The Annals of Statistics
1973, Vol. 1, No. 1, 175-179

AN ASYMPTOTICALLY OPTIMAL SEQUENTIAL PROCEDURE FOR THE ESTIMATION OF THE LARGEST MEAN

By Yung Liang Tong

University of Nebraska

Interval estimation of the largest mean of k normal populations ($k \geqq 1$) with a common variance σ^2 is considered. When σ^2 is known the optimal fixed-width interval is given so that, to have the probability of coverage uniformly lower bounded by γ (preassigned), the sample size needed is minimized. This optimal interval is unsymmetric for $k > 2$. When σ^2 is unknown a sequential procedure is proposed and its behavior is studied. It is shown that the confidence interval obtained, which is also unsymmetric for $k > 2$, behaves asymptotically as well as the optimal interval. This represents an improvement of the procedure of symmetric intervals considered by the author previously; the improvement is significant, especially when k is large.

1. Introduction. Let there be k normal populations ($k \geqq 1$) with unknown means $\mu_1, \cdots, \mu_k$, respectively, and a common variance σ^2. In some statistical problems it is required to estimate the largest mean $\mu^* = \max_{1 \leqq i \leqq k} \mu_i$ when there is no prior knowledge regarding the order of the μ_i's. The point and interval estimations of μ^* based on the sample means have been considered by several authors; and in a recent paper Dudewicz and Tong [3] showed that, to maximize the probability of coverage when σ^2 is known, symmetric intervals are optimal if and only if $k \leqq 2$. When $k > 2$, due to the fact that the largest sample mean always overestimates μ^*, the optimal interval is always unsymmetric and its performance is significantly better than that of the symmetric intervals for large k.

In this paper we consider a sequential procedure for the construction of fixed-width confidence intervals for μ^* when σ^2 is unknown. It is shown that when the width of the interval is small, the confidence interval obtained behaves approximately as well as the optimal confidence interval which could be obtained only if σ^2 were known. This procedure represents an improvement over a procedure of symmetric intervals considered by the author in a previous paper [6], and the improvement is significant, especially when k is large.

In Section 2 we first construct an optimal confidence interval under the single-stage procedure when σ^2 is known; it minimizes the number of observations needed to achieve the probability requirement. The sequential procedure is then given for the case of unknown σ^2 and its probability of coverage is derived. In Section 3 asymptotically optimal properties of this procedure are proved. A comparison of this procedure with the procedure of symmetric intervals and an extension of the solution to the estimation of the smallest mean are provided in Section 4.

Received October 26, 1971; revised April 25, 1972.

2. The procedure and its probability of coverage. Let $\bar{X}_{1n}, \cdots, \bar{X}_{kn}$ denote the sample means with a common variance σ^2/n after n observations are taken from each of the k populations, and let $\bar{X}_n^* = \max_{1\leq i\leq k} \bar{X}_{in}$ denote the largest sample mean. For preassigned $L > 0$ let the interval for μ^* be given by $I = (\bar{X}_n^* - (L-d), \bar{X}_n^* + d)$. Since there is no prior knowledge about the mean vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_k)$, it is a natural thing to require $\inf_\mu P_\mu[\mu^* \in I] \geqq \gamma$ where the probability of coverage $\gamma \in (0, 1)$ is arbitrary but preassigned. Following from Theorem 1 of [3] for every L, d, σ^2, n and k we have

$$\inf_\mu P_\mu[\mu^* \in I] = \min\{\Phi^k(c-x) - \Phi^k(-x), \Phi(c-x) - \Phi(-x)\} = \alpha_k(c, x) \quad \text{(say)}, \tag{1}$$

where $c = n^{\frac{1}{2}}L/\sigma$, $x = n^{\frac{1}{2}}d/\sigma$ and Φ is the $N(0, 1)$ cdf. Clearly for every c and k there is an optimal value of x (which depends on c), $x_0 = x_0(c)$ (say), such that $\alpha_k(c, x_0) = \sup_x \alpha_k(c, x)$. Let c_0 and n_0 satisfy

