

ON A THEOREM OF BAHADUR AND GOODMAN¹

BY E. L. LEHMANN

University of California, Berkeley

1. Introduction. This note is concerned with the problem of selecting the best one (or any other specified number) of several populations. It is restricted to the symmetric case where typically the observations consist of samples of equal size from the different populations. For certain families of distributions, Bahadur (1950) and Bahadur and Goodman (1952) have proved that the natural selection procedure uniformly minimizes the risk among all symmetric procedures for a large class of loss functions. In Section 2 we give an alternative proof of this theorem, and in Section 3 show that the theorem implies many other optimum properties including one obtained in a different manner by Hall (1959).

The problem of selecting the best one of s populations is a finite decision problem with s possible decisions. Let us more generally consider any finite decision problem with possible decisions d_1, \dots, d_s . A (randomized) decision procedure is a vector $\varphi = (\varphi_1, \dots, \varphi_s)$ where $\varphi_i(x)$ denotes the probability of taking decision d_i when the value of the random observable X is x , and where $\sum \varphi_i = 1$ for all x . We suppose that the distribution P_θ of X depends on the parameter θ and that the loss resulting from decision d_i when θ is the true parameter value is $L(\theta, d_i) = L_i(\theta)$.

Corresponding to the symmetry assumed for the selection problem, we shall assume that the problem is invariant under the finite transformation group $G = \{g_1, \dots, g_N\}$: if the distribution of X is $d[X] = P_\theta$, the random variable $g_i X$ has distribution $d[g_i X] = P_{\bar{g}_i \theta}$ where g_i and \bar{g}_i are 1:1 mappings respectively of the sample space and of the parameter space onto themselves; furthermore there exist transformations g_1^*, \dots, g_N^* of the decision space (i.e. permutations of d_1, \dots, d_s) such that for any i, j and θ

$$(1) \quad L(\bar{g}_i \theta, g_i^* d_j) = L(\theta, d_j).$$

A procedure φ is then said to be invariant if

$$(2) \quad g^* \varphi(x) = \varphi(gx) \text{ for all } x \text{ and } g.$$

The procedure taking on the value $\varphi(gx)$ at the point x will be denoted by φg , and (2) can then be written as $g^* \varphi g^{-1} = \varphi$.

To prove that their procedure uniformly minimizes the risk among all invariant procedures, Bahadur and Goodman first characterize the totality of invariant procedures. An alternative proof can be based on the following lemma concerning general finite invariant decision problems.

Received 12 August 1965.

¹With the partial support of the Office of Naval Research under contract number NONR-222(43).

LEMMA 1. A necessary and sufficient condition for an invariant procedure $\varphi^{(0)}$ to minimize the risk $R(\theta, \varphi)$ among all invariant procedures is that it minimizes the average risk

$$(3) \quad r(\theta, \varphi) = \sum_{i=1}^N R(\bar{g}_i \theta, \varphi) / N$$

among all procedures.

PROOF. (i) Let $\varphi^{(0)}$ be an invariant procedure which minimizes $r(\theta, \varphi)$ among all procedures. If φ' is any other invariant procedure, it follows from the fact that the risk function of any invariant procedure is constant over each orbit $\{\bar{g}_i \theta: i = 1, \dots, N\}$ that

$$R(\theta, \varphi') = r(\theta, \varphi') \geq r(\theta, \varphi^{(0)}) = R(\theta, \varphi^{(0)})$$

Hence $\varphi^{(0)}$ minimizes $R(\theta, \varphi)$ among all invariant procedures.

(ii) Suppose conversely that $\varphi^{(0)}$ minimizes $R(\theta, \varphi)$ among all invariant procedures and let φ' be any procedure. Then there exists an invariant procedure φ'' such that $r(\theta, \varphi') = r(\theta, \varphi'')$. (This follows from Lemma 2 in Section 3, or can be verified directly by taking $\varphi'' = \sum g_i^* \varphi' g_i^{-1} / N$). We now have

$$r(\theta, \varphi') = r(\theta, \varphi'') = R(\theta, \varphi'') \geq R(\theta, \varphi^{(0)}) = r(\theta, \varphi^{(0)}),$$

2. The Bahadur-Goodman theorem. Let the random observable X have a density of the form

$$(4) \quad h_{\theta}(t) = C(\theta) f_{\theta_1}(t_1) \cdots f_{\theta_s}(t_s)$$

with respect to some σ -finite measure μ , where $\theta = (\theta_1, \dots, \theta_s)$, $t = (t_1, \dots, t_s)$ and where $t_i = t_i(x)$, $i = 1, \dots, s$, are real-valued statistics. Suppose that f_{θ_i} has monotone (non-decreasing) likelihood ratio in t_i for each $i = 1, \dots, s$. Consider the problem of selecting the largest among $\theta_1, \dots, \theta_s$ and let d_i denote the selection of θ_i .

The following are some typical examples of (4).

(i) X_{ij} ($j = 1, \dots, n; i = 1, \dots, s$) are independently normally distributed with mean θ_i and variance σ^2 .

(ii) X_i are independently distributed according to the binomial distribution $b(\theta_i, n)$.

