ON MULTIPLE DECISION METHODS FOR RANKING POPULATION MEANS

By Václav Fabian

Research Institute for Mathematical Machines, Prague

1. Summary. (i) In the case of the choice of the largest among k population means (special case $\mathfrak{A}$ of the use of ranking methods) the assertion of Bechhofer's method ([1], [2], [3]) can be strengthened without decreasing the probability of a correct decision (Sec. 3).

(ii) Bechhofer's concept of the "least favorable configuration of the population means" is studied (Sec. 4). The result suggests that the concept is not always in accord with the underlying practical problem, but that it is in accord in the important case $\mathfrak{A}$.

(iii) An approximation is suggested for use in the case of normal populations with a common unknown variance (Sec. 5).

2. Introduction. R. E. Bechhofer, in his pioneering paper [1], pointed out plainly the inappropriateness of testing traditional null hypotheses in ranking problems and stated the basic concepts of his multiple decision ranking methods. These methods for ranking, or partially ranking, a group of populations on the basis of an experiment are of great practical importance especially in connection with the problem of selecting the best from a set of possible alternatives. It is because of this importance that the following mathematically simple remarks are offered.

Now we shall give a mathematical formulation of the problem which includes the situations studied in [1], [2] and [3].

Let $I = \{1, 2, \cdots, k\}$, let $\xi_i (i \varepsilon I)$ be random variables satisfying the following condition of permutability:

$$P\{[\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}] \varepsilon A\} = P\{[\xi_1, \xi_2, \cdots, \xi_k] \varepsilon A\}, \tag{2.1}$$

for every permutation $i_1, i_2, \cdots, i_k$ of $1, 2, \cdots, k$ and every k-dimensional Borel set A. Suppose the ξ_i's are continuous.

For every $\mu = [\mu_1, \mu_2, \cdots, \mu_k]$ in the k-dimensional Euclidean space E_k let us define numbers $[i, \mu] \varepsilon I$, random variables $X_{i,\mu}$ and random variables $\langle i, \mu \rangle$ with values in I, such that

$$\mu_{[1,\mu]} \leqq \mu_{[2,\mu]} \leqq \cdots \leqq \mu_{[k,\mu]}, \qquad [i, \mu] \neq [j, \mu] \qquad \text{for } i \neq j; \tag{2.2}$$

$$X_{i,\mu} = \xi_i + \mu_i, \qquad i \varepsilon I, \tag{2.3}$$

$$X_{\langle 1,\mu \rangle,\mu} \leqq X_{\langle 2,\mu \rangle,\mu} \leqq \cdots \leqq X_{\langle k,\mu \rangle,\mu}, \qquad \langle i, \mu \rangle \neq \langle j, \mu \rangle \qquad \text{for } i \neq j. \tag{2.4}$$

When there is no danger of confusion, we shall use the abbreviated symbols $[i]$, X_i, $\langle i \rangle$ for $[i, \mu]$, $X_{i,\mu}$, $\langle i, \mu \rangle$.

Received July 2, 1960; revised August 15, 1961.

Now let $0 = k_0 < k_1 < \cdots < k_s = k$ be given numbers. By a ranking we mean a sequence

$$(2.5) \qquad V = [V_1, V_2, \cdots, V_s],$$

where V_α are disjoint subsets of I containing $k_\alpha - k_{\alpha-1}$ elements respectively.

The experimenter who can study μ by observing only the X_i's, wishes to determine a correct ranking, i.e., a ranking V satisfying the following implication:

$$(2.6) \qquad i \,\varepsilon\, V_\alpha, j \,\varepsilon\, V_{\alpha+1}, 1 \leqq \alpha < s \Rightarrow \mu_i \leqq \mu_j .$$

Obviously the ranking $Z = [Z_1, Z_2, \cdots, Z_s]$ defined by

$$(2.7) \qquad Z_\alpha = \{[k_{\alpha-1} + 1], [k_{\alpha-1} + 2], \cdots, [k_\alpha]\}$$

is correct.

