}_{2m} \ldots \overline{X}_{km})$ and the ordered sample means by $\overline{X}_{[1]m} \leq \overline{X}_{[2]m} \ldots \leq \overline{X}_{[k]m}$. Define

$$\hat{n}(m) = \inf\left\{n \geq n_0 \colon H\left(\frac{n^{\frac{1}{2}}\overline{X}_{[1]m}}{\sigma}, \ldots \frac{n^{\frac{1}{2}}\overline{X}_{[k]m}}{\sigma}\right) \leq nr/\sigma^2\right\} .$$

Then the stopping time is

$$N_B(k) = \inf\{m \geq m_0 \colon \hat{n}(m) \leq m\} .$$

Carroll (1978) has shown that Blumenthal's procedure N_B is inefficient in that it does not make use of all the information available in the data. In particular, it does not recognize cases when one population is obviously associated with the smaller mean. Carroll (1978) defined a procedure which attempts to recognize this situation and stop sampling (early in the experiment) for populations which provide information about μ_*. The idea is based on a technique of Swanepoel and Geertsema (1976) and can be described fully as follows. We take $\sigma^2 = 1$ throughout.

Step #1. Choose a small value α, which is the probability of falsely eliminating the population associated with the larger mean. Letting $\Phi(\phi)$ be the standard normal distribution (density) function, define $b = b(\alpha)$ by

$$1 - \Phi(b) + b\phi(b) + \phi^2(b)/\Phi(b) = \alpha/(k-1) \quad (= \alpha/2 \text{ when } k = 3) \; .$$

Note that $\alpha = 0$ implies $b = \infty$.

Step #2. Take n_0 observations on each population. Initialize $k_* = k$, $m = n_0$.

Step #3. If $N_B(k_*) = m$, stop and go to Step #7.

Step #4. Otherwise, go to Step #6 unless

$$\text{(1)} \qquad \overline{X}_{[k_*]m} - \overline{X}_{[1]m} \geq 2^{\frac{1}{2}}((b^2+\log n)/n)^{\frac{1}{2}} \; .$$

If (1) occurs, go to Step #5.

Step #5. Eliminate from all further consideration the population associated with the smallest sample mean. Set $k_* = k_* - 1$ and take the sample means of the remaining populations as $\overline{X}_{1m},\ldots\overline{X}_{k_*m}$. Go to Step #6.

Step #6. Take another m_0 observations from each of the remaining populations, set $m = m + m_0$, and return to Step #3.

Step #7. Estimate μ_* by the largest sample mean of the populations.

Note that when $\alpha = 0$, $b = \infty$ and (1) cannot happen, so $\alpha = 0$ is Blumenthal's original rule.

The small sample performance of our stopping times was investigated in a Monte-Carlo study. We took $k = 3$, $n_0 = 5$, $m_0 = 3$ and studied the following configurations of means:

$$\mu_1 = \mu_2 = \mu_3 = 0 \qquad \text{(Table 1)}$$

$$\mu_1 = \mu_2 = 0,\ \mu_3 = \tfrac{1}{4} \qquad \text{(Table 2)}$$

$$\mu_1 = \mu_2 = 0,\ \mu_3 = 1 \qquad \text{(Table 3)}$$

$$\mu_1 = 0,\ \mu_2 = 1,\ \mu_3 = 2 \qquad \text{(Table 4)}$$

$$\mu_1 = 0,\ \mu_2 = 2,\ \mu_3 = 6 \qquad \text{(Table 5)}\ .$$

For most cases we took $\alpha = .00$ (Blumenthal's procedure), .01 and .05; while $r = .10, .05, .02$.

The tables are based on 200 observations. Each table lists the following information:

(i) Bias of the estimate

(ii) MSE/r (we want MSE/r ≤ 1 as in the introduction)

(iii) rN_1, rN_2, rN_3, where N_i is the average number of observations taken on the population with mean μ_i $(i = 1,2,3)$.

The conclusion one can make from the information in Tables 1-5 is obvious; using elimination results in smaller (sometimes much smaller) sample sizes with no real increase in bias or mean square error.

