Ann. Inst. Statist. Math.
29 (1977), Part A, 59-66

ON CORRECT SELECTION FOR A RANKING PROBLEM

W. K. CHIU

(Received Dec. 17, 1973; revised June 9, 1975)

Abstract

Given k populations which belong to the exponential class, and having specified two positive constants (δ^*, P^*), an experimenter wishes to select t populations which (a) exclude all those populations with a parameter value not greater than the tth largest parameter value minus δ^*, and (b) include all those populations with a parameter value not smaller than the $(t+1)$th largest parameter value plus δ^*. This paper shows that the probability of successfully making such a selection, called a δ^*-correct selection, is at least P^* if the basic sequential procedure, P_B^*, of Bechhofer, et al. [3] is used. This result includes the corresponding old result of their book (p. 129) as a particular case.

1. Introduction

Selecting the best t from among k populations is one of the most important ranking problems and has wide practical application. Bechhofer, Kiefer and Sobel ([3], p. 255) have formulated the problem in the following way: An experimenter is given k ($\geqq 2$) populations $\pi_1, \cdots, \pi_k$, whose density functions are members of the same univariate exponential (Koopman-Darmois) family:

$$f(x_i)=\exp\{P(x_i)\tau_i+R(x_i)+V(\tau_i)\} \qquad (i=1,\cdots,k)$$

where $P(x)$, $R(x)$, and $V(\tau)$ are known functions. The populations are assumed to differ at most in the unknown parameter values τ_i ($i=1, \cdots, k$). Let the ranked τ-values be denoted by

$$\tau_{[1]}\leqq\tau_{[2]}\leqq\cdots\leqq\tau_{[k]}$$

and the differences between any pair of them be denoted by:

$$\delta_{i,j}=\tau_{[i]}-\tau_{[j]} \qquad (i\geqq j;\ i, j=1,\cdots,k)\,.$$

It is assumed that the true pairing of π_i with $\tau_{[j]}$ ($i, j=1,\cdots,k$)

is unknown to the experimenter, who is interested in selecting an unordered set of t ($\geqq 1$) 'best' populations, which are associated with the t largest parameter values $\tau_{[k-t+1]}, \cdots, \tau_{[k]}$. Before the experimentation, the experimenter has to specify a pair of constants, as below:

$$(1) \qquad \{\delta^*, P^*\}, \text{ with } 0<\delta^*<\infty \text{ and } 1\Big/\binom{k}{t}<P^*<1 ,$$

with the intention to achieve the following probability requirement:

$$(2) \qquad \Pr\{\text{Correct selection} \mid \delta_{k-t+1,k-t} \geqq \delta^*\} = P^* .$$

Bechhofer, et al. [3] have proposed an interesting sequential sampling procedure, P_B^*, and have proved that their procedure guarantees the probability requirement (2) (p. 129). We shall describe the procedure in the next section. Two outstanding features of the procedure are: (a) The same procedure is applicable to a large number of distributions, including the normal, the exponential, Poisson, Bernoulli, and the negative binomial distributions, which belong to the exponential family; (b) The average sample number is, in general, considerably smaller than the sample size for a single-sample procedure.

However, the formulation of the problem has a fundamentally restrictive character in that its goal as expressed in (2) requires the condition: $\delta_{k-t+1,k-t} \geqq \delta^*$. In practice, the experimenter would be unlikely to know that this can be satisfied. If $\delta_{k-t+1,k-t}<\delta^*$, then he does not know what alternate probability the sequential procedure can guarantee him. Further, in specifying the constants $\{\delta^*, P^*\}$ the experimenter usually has some economic consideration in mind so that δ^* is the smallest difference in parameter values that he wishes to detect, and he is indifferent to incorrect selections made in the zone $\tau_{[k-t+1]}-\delta^* <\tau_{[k-t]}+\delta^*$. Indeed he must be more interested in the probability of making such a selection than the conditional probability of (2).

