

AN OPTIMUM PROPERTY OF SOME BECHHOFFER AND SOBEL
NON-SEQUENTIAL MULTIPLE-DECISION RULES

by

Wm. Jackson Hall
University of North Carolina

This research was supported by the United States Air Force through the 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.

Institute of Statistics
Mimeograph Series No. 210
November, 1958

AN OPTIMUM PROPERTY OF SOME BECHHOFFER AND SOBEL
NON-SEQUENTIAL MULTIPLE-DECISION RULES ¹

by Wm. Jackson Hall

University of North Carolina

1. Introduction. R. E. Bechhofer [1] has proposed 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 treated.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].

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 ratio³ and the parameter is a location or scale (but not range) parameter³ in the distribution of the statistic. The theorem is applicable to

¹ This research was supported by the United States Air Force through the 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 [5], for example.

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) In section 4, the result is further extended to problems 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 [6] for choosing the largest of several binomial parameters are proved to be most economical. In section 6, the broader optimality of this latter procedure is discussed.

These results, some of which appeared in [7], 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 $\{f_{\theta}\}$, $\theta \in \Gamma \subset R_1$, be a homogeneous class of density functions⁴ w.r.t. a fixed measure. Let $\{X_{ij}\}$ ($i = 1, \dots, m$; $j = 1, \dots, n$) denote mn independent random variables where X_{ij} has the density function f_{θ_i} , $\theta_i \in \Gamma$, $i = 1, \dots, m$, and let $\theta_{[1]} \leq \dots \leq \theta_{[m]}$ be the ordered values of the θ_i 's. Denote $\underline{\theta} = (\theta_1, \dots, \theta_m)$.

(ii) Suppose $t_i = t_i(x_{i1}, \dots, x_{in})$ is a numerical sufficient statistic for θ_i , that t_i has a monotone likelihood ratio, and that θ_i is a location parameter in

⁴ The region of positive density is independent of θ .

the induced distribution of t_i ($i = 1, \dots, m$).

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

$$(1) \quad \int F_n^{m-1}(t + \delta) dF_n(t) \geq \gamma \quad (\delta > 0, 0 < \gamma < 1)$$

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) $\Pr \left\{ \text{correct decision using } D_n \mid \underline{\theta} \right\} \geq \gamma$ for all $\underline{\theta}$ for which $\theta_{[m]} - \theta_{[m-1]} \geq \delta$ and

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

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

3. Proof of theorem. Denote $\underline{\theta} = (\theta_1, \dots, \theta_m)$, $\omega_i = \{ \underline{\theta} \mid \theta_i = \theta_{[m]} - \theta_{[m-1]} \geq \delta \}$, and $p_i(\underline{\theta}) = \Pr \left\{ \text{choosing } \theta_i \text{ using } D_n \mid \underline{\theta} \right\}$, $i = 1, \dots, m$. Then (a) is equivalent to $p_i(\underline{\theta}) \geq \gamma$ for $\underline{\theta} \in \omega_i$ ($i = 1, \dots, m$).

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

We first show that, for n fixed, D_n^0 is minimax for choosing among $\theta_1, \dots, \theta_m$ where the loss function is $-1/\gamma$ if a correct decision is made and zero otherwise, and that, when using D_n^0 , $p_i(\underline{\theta}_i) = \text{L.H.S. of (1) for all } i$. Secondly, we show that the λ_i 's are least favorable in the sense that $\inf_{\omega_i} p_i(\underline{\theta}) = p_i(\underline{\theta}_i) = \int_{\omega_i} p_i(\underline{\theta}) d\lambda_i$, 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 section 1.B of [7], a minimax rule for choosing among the $\underline{\theta}_i$'s with the specified loss is one which chooses $\underline{\theta}_i$ as the largest $\underline{\theta}$ if $a_i h_i \geq a_j h_j$ for all j where $h(\underline{t}, \underline{\theta})$ is the joint density of t_1, \dots, t_m when the parameter is $\underline{\theta}$, $h_j = h(\underline{t}, \underline{\theta}_j)$, and a_1, \dots, a_m are positive constants chosen so that $p_1(\underline{\theta}_1) = \dots = p_m(\underline{\theta}_m)$. Denoting the density of F by g and pf $F_{\underline{\theta}}$ by $g_{\underline{\theta}}$ (dropping the subscript n assumed fixed),

$$h(\underline{t}, \underline{\theta}) = g_{\underline{\theta}_1}(t_1) g_{\underline{\theta}_2}(t_2) \dots g_{\underline{\theta}_m}(t_m)$$

so that $a_i h_i \geq a_j h_j$ implies

$$(2) \quad a_i g_{\underline{\theta}_0 + \delta}(t_i) g_{\underline{\theta}_0}(t_j) \geq a_j g_{\underline{\theta}_0}(t_i) g_{\underline{\theta}_0 + \delta}(t_j)$$

