THE MOST-ECONOMICAL CHARACTER OF SOME BECHHOFER AND SOBEL DECISION RULES[1]

By Wm. Jackson Hall

University of North Carolina

1. Introduction. R. E. Bechhofer [1] has considered a single-sample multiple-decision procedure for choosing, among a group of normal populations with common known variances, that population with the largest mean, and, with M. Sobel [2], a procedure for choosing the normal population with the smallest variance. Several other analogous problems have also been considered.[2] They suggest, with only intuitive justification, choosing the population with the largest (smallest) sample mean (variance), and give tables for finding the minimum sample size (assumed equal for all populations) which will guarantee a correct decision with prescribed probability when the extreme population parameter is sufficiently distinct from the others. This paper gives justification for a wide class of such procedures, proving that no other rules can meet this guarantee with a smaller (fixed) sample size; that is, such rules are *most economical* [4].

Proof of the most-economical character of these rules is achieved by proving their minimax character when a suitable loss function is introduced. R. R. Bahadur and L. A. Goodman [5] have considered a class of multiple-decision rules which they have called *impartial* (invariant under permutations of the populations). Their results are applicable to such problems of choosing the best population and imply that Bechhofer and Sobel's rules are minimax rules (in fact, uniformly minimum risk rules) among the class of impartial decision rules. The present paper removes this restriction of impartiality. Thus, in the present context, impartiality is no restriction when looking for minimax rules, as is well-known to be the case for certain other kinds of invariance.

The main result is stated in Section 2 and proved in Section 3. It is applicable to any analogous problem of choosing the population with the most extreme parameter when, for each sample, there is a numerical sufficient statistic with a *monotone likelihood ratio*[3] and the (numerical) parameter is a *location* or *scale* (but not range) *parameter*[3] in the distribution of the statistic. The theorem is applicable to Bechhofer's procedure and the corollary to Bechhofer and Sobel's. (In the latter example, if the means are unknown, it will be necessary to invoke invariance under changes in scale.) In Section 4, the result is further extended to problems

Received November 23, 1958; revised March 26, 1959.

[1] This research was supported by the United States Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] For a review, see the introduction in [3].

[3] For definition, see [6], for example.

of ranking the populations according to the parameter values, or of grouping them by ranks, as formulated by Bechhofer [1].

The requirement that the parameter be one of location or scale is dropped in Section 5. Then the *guarantee* holds only at a *specified location*; for many problems, a *least favorable location* can be determined so that the guarantee can be made to hold irrespective of location. For example, the procedures of M. Sobel and M. J. Huyett [7] for choosing the largest of several binomial parameters are proved to be most economical. In Section 6, the broader optimality of these latter procedures is discussed.

These results, some of which appeared in [8], are obtained from application of *most economical decision theory* [4]. As indicated by Bechhofer [1], if the populations differ in a known way (normal populations with different known variances, for example), optimal allocation of the sample sizes is apparently exceedingly complex; such problems are not treated here.

2. Theorem.

(i) *Let* $\{f_\theta\}$, $\theta \varepsilon \Omega \subset R_1$, *be a homogeneous class of density functions*[4] *w.r.t. a fixed measure. Let* $\{X_{ij}\}$ $(i = 1, \cdots, m; j = 1, \cdots, n)$ *denote mn independent random variables where* X_{ij} *has the density function* f_{θ_i}, $\theta_i \varepsilon \Omega$, $i = 1, \cdots, m$, *and let* $\theta_{[1]} \leqq \cdots \leqq \theta_{[m]}$ *be the ordered values of the* θ_i*'s.* Set $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)$.

(ii) *Suppose* $t_i = t_i(x_{i1}, \cdots, x_{in})$ *is a numerical sufficient statistic for* $(X_{i1}, \cdots, X_{in})$, *that* t_i *has a monotone likelihood ratio, and that* θ_i *is a location parameter in the induced distribution of* t_i $(i = 1, \cdots, m)$.

