The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 88–92

A SEQUENTIAL PROCEDURE FOR SELECTING THE LARGEST OF k MEANS[1]

BY H. ROBBINS, M. SOBEL, AND N. STARR

University of California, Berkeley, University of Minnesota, and Carnegie-Mellon University

1. Introduction. Let $\pi_1, \pi_2, \cdots, \pi_k$ denote k populations ($k \geqq 2$) in which we may observe the independent random variables $x_1, x_2, \cdots, x_k$, respectively, where x_j is $N(\mu_j, \sigma^2)$ for $j = 1, \cdots, k$. The $k + 1$ parameters $\mu_1, \cdots, \mu_k, \sigma^2$ are assumed unknown. We denote the ordered μ-values by $\mu_{[1]} \leqq \mu_{[2]} \leqq \cdots \leqq \mu_{[k]}$, and if $\mu_{[k]} > \mu_{[k-1]}$ we refer to the population π with $\mu = \mu_{[k]}$ as the best population. Our goal is to select the best population with probability at least P^* whenever $\mu_{[k]} - \mu_{[k-1]} \geqq \delta^*$; here P^* and δ^* are preassigned constants with $1/k < P^* < 1$ and $\delta^* > 0$. In other words, letting CS denote correct selection and letting Ω_1 be the set of all vectors $\omega = (\mu_1, \cdots, \mu_k, \sigma^2)$ with $\mu_{[k]} - \mu_{[k-1]} \geqq \delta^*$, we wish to obtain a procedure for which

$$P(\text{CS}) \geqq P^* \quad \text{for all} \quad \omega \in \Omega_1 . \tag{1}$$

If σ^2 were known we could proceed as in [1]. Take a fixed number n of independent observations on each of the k random variables, and denote by $\bar{x}_j(n)$ the sample mean of the n observations on x_j ($j = 1, \cdots, k$). If $\bar{x}_\alpha(n) \geqq \bar{x}_i(n)$ for all $i = 1, \cdots, k$, select π_α.

Let Φ denote the cdf of a $N(0, 1)$ random variable. From [1], for all $\omega \in \Omega_1$ and $n \geqq 1$

$$P_n(\text{CS}) \geqq \int_{-\infty}^{\infty} \Phi^{k-1}(y + n^{\frac{1}{2}}\delta^*/\sigma)\, d\Phi(y), \tag{2}$$

with equality when $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. Define for $\delta^* > 0$

$$c \equiv c(k, \delta^*) = (\delta^*)^2/h^2, \tag{3}$$

where for given k, $h = h(k, P^*)$ is the solution of

$$\int_{-\infty}^{\infty} \Phi^{k-1}(y + h)\, d\Phi(y) = P^*;$$

for $P^* > 1/k$ it is clear that $h > 0$ and hence $c > 0$. Then from (2) it is easily seen that (1) is satisfied provided n is chosen so that

$$\sigma^2 \leqq cn. \tag{4}$$

Treating n as a continuous variable we denote by $n^* = \sigma^2/c$ the minimal fixed number of k-tuples of observations. Tables of h for various values of P^* and k are given in [1].

Received 3 May 1967; revised 10 July 1967.

[1] This research was partially supported by a grant from the R. K. Mellon Foundation at Carnegie-Mellon University and by N. S. F. Grant GP-3813 at the University of Minnesota.

Clearly, when σ^2 is unknown no fixed sample size procedure will work for all $\omega \varepsilon \Omega_1$. A two-stage procedure for this problem has been studied [2]. This procedure, while guaranteeing (1) for all $\omega \varepsilon \Omega_1$, is inefficient, since it utilizes only part of the sample to estimate σ^2. Accordingly, we consider here a more fully

Sequential Procedure. Let x_{ij} denote the ith observation on x_j ($j = 1, \cdots, k$), and define for $r \geqq 2$

$$u_r = (k(r-1))^{-1} \textstyle\sum_{j=1}^{k} \sum_{i=1}^{r} (x_{ij} - \bar{x}_j(r))^2.$$

