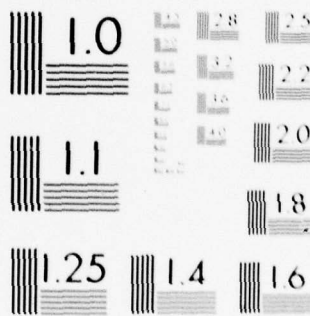


AD-A072 599 NORTH CAROLINA UNIV AT CHAPEL HILL INST OF STATISTICS F/G 12/1
ON SEQUENTIAL CONFIDENCE INTERVALS FOR THE LARGEST NORMAL MEAN.--ETC(U)
AUG 78 R J CARROLL AFOSR-75-2796
UNCLASSIFIED Mimeo SER-1191 AFOSR-TR-78-1437 NL

1 OF 1
AD
A072599





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 78-1437

LEVEL II

B.S. 2

AD A 072599



DDC FILE COPY

On Sequential Confidence Intervals for the Largest Normal Mean

Raymond J. Carroll

Institute of Statistics Mimeo Series #1191

August 1978

DDC
RECEIVED
AUG 13 1979
D

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

Approved for public release;
distribution unlimited.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
18/ AFOSR TR-78-1437			
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED	
ON SEQUENTIAL CONFIDENCE INTERVALS FOR THE LARGEST NORMAL MEAN.		9/ Interim rept.	
7. AUTHOR(s)		6. PERFORMING ORG. REPORT NUMBER	
10/ Raymond J. Carroll			
9. PERFORMING ORGANIZATION NAME AND ADDRESS		8. CONTRACT OR GRANT NUMBER(s)	
University of North Carolina Department of Statistics Chapel Hill, North Carolina 27514		15/ AFOSR-75-2796	
11. CONTROLLING OFFICE NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		16/ 61102F 2304 A5 17/ A5	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE	
12/ 19p.		11/ August 1978	
		13. NUMBER OF PAGES	
		17	
		15. SECURITY CLASS. (of this report)	
		UNCLASSIFIED	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)			
Approved for public release; distribution unlimited.			
14/ MIMED SER-1191			
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, Largest Mean, Ranking and Selection, Sequential Analysis, Elimination			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)			
We construct a competitor to Tong's method for defining a fixed-width confidence interval for the largest normal mean. This competitor eliminates inferior populations early in the experiment; a Monte-Carlo experiment shows that it can use significantly fewer observations than Tong's method without any real loss in observed coverage probability.			

Accession For	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/ _____	
Availability Codes	
Dist.	Availand/or special
A	

LEVEL ~~II~~

2

On Sequential Confidence Intervals for the Largest
Normal Mean

by

Raymond J. Carroll*
University of North Carolina at Chapel Hill

Abstract

We construct a competitor to Tong's method for defining a fixed-width confidence interval for the largest normal mean. This competitor eliminates inferior populations early in the experiment; a Monte-Carlo experiment shows that it can use significantly fewer observations than Tong's method without any real loss in observed coverage probability.

Key Words and Phrases: Monte-Carlo, Largest Mean, Ranking and Selection, Sequential Analysis, Elimination

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

DDC
RECEIVED
AUG 13 1979
RECEIVED
D

1. Introduction

Tong (1975) considered the problem of constructing a fixed-width confidence interval for the largest mean from k normal populations with unknown variance. His procedure is a sequential procedure and is based on the ideas of Chow and Robbins (1965). However, the very nature of this k population problem is qualitatively different from the simpler problem constructing a confidence interval for a mean because the user has the flexibility (not found in Tong's procedure) of sampling selectively from the populations. One method of such selective sampling is *elimination*, where a population is eliminated from further consideration when the data indicate said population is unlikely to be associated with the largest mean.

The purpose of this note is to show that, using the ideas of Swanepoel and Geertsema (1976), it is easy to construct a sequential competitor to Tong's procedure which possesses the elimination option, achieves its intended coverage probability, and can lead to great savings in sample size when the population means are not identical. When the population means are nearly identical, the procedure will take approximately 10% more observations than Tong's procedure, a small price to pay for possible large savings. To this end, in Section 2 we present a Monte-Carlo study of Tong's procedure. In Section 3, we introduce the elimination procedure and study its small sample behavior in a Monte-Carlo study.