Since, however, the experimenter does not know the $[i]$'s, it is natural to substitute for them the random variables $\langle i \rangle$ and choose the ranking S defined by

$$(2.8) \qquad S_\alpha = \{\langle k_{\alpha-1} + 1\rangle, \cdots, \langle k_\alpha \rangle\} \qquad \text{for } \alpha = 1, 2, \cdots, s.$$

Example 1. Suppose $k = 3$, $s = 3$, $k_0 = 0$, $k_1 = 1$, $k_2 = 2$, $k_3 = 3$, $\mu = [1, 0, 10]$. Then $[1] = 2$, $[2] = 1$, $[3] = 3$, since $\mu_{[1]} = \mu_2 = 0 < \mu_{[2]} = \mu_1 = 1 < \mu_{[3]} = \mu_3 = 10$. There are six different rankings; among these only one, namely $Z = [\{2\}, \{1\}, \{3\}]$, is correct. The random variables $\langle i \rangle$ are defined by (2.4) almost everywhere since the ξ_i's are supposed to be continuous. If, for example, the observed values of X_1, X_2, X_3 are 3, 2, 8, the observed values of $\langle 1 \rangle$, $\langle 2 \rangle$, $\langle 3 \rangle$ are 2, 1, 3 and the (non-numerical) value of the random ranking S is $[\{2\}, \{1\}, \{3\}]$.

After defining S by (2.8) the question arises of the probability of a correct decision, i.e., of the probability that S is a correct ranking. For a fixed goal, i.e., for fixed s, $k_0, \cdots, k_s$ and for a fixed probability distribution F of the random variables ξ_i, this probability is a function of μ. Let us denote it by $P\{CD; \mu\}$. We do not know μ; at most we know that μ is in a subset M of E_k. In such a case we can guarantee that the probability of a correct decision is not less than a prescribed number P^*, if and only if

$$(2.9) \qquad p(M) = \inf \{P\{CD; \mu\}; \mu \,\varepsilon\, M\} \geqq P^*.$$

(If F is also unknown but we know that $F \,\varepsilon\, \mathfrak{F}$, $\mu \,\varepsilon\, M_F$, where $M_F \subset E_k$ for every $F \subset \mathfrak{F}$, we require that $p(M_F) \geqq P^*$ for every $F \,\varepsilon\, \mathfrak{F}$. In all our assertions, however, we can limit our considerations to the case of a fixed F, the generalization to an unknown F in some class $\mathfrak{F}$ being trivial.)

Now it is easy to see that $p(M)$ says little or nothing about the properties of the ranking methods if M is ample enough, for example if for every $\epsilon > 0$ there exists a $\mu \,\varepsilon\, M$ such that $0 < \mu_{i+1} - \mu_i < \epsilon$ for $i = 1, 2, \cdots, k - 1$. In such a case (we have supposed that the ξ_i's are continuous)

$$p(M) = k_1!(k_2 - k_1)! \cdots (k_s - k_{s-1})!/k!$$

which is the probability of a correct decision for the procedure that ignores the observed values of the X_i's and selects the decomposition of I into S_α entirely by random.

This inconvenience is a consequence of neglecting the differences in the practical importance of incorrect decisions for various μ. There are two (among other) ways of avoiding this situation. We may restrict the set M (this is what Bechhofer does) or we may modify the concept of the correct decision, which seems to us to be a more straightforward approach.

In the papers [1], [2], [3], Bechhofer and his co-authors give sufficient conditions for the inequality (2.9) in the case $M = M_\Delta$, where

$$\Delta = [\Delta_1, \Delta_2, \cdots, \Delta_{s-1}] \varepsilon E_{s-1}, \qquad \Delta_\alpha > 0$$

and

$$(2.10) \quad M_\Delta = \{\mu; \mu \varepsilon E_k, \mu_{[k_\alpha+1]} - \mu_{[k_\alpha]} \geqq \Delta_\alpha \quad \text{for} \quad \alpha = 1, 2, \cdots, s-1\}.$$

The numbers Δ_α are chosen by the experimenter as "the smallest values of the parameters $\delta_\alpha = \mu_{[k_\alpha+1]} - \mu_{[k_\alpha]}$ which are 'worth detecting'" ([1], p. 23).