$$c_0 = \inf\{c : \alpha_k(c, x_0) \geqq \gamma\}, \qquad n_0 = \text{ the smallest integer } \geqq (c_0\lambda)^2 \tag{2}$$

where $\lambda = \sigma/L$. Then with $n = n_0$ we observe $\bar{X}_{n_0}^*$ and construct

$$I_0 = (\bar{X}_{n_0}^* - (L - d_0), \bar{X}_{n_0}^* + d_0)$$

with $d_0 = x_0\sigma/n_0^{\frac{1}{2}}$, I_0 is the optimal confidence interval for μ^* and n_0 is the smallest sample size needed to satisfy the probability requirement. The values of (c_0, x_0) are tabulated in the attached table for $\gamma = 0.75$, 0.90, 0.95, 0.975, 0.99 and $k = 2(1)14$. We note that σ^2 is needed to determine both the sample size n_0 and the location of the interval through d_0, and that $d_0 = L/2$ for $k \leqq 2$ and $d_0 < L/2$ for $k > 2$ as proved in [3].

We now consider the following sequential procedure when σ^2 is unknown. For given k and for arbitrary but preassigned γ and L, let (c_0, x_0) be the corresponding optimal values of c and x given in Table 1.

Procedure R. (a) Take n^* observations from each of the k populations where $n^* \geqq 2$ is arbitrary but preassigned.

(b) After n observations $\{X_{ij}\}\, j = 1, \cdots, n;\ i = 1, \cdots, k$ are taken from each of the k populations compute

$$S_n^2 = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n \left(X_{ij} - \frac{1}{n}\sum_{j=1}^n X_{ij}\right)^2, \tag{3}$$

and stop with stopping variable N where

$$N = \text{the smallest integer } n \text{ such that } n \geqq c_0^2 S_n^2/L^2. \tag{4}$$

(c) When sampling is stopped observe $\bar{X}_N^*$ and S_N^2, and construct

$$I_0' = (\bar{X}_N^* - (L - x_0 S_N/N^{\frac{1}{2}}), \bar{X}_N^* + x_0 S_N/N^{\frac{1}{2}}). \tag{5}$$

We first give a lower bound on the probability of coverage under R. It is easy to see from [4] that for every $\boldsymbol{\mu}$ and σ^2 and for every given $(N, S_N) = (n, s)$

the conditional probability that I_0' will cover μ^* is

$$P_{\mu,\sigma^2}[\mu^* \in I_0' \mid (n, s)] = \prod_{i=1}^k \Phi(n^{\frac{1}{2}}\delta_i/\sigma + n^{\frac{1}{2}}/\lambda - x_0 s/\sigma) - \prod_{i=1}^k \Phi(n^{\frac{1}{2}}\delta_i/\sigma - x_0 s/\sigma) \tag{6}$$

where $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_k)$ is such that $\delta_i = \mu^* - \mu_i \geqq 0$ $(i = 1, \cdots, k)$ and at least one of the δ_i's is 0. Applying Theorem 1 of [3] it follows that for every (n, s) the r.h.s. of (6) is lower bounded by

$$g_k(\sigma, L, n, s) = \min\{\Phi^k(n^{\frac{1}{2}}/\lambda - x_0 s/\sigma) - \Phi^k(-x_0 s/\sigma), \Phi(n^{\frac{1}{2}}/\lambda - x_0 s/\sigma) - \Phi(-x_0 s/\sigma)\}\,. \tag{7}$$

Since the probability of coverage is the expection of the r.h.s. of (6) taken over the distribution of (N, S_N), for every μ, σ^2 we have

$$P_{\mu,\sigma^2}[\mu^* \in I_0'] \geqq Eg_k(\sigma, L, N, S_N)\,. \tag{8}$$

3. Asymptotically optimal properties of the Procedure. We now show that the confidence interval I_0' obtained under the Procedure R behaves asymptotically as well as the optimal interval I_0 which cannot be obtained when σ^2 is unknown. We first observe the asymptotic behavior of the random sample size N.