2. Tong's Procedure

Suppose we have k populations and take independent and identically distributed observations $X_{i1}, X_{i2}, \dots$ from population i . Define

$$s_{kn}^2 = (k(n-1))^{-1} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i^{(n)})^2, \text{ (pooled sample variance)}$$

$$\bar{x}_i^{(n)} = \frac{1}{n} \sum_{j=1}^n x_{ij}.$$

We will assume that the observations from the i th population are normally distributed with mean μ_i and common variance σ^2 . If $\mu_{[1]} \leq \dots \leq \mu_{[k]}$ denote the unknown ordered means, the goal is to estimate the largest mean $\mu_{[k]}$ by a confidence interval with coverage probability γ and prescribed fixed length L . Tong's (1973) sequential procedure is of the following form. Let Φ be the standard normal distribution function and define

$$\alpha_k(c, x) = \min\{\Phi^k(c-x) - \Phi^k(-x), \Phi(c-x) - \Phi(-x)\}$$

$$\sup_x \alpha_k(c, x) = \alpha_k(c, x_0(c))$$

$$c_0(\gamma) = \inf\{c: \alpha_k(c, x_0(c)) \geq \gamma\}.$$

He then takes N_T observations from each population, where

$$N_T(L) = \text{smallest integer } n \geq 5 \text{ such that } n \geq (c_0(\gamma) s_{kn}/L)^2$$

and announces the following confidence interval of length L :

$$I(N_T(L)) = (\bar{X}_{N_T} - (L - x_0 s_{N_T}/N_T^{1/2}), \bar{X}_{N_T} + x_0 s_{N_T}/N_T^{1/2}),$$

where $x_0 = x_0(c_0(\gamma))$. Tong shows that

$$\lim_{L \rightarrow 0} \frac{N_T(L)}{(c_0(\gamma) \sigma/L)^2} \rightarrow 1 \quad (\text{almost surely})$$

$$\lim_{L \rightarrow 0} P\{\mu_{[k]} \in I(N_T(L))\} \geq \gamma \quad \text{for all } \mu_1, \dots, \mu_k \text{ and } \sigma.$$

In Table 1 we present values of $c_0 = c_0(\gamma)$ and $x_0 = x_0(c_0(\gamma))$ for various values of γ and k .

As part of this study we decided to investigate the small sample behavior of Tong's rule by means of a Monte-Carlo study. We studied the following configurations of means

$$\mu_1 = \mu_2 = \dots = \mu_{k-1} = \mu_k - d$$

$$\mu_{i+1} - \mu_i = d \quad (i = 1, \dots, k-1)$$

with $d = 0.0, 0.5, 2.0$. We chose $\gamma = .90$ and $L = 0.50, 1.00$, with $k = 2, 3, 5$ and 10. The results are reported in Tables 2 and 3. It appears that Tong's procedure does indeed approximately achieve its prescribed coverage probability. Note however that the average total number of observations is independent of d , the spacing of the means. It is this undesirable feature of Tong's procedure which we will attempt to improve upon in the next section.

3. Elimination Procedures

The previous section makes clear that while Tong's procedure achieves its coverage probability in small samples, it is "data-blind" in the sense that it takes no account of information available from the data about the relative differences among the means. In order to begin to design a procedure which will take the data more fully into account we investigate a procedure which attempts to eliminate early in the experiment populations which are obviously not associated with the largest mean.

The idea is based upon a technique due to Swanepoel and Geertsema (1976). Essentially, if we desire a coverage probability γ and set $(1-\gamma) = (1-\gamma_0) + (1-\beta)$, we will use Swanepoel and Geertsema's technique (with a minimal sample size of 5), to eliminate populations with error probability at most $(1-\beta)$, and we will use Tong's procedure with coverage probability γ_0 and the remaining populations.

