The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 2198–2202

A COMPARISON OF THE ASYMPTOTIC EXPECTED SAMPLE SIZES OF TWO SEQUENTIAL PROCEDURES FOR RANKING PROBLEM

BY S. K. PERNG

Kansas State University

0. Summary. The purpose of this paper is to compare the asymptotic expected sample sizes of two sequential procedures for ranking k normal populations with known variance and unknown means for the cases (i) $\mu_1 \leqq \mu_2 \leqq \cdots \leqq \mu_{k-1} < \mu_k$ and (ii) $\mu_k - \mu_{k-1} = \delta^* > 0$. The procedures are: (1) the Bechhofer-Kiefer-Sobel (BKS) sequential procedure [1], and (2) Paulson's (P) sequential procedure [2].

1. Introduction. Let X_{ij} $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, N)$ be independent normally distributed random variables with population means μ_i and a common variance σ^2. The μ_i are assumed unknown and σ^2 is known. Let Π_i denote the population associated with μ_i, and $\mu_{[1]} \leqq \cdots \leqq \mu_{[k]}$ denote the ranked values of the μ_i. We would like to select the population associated with $\mu_{[k]}$. Two constants δ^*, P^* must be specified before the experimentation starts: (a) $\delta^* > 0$, the smallest value of the difference $\mu_{[k]} - \mu_{[k-1]}$ that the experimenter is interested in detecting (b) P^* the smallest acceptable value of the probability that the population associated with $\mu_{[k]}$ is selected.

Now we describe the two procedures. First, the BKS-procedure; at the mth stage of the experiment $(m = 1, 2, \cdots)$ take an observation from each of the k populations and compute the sample totals $Y_{im} = \sum_{j=1}^{m} X_{ij}$ $(i = 1, 2, \cdots, k)$ based on first m observations. Next compute the statistic

$$W_m = \sum_{i=1}^{k-1} \exp\{-\delta^* D_{im}/\sigma^2\}$$

where $D_{im} = Y_{[k]m} - Y_{[i]m}$ $(i = 1, 2, \cdots, k-1)$ and $Y_{[1]m} \leqq Y_{[2]m} \leqq \cdots \leqq Y_{[k]m}$ denote the ranked values of the Y_{im}. If $W_m \leqq (1 - P^*)/P^*$, stop experimentation and choose the population yielding $Y_{[k]m}$ as the one with the largest population mean. If $W_m > (1 - P^*)/P^*$, take another observation from each of the k populations and compute W_{m+1}. Continue in this manner until the rule calls for stopping.

Next, we describe P-procedure. The P-procedure is actually a family of procedures depending on a parameter λ that must be specified before starting experimentation. The P-procedure proceeds as follows. Choose a value λ $(0 < \lambda < \delta^*)$. At the first stage of the experiment take an observation from each of the k populations, and eliminate from any further consideration any population for which $Y_{[i]1} < Y_{[k]1} + \lambda - a$, where $a = [\sigma^2/(\delta^* - \lambda)] \log[(k-1)/(1 - P^*)]$. If all populations but one are eliminated at the first stage, stop the experiment and select the remaining population as the best one. Otherwise proceed to the second stage and take one observation from each of the populations not eliminated.

Received 10 January 1969.

Thus at the nth stage take one observation from each of the populations not eliminated after the $(n-1)$th stage and eliminate from further consideration any population Π_i for which

$$\max_j \{Y_{jn} - Y_{in}\} > a - n\lambda$$

where max is taken over all populations left after the $(n-1)$ stage. If only one population is left after the nth stage, the experiment is terminated and the remaining population is selected. Otherwise, go on to the $(n+1)$th stage. If more than one population remains after the W_λ stage, where W_λ equals the largest integer less than a/λ, the experiment is terminated at the $(W_\lambda + 1)$ stage by selecting the remaining population for which the sum of the $W_\lambda + 1$ observations is a maximum.

Both sequential procedures satisfy the following probability requirement:

$$P[\Pi_{[k]} \text{ is selected} \mid \mu_{[k]} - \mu_{[k-1]} \geqq \delta^*] \geqq P^*$$

where $\Pi_{[k]}$ is the population associated with $\mu_{[k]}$.

2. Results. We now give the asymptotic total expected sample size for both sequential procedures. First we quote a result given in [1, page 161] concerning the total expected stages of BKS-procedure.

THEOREM 2.1. (Bechhofer-Kiefer-Sobel). *Under the assumptions and with the notation of* BKS-*procedure in Section* 1, *if* $\mu_1 \leqq \mu_2 \leqq \cdots \leqq \mu_{k-1} < \mu_k$ *and* $\mu_k - \mu_{k-1} = \delta^* > 0$, *then the expected number of stages needed to terminate the experiment is*

$$E(N) = \sigma^2(\delta^*)^{-2} \log (1 - P^*)^{-1} + o(\log (1 - P^*))$$

as $P^* \to 1$.