On this intuitive meaning of Δ we shall base a definition of a Δ-correct ranking, thus proceeding in the second of the indicated ways. Any ranking can be regarded as k classification-statements of the form "i belongs to the αth group", and we shall first give the definition of Δ-correctness of pairs of such statements.

In direct accord with the meaning of Δ_α we say that

(i) the classification of i and j in the groups ρ and $\rho + 1$ respectively, is Δ-correct if

$$\mu_j > \mu_i - \Delta_\rho .$$

Further, since the classification of i and j in the groups α and β with $\alpha \leqq \rho < \beta$ can naturally be regarded as stronger than the classification of these indices in the ρth and $(\rho + 1)$th groups respectively, we say that

(ii) the classification of i and j in the groups α and β respectively, is Δ-correct, if for every $\rho = \alpha, \alpha + 1, \cdots, \beta - 1$ the classification of these indices in the groups ρ and $\rho + 1$ respectively, is Δ-correct, i.e., if

$$\mu_j > \mu_i - \min\{\Delta_\rho; \rho = \alpha, \alpha + 1, \cdots, \beta - 1\}.$$

Now we require that the k classification-statements of a Δ-correct ranking V be Δ-correct in the sense of (i) and (ii). We do not add further requirements and therefore we say that

(iii) a ranking V is Δ-correct, if for every $i \varepsilon V_\alpha, j \varepsilon V_\beta, 1 \leqq \alpha < \beta \leqq s$, the classification of i and j in the groups α and β respectively, is Δ-correct.

This definition is the weakest that can be derived from the indicated point of view (and yet it is, in general, stronger than Bechhofer's approach, see Theorem 2).

Denote by $P\{\Delta\text{-}CD; \mu\}$ the probability of Δ-correct decision, i.e., the probability that S is a Δ-correct ranking. According to the definition, S is Δ-correct if and only if the following implication holds:

$$(2.11)\quad 1 \leqq \alpha < \beta \leqq s, \quad i \,\varepsilon\, S_\alpha, \quad j \,\varepsilon\, S_\beta \Rightarrow \mu_j > \mu_i - \min\{\Delta_\rho \,;\, \rho = \alpha, \alpha + 1, \cdots, \beta - 1\}.$$

EXAMPLE 2. If, in Example 1, we have $\Delta = [11, 2]$, then the rankings $V^{(1)} = [\{1\}, \{2\}, \{3\}]$, $V^{(2)} = [\{2\}, \{1\}, \{3\}]$ are Δ-correct, while the remaining are not. Suppose, for a moment, that a formal analogy with (2.6) has led us to another definition and that we have said that V is Δ-correct if the implication

$$(2.12)\quad i \,\varepsilon\, V_\alpha, j \,\varepsilon\, V_{\alpha+1}, 1 \leqq \alpha < s \Rightarrow \mu_j > \mu_i - \Delta_\alpha$$

holds. Then $V^{(3)} = [\{3\}, \{1\}, \{2\}]$ would be Δ-correct. But, if I would change the ranking $V^{(3)}$ into a "better" one $V^{(4)} = [\{1\}, \{3\}, \{2\}]$, realizing that in fact $\mu_3 = 10 > \mu_1 = 1$, this better ranking would not be Δ-correct since $\mu_1 = 1$, $\mu_3 = 10$, $\mu_2 = 0$ and 10 is greater than $2 = \Delta_2$.