When $\gamma = .90$, we will choose $\gamma_0 = .92$ and $\beta = .98$. This simple grafting of two techniques will result in a procedure which is slightly conservative when the means are all nearly the same but is very efficient when some observations should be eliminated. We outline the steps in the grafting as follows:

Step #1. For any β (typically $.90 \leq \beta \leq .99$) and k , choose values of (a, t) , where

$$t = .2(1 + a^2/4)^5$$

$$1 - F_4(a) + a f_4(a) = (1-\beta)/(k-1) ,$$

and $F_4(f_4)$ is the distribution (density) function of a t distribution with four degrees of freedom. The values of (a, t) are given in Table 4.

Step #2. Define

$$H^2(i, j, n) = \frac{1}{n-1} \sum_{p=1}^n (x_{ip} - x_{jp} - \bar{x}_i^{(n)} + \bar{x}_j^{(n)})^2$$

$$h(\beta, n) = [(tn)^{1/n} - 1]^{1/2} .$$

We say that population i is *eliminated* at stage $n \geq 5$ if it has not been eliminated before stage n and if

$$\bar{x}_j^{(n)} - \bar{x}_i^{(n)} > h(\beta, n) H(i, j, n)$$

for some population j which has yet to be eliminated at stage n .

Step #3. Choose γ as the intended coverage probability. Let

$(1-\beta) + (1-\gamma_0) = 1-\gamma$. Take five observations from each population. Use the Tong procedure with γ_0 if $N_T = 5$.

Step #4. If $N_T \neq 5$, eliminate whatever populations you can. Suppose there are k_6 populations left.

Step #5. Take another observation on each remaining population, so that there are now n observations on k_n populations. Compute c_0 and x_0 as in Section 2 of this paper based on γ_0 and k_n . Compute $s^2(n, k_n)$ = pooled sample variance on the k_n populations.

Step #6. If $n \geq (c_0 s(n, k_n)/L)^2$, discontinue sampling and announce the Tong confidence interval with k_n populations.

Step #7. Otherwise, set $n = n+1$ and see if any more populations can be eliminated. There are now k_n populations left. Return to step #5.

In Tables 5-8 we present the results of a simulation (with 200 iterations) of the elimination rule for a fixed-width confidence interval of length L for the largest normal mean. The prescribed confidence level is $\gamma = .90$, and we chose $\gamma_0 = .92$, $\beta = .98$, so that only obviously inferior populations were eliminated from consideration. In Tables 9 and 10 we present values of the ratio

$$\frac{\text{total observations used by Tong's procedure}}{\text{total observations used by the elimination procedure}}$$

The tables make clear the following conclusions:

- (1) The elimination rule achieves its intended coverage probability.
- (2) The elimination rule results in approximately 10% more observations in the case that the means are relatively close and k is small.
- (3) When the means differ to any appreciable degree and for larger k , the elimination rule can result in substantial savings in the number of observations taken. For example if $k = 10$, $L = .50$, $d = .50$ and

$$\begin{aligned}\mu_2 - \mu_1 &= d = .50 \\ \mu_3 - \mu_2 &= d = .50 \\ &\vdots \\ \mu_{10} - \mu_9 &= d = .50 ,\end{aligned}$$

the elimination procedure takes only 35% of the total observations needed by Tong's procedure, a dramatic savings.

4. Conclusion

We have discovered by a simple grafting technique a procedure which eliminates obviously inferior populations early in the experiment, thus leading to possibly dramatic savings in sample size over the conventional procedure. We argue that elimination methodology is easy to use and easy to study, and that the savings in sample size argue for their implementation.

5. Acknowledgement

Mr. Robert Smith prepared the tables, and his help is gratefully acknowledged.

References

- Chow, Y.S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* 36, 457-462.
- 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, 9-36.
- Tong, Y.L. (1973). An asymptotically optimal sequential procedure for the estimation of the largest mean. *Ann. Statist.* 1, 175-179.