Bechhofer, et al. [3] are well aware of these conditions and have proposed a related 'unsolved problem' in their book (p. 337), which we will restate with minor modifications. For any configuration of the parameter values, denoted as a k-vector: $\vec{\tau}=(\tau_{[1]}, \cdots, \tau_{[k]})$, there are unique integers r and s satisfying $0 \leqq r \leqq k-t$, $0 \leqq s \leqq t$, and such that

$$(3) \qquad \tau_{[r]} \leqq \tau_{[k-t+1]}-\delta^* < \tau_{[r+1]} \quad \text{and} \quad \tau_{[k-s]} < \tau_{[k-t]}+\delta^* \leqq \tau_{[k-s+1]} .$$

Here it is understood that $\tau_{[0]}=-\infty$ and $\tau_{[k+1]}=\infty$. Thus there are r populations with τ-values $\leqq \tau_{[k-t+1]}-\delta^*$, and s populations with τ-values $\geqq \tau_{[k-t]}+\delta^*$. In this situation, the experimenter might reasonably still want the t selected populations (a) to exclude all those with a parameter value not greater than the tth largest parameter value minus δ^*, namely, the r populations corresponding to the parameter values $\tau_{[1]}$,

$\cdots, \tau_{[r]}$, and (b) to include all those with a parameter value not smaller than the $(t+1)$th largest parameter value plus δ^*, namely, the s populations corresponding to the parameter values $\tau_{[k-s+1]}, \cdots, \tau_{[k]}$. To clarify some notational problem, we should note that when $r=0$ the experimenter would not specifically intend to exclude any population because then every population mean is within δ^* from $\tau_{[k-t+1]}$. Likewise, when $s=0$ the experimenter would not specifically intend to include any particular population because then no population mean is large enough to exceed $\tau_{[k-t]}+\delta^*$. Let us call any such selection a 'δ^*-correct selection'. (Bechhofer, et al. have not named it for the $\{\delta^*, P^*\}$-approach, although they have called it a 'correct decision' for a decision theoretic approach, p. 47.) Bechhofer, et al. conjecture that their sequential procedure, P_B^*, still guarantees that the probability of making a δ^*-correct selection is at least P^*. This paper proves their conjecture.

2. The sequential procedure P_B^*

We shall refer to pp. 125 and 256 of Bechhofer, et al. [3] for the description of the sequential procedure P_B^*. At each stage of sampling, one observation is taken from each of the k populations. Let X_{ij} denote the chance variable corresponding to the jth observation taken from the populations π_i $(i=1,\cdots,k;\ j=1,2,\cdots)$. They are assumed to be mutually independent. At the mth stage of experimentation $(m=1,2,\cdots)$ the values, y_{im}, of the sufficient statistics:

$$Y_{im}=\sum_{j=1}^{m} P(X_{ij}) \qquad (i=1,\cdots,k)$$

will be used for decision making. Denoting by $y_{[1]m}\leqq\cdots\leqq y_{[k]m}$ and $Y_{[1]m}\leqq\cdots\leqq Y_{[k]m}$ the ranked y_{im} and the ranked Y_{im} $(i=1,\cdots,k;\ m=1,2,\cdots)$, respectively, we may introduce two k-vectors:

$$\vec{y}_m=(y_{[1]m},\cdots,y_{[k]m}), \qquad \vec{Y}_m=(Y_{[1]m},\cdots,Y_{[k]m}).$$

Let C denote the set of $k!/\{(k-t)!t!\}$ ways of choosing t integers from the set $\{1,2,\cdots,k\}$. Any element of C can be represented by $\gamma=\{\gamma_1,\cdots,\gamma_t\}$ where γ_i $(i=1,\cdots,t)$ are different positive integers not larger than k. The particular element $\{k-t+1,\cdots,k\}$ will be denoted by γ_e. We define, for any $\gamma\in C$, the product:

$$\gamma\vec{y}_m=\sum_{i=1}^{t} y_{[\gamma_i]m}.$$

The rules for the procedure P_B^* are as follows. After the mth stage of observation, the experimenter calculates the quantity

$$(4)\qquad Q^*(\vec{y}_m)=\frac{\exp\{\delta^*\gamma_e\vec{y}_m\}}{\sum_{\gamma\in C}\exp\{\delta^*\gamma\vec{y}_m\}}\,.$$

If $Q^*(\vec{y}_m)<P^*$, he continues to the $(m+1)$th stage of the experimentation. If $Q^*(\vec{y}_m)\geqq P^*$, he terminates the experimentation and selects the t populations associated with the t largest observed values $y_{[k-t+1]m},\cdots,y_{[k]m}$. Bechhofer, et al. ([3], p. 129) have proved that P_B^* guarantees the probability of a correct selection to be not smaller than P^* whenever $\delta_{k-t+1,k-t}\geqq\delta^*$, namely,

$$(5)\qquad \Pr\{\text{Correct selection}\,|\,\delta_{k-t+1,k-t}\geqq\delta^*\}=\mathrm{E}\,\{Q^*(\vec{Y}_n)\}\geqq P^*\,,$$

where $\vec{Y}_n$ is the terminal decision vector for the procedure.