Sampling proceeds sequentially, where at the rth stage we make a single observation on each of the k random variables x_j and compute a fresh estimate u_r of σ^2. We terminate sampling at the Nth stage, where the random variable N is the least *odd* integer $n \geqq 5$ such that

$$u_n \leqq cn. \qquad \text{(cf. (4).)} \tag{5}$$

If α is the smallest j such that $\bar{x}_j(N) \geqq \bar{x}_i(N)$ for all $i = 1, \cdots, k$, we select π_α. (We require that sampling may be terminated only with an odd number of k-tuples of observations in order to simplify the later computations of the probability distribution of N.) Certain aspects of a similar procedure have been considered by Srivastava [8].

2. Some properties of the procedure. As in [6], [9] we note that for each fixed n the vector $\bar{\mathbf{x}}(n) = (\bar{x}_1(n), \cdots, \bar{x}_k(n))$ is independent of the event $\{N = n\}$. Accordingly, since the selection procedure depends only on $\bar{\mathbf{x}}(N)$, we have from (2) for all fixed $\delta^* > 0$, $\omega \varepsilon \Omega_1$

$$\begin{aligned} P(\text{CS}) &= \textstyle\sum_n P(CS \mid N = n)P(N = n) \\ &= \textstyle\sum_n P_n(\text{CS})P(N = n) \\ &\geqq \textstyle\sum_n \int_{-\infty}^{\infty} \Phi^{k-1}(y + n^{\frac{1}{2}}\delta^*/\sigma)\, d\Phi(y) \cdot P(N = n) \equiv \beta(\lambda), \quad \text{say,} \end{aligned} \tag{6}$$

where

$$\lambda = \sigma/\delta^*. \tag{7}$$

That the procedure is asymptotically satisfactory as $\delta^* \to 0$ follows from [3], [5]. Let ω be fixed and let $\delta^* \to 0$. Then $n^* \to \infty$, $N \to \infty$ a.s., $N/n^* \to 1$ a.s., $EN/n^* \to 1$. Since therefore

$$N^{\frac{1}{2}}\delta^*/\sigma \to h \quad \text{in probability,} \tag{8}$$

from (6) it follows that

$$\liminf P(CS) \geqq \int_{-\infty}^{\infty} \Phi^{k-1}(y + h)\, d\Phi(y) = P^*, \qquad \omega \varepsilon \Omega_1. \tag{9}$$

In fact, all these results hold without the assumption of normality provided only that the x_j have $0 < \sigma^2 < \infty$.

As a measure of the cost of ignorance of σ^2 we define

$$I = I(\lambda) = EN - n^*.$$

This cost is negligible, as may be seen from the following

THEOREM. *For all* $\delta^* > 0, 0 < \sigma^2 < \infty$,

$$I \leqq 5. \tag{10}$$

PROOF. Define

$$U_r = \sum_{j=1}^{k} \sum_{i=1}^{r} (x_{ij} - \mu_j)^2;$$

then for all $r \geqq 1$,

$$k(r-1)u_r \leqq U_r \leqq U_{r+2}. \tag{11}$$

Since from (5) for all odd $N > 5$, $c(N-2) < u_{N-2}$, we have from (11) for odd $N > 5$, $c(N-2) < U_N/k(N-3)$. Thus for all odd $N \geqq 5$,

$$ckN(N-5) < U_N. \tag{12}$$

Taking expectations on both sides of (12) we obtain by Wald's lemma $ckE(N(N-5)) < \sigma^2\, kEN$; hence $(EN)^2 - 5EN < EN^2 - 5EN < (\sigma^2/c)EN$, and $EN < \sigma^2/c + 5 = n^* + 5$.

In our proof we tacitly assumed that $EN < \infty$ so as to be able to apply Wald's lemma. We may eliminate this assumption by the following device. Define for each $m > 5$ the random variable $N_m = \min(m, N)$ and deduce as above that $EN_m < n^* + 5$. Letting $m \to \infty$ we have $EN_m \uparrow EN \leqq n^* + 5$. Thus (10) holds whatever the distribution of the x_j, provided only that $\sigma^2 < \infty$.