This is a very strong indication that (2.12) is a too weak condition; by strengthening it so that it would be free of such an inconvenience, we obtain our definition exactly. In fact it can be shown that a ranking V is Δ-correct[1] if and only if it, together with every better ranking, satisfies (2.12). By a ranking better than V we mean a ranking that can be obtained by a finite number of successive changes consisting of transferring indices i and j — if they satisfy $i \,\varepsilon\, V_\alpha, j \,\varepsilon\, V_\beta$, $\mu_i > \mu_j$, $\alpha < \beta$ — from V_α to V_β and from V_β to V_α respectively.

3. The special case $\mathcal{A}$. This case deals with the problem of selecting the greatest among the numbers μ_i and is a special case of the general problem, characterized by $s = 2$, $k_1 = k - 1$, so that the decomposition of I into the sets Z_α and S_α is determined by the unique element $\mu_{[k]}$ and $\mu_{\langle k\rangle}$ belonging to Z_2 and S_2 respectively.

Now $\Delta \,\varepsilon\, E_{s-1} = E_1$ is a number and the relation (2.9) (with $M = M_\Delta$ defined by (2.10)), for which sufficient conditions are given by Bechhofer, becomes

$$(3.1)\quad P\{\mu_{\langle k\rangle} = \mu_{[k]}\} \geqq P^* \quad \text{for every} \quad \mu \,\varepsilon\, E_k \quad \text{such that} \quad \mu_{[k]} \geqq \mu_{[k-1]} + \Delta.$$

It is of interest, however, how the method behaves when μ is not in M_Δ. As it is easy to show (we shall do it later), (3.1) is equivalent to

$$(3.2)\quad P\{\mu_{\langle k\rangle} > \mu_{[k]} - \Delta\} \geqq P^* \qquad \text{for every } \mu \,\varepsilon\, E_k.$$

This means that in case $\mathcal{A}$, the infimum of the probability of a correct decision for μ restricted to M_Δ equals the unrestricted infimum of the probability of a Δ-correct decision. We shall show, however, that the assertion in brackets { } in (3.2) can be strengthened without any decrease of the infimum of the probability.

THEOREM 1. *Put*

$$(3.3)\quad d = \max\{0, \Delta - (X_{\langle k\rangle} - X_{\langle k-1\rangle})\},$$

[1] From here on Δ-correctness is as originally defined.

$$(3.4)\qquad I(d) = \begin{cases}\langle \mu_{[k]} - d, \mu_{[k]}\rangle & \textit{for } d > 0,\\ \langle \mu_{[k]}, \mu_{[k]}\rangle & \textit{for } d = 0.\end{cases}$$

Then (3.1), (3.2) *and*

$$(3.5)\qquad P\{\mu_{\langle k\rangle} \varepsilon I(d)\} \geqq P^* \qquad \textit{for every } \mu \varepsilon E_k$$

are mutually equivalent.

Remark. The assertion $\mu_{\langle k\rangle} \varepsilon I(d)$ is stronger than the assertion $\mu_{\langle k\rangle} > \mu_{[k]} - \Delta$, since $d \leqq \Delta$. The surplus in the information can be considerable if $\mu_{[k]} - \mu_{[k-1]}$ is large, in which case the assertions in (3.1) and (3.2) are unnecessary weak.

Proof. Since $d \leqq \Delta$, we have $\{\mu_{\langle k\rangle} \varepsilon I(d)\} \subset \{\mu_{\langle k\rangle} > \mu_{[k]} - \Delta\}$ and hence (3.5) $\Rightarrow$ (3.2). If $\mu \varepsilon M_\Delta$, i.e., if $\mu_{[k-1]} \leqq \mu_{[k]} - \Delta$, we get

$$\{\mu_{\langle k\rangle} > \mu_{[k]} - \Delta\} \subset \{\mu_{\langle k\rangle} = \mu_{[k]}\}.$$

Thus (3.2) $\Rightarrow$ (3.1) and it remains to prove (3.1) $\Rightarrow$ (3.5).