(iii) *Let* D_n *denote any decision rule for choosing which* θ_i *is* $\theta_{[m]}$ *based on an observation on the mn random variables* $\{X_{ij}\}$, *and let* D_n^0 *denote that* D_n *which chooses as* $\theta_{[m]}$ *that* θ_i *corresponding to the largest of the* t_i*'s with ties broken by randomization. Suppose N is the least n for which*

$$(1)\quad \begin{aligned} &\int F_n^{m-1}(t + \delta - 0)\, dF_n(t) + \sum_{r=2}^{m} \frac{1}{r}\binom{m-1}{r-1} \\ &\cdot \int [F_n(t+\delta) - F_n(t+\delta-0)]^{r-1}[F_n(t+\delta-0)]^{m-r}\, dF_n(t) \geqq \gamma \end{aligned}$$

$$(\delta > 0, 0 < \gamma < 1)$$

where $F_{\theta,n}(t) = F_n(t - \theta)$ *is the c.d.f. of t with parameters* θ *and n.*

Then D_N^0 *satisfies*

(a) $Pr\{$*correct decision using* $D_n \mid \boldsymbol{\theta}\} \geqq \gamma$ *for all* $\boldsymbol{\theta}$ *for which* $\theta_{[m]} - \theta_{[m-1]} \geqq \delta$ *and*

(b) *there does not exist a decision rule* D_n *satisfying* (a) *with* $n < N$.

COROLLARY: *Replace in* (i) "R_1" *by* "*positive* R_1"; *replace in* (ii) "*location*" *by* "*scale*"; *replace in* (iii) "$t + \delta$" *by* "$t\delta$", "$\delta > 0$" *by* "$\delta > 1$", "$F_n(t - \theta)$" *by* "$F_n(t/\theta)$"; *replace in* (a) "$\theta_{[m]} - \theta_{[m-1]}$" *by* "$\theta_{[m]}/\theta_{[m-1]}$".

Note: The summation term in (1) accommodates the possibility of ties—when $r = 2, \cdots, m$ *t*-values may be largest—and drops out if $F_{\theta,n}$ is absolutely con-

[4] The region of positive density is independent of θ.

tinuous; hereafter, for simplicity of presentation, we make this assumption and thereby replace (1) by

$$(1') \qquad \int F_n^{m-1}(t+\delta)\, dF_n(t) \geqq \gamma.$$

3. Proof of theorem. Set $\omega_i = \{\boldsymbol{\theta} \mid \theta_i = \theta_{[m]}, \theta_{[m]} - \theta_{[m-1]} \geqq \delta\}$, and $p_i(\boldsymbol{\theta}) = \Pr\{\text{choosing } \theta_i \text{ using } D_n \mid \boldsymbol{\theta}\}$, $i = 1, \cdots, m$. Then (a) is equivalent to: $p_i(\boldsymbol{\theta}) \geqq \gamma$ for $\boldsymbol{\theta} \varepsilon \omega_i$ $(i = 1, \cdots, m)$.

Let λ_i be a distribution over ω_i which assigns probability one to the $\boldsymbol{\theta}$-point with all coordinates equal to θ_0 (arbitrary) except the ith coordinate which equals $\theta_0 + \delta$; i.e., $\theta_{[1]} = \theta_{[m-1]} = \theta_0 = \theta_i - \delta$. Denote this point $\boldsymbol{\theta}_i$.

We first show that, for n fixed, D_n^0 is minimax for choosing among $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m$ where the loss function is $-1/\gamma$ if a correct decision is made and zero otherwise, and that, when using D_n^0, $p_i(\boldsymbol{\theta}_i) = \int F_n^{m-1}(t+\delta)\, dF_n(t)$ for all i. Secondly, we show that the λ_i's are least favorable in the sense that $\inf_{\omega_i} p_i(\boldsymbol{\theta}) = p_i(\boldsymbol{\theta}_i) = \int_{\omega_i} p_i(\boldsymbol{\theta})\, d\lambda_i$, as shown in the special case of Bechhofer in [1]. Application of Theorems 7 and 9 from [4] completes the proof of the theorem. The corollary may be proved by applying a log transform to t, θ, and δ.

1.° According to well-known results of Wald (e.g., see Section 1.B of [4]), a minimax rule for choosing among the $\boldsymbol{\theta}_i$'s with the specified loss is one which chooses θ_i as the largest θ if $a_i h_i \geqq a_j h_j$ for all j where $h(\mathbf{t}, \boldsymbol{\theta})$ is the joint density of $t_1, \cdots, t_m$ when the parameter is $\boldsymbol{\theta}$, $h_j = h(\mathbf{t}, \boldsymbol{\theta}_j)$, and $a_1, \cdots, a_m$ are positive constants chosen so that $p_1(\boldsymbol{\theta}_1) = \cdots = p_m(\boldsymbol{\theta}_m)$. Denoting the density of F by g and of F_θ by g_θ (dropping the subscript n assumed fixed), $h(\mathbf{t}, \boldsymbol{\theta}) = g_{\theta_1}(t_1) g_{\theta_2}(t_2) \cdots g_{\theta_m}(t_m)$ so that $a_i h_i \geqq a_j h_j$ implies