(iii) X_i are independently distributed according to the Poisson distribution with mean θ_i .

(iv) (X_1, \dots, X_s) have a multinomial distribution with success probabilities $(\theta_1, \dots, \theta_s)$.

(v) A population consists of N items of s different types with $r_i = \theta_i N$ items of type i . A sample of n items is drawn at random. If X_i denotes the number of items of type i in the sample, then the joint distribution of (X_1, \dots, X_s) is the multiple hypergeometric distribution

$$P(X_1 = x_1, \dots, X_s = x_s) = \binom{r_1}{x_1} \cdots \binom{r_s}{x_s} / \binom{N}{n}, \quad (\sum r_i = N, \quad \sum x_i = n).$$

In all these examples, the distribution is of form (4) with $T_i = \sum X_{ij}$ in (i), and $T_i = X_i$ in the other cases.

Invariance can be introduced by assuming that $x = (x_1, \dots, x_N)$ and that there exist permutations of the x 's which leave μ invariant. It is perhaps simpler instead to reduce the problem first to the sufficient statistics $t = (t_1, \dots, t_s)$ whose joint density is given by (4) with respect to a σ -finite measure ν in t -space which we assume to be invariant under the group G of all permutations of (t_1, \dots, t_s) . (For a discussion of the relation between these approaches, see [9]; for the particular problem at hand it is also discussed in [2]).

THEOREM (Bahadur-Goodman). *Let the distribution of the sufficient statistics $T = (T_1, \dots, T_s)$ have density (4) with respect to a σ -finite measure ν which is invariant under G . For any permutation g of (t_1, \dots, t_s) define \bar{g} and g^* as the same permutation of $(\theta_1, \dots, \theta_s)$ and (d_1, \dots, d_s) respectively, and suppose that the loss function L satisfies (1) and*

$$(5) \quad \theta_i < \theta_j \Rightarrow L_i(\theta) \geq L_j(\theta).$$

Let $\varphi^{(0)}$ be the procedure which takes decision d_i when t_i is the unique largest among (t_1, \dots, t_s) , and which takes decisions d_{i_1}, \dots, d_{i_r} each with probability $1/r$ if $(t_{i_1}, \dots, t_{i_r})$ is the set of t -values equal to $\max t_j$ (i.e. which breaks ties at random). Then $\varphi^{(0)}$ uniformly minimizes the risk among all procedures based on t which are invariant under G .

PROOF. To prove this theorem, it is by Lemma 1 enough to show that the procedure $\varphi^{(0)}$, which is clearly invariant under G , uniformly minimizes $r(\theta, \varphi)$ among all procedures. Now

$$(6) \quad r(\theta, \varphi) = \sum_{i=1}^s \int A_i \varphi_i(t) d\nu(t)$$

where

$$(7) \quad A_i = \sum_{j=1}^N L_i(\bar{g}_j \theta) h_{\bar{g}_j \theta}(t).$$

Suppose without loss of generality that $\theta_1 < \dots < \theta_s$ and consider the difference $A_i - A_k$ for $i < k$. For any permutation g of $(1, \dots, s)$, say $g = (j_1, \dots, j_s)$, with $j_i < j_k$ let g' denote the permutation obtained from g by interchanging j_i and j_k . The contribution to $A_i - A_k$ from the terms (in (7)) corresponding to g and g' is

$$(8) \quad [L_i(\bar{g}\theta) - L_k(\bar{g}\theta)]h_{\bar{g}\theta}(t) + [L_i(\bar{g}'\theta) - L_k(\bar{g}'\theta)]h_{\bar{g}'\theta}(t).$$

Since the loss function is invariant, we have

$$L_i(\bar{g}\theta) = L_k(\bar{g}'\theta) \quad \text{and} \quad L_k(\bar{g}\theta) = L_i(\bar{g}'\theta)$$

so that (8) reduces to

$$(9) \quad [L_i(\bar{g}\theta) - L_k(\bar{g}\theta)][h_{\bar{g}\theta}(t) - h_{\bar{g}'\theta}(t)].$$

The second factor in (9) is \geq or ≤ 0 as

$$f_{\theta_{j_i}}(t_i) f_{\theta_{j_k}}(t_k) \quad \text{is} \quad \geq \quad \text{or} \quad \leq \quad f_{\theta_{j_k}}(t_i) f_{\theta_{j_i}}(t_k)$$

and hence, since $\theta_{j_i} < \theta_{j_k}$ and f_θ has monotone likelihood ratio, as t_i is \leq or $\geq t_k$.

Hence, for a given point t , the quantity A_i is a minimum for all values of i for which $t_i = \max t_j$. The average risk $r(\theta, \varphi)$ is therefore minimized for any procedure φ which puts $\varphi_i(t) = 0$ whenever $t_i < \max t_j$. Since Lemma 2 requires the procedure to be invariant under G , a minimum risk procedure is obtained by selecting a population with maximum t -value, where ties between maxima are broken at random.

We conclude this section with a discussion of two extensions of the theorem.