If (3.1) holds, we have for every $\mu \varepsilon M_\Delta$

$$P^* \leqq P\{\mu_{\langle k\rangle} = \mu_{[k]}\} = P\{\langle k\rangle = [k]\} = P\bigcap_{i=1}^{k-1}\{X_{[k]} > X_{[i]}\}$$

$$= P\bigcap_{i=1}^{k-1}\{\xi_{[k]} > \xi_{[i]} - (\mu_{[k]} - \mu_{[i]})\}$$

(the equality of the third and fourth terms follows from the continuity of the ξ_i's).

Especially for $\mu_{[1]} = \cdots = \mu_{[k-1]} = 0$, $\mu_{[k]} = \Delta$ we have $\mu \varepsilon E_\Delta$ and

$$P\bigcap_{i=1}^{k-1}\{\xi_{[k]} > \xi_{[i]} - \Delta\} \geqq P^*.$$

Hence and from the condition of permutability (2.1) we get

$$(3.6)\qquad P\bigcap_{i \varepsilon I-\{j\}}\{\xi_j > \xi_i - \Delta\} \geqq P^* \qquad \text{for every } j \varepsilon I.$$

Now let us choose an arbitrary $\mu \varepsilon E_k$, which determines the values of $[i] = [i, \mu]$. We have

$$\bigcap_{i=1}^{k-1}\{\xi_{[k]} > \xi_{[i]} - \Delta\} = \bigcap_{i=1}^{k-1}\{X_{[k]} - \mu_{[k]} > X_{[i]} - \mu_{[i]} - \Delta\}$$

$$\subset \{\langle k\rangle = [k]\} \cup \{X_{[k]} - \mu_{[k]} > X_{\langle k\rangle} - \mu_{\langle k\rangle} - \Delta\}$$

$$\subset \{\mu_{\langle k\rangle} = \mu_{[k]}\} \cup \{\mu_{\langle k\rangle} > -X_{[k]} + X_{\langle k\rangle} + \mu_{[k]} - \Delta\}$$

$$\subset \{\mu_{\langle k\rangle} = \mu_{[k]}\} \cup \{\mu_{\langle k\rangle} > \mu_{[k]} - d\} \subset \{\mu_{\langle k\rangle} \varepsilon I(d)\},$$

which with (3.6) implies (3.5). Thus (3.1) $\Rightarrow$ (3.5) and the theorem is proved.

4. The general case. Let us denote by p_Δ the infimum of the probability of a Δ-correct decision, i.e., denote

$$(4.1)\qquad p_\Delta = \inf\{P\{\Delta\text{-}CD; \mu\}; \mu \varepsilon E_k\}.$$

THEOREM 2. *For every* $\Delta = [\Delta_1, \cdots, \Delta_{s-1}] \varepsilon E_{s-1}$, $\Delta_\alpha > 0$ *we have*

$$p_\Delta \leqq p(M_\Delta); \tag{4.2}$$

with equality holding in case $\mathfrak{A}$ *but not in general.*

PROOF. To prove (4.2) it suffices to show that for $\mu \varepsilon M_\Delta$, Δ-correctness implies correctness and for this it suffices to show that if $S_1, \cdots, S_s$ are Δ-correct, then $S_\alpha = Z_\alpha$ for $\alpha = 1, \cdots, s$. Suppose that this is not true and that (2.11) holds but $S_1 = Z_1, \cdots, S_{\gamma-1} = Z_{\gamma-1}$, $S_\gamma \neq Z_\gamma$. Then there are two indices i and j such that

$$i \varepsilon Z_\gamma, \qquad j \varepsilon Z_{\beta_1} \qquad \text{for } \beta_1 > \gamma \tag{4.3}$$

and

$$j \varepsilon S_\gamma, \qquad i \varepsilon S_{\beta_2} \qquad \text{for } \beta_2 > \gamma. \tag{4.4}$$

However, from (2.11) and (4.4) it follows that $\mu_i > \mu_j - \Delta_\gamma$ and from (4.3) and the relation $\mu \varepsilon M_\Delta$ it follows that $\mu_j \geqq \mu_i + \Delta_\gamma$. This contradiction proves (4.2).