TABLE I
VALUES OF C0 (UPPER ENTRY) AND X0 (LOWER ENTRY) FOR THE LARGEST NORMAL MEAN

Y: K:	0.900	0.905	0.910	0.915	0.920	0.925	0.930	0.935	0.940	0.945	0.950	0.955	0.960	0.965	0.970	0.975	0.980	0.985	0.990
2	3.290	3.339	3.391	3.445	3.501	3.561	3.624	3.691	3.762	3.833	3.920	4.009	4.107	4.217	4.340	4.483	4.653	4.865	5.152
	1.645	1.670	1.695	1.722	1.751	1.780	1.812	1.845	1.881	1.919	1.960	2.005	2.054	2.108	2.170	2.241	2.326	2.432	2.576
3	3.329	3.379	3.430	3.484	3.541	3.600	3.663	3.730	3.801	3.877	3.959	4.043	4.146	4.255	4.378	4.521	4.690	4.901	5.187
	1.509	1.535	1.562	1.590	1.620	1.651	1.684	1.718	1.755	1.795	1.838	1.884	1.935	1.992	2.056	2.130	2.218	2.327	2.475
4	3.390	3.439	3.491	3.545	3.601	3.660	3.723	3.789	3.860	3.936	4.017	4.106	4.204	4.312	4.435	4.576	4.744	4.954	5.239
	1.447	1.473	1.501	1.530	1.560	1.592	1.626	1.661	1.699	1.740	1.783	1.831	1.883	1.941	2.006	2.082	2.171	2.283	2.433
5	3.447	3.496	3.548	3.601	3.658	3.717	3.779	3.845	3.915	3.991	4.072	4.160	4.257	4.365	4.487	4.627	4.795	5.004	5.286
	1.411	1.438	1.466	1.496	1.526	1.559	1.593	1.629	1.668	1.709	1.753	1.801	1.854	1.913	1.979	2.055	2.145	2.258	2.409
6	3.490	3.547	3.605	3.662	3.720	3.780	3.842	3.904	3.964	4.039	4.120	4.208	4.304	4.411	4.533	4.672	4.839	5.047	5.328
	1.388	1.415	1.444	1.474	1.505	1.537	1.572	1.609	1.647	1.689	1.734	1.782	1.835	1.894	1.961	2.038	2.129	2.242	2.394
7	3.544	3.592	3.643	3.696	3.752	3.810	3.872	3.938	4.007	4.082	4.162	4.250	4.346	4.452	4.573	4.712	4.878	5.085	5.365
	1.372	1.399	1.426	1.453	1.480	1.523	1.557	1.594	1.633	1.675	1.720	1.769	1.823	1.882	1.949	2.026	2.117	2.231	2.384
8	3.594	3.632	3.683	3.736	3.791	3.850	3.911	3.976	4.046	4.120	4.200	4.287	4.383	4.489	4.609	4.748	4.913	5.119	5.398
	1.360	1.388	1.417	1.447	1.479	1.512	1.547	1.584	1.623	1.665	1.710	1.759	1.813	1.873	1.940	2.017	2.109	2.223	2.376
9	3.620	3.668	3.719	3.771	3.827	3.885	3.946	4.011	4.080	4.154	4.234	4.321	4.416	4.522	4.642	4.780	4.944	5.150	5.428
	1.351	1.379	1.408	1.438	1.470	1.503	1.538	1.576	1.615	1.657	1.703	1.752	1.806	1.866	1.933	2.010	2.102	2.217	2.370
10	3.652	3.701	3.751	3.804	3.859	3.916	3.978	4.042	4.111	4.185	4.264	4.351	4.446	4.552	4.671	4.809	4.973	5.177	5.455
	1.344	1.372	1.401	1.431	1.463	1.497	1.532	1.569	1.609	1.651	1.697	1.746	1.800	1.860	1.928	2.005	2.097	2.212	2.366

Table 2

Average number of correct decisions in 200 simulations of Tong's procedure,

$$\mu_1 = \dots = \mu_{k-1} = \mu_{k-d}$$

k	2	3	5	10
d = 0.0, L = .50	.900	.900	.930	.905
d = 0.0, L = 1.0	.905	.910	.915	.890
d = .50, L = .50	.890	.930	.930	.920
d = .50, L = 1.0	.930	.925	.975	.980
d = 2.0, L = .50	.890	.925	.920	.900
d = 2.0, L = 1.0	.900	.865	.910	.890