THEOREM 1. *Under the Procedure R we have,*

$$P_{\mu,\sigma^2}[N < \infty] = 1 \qquad \textit{for every } \mu \textit{ and } \sigma^2\,, \tag{9}$$

$$\lim_{L\to 0} N = \infty \quad \text{a.s.}\,, \qquad \lim_{L\to 0} N/(c_0\lambda)^2 = 1 \quad \text{a.s.}\,, \qquad \lim_{L\to 0} EN/(c_0\lambda)^2 = 1\,. \tag{10}$$

PROOF. (9) follows from the a.s. convergence of the sample variances to σ^2. (10) follows from Lemmas 1 and 2 of Chow and Robbins [2] with $y_n = S_n^2/\sigma^2$, $f(n) = n$ and $t = (c_0\lambda)^2$. The additional condition in Lemma 2 on the existence of $E\sup_n y_n$ can be justified by an argument similar to that given in the proof of Theorem 3.3 of [5].

To prove the convergence of the probabilities of coverage we first observe a lemma on the convergence of a sequence of estimators based on a random number of observations. Since the proof is elementary it is not given here. Let $\{T_n\}$ be a sequence of estimators of θ so that T_n is based on the first n observations only, let N be any stopping variable which depends on a parameter λ, and let λ_0 be either a finite real number or infinity.

LEMMA. (a) *If* $T_n \to \theta$ a.s. *as* $n \to \infty$ *and* $N \to \infty$ a.s. *as* $\lambda \to \lambda_0$, *then* $T_N \to \theta$ a.s. *as* $\lambda \to \lambda_0$. (b) *If* $T_n \to_P \theta$ *as* $n \to \infty$ *and* $N \to_P \infty$ *as* $\lambda \to \lambda_0$, *then* $T_N \to_P \theta$ *as* $\lambda \to \lambda_0$.

Note that T_N and N are not necessarily independent random variables.

THEOREM 2. *Under the Procedure R, for every* μ, σ^2 *we have*

$$\lim_{L\to 0} P_{\mu,\sigma^2}[\mu^* \in I_0'] \geqq \gamma\,. \tag{11}$$

PROOF. From (8) it suffices to show the convergence of the expections of $g_k(\sigma, L, N, S_N)$ taken over the joint distributions of (N, S_N), which depend on λ. From (10) and the lemma we have $N^{\frac{1}{2}}/\lambda \to c_0$ a.s. and $S_N/\sigma \to 1$ a.s. as $L \to 0$ $(\lambda \to \infty)$. Since a.s. convergence is preserved by continuous mappings, $g_k(\sigma, L, N, S_N)$ converges to

$$(12) \qquad \min\{\Phi^k(c_0 - x_0) - \Phi^k(-x_0), \Phi(c_0 - x_0) - \Phi(-x_0)\} = \alpha_k(c_0, x_0)$$

a.s. as $L \to 0$. Therefore for every μ we have

$$\begin{aligned} \lim_{L\to 0} P_{\mu,\sigma^2}[\mu^* \in I_0'] &\geqq \lim_{L\to 0} Eg_k(\sigma, L, N, S_N) \\ &= E \lim_{L\to 0} g_k(\sigma, L, N, S_N) = \alpha_k(c_0, x_0) = \gamma \end{aligned}$$

where the first equality follows from the fact that $g_k(\sigma, L, N, S_N)$ is uniformly bounded and the second equality follows from the definition of (c_0, x_0). This completes the proof of the Theorem.

4. Comparison and extension. We can now compare the Procedure R with the procedure of symmetric intervals considered in [6]. Let N be defined in (4) and N' be the random sample size under the sequential procedure defined in [6]. It follows that for every fixed γ we have

$$(13) \qquad \lim_{L\to 0} EN'/EN = (2z/c_0)^2 = \beta_k(\gamma) \quad \text{(say)},$$

where z satisfies, $\Phi^k(z) - \Phi^k(-z) = \gamma$. Some of the $\beta_k(\gamma)$ values are computed and an excerpt is given below:

TABLE 1
Values of $\beta_k(\gamma)$

k \ γ	0.75	0.90	0.95
3	1.297	1.196	1.147
8	1.941	1.535	1.406
12	2.098	1.647	1.484

Clearly the improvement is significant, especially when k is large.