That the equality $p_\Delta = p(M_\Delta)$ holds in the case $\mathfrak{A}$ has been proved in the preceding section and it remains to give an example in which this equality does not hold.

Let us choose $k = 3$, $s = 3$, $k_0 = 0$, $k_1 = 1$, $k_2 = 2$, $k_3 = 3$, $\Delta = [\Delta_1, 1]$, let ξ_1, ξ_2, ξ_3 be independent normal (0, 1) random variables. The infimum $p(M_\Delta)$ is attained for $\mu = [-\Delta_1, 0, 1]$ so that $p(M_\Delta) = P\{\xi_1 - \Delta_1 < \xi_2 < \xi_3 + 1\}$ and we see that

$$p(M_\Delta) \to P\{\xi_2 < \xi_3 + 1\} \qquad \text{for } \Delta_1 \to +\infty. \tag{4.5}$$

On the other hand $p_\Delta \leqq P\{\Delta\text{-}CD; [0, 0, 1]\} = P\{\xi_2 < \xi_3 + 1, \xi_1 < \xi_3 + 1\} < P\{\xi_2 < \xi_3 + 1\}$, whence for sufficiently large Δ_1 we get

$$p(M_\Delta) > p_\Delta, \tag{4.6}$$

which completes the proof.

This result can be reinterpreted as follows: According to the definition of Δ-correctness, derived in Section 2, the least favorable configuration of means μ_i (for which $\inf P\{\Delta\text{-}CD; \mu\}$ is attained) does not necessarily lie in the set M_Δ. Thus this least favorable configuration can be different from that of Bechhofer, who limits his considerations to M_Δ. (For $\mu \varepsilon M_\Delta$ both $P\{\Delta\text{-}CD; \mu\}$ and $P\{CD; \mu\}$ coincide.)

5. An approximation. In [2] and [1] one of the sufficient conditions for

$$p(M_\Delta) \geqq P^*$$

is of the form $N \geqq \kappa(k, f, P^*)\hat{\sigma}$, where $\hat{\sigma}^2$ is an estimate of the unknown variance σ^2, derived from a χ^2-random variable with f degrees of freedom and N determines the size of the experiment necessary to guarantee that the probability of correct ranking be at least P^* for case $\mathfrak{A}$ with $\mu \varepsilon M_\Delta \subset E_k$. The values $\kappa(k, f, P^*)$ can be determined ([1], [2], [5]) from existing tables for $k = 2$ and 3 and various f and also for $k > 3$ and $f = +\infty$ (the last case corresponding to the exact knowledge of σ^2). The approximation

$$\kappa(k, f, P^*) = \kappa(k, +\infty, P^*)q_{k,f}(1 - P^*)/q_{k,+\infty}(1 - P^*) \tag{5.1}$$

where $q_{k,f}(r)$ is the (double sided) r-critical point of the Studentized range for k variables with f degrees of freedom, was found to give good results even for small f in the cases where it could be checked by existing tables, i.e., for $k = 2, 3$. A more detailed numerical study for $k > 2$ seems to be worth-while.

REFERENCES

[1] Robert E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–39.

[2] Robert E. Bechhofer, Charles W. Dunnett and Milton Sobel, "A two-sample multiple decision procedure for ranking means of populations with a common unknown variance," *Biometrika*, Vol. 41 (1954), pp. 170–176.

[3] Robert E. Bechhofer and Milton Sobel, "A single-sample multiple procedure for ranking variances of normal populations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 273–289.

[4] Robert E. Bechhofer, "A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance and its use in various experimental designs," *Biometrics*, Vol. 14 (1958), pp. 408–429.

[5] Charles W. Dunnett and Milton Sobel, "A bivariate generalization of Student's t-distribution, with tables for certain special cases," *Biometrika* Vol. 41 (1954), pp. 153–169.