Next we give the expected sample size needed to eliminate Π_i of P–procedure for $i = 1, 2, \cdots k - 1$. Define N_i to be the smallest $n \geqq 1$ such that

$$\max_{j\varepsilon I_{n-1}} (Y_{jn} - Y_{in}) > a - n\lambda$$

where I_{n-1} is the set of all populations which have not been eliminated after the $(n-1)$th stage. (i.e. N_i equals the number of observations needed to eliminate Π_i).

THEOREM 2.2. *Under the assumptions and notation of the* P-*procedure in Section* 1, *if* $\mu_1 \leqq \mu_2 \leqq \cdots \leqq \mu_{k-1} < \mu_k$ *and* $\mu_k - \mu_{k-1} = \delta^* > 0$, *then*

$$E(N_i) = \sigma^2(\lambda + \mu_k - \mu_i)^{-1}(\delta^* - \lambda)^{-1} \log [(k-1)/(1-P^*)] + o(\log [(k-1)/(1-P^*)])$$

as $P^* \to 1$, *for* $i = 1, \cdots, k-1$.

PROOF. Let $(\Omega, \mathcal{B}, P)$ be the underlying probability space and let

$$\Omega_0 = \{\omega \mid \Pi_k \text{ is selected}\}, \tag{1}$$

$$\Omega_1 = \Omega - \Omega_0. \tag{2}$$

Throughout the proof, i is fixed and $j = 1, 2, \cdots, k$, unless stated otherwise. By the Strong Law of Large Numbers the random variable $(Y_{jn} - Y_{in})/n$ converges to $(\mu_j - \mu_i)$ with probability one. On the set Ω_0 and by Egoroff's theorem for every $\delta > 0$ there exist disjoint sets A_j, B_j such that $A_j \cup B_j = \Omega_0$, $P(B_j) < \delta$ and $(Y_{jn} - Y_{in})/n$ converges uniformly on A_j. Hence there exists a positive integer $n_j(\delta)$ such that for every $\omega \varepsilon A_j$ and $n > n_j(\delta)$

$$n(\mu_j - \mu_i - \delta) \leqq Y_{jn} - Y_{in} \leqq n(\mu_j - \mu_i + \delta). \tag{3}$$

Let $n_0(\delta) = \max_j n_j(\delta)$ and

$$B = \textstyle\bigcup_{j=1}^{k} B_j, \qquad \Omega_2 = \Omega_0 - B. \tag{4}$$

Then for every $n > n_0(\delta)$ and every $\omega \varepsilon \Omega_2$

$$n(\mu_j - \mu_i - \delta) \leqq Y_{jn} - Y_{in} \leqq n(\mu_j - \mu_i + \delta). \tag{5}$$

Hence

$$\max_j n(\mu_j - \mu_i - \delta) \leqq \max_j (Y_{jn} - Y_{in}) \leqq \max_j n(\mu_j - \mu_i + \delta) \tag{6}$$

where max is taken over all populations left after the $(n - 1)$th stage. Since, on the set Ω_2, Π_k is not eliminated at any stage and $\mu_k \geqq \mu_i$ for all i, so for every $n > n_0(\delta)$ and every $\omega \varepsilon \Omega_2$ we have

$$n(\mu_k - \mu_i - \delta) \leqq \max_j (Y_{jn} - Y_{in}) \leqq n(\mu_k - \mu_i + \delta). \tag{7}$$

Hence on Ω_2, we have either

$$N_i \leqq n_0(\delta), \tag{8}$$

or

$$a(\lambda + \mu_k - \mu_i + \delta)^{-1} \leqq N_i \leqq a(\lambda + \mu_k - \mu_i - \delta)^{-1} + 1 \tag{9}$$

This follows from the definition of N_i and (7). However, for a given $\delta > 0$, $n_0(\delta)$ is fixed. And according to the definition of a, $a \to \infty$ as $P^* \to 1$ hence we can make $a/(\lambda + \mu_k - \mu_i + \delta) \geqq n_0(\delta)$ by taking P^* close to one. Thus, on Ω_2, we have either

$$N_i \leqq n_0(\delta), \tag{10}$$

or

$$n_0(\delta) \leqq a(\lambda + \mu_k - \mu_i + \delta)^{-1} \leqq N_i \leqq a(\lambda + \mu_k - \mu_i - \delta)^{-1} + 1 \tag{11}$$

as P^* is close to one. We define

$$\Omega_2^* = \{\omega \mid \omega \varepsilon \Omega_2 \text{ and } N_i(\omega) \leqq n_0(\delta)\}. \tag{12}$$

Now we are ready to find the asymptotic expected sample size of eliminating Π_i. Clearly

$$E(N_i) = \int_{\Omega_1} N_i(\omega)\, dP(\omega) + \int_{\Omega_2^*} N_i(\omega)\, dP(\omega) + \int_{\Omega_2 - \Omega_2^*} N_i(\omega)\, dP(\omega) + \int_B N_i(\omega)\, dP(\omega). \tag{13}$$