or equivalently, since $\underline{\theta}$ is a location parameter, the subscripts on the g 's can be subtracted from the arguments. Denoting $r(t) = g_{\underline{\theta}_0 + \delta}(t) / g_{\underline{\theta}_0}(t)$ for fixed $\underline{\theta}_0$ and δ , defined throughout the region of positive density for t , (2) implies $r(t_j) \leq r(t_i) a_i / a_j$. Since t has a monotone likelihood ratio, $r(t)$ increases with t , the inverse function exists, and (2) may be written $t_j \leq r^{-1} [r(t_i) a_i / a_j]$. Therefore, the probability that the minimax rule chooses $\underline{\theta}_i$ as largest when $\underline{\theta} = \underline{\theta}_1$ is

$$(3) \quad \begin{aligned} p_1(\underline{\theta}_1) &= \Pr \{ a_i h_i \geq a_j h_j \text{ for all } j \mid \underline{\theta} = \underline{\theta}_1 \} \\ &= \Pr \{ a_i h_i \geq a_j h_j \text{ for all } j \mid t_i = y, \underline{\theta} = \underline{\theta}_1 \} \\ &= \int \prod_{j \neq i} F_{\underline{\theta}_0, n} \{ r^{-1} [r(y) a_i / a_j] \} dF_{\underline{\theta}_0 + \delta, n}(y) \end{aligned}$$

This is independent of i if $a_1 = a_2 = \dots = a_m$, in which case $\underline{\theta}_1$ is chosen if h_1 is largest. Because of the monotone property of h_1 , the minimax rule is thus D_n^0 .

Upon setting the a_i 's equal and transforming $t = y - \underline{\theta}_0 - \delta$, (3) becomes $p_1(\underline{\theta}_1) =$ L.H.S. of (1), and is thus independent of the choice of $\underline{\theta}_0$.

2.° Similarly to (3) above, for D_n^0 we have $p_i(\underline{\theta}) = \Pr \{ t_i \geq t_j \text{ for all } j | \underline{\theta} \} = \int \prod_{j \neq i} F_{\theta_j, n}(y) dF_{\theta_i, n}(y) = \int \prod_{j \neq i} F_n(u + \theta_i - \theta_j) dF_n(u)$ which increases with $\theta_i - \theta_j$ for each $j \neq i$. For $\underline{\theta} \in \omega_i$, $\theta_i - \theta_j \geq \delta$ 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 $\underline{\theta} = \underline{\theta}_i$.

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, \dots, m_s ($s \leq m, \sum m_i = m$) -- the "general goal" expressed by Bechhofer in section 3.B of [1_7]. The rule D_N^0 is to rank according to t-values, choose N by a rule analogous to (1) (see [1_7]), 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 $m!/(m_1!m_2! \dots m_s!)$ 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, if $\theta_1 < \theta_2$ then $\Pr \{ T > t | \theta_1 \} \leq \Pr \{ T > t | \theta_2 \}$ for all t, in which case the one random variable is said to be stochastically larger than the other. E. L. Lehmann has shown (theorem 1 in [8_7]) 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 replacing (a) by (a') in which the inequality is required to hold for all θ for which $\theta_{[m]} = \theta_0$ and $\theta_{[m]} - \theta_{[m-1]} \geq \delta$ 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 $\int F_{\theta_0 - \delta}^{m-1}(t) dF_{\theta}(t)$, in which case (a) need not be replaced by (a'). Sufficient conditions are that \int 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 [6], 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 [9].

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 $0 < \mu(\{x \geq a\}) < 1$ for some specified a . The $\{X_{ij}\}$ are assumed independent with $\theta_i \equiv \Pr\{X_{ij} \geq a\}$, constant over $j = 1, \dots, n$ ($i = 1, \dots, m$). 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 [6]. 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.

The author would like to acknowledge several helpful comments by Professor Wassily Hoeffding.

REFERENCES

- [1] R. E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 16-39.
- [2] R. E. Bechhofer and M. Sobel, "A single-sample multiple decision procedure for ranking variances of normal populations," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 273-289.
- [3] R. 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] W. J. Hall, "Most economical multiple-decision rules," Annals of Mathematical Statistics, Vol. 29 (1958) pp.
- [5] D. Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions, John Wiley and Sons, 1954.
- [6] M. Sobel and M. J. Huyett, "Selecting the best one of several binomial populations," Bell System Technical Journal, Vol. 36 (1957), pp. 537-576.
- [7] W. J. 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.
- [8] E. L. Lehmann, "Ordered families of distributions," Annals of Mathematical Statistics, Vol. 26 (1955), pp. 399-419.
- [9] W. Hoeffding, "'Optimum' nonparametric tests," Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951, pp. 83-92.