We further observe that the inequality (10) cannot be sharpened since our procedure requires that $N \geqq 5$, and accordingly (from (3)) $I \to 5$ as $\delta^* \to \infty$.

3. Small sample performance. We have established that the sequential procedure is asymptotically consistent and efficient (in the sense of [3]) and that the cost of ignorance of σ^2 is of little consequence when the sequential procedure is used, for all $\delta^* > 0$, $0 < \sigma^2 < \infty$. It remains to verify that $P(CS)$ is approximately $\geqq P^*$ for all $\omega \in \Omega_1$.

Let $\{\nu_i\}$ $(i = 1, 2, \cdots)$ be independent random variables, each with a chi-squared distribution with k degrees of freedom. By Helmert's transformation we can write for all $n \geqq 2$

$$k(n-1)u_n/\sigma^2 = \sum_{i=1}^{n-1} \nu_i. \tag{13}$$

Let $n = 2m + 1$; then from (5) N is the least integer $m \geqq 2$ such that

$$u_{2m+1} \leqq c(2m+1). \tag{14}$$

From (13) (recalling the definition (3) of c) we may rewrite (14) in the form

$$\sum_{n=1}^{2m} \nu_i \leqq k(2m+1)2m/h^2\lambda^2.$$

The random variable $w = (\nu_1 + \nu_2)/2$ is the sum of k independent standardized exponential random variables, so N is the least integer $m \geqq 1$ such that $w_1 + \cdots + w_m \leqq a_{m+1}$, where the w's are independent and distributed as the sum of k standardized exponential random variables and where the constants

$\{a_m\}$ are given by

$$a_2 = 0, \qquad a_m = k(2m - 1)(m - 1)/h^2\lambda^2 \quad \text{for} \quad m = 3, 4, \cdots .$$

Thus for values of the parameter λ, $0 < \lambda < \infty$, the probability distribution p_m of N is defined for $m = 1, 2, \cdots$ by

$$p_m(\lambda) = P_\lambda(N = 2m + 1) = P_\lambda(z_1 > a_2, z_2 > a_3, \cdots z_{m-1} > a_m, z_m \leqq a_{m+1})$$

where for $m \geqq 1$, $z_m = w_1 + \cdots + w_m$.

The following recursive scheme, which generalizes [6], [9], gives a method for computing $p_m(\lambda)$ $(m = 1, 2, \cdots)$ for given values of the parameter λ.

Define for $\alpha = 0, 1, \cdots, k - 1$,

$$h_k^{(\alpha)}(x) = x^{k-1-\alpha}/(k - 1 - \alpha)!$$

and for $m = 2, 3, 4, \cdots$,

$$h_{mk}^{(\alpha)}(x) = \sum_{j=1}^{m-1} \sum_{\beta=0}^{k-1} (x - a_m)^{jk+\beta-\alpha}((jk + \beta - \alpha)!)^{-1} h_{(m-j)k}^{(\beta)}(a_m).$$

Let $c_1 = c_2 = 1$, and for $m = 3, 4, \cdots$ define

$$c_m = e^{-a_m} \sum_{j=0}^{m-1} \sum_{\alpha=0}^{k-1} h_{(m-j)k}^{(\alpha)}(a_m);$$