Since $0 \leqq N_i < a/\lambda$ on Ω, $P(\Omega_1) \leqq 1 - P^*$ and $P(B) \leqq k\delta$, hence for P^* sufficiently close to one and using (10), (11), (13), we have

$$(14)\quad a(\lambda + \mu_k - \mu_i + \delta)^{-1}P(\Omega_2 - \Omega_2^*) \leqq E(N_i) \leqq a\lambda^{-1}(1 - P^*) + a(\lambda + \mu_k - \mu_i - \delta)^{-1} + 1 + a\lambda^{-1}k\delta.$$

Dividing all terms in (14) by $\log[(k-1)/(1-P^*)]$ and taking limits as $P^* \to 1$, we have

$$(15)\quad \begin{aligned} &\sigma^2(\delta^* - \lambda)^{-1}(\lambda + \mu_k - \mu_i + \delta)^{-1}(1 - k\delta)^{-1} \\ &\qquad \leqq \lim_{P^* \to 1}\{E(N_i)/\log[(k-1)/(1-P^*)]\} \\ &\qquad \leqq \sigma^2(\delta^* - \lambda)^{-1}((\lambda + \mu_k - \mu_i - \delta)^{-1} + k\delta\lambda^{-1}). \end{aligned}$$

This follows from the definition of a and the fact $\lim_{P^* \to 1} P(\Omega_2 - \Omega_2^*) = 1 - P(B) \geqq 1 - k\delta$. Since δ is arbitrary in (15), we have

$$(16)\quad \begin{aligned} &\sigma^2(\delta^* - \lambda)^{-1}(\lambda + \mu_k - \mu_i)^{-1} \\ &\qquad \leqq \lim_{P^* \to 1} E(N_i)/\log[(k-1)/(1-P^*)] \\ &\qquad \leqq \sigma^2(\delta^* - \lambda)^{-1}(\lambda + \mu_k - \mu_i)^{-1}. \end{aligned}$$

This completes the proof.

COROLLARY. *Under the assumptions and the notation given in Theorem 2.2, except $\mu_i < \mu_{i+1}$ for $i = 1, 2, \cdots, k-1$; then*

$$(17)\quad P\{N_1 < N_2 < \cdots < N_{k-1}\} \to 1$$

as $P^* \to 1$.

We omit the proof.

Using Theorem 2.1, we have the asymptotic expected total number of observations of BKS-procedure:

$$(18)\quad E(T_1) = (k\sigma^2(\delta^*)^{-2}\log(1 - P^*)^{-1} + o(\log(1 - P^*))$$

as $P^* \to 1$.

And using Theorem 2.2, we have the asymptotic expected total number of observations of P-procedure:

$$(19)\quad \begin{aligned} E(T_2) &= \sigma^2(\delta^* - \lambda)^{-1}\log[(k-1)/(1-P^*)] \\ &\quad \cdot [\textstyle\sum_{i=1}^{k-2}(\lambda + \mu_k - \mu_i)^{-1} + 2(\lambda + \delta^*)^{-1}] \\ &\quad + o(\log[(k-1)/(1-P^*)] \end{aligned}$$

as $P^* \to 1$.

If we define the asymptotic relative efficiency to be the ratio of asymptotic expected total number of observations of two procedures, then the asymptotic relative efficiency of the two procedures is:

$$(20)\quad \lim_{P^* \to}[E(T_1)/E(T_2)] = k(\delta^* - \lambda)(\delta^*)^{-2}[\textstyle\sum_{i=1}^{k-2}(\lambda + \mu_k - \mu_i)^{-1} + 2(\lambda + \delta^*)^{-1}]^{-1}.$$

If $k = 2$, or $\mu_1 = \mu_2 = \cdots = \mu_{k-1} < \mu_k$ then (20) becomes

$$\lim_{P^*\to 1} E(T_1)/E(T_2) = [(\delta^*)^2 - \lambda^2](\delta^*)^{-2} < 1. \tag{21}$$

From (21) we may conclude that if $k = 2$, or $\mu_i = \mu_{i+1}$ for $i = 1, \cdots, k-2$; then the BKS-procedure is more efficient asymptotically. From (20) we may conclude

(a) if $\sum_{i=1}^{k-2} (\lambda + \mu_k - \mu_i)^{-1} < k(\delta^* - \lambda)(\delta^{*2})^{-1} - 2(\lambda + \delta^*)^{-1}$ then the P-procedure is more efficient asymptotically.

(b) $\lim_{P^*\to 1} E(T_1)/E(T_2)$ is a decreasing function of λ.

Intuitively speaking these imply that (1) P-procedure is better when k and the differences $(\mu_k - \mu_i)$ are large, (2) to maximize the asymptotic relative efficiency of the P-procedure, we should take a small value for λ.

Acknowledgment. The author is deeply grateful to Professor V. Fabian for his guidance.

REFERENCES

[1] Bechhofer, R. E., Kiefer, J., and Sobel, M. (1968). *Sequential Identification and Ranking Procedures*. Univ. of Chicago Press.

[2] Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Statist.* **35** 174–180.