$$(2) \qquad a_i g_{\theta_0+\delta}(t_i) g_{\theta_0}(t_j) \geqq a_j g_{\theta_0}(t_i) g_{\theta_0+\delta}(t_j),$$

or equivalently, since θ is a location parameter, the subscripts on the g's can be subtracted from the arguments. Denoting $r(t) = g_{\theta_0+\delta}(t)/g_{\theta_0}(t)$ for fixed θ_0 and δ, defined throughout the region of positive density for t, (2) implies $r(t_j) \leqq r(t_i) a_i/a_j$. Since t has a monotone likelihood ratio, $r(t)$ increases with t, the inverse function exists, and (2) may be written $t_j \leqq r^{-1}[r(t_i) a_i/a_j]$. Therefore, the probability that the minimax rule chooses θ_i as largest when $\boldsymbol{\theta} = \boldsymbol{\theta}_i$ is

$$(3) \qquad \begin{aligned} p_i(\boldsymbol{\theta}_i) &= \Pr\{a_i h_i \geqq a_j h_j \quad \text{for all} \quad j \mid \boldsymbol{\theta} = \boldsymbol{\theta}_i\} \\ &= \mathcal{E} \Pr\{a_i h_i \geqq a_j h_j \quad \text{for all} \quad j \mid t_i = y, \boldsymbol{\theta} = \boldsymbol{\theta}_i\} \\ &= \int \prod_{j \neq i} F_{\theta_0, n}\{r^{-1}[r(y) a_i/a_j]\}\, dF_{\theta_0+\delta, n}(y). \end{aligned}$$

This is independent of i if $a_1 = a_2 = \cdots = a_m$, in which case θ_i is chosen if h_i is largest. Because of the monotone property of h_i, the minimax rule is thus D_n^0. Upon setting the a_i's equal and transforming $t = y - \theta_0 - \delta$, (3) becomes $p_i(\boldsymbol{\theta}_i) = \int F_n^{m-1}(t+\delta)\, dF_n(t)$, and is thus independent of the choice of θ_0.

2.° Similarly to (3) above, for D_n^0 we have $p_i(\boldsymbol{\theta}) = \Pr\{t_i \geqq t_j \text{ for all } j \mid \boldsymbol{\theta}\} =$

$\int \Pi_{j\neq i} F_{\theta_j,n}(y)\, dF_{\theta_i,n}(y) = \int \Pi_{j\neq i} F_n(u + \theta_i - \theta_j)\, dF_n(u)$ which increases with $\theta_i - \theta_j$ for each $j \neq i$. For $\boldsymbol{\theta} \varepsilon \omega_i$, $\theta_i - \theta_j \geqq \delta$ so that the infimum over ω_i of p_i is attained at $\theta_i - \theta_j = \delta$ for $j \neq i$ and, in particular, at $\boldsymbol{\theta} = \boldsymbol{\theta}_i$.

4. Extension to procedures for grouping by ranks. The results in Section 2 can be extended to the problem of ranking the populations according to their θ-values, or more generally, of selecting the m_s "best" populations, the m_{s-1} "second best", etc., the m_1 "worst" populations, given $m_1, \cdots, m_s (s \leqq m, \sum m_i = m)$—the "general goal" expressed by Bechhofer in Section 3.B of [1]. The rule D_N^0 is to rank according to t-values, choose N by a rule analogous to (1) or (1′) (see [1]), and then the probability of a correct grouping will be at least γ when the groups are sufficiently far apart. Most economical theory is used for discriminating among the $m!/(m_1!\, m_2! \cdots m_s!)$ possible alternative decisions. The proof differs little except for the notational complexities.