The procedure can also be extended to obtain a confidence interval for the smallest mean $\mu_* = \min_{1\le i\le k} \mu_i$ in the following way: for fixed k, γ and L let (c_0, x_0) be given in the attached table and consider

Procedure R': (a) Apply the same stopping rule given in Procedure R.

(b) When sampling is stopped observe $S_N{}^2$ and $\bar{X}_{*N} = \min_{1\le i\le k} \bar{X}_{iN}$, and construct

$$(14) \qquad I_0'' = (\bar{X}_{*N} - x_0 S_N/N^{\frac{1}{2}}, \bar{X}_{*N} + (L - x_0 S_N/N^{\frac{1}{2}}))\,.$$

Then following from arguments similar to that given in Section 5 of [6] and in the previous section of this paper it is easy to see that the procedure R' is asymptotically optimal for the estimation of the smallest mean μ_*.

TABLE 2
Values of c_0 (upper entry) and x_0 (lower entry) for the largest normal mean

k \ γ	0.75	0.90	0.95	0.975	0.99
2	2.3007 1.1503	3.2897 1.6448	3.9200 1.9600	4.4830 2.2415	5.1510 2.5755
3	2.3357 0.9936	3.3290 1.5094	3.9590 1.8375	4.5204 2.1295	5.1870 2.4752
4	2.3946 0.9118	3.3899 1.4466	4.0173 1.7833	4.5760 2.0816	5.2380 2.4325
5	2.4534 0.8627	3.4473 1.4108	4.0717 1.7530	4.6270 2.0547	5.2860 2.4092
6	2.5070 0.8303	3.4983 1.3878	4.1200 1.7338	4.6722 2.0376	5.3280 2.3942
7	2.5550 0.8073	3.5435 1.3718	4.1622 1.7201	4.7121 2.0257	5.3650 2.3837
8	2.5981 0.7903	3.5837 1.3600	4.2000 1.7102	4.7480 2.0172	5.3980 2.3761
9	2.6369 0.7771	3.6197 1.3509	4.2340 1.7027	4.7800 2.0106	5.4280 2.3704
10	2.6722 0.7667	3.6524 1.3437	4.2644 1.6966	4.8083 2.0050	5.4550 2.3659
11	2.7044 0.7581	3.6821 1.3379	4.2923 1.6917	4.8350 2.0010	5.4790 2.3620
12	2.7340 0.7511	3.7094 1.3330	4.3180 1.6877	4.8590 1.9974	5.5020 2.3592
13	2.7613 0.7450	3.7346 1.3289	4.3416 1.6843	4.8810 1.9942	5.5220 2.3562
14	2.7868 0.7400	3.7580 1.3255	4.3635 1.6813	4.9020 1.9919	5.5410 2.3538

Acknowledgment. I wish to thank Professor S. K. Perng for a discussion.

REFERENCES

[1] BLUMENTHAL, S. and COHEN, A. (1968). Estimation of the larger translation parameter. *Ann. Math. Statist.* **39** 502–516.

[2] 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.

[3] DUDEWICZ, E. J. and TONG, Y. L. (1971). Optimal confidence intervals for the largest location parameter. *Statistical Decision Theory and Related Topics*, ed. S. S. Gupta and J. Yackel. Academic Press, New York, 363–376.

[4] SAXENA, K. M. L. and TONG, Y. L. (1969). Interval estimation of the largest mean of k normal population with known variances. *J. Amer. Statist. Assoc.* **64** 296–299.

[5] TONG, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. *Ann. Math. Statist.* **40** 1300–1324.

[6] TONG, Y. L. (1970). Multi-stage interval estimations of the largest mean of k normal populations. *J. Roy. Statist. Soc. Ser. B* **32** 272–277.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NEBRASKA
LINCOLN, NEBRASKA 68508