3. δ^*-correct selection

Let us consider the general configuration $\vec{\tau}=(\tau_{[1]},\cdots,\tau_{[k]})$ in which $\delta_{k-t+1,k-t}$ is not necessarily $\geqq\delta^*$, as explained in Section 1. Our goal is to select t populations which exclude all those populations with parameter values $\leqq\tau_{[k-t+1]}-\delta^*$ and include all those populations with parameter values $\geqq\tau_{[k-t]}+\delta^*$. Let S_k be the set (which is in fact a symmetric group) of $k!$ permutations on the integers $1,2,\cdots,k$. If $\alpha=(\alpha_1,\cdots,\alpha_k)$ is any element of S_k we define $\alpha\vec{y}_m=(y_{[\alpha_1]m},\cdots,y_{[\alpha_k]m})$, which is a k-vector; also, we define the inner product

$$\vec{\tau}\cdot\alpha\vec{y}_m=\sum_{i=1}^{k}\tau_{[i]}y_{[\alpha_i]m}\,.$$

Let two subsets of S_k be defined as follows:

$$S=\{\alpha:\ \alpha_i>k-t\ \text{for some}\ i\leqq r\ \text{or}\ \alpha_j\leqq k-t\ \text{for some}\ j\geqq k-s+1\}\,,$$

$$T=\{\alpha:\ \alpha_i\leqq k-t\ \text{for all}\ i\leqq r\ \text{and}\ \alpha_j>k-t\ \text{for all}\ j\geqq k-s+1\}\,.$$

Clearly, S and T are disjoint and they exhaust S_k. Assume that procedure P_B^* is used for the selection. Following the arguments of Sections 4.2, 5.4, and 6.1.1 of Bechhofer, et al. [3] we have the following expression:

$$(6)\qquad \Pr\{\delta^*\text{-correct selection}\}=\mathrm{E}\,\{Q(\vec{Y}_n)\}\,,$$

where $\vec{Y}_n$ is the terminal decision vector of the procedure and the quantity $Q(\vec{y}_m)$, at the mth stage of sampling, is defined to be:

$$(7)\qquad Q(\vec{y}_m)=\frac{\sum_{\alpha\in T}\exp\{\vec{\tau}\cdot\alpha\vec{y}_m\}}{\sum_{\alpha\in S_k}\exp\{\vec{\tau}\cdot\alpha\vec{y}_m\}}\,.$$

We are now in a position to prove the following theorem.

THEOREM. *Let two constants* $\{\delta^*, P^*\}$ *be specified as in* (1), *and let* $\vec{\tau}=(\tau_{[1]},\cdots,\tau_{[k]})$ *be any configuration of the parameter values. For the selection of* t *populations which exclude all those populations with a* τ-*value* $\leqq\tau_{[k-t+1]}-\delta^*$ *and include all those populations with a* τ-*value* $\geqq\tau_{[k-t]}+\delta^*$ *by the sequential procedure* P_B^*, *we have*:

$$\Pr\{\delta^*\text{-correct selection}\}\geqq P^* . \tag{8}$$

PROOF. Let the integers r and s be defined as in (3). The theorem is obviously true if $r=s=0$, because then the selection of *any* t populations is δ^*-correct (see the notational clarification at the end of the introductory section) yielding $\Pr\{\delta^*\text{-correct selection}\}=1$. Thus without loss of generality, we may assume that $r\geqq1$. Because of (5) and (6), we only need to prove, without considering any stopping rule, that the inequality

$$Q(\vec{y}_m)\geqq Q^*(\vec{y}_m)$$

holds for any possible value $\vec{y}_m$ of $\vec{Y}_m$ at the mth stage of sampling $(m=1,2,\cdots)$, where $Q^*(\vec{y}_m)$ and $Q(\vec{y}_m)$ are given in (4) and (7), respectively. After subtracting 1 from the reciprocal of each side and rearranging the terms, we have the following equivalent inequality:

$$\exp\{\delta^*\gamma_e\vec{y}_m\}\Big(\sum_{\alpha\in S}\exp\{\vec{\tau}\cdot\alpha\vec{y}_m\}\Big)\leqq\Big(\sum_{\gamma\in D}\exp\{\delta^*\gamma\vec{y}_m\}\Big)\Big(\sum_{\alpha\in T}\exp\{\vec{\tau}\cdot\alpha\vec{y}_m\}\Big) \tag{9}$$