5. Extension to other ordering parameters. Intuitively, a procedure which ranks the θ's according to the values of the sufficient statistic t should be optimal whenever θ is some kind of *ordering parameter* in the distribution of t. That t should have a monotone likelihood ratio is such an ordering requirement. A less stringent requirement is that the c.d.f. of t be monotone in θ for all t; that is, denoting by T_i a random variable with distribution parameter $\theta_i (\theta_1 < \theta_2)$, $\Pr\{T_2 > t\} \geqq \Pr\{T_1 > t\}$ for all t, in which case T_2 is said to be *stochastically larger* than T_1. E. L. Lehmann has shown (Theorem 1 in [9]) that a monotone likelihood ratio assumption implies the latter type of ordering. That θ be a location parameter is an additional ordering requirement—that F_θ be a particular kind of monotone function, namely $F(t - \theta)$. It was required in the theorem so that the probability in (a) could be computed on the condition that the best population was sufficiently *distant* from the second best without regard to the *location* of the best population; that F_θ be monotone was also required, but this follows from the monotone likelihood ratio assumption. Thus, the location requirement in (ii) can be removed by adding it in (a), so that replacing $F(t + \delta)$ by $F_{\theta_0 - \delta}(t)$ and dF by dF_{θ_0} in (1) and (1′) and replacing (a) by (a′) in which the inequality is required to hold for all $\boldsymbol{\theta}$ for which $\theta_{[m]} = \theta_0$ and $\theta_{[m]} - \theta_{[m-1]} \geqq \delta$ for some specified value θ_0 of the parameter, the theorem and proof remain valid. The guarantee of a correct decision is only calculated at one location, specified by θ_0.

In many such problems, it will be possible to find a least favorable location θ_0, i.e., a value of θ which minimizes $\int F_{\theta-\delta}^{m-1}(t)\, dF_\theta(t)$, in which case (a) need not be replaced by (a′). Sufficient conditions are that Ω be bounded and closed. If not, it may be possible to find a *least favorable sequence*, applying Theorem 8 of [4].

In this revised form, the theorem applies to all such problems of choosing the best population whenever there is a numerical sufficient statistic with a monotone likelihood ratio, and therefore, in particular, if its distribution is in the exponential family. Thus, it holds for Sobel and Huyett's procedures [7], the condition (a) corresponding to their "original specification" using a least favorable θ_0, and (a′) to their "alternative specification."

6. A distribution-free extension. It can be shown that Sobel and Huyett's procedure is optimal not only for choosing the best binomial population but for the more general problem they describe of choosing the population with the largest "survival probability", with no parametric specification of the underlying distributions. If the distributions differ only in location, then the problem is equivalent to that of choosing the population with the largest median. This application is adapted from an example by W. Hoeffding [10].

Let the class of density functions under consideration include all densities, f, w.r.t. a fixed measure μ on the real line such that $0 < \mu(\{x \geqq a\}) < 1$ for some specified a. The $\{X_{ij}\}$ are assumed independent with $\theta_i \equiv \Pr\{X_{ij} \geqq a\}$, constant over $j = 1, \cdots, n(i = 1, \cdots, m)$. If X_{ij} represents a lifetime, then θ_i is the probability of survival to age a, and none of the distributions need coincide except in their θ-values.

Extension of the theorem can be accomplished as indicated briefly here: Subsets $\{\omega_i\}$ of density functions are specified in terms of the θ's as in Section 3; *a priori* distributions over these sets are specified, somewhat as in example 5 in [4], which reduces the problem to that of choosing the best of m binomial distributions. The decision procedure is to choose θ_k as the largest of the θ_i's if more of the x_{kj}'s exceed a than do the x_{ij}'s for any other i. That these *a priori* distributions are least favorable follows as in Section 3, using Theorem 7 from [4], and noting that the probability of a correct decision depends only on the θ-values. A least favorable θ_0 can be chosen, if desired, as in [7]. Thus, the procedure of Sobel and Huyett is most economical for this distribution-free problem in the sense that (a), or (a′), and (b) are satisfied.

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 AND MILTON SOBEL, "A single-sample multiple decision procedure for ranking variances of normal populations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 273–289.

[3] 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 with various experimental designs," *Biometrics*, Vol. 14 (1958), pp. 408–429.

[4] WM. JACKSON HALL, "Most economical multiple-decision rules," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1079–1094.

[5] RAGHU RAJ BAHADUR AND LEO A. GOODMAN, "Impartial decision rules and sufficient statistics," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 553–562.

[6] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[7] MILTON SOBEL AND MARILYN J. HUYETT, "Selecting the best one of several binomial populations," *Bell System Tech. Jour.*, Vol. 36 (1957), pp. 537–576.

[8] WM. JACKSON HALL, "An optimum property of Bechhofer's single-sample multiple-decision procedure for ranking means and some extensions," *Institute of Statistics Mimeograph Series,* No. 118, University of North Carolina, September 1954.

[9] E. L. LEHMANN, "Ordered families of distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 399–419.

[10] WASSILY HOEFFDING, " 'Optimum' nonparametric tests," *Second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, Berkeley, 1951, pp. 83–92.