then for all $m \geqq 1$

$$p_m = c_m - c_{m+1}.$$

TABLE

Values of β and EN for $P^* = .95$

n^*	$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	EN	β	*EN*	β	*EN*	β	*EN*	β
2	5.02	.99539	5.01	.99763	5.00	.99840	5.10	.99879
4	5.63	.97223	5.45	.97509	5.46	.97636	5.42	.97710
6	6.88	.95503	6.87	.95598	6.86	.95682	6.86	.95759
8	8.44	.94620	8.56	.94734	8.64	.94918	8.70	.95099
10	10.17	.94217	10.42	.94431	10.57	.94713	10.67	.94956
12	12.01	.94065	12.37	.94376	12.56	.94708	12.67	.94964
14	13.92	.94043	14.25	.94420	14.57	.94760	14.69	.95000
16	15.87	.94084	16.36	.94497	16.59	.94821	16.70	.95033
18	17.86	.94155	18.38	.94576	18.60	.94872	18.71	.95055
20	19.87	.94236	20.40	.94648	20.62	.94911	20.73	.95068
25	25.00	.94428	25.46	.94779	25.65	.94969	25.74	.95079
30	30.00	.94576	30.49	.94853	30.67	.94993	30.75	.95077
35	35.07	.94680	35.52	.94896	35.68	.95003	35.76	.95072
40	40.12	.94751	40.53	.94921	40.68	.95009	40.76	.95066
45	45.16	.94801	45.55	.94937	45.69	.95011	45.76	.95061
50	50.20	.94835	50.55	.94947	50.69	.95012	50.76	.95057
60	60.24	.94879	60.56	.94961	60.70	.95012	60.77	.95049
70	70.26	.94904	70.57	.94969	70.70	.95012	70.77	.95043
80	80.28	.94921	80.57	.94974	80.70	.95011	80.77	.95039

Thus from (6), for any fixed value of the parameter λ,

$$(15)\quad P(\text{CS};\lambda) \geqq \textstyle\sum_{m=1}^{\infty} \int_{-\infty}^{\infty} \Phi(y + (2m+1)^{\frac{1}{2}}/\lambda)\, d\Phi(y)\cdot p_m(\lambda) = \beta(\lambda)$$

for all $\mu_1, \cdots, \mu_k$ for which $\omega \varepsilon \Omega_1$. Moreover, the expected sample size $EN = E_\lambda N$ is defined by

$$(16)\qquad EN = \textstyle\sum_{m=1}^{\infty} (2m+1)p_m(\lambda) = 2\sum_{m=1}^{\infty} mp_m(\lambda) + 1.$$

Exact computations of the functions β and EN defined by (15) and (16) have been carried out (for a number of values of λ) when $P^* = .95$, $k = 2, 3, 4, 5$, and are presented in the accompanying table. The values of the fixed sample size $n^* = \sigma^2/c = h^2\lambda^2$ which would be used if σ^2 were known are included for comparison with EN.

We have no proof that the minimum value of the lower bound β of $P(\text{CS})$ is attained in the computed range $n^* = 2, \cdots, 80$, but this appears to be the case.

REMARKS. 1. The procedure discussed above is not fully sequential in the sense that it does not discard obviously unwanted populations (i.e. those with small means) as sampling proceeds. We are now studying a procedure based on results in [4] which has this desirable feature.

2. The method used in proving (10) can also be applied to obtain an upper bound on the expected sample size for the problems considered in [3], [5], [7], and [9].

REFERENCES

[1] BECHHOFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16–39.

[2] BECHHOFER, R. E., DUNNETT, C. W. and SOBEL, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41** 170–176.

[3] CHOW, Y. S. and ROBBINS, H. (1965). On the asympototic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 463–467.

[4] DARLING, D. A. and ROBBINS, H. (1967). Confidence sequences for mean, variance, and median. *Proc. Nat. Acad. Sci.* **58** 66–68.

[5] GLESER, L. J., ROBBINS, H. and STARR, N. (1964). Some asymptotic properties of fixed-width sequential confidence intervals for the mean of a normal population with unknown variance. Columbia Univ. Report.

[6] ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics—The Harald Cramér Volume*, 235–245. Almqvist and Wiksell, Uppsala.

[7] ROBBINS, H. and STARR, N. (1965). Remarks on sequential hypothesis testing. Univ. of Minnesota Tech. Report #68.

[8] SRIVASTAVA, M. S. (1966). Some asymptotically efficient procedures for ranking and slippage problems. *J. Roy. Statist. Soc. Ser. B* **28** 370–380.

[9] STARR, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.