References

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

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

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

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

Table 1

$\mu_1 = \mu_2 = \mu_3 = 0$

α \ r		.10	.05	.02
.05	Bias	.32	.21	
	MSE/r	1.60	1.45	
	rN_1	.97	.91	
	rN_2	.97	.91	
	rN_3	.97	.91	
.01	Bias	.32	.21	
	MSE/r	1.60	1.49	
	rN_1	.97	.91	
	rN_2	.97	.91	
	rN_3	.97	.91	
Blumenthal's Procedure (α = .00)	Bias	.32	.21	
	MSE/r	1.60	1.49	
	rN_1	.97	.91	
	rN_2	.97	.91	
	rN_3	.97	.91	

Table 2

$\mu_1 = \mu_2 = 0, \ \mu_3 = \frac{1}{4}$

α \ r		.10	.05	.02
.05	Bias	.15	.07	.02
	MSE/r	.80	.68	.80
	rN_1	.98	.90	.85
	rN_2	.98	.90	.85
	rN_3	.98	.92	.89
.01	Bias	.15	.07	.02
	MSE/r	.80	.68	.81
	rN_1	.98	.91	.87
	rN_2	.98	.91	.87
	rN_3	.99	.91	.88
Blumenthal's Procedure (α = .00)	Bias	.15	.07	.02
	MSE/r	.80	.68	.81
	rN_1	.99	.91	.88
	rN_2	.99	.91	.88
	rN_3	.99	.91	.88

Table 3

$\mu_1 = \mu_2 = 0, \mu_3 = 1$

α \ r		.10	.05	.02
.05	Bias	.00	-.01	.00
	MSE/r	.86	.85	1.01
	rN_1	.94	.76	.41
	rN_2	.97	.77	.41
	rN_3	1.04	1.01	1.02
.01	Bias	.00	-.01	.00
	MSE/r	.86	.85	1.01
	rN_1	1.01	.87	.56
	rN_2	1.01	.89	.55
	rN_3	1.04	1.00	1.02
Blumenthal's Procedure (α = .00)	Bias	.00	-.01	.00
	MSE/r	.86	.86	1.02
	rN_1	1.04	.99	1.00
	rN_2	1.04	.99	1.00
	rN_3	1.04	.99	1.00

Table 4

$\mu_1 = 0,\ \mu_2 = 1,\ \mu_3 = 2$

α \ r		.10	.05	.02
.05	Bias	.01	.00	.00
	MSE/r	.82	.83	1.00
	rN_1	.62	.32	.13
	rN_2	1.01	.78	.41
	rN_3	1.09	1.02	1.02
.01	Bias	.01	-.01	.00
	MSE/r	.83	.83	1.01
	rN_1	.72	.40	.16
	rN_2	1.06	.90	.55
	rN_3	1.09	1.01	1.01
Blumenthal's Procedure (α = .00)	Bias	.01	-.01	.00
	MSE/r	.83	.84	1.02
	rN_1	1.09	1.00	1.00
	rN_2	1.09	1.00	1.00
	rN_3	1.09	1.00	1.00

Table 5

$\mu_1 = 0,\ \mu_2 = 2,\ \mu_3 = 6$

α \ r		.10	.05	.02
.05	Bias	.01	.00	.00
	MSE/r	.85	.80	1.00
	rN_1	.50	.25	.10
	rN_2	.50	.25	.10
	rN_3	1.10	1.05	1.02
.01	Bias	.01	.00	.00
	MSE/r	.85	.80	1.00
	rN_1	.50	.25	.10
	rN_2	.50	.25	.10
	rN_3	1.10	1.05	1.02
Blumenthal's Procedure (α = .00)	Bias	.01	.00	.00
	MSE/r	.85	.83	1.02
	rN_1	1.10	1.00	1.00
	rN_2	1.10	1.00	1.00
	rN_3	1.10	1.00	1.00

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) A Study of Sequential Procedures for Estimating the Largest of Three Normal Means	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1254
7. AUTHOR(s) Raymond J. Carroll and Robert Smith	8. CONTRACT OR GRANT NUMBER(s) Contract AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, DC 20332	12. REPORT DATE September 1979
	13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

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)

Monte-Carlo, sequential analysis, ranking and selection, estimating the larger mean, elimination, mean square error

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

We study the problem of estimating the largest mean from three populations when data are normally distributed. A Monte-Carlo study is performed to compare two sequential procedures, one of which eliminates populations during the experiment, while the other does not. The elimination procedure is shown to be preferable.