where $D=C-\{\gamma_e\}$. To establish (9) we shall show that for every term, $\exp\{\delta^*\gamma_e\vec{y}_m+\vec{\tau}\cdot\alpha\vec{y}_m\}$, on the left-hand side, we can uniquely define an equal or larger term, $\exp\{\delta^*\gamma'\vec{y}_m+\vec{\tau}\cdot\alpha'\vec{y}_m\}$, on the right-hand side. In other words, we shall show that:

$$\delta^*\gamma_e\vec{y}_m+\vec{\tau}\cdot\alpha\vec{y}_m\leqq\delta^*\gamma'\vec{y}+\vec{\tau}\cdot\alpha'\vec{y}_m , \tag{10}$$

where $(\alpha',\gamma')=g(\alpha)$ is the image of $\alpha\in S$, under an injective mapping $g: S\rightarrow T\times D$, to be defined below.

Consider any $\alpha\in S$. Let p be the total number of components which have a value $>k-t$, among the first r components of α. If $p=0$, we define $a_0=0$, $b_0=k-s$, and go to step (A) to continue our arguments therefrom. If $p\geqq1$, then there are p integers $a_1,\cdots,a_p$ such that

$$\alpha_{a_i}>k-t \qquad (i=1,\cdots,p;\ 1\leqq a_1<\cdots<a_p\leqq r) .$$

Among the last t components of α, there must exist a set of p smallest possible integers $b_1,\cdots,b_p$ such that

$$\alpha_{b_i}\leqq k-t \qquad (i=1,\cdots,p;\ k-t+1\leqq b_1<\cdots<b_p\leqq k) .$$

Step (A). Let q be the total number of components which have a value $\leqq k-t$, among the last $\min\{s, k-b_p\}$ components of α. If $q=0$, then we go to step (B) and continue our arguments therefrom. If $q\geqq 1$, then there are q integers $b_{p+1},\cdots,b_{p+q}$ such that

$$\alpha_{b_i}\leqq k-t \qquad (i=p+1,\cdots,p+q;\ \max\{k-s, b_p\}<b_{p+1}<\cdots<b_{p+q}\leqq k).$$

There must exist a set of q smallest possible integers $a_{p+1},\cdots,a_{p+q}$ such that

$$\alpha_{a_i}>k-t \qquad (i=p+1,\cdots,p+q;\ a_p<a_{p+1}<\cdots<a_{p+q}\leqq k-t).$$

Step (B). We may now define an $\alpha'\in T$ corresponding to α, in the following manner:

$$\alpha'_{a_i}=\alpha_{b_i},\quad \alpha'_{b_i}=\alpha_{a_i},\qquad i=1,\cdots,p+q;$$
$$\alpha'_j=\alpha_j,\qquad \text{for all other } j\leqq k.$$

Three facts may be noticed. First, $p+q\geqq 1$, otherwise α would not have been in S. Secondly, there are as many (in fact, $p+q$) components in α which are different from those in α' among the first $k-t$ components as there are among the last t components. Thirdly, $\tau_{[b_i]}-\tau_{[a_i]}\geqq\delta^*$ because of (3), and $y_{[\alpha_{a_i}]m}\geqq y_{[\alpha_{b_i}]m}$ because $\alpha_{a_i}>\alpha_{b_i}$ $(i=1,\cdots,p+q)$.

We may further define a $\gamma'\in D$, corresponding to α, by

$$(11)\quad \gamma'=\{\gamma'_1,\cdots,\gamma'_t\}=\{k-t+1,\cdots,k,\alpha_{k-t+1},\cdots,\alpha_k\}-\{\alpha'_{k-t+1},\cdots,\alpha'_k\}.$$

Thus, a mapping $g: S\to T\times D$ has been defined by $g(\alpha)=(\alpha',\gamma')$. At this point, it may be profitable to give an example. Suppose $k=10$, $t=5$, $r=2$, and $s=3$. If $\alpha=(9, 5, 6, 10, 8, 2, 3, 1, 7, 4)$, then $\alpha'=(2, 5, 1, 4, 8, 9, 3, 6, 7, 10)$ and $\gamma'=\{1, 2, 4, 7, 8\}$.