N.B. The average sample size $\times k$ (the total number of observations) upon stopping is

k	2	3	5	10
L = 1.00	21	33	59	137
L = .50	84	132	238	538

Table 3

Average number of correct decisions in 200 simulations of Tong's procedure,

$$\mu_{i+1} - \mu_i = d \quad (i = 1, \dots, k-1) .$$

k	2	3	5	10
d = 0.0 L = .50	.900	.900	.930	.905
d = 0.0 L = 1.0	.905	.910	.915	.890
d = .50 L = .50	.890	.925	.925	.900
d = .50 L = 1.0	.930	.890	.955	.930
d = 2.0 L = .50	.890	.925	.920	.900
d = 2.0 L = 1.0	.900	.865	.910	.890

Table 4

SEQUENTIAL ESTIMATION OF LARGEST MEAN: ELIMINATION PROCEDURE
VALUES OF A (UPPER ENTRY) AND T (LOWER ENTRY) FOR LARGEST K NORMAL MEANS

[illegible]

Table 5

Coverage probabilities for elimination rule with $\gamma = .90$ and

$$\mu_1 = \dots = \mu_{k-1} = \mu_k - d.$$

k	2	3	5	10
d = 0.0 L = .50	.930	.930	.915	.905
d = 0.0 L = 1.0	.920	.890	.915	.920
d = .50 L = .50	.895	.950	.910	.925
d = .50 L = 1.0	.950	.940	.980	.980
d = 2.0 L = .50	.880	.935	.895	.890
d = 2.0 L = 1.0	.880	.880	.915	.910

Table 6

Average total number of observations for elimination rule with $\gamma = .90$ and

$$\mu_1 = \dots = \mu_{k-1} = \mu_k - d.$$

k	2	3	5	10
k = 0.0 L = .50	94	147	264	593
d = 0.0 L = 1.0	23	36	67	152
d = .50 L = .50	88	141	253	575
d = .50 L = 1.0	23	36	67	152
d = 2.0 L = .50	55	67	90	160
d = 2.0 L = 1.0	19	30	53	120

Table 7

Coverage probabilities for elimination rule with $\gamma = .90$ and

$$\mu_{i+1} - \mu_i = d, i = 1, \dots, k-1.$$

k	2	3	5	10
d = 0.0 L = .50	.930	.930	.915	.905
d = 0.0 L = 1.0	.920	.890	.915	.920
d = .50 L = .50	.895	.940	.895	.910
d = .50 L = 1.0	.950	.946	.960	.930
d = 2.0 L = .50	.880	.935	.900	.885
d = 2.0 L = 1.0	.880	.875	.875	.910

Table 8

Average total number of observations for elimination rule with $\gamma = .90$ and

$$\mu_{i+1} - \mu_i = d, i = 1, \dots, k-1.$$

k	2	3	5	10
d = 0.0 L = .50	94	147	264	593
d = 0.0 L = 1.0	23	36	67	152
d = .50 L = .50	88	120	150	189
d = .50 L = 1.0	23	36	61	97
d = 2.0 L = .50	55	62	73	98
d = 2.0 L = 1.0	19	25	36	62

Table 9

Ratio of Tong's total sample size to elimination rule's total sample size for

$$\mu_1 = \dots = \mu_{k-1} = \mu_k - d.$$

k	2	3	5	10
d = 0.0 L = .50	.89	.90	.90	.91
d = 0.0 L = 1.0	.91	.92	.88	.90
d = .50 L = .50	.95	.94	.94	.94
d = .50 L = 1.0	.91	.92	.88	.90
d = 2.0 L = .50	1.53	1.97	2.64	3.36
d = 2.0 L = 1.0	1.11	1.10	1.11	1.14

Table 10

Ratio of T Ong's total sample size to elimination rule's total sample size for

$$\mu_{i+1} - \mu_i = d, i = 1, \dots, k-1.$$

k	2	3	5	10
d = 0.0 L = .50	.89	.90	.90	.91
d = 0.0 L = 1.0	.91	.92	.88	.90
d = .50 L = .50	.95	1.1	1.59	2.85
d = .50 L = 1.0	.91	.92	.97	1.41
d = 2.0 L = .50	1.53	2.13	3.26	5.49
d = 2.0 L = 1.0	1.11	1.32	1.64	2.21