To show that g is injective, consider any (α',γ') which is in the image set, $g(S)$, of g. We shall find a unique $\alpha\in S$ such that $g(\alpha)=(\alpha',\gamma')$. Hinted by (11), we have

$$\{\alpha_{k-t+1},\cdots,\alpha_k\}=\{\gamma'_1,\cdots,\gamma'_t,\alpha'_{k-t+1},\cdots,\alpha'_k\}-\{k-t+1,\cdots,k\}.$$

Let $\{\alpha'_{b_i}: i=1,\cdots,h;\ k-t+1\leqq b_1<\cdots<b_h\leqq k\}$ be the set of all integers in the set $\{\alpha'_{k-t+1},\cdots,\alpha'_k\}$ which are not found in the set $\{\alpha_{k-t+1},\cdots,\alpha_k\}$. Then the elements of the set

$$\{\alpha_{k-t+1},\cdots,\alpha_k\}-(\{\alpha'_{k-t+1},\cdots,\alpha'_k\}-\{\alpha'_{b_1},\cdots,\alpha'_{b_h}\})$$

must appear in the first $k-t$ components of α' with their positions identified. Suppose these are $\alpha'_{a_1},\cdots,\alpha'_{a_h}$, $(1\leqq a_1<\cdots<a_h\leqq k-t)$. We may now define α in a unique manner:

$$\alpha_{a_i}=\alpha'_{b_i}\,,\quad \alpha_{b_i}=\alpha'_{a_i}\,,\qquad i=1,\cdots,h\,;$$

$$\alpha_j=\alpha'_j\,,\qquad\qquad \text{for all other } j\leqq k\,.$$

Obviously $g(\alpha)=(\alpha',\gamma')$, and the injectiveness of g is proved.

Finally, we may prove (10), which is equivalent to

$$\delta^*(\gamma_e\vec{y}_m-\gamma'\vec{y}_m)\leqq\vec{\tau}\cdot(\alpha'\vec{y}_m-\alpha\vec{y}_m)\,,$$

which is further equivalent to

$$\sum_{i=1}^{p+q}(y_{[\alpha_{a_i}]m}-y_{[\alpha_{b_i}]m})\delta^*\leqq\sum_{i=1}^{p+q}(y_{[\alpha_{a_i}]m}-y_{[\alpha_{b_i}]m})(\tau_{[b_i]}-\tau_{[a_i]})\,.$$

Now, this last inequality is true because

$$\delta^*\leqq\tau_{[b_i]}-\tau_{[a_i]}\quad\text{and}\quad 0\leqq y_{[\alpha_{a_i}]m}-y_{[\alpha_{b_i}]m}\qquad (i=1,\cdots,p+q)\,.$$

This completes the proof.

4. Discussion

The result of the theorem in the last section is interesting because (i) the experimenter need not modify his sampling procedure, but just uses P_B^*, and (ii) when $\delta_{k-t+1,k-t}\geqq\delta^*$, result (8) automatically reduces to result (5).

A similar situation exists for the ranking problem of selecting the best t from among k *normal* populations, using a single-sample procedure (e.g., Bechhofer, [1]) or a two-sample procedure (Bechhofer, et al. [2]; Maurice, [5]; Ofosu, [6]). A result analogous to (8) has been obtained by Chiu [4] using entirely different mathematical techniques.

Acknowledgement

The author is grateful to Dr. C. Chatfield for some helpful suggestions and to Dr. J. B. Ofosu for some interesting discussion. The work was supported by the Commonwealth Scholarship Commission in Britain when the author was visiting the University of Bath.

UNIVERSITY OF HONG KONG

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] Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). *Sequential Identification and Ranking Procedures*, University of Chicago Press, Chicago.
[4] Chiu, W. K. (1974). The ranking of means of normal populations for a generalized selection goal, *Biometrika*, **61**, 579-584.
[5] Maurice, R. J. (1958). Ranking means of two normal populations with unknown variances, *Biometrika*, **45**, 250-252.
[6] Ofosu, J. B. (1973). A two-sample procedure for selecting the population with the largest mean from several normal populations with unknown variances, *Biometrika*, **60**, 117-124.

Ann. Inst. Statist. Math.
30 (1978), Part A, 477

ACKNOWLEDGEMENT OF PRIORITY

W. K. CHIU

The result of the paper "On correct selection for a ranking problem" (this Annals **29**(1977), A, pp. 59–66) was proved by a different method in Dr. S. Sievers' unpublished Ph.D. dissertation and was abstracted in Annals of Mathematical Statistics **42**(1971), p. 1795, "A solution to an open problem of Bechhofer, Kiefer, and Sobel." The author is grateful to Dr. S. Sievers for drawing his attention to this priority.

UNIVERSITY OF HONG KONG