(i) One extension is indicated by the example of samples X_{ij} ($j = 1, \dots, n$; $i = 1, \dots, s$) from the normal distributions $N(\xi_i, \sigma_i^2)$. Suppose we wish to select the population with the smallest variance, with a loss function independent of the ξ 's, satisfying (1), and

$$(10) \quad \sigma_i^2 < \sigma_j^2 \Rightarrow L_i(\sigma_1^2, \dots, \sigma_s^2) \leq L_j(\sigma_1^2, \dots, \sigma_s^2).$$

If the ξ_i are known, the theorem applies with $t_i = \sum (x_{ij} - \xi_i)^2$. If they are unknown, the problem remains invariant not only under permutations of the samples but also under the translations

$$(11) \quad X'_{ij} = X_{ij} + c_i.$$

These induce in the parameter space the transformations

$$\xi_{i'} = \xi_i + c_i, \quad \sigma_{i'} = \sigma_i$$

and leave the decision space unchanged (i.e. $d_{i'} = d_i$).

A necessary and sufficient condition for a procedure to be invariant under these translations is that it only depend on the differences $X_{ij} - X_{ik}$ within samples, and for these differences the statistics $T_i = \sum (X_{ij} - X_{ik})^2$, $i = 1, \dots, s$, form a set of sufficient statistics. Application of the Bahadur-Goodman theorem to the distribution of the T 's now proves the selection of the population for which $T_i = \min T_j$, to uniformly minimize the risk among all procedures which are invariant both under the translations (11) and under permutation of the T_i (or of the samples).

(ii) The theorem extends easily to the ordered or unordered selection of a fixed number of populations greater than one. The situation is sufficiently well illustrated with the selection of two populations. A selection procedure may now be denoted by $\varphi = \{\varphi_{ij}\}$, the loss resulting from the selection of the i th and j th population by $L_{ij}(\theta)$. In the unordered case, (in which we do not specify which of the two selected populations is considered the best and which the second best) condition (5) is replaced by

$$(12) \quad \theta_i < \theta_j \Rightarrow L_{ik}(\theta) \geq L_{jk}(\theta) \quad \text{and} \quad L_{ki}(\theta) \geq L_{kj}(\theta).$$

The optimum procedure selects the two populations with the largest t -values, with ties being broken at random. This procedure uniformly minimizes the risk among all procedures that are invariant under G . If in the proof of the original theorem, A_i in Equations (6) and (7) is replaced by A_{ik} we must show that for fixed t , $A_{ik} = A_{ik}(t)$ is minimized by taking for i and k the subscripts of the two

two largest t 's. But this follows from the earlier argument by first holding i fixed and minimizing with respect to $k \neq i$ and then, having determined k , minimizing with respect to i .

In the ordered case, in which d_{ij} corresponds to the selection of θ_i as the largest and θ_j as the second largest value, we must add to (12) the condition

$$(13) \quad \theta_i < \theta_j \Rightarrow L_{ij}(\theta) \geq L_{ji}(\theta).$$

Then it is seen as in the original proof that

$$A_{ik} \leq A_{ki} \quad \text{if} \quad t_i \geq t_k,$$

and this, in conjunction with the result for the unordered case, establishes the optimum property of the obvious procedure in the ordered case.

3. Other optimum properties of $\varphi^{(0)}$. The fact that $\varphi^{(0)}$ uniformly minimizes the risk among all invariant procedures entails a number of other optimum properties. The approach which is essentially that of [10], is most simply given in terms of the general finite decision problem discussed in Section 1. In such a problem, which is assumed to remain invariant under a finite group G , suppose an ordering \leq is introduced among the procedures (read $\varphi \leq \varphi'$ as φ' is at least as good as φ).

LEMMA 2. (a) *If the ordering is such that*

$$(i) \quad \varphi \leq \varphi' \Rightarrow g^* \varphi g^{-1} \leq g^* \varphi' g^{-1}, \text{ and}$$

$$(ii) \quad \varphi \leq \varphi^{(i)}, \quad i = 1, \dots, r \Rightarrow \varphi \leq \sum_{i=1}^r \varphi^{(i)} / r$$

then given any procedure φ there exists an invariant procedure φ' such that $\varphi \leq \varphi'$.

(b) *Suppose that in addition to (i) and (ii) the ordering satisfies*

$$(iii) \quad R(\theta, \varphi') \leq R(\theta, \varphi) \text{ for all } \theta \Rightarrow \varphi \leq \varphi'.$$

Then, if there exists a procedure $\varphi^{(0)}$ that uniformly minimizes the risk among all invariant procedures, $\varphi^{(0)}$ is optimum with respect to the ordering \leq , i.e.

$$\varphi \leq \varphi^{(0)} \quad \text{for all } \varphi.$$

PROOF. (a) The argument is familiar from many special cases, for example Theorem 8.6.4. of Blackwell and Girshick (1954), and consists in noting that

$$\varphi' = \sum_{i=1}^r g_i^* \varphi g_i^{-1} / r$$

is invariant and, by (i) and (ii), at least as good as φ .

(b) This is an immediate consequence of (a).

Before applying the lemma we note that it remains valid if instead of considering all procedures φ , attention is restricted to some class \mathcal{C} of procedures which is closed under convex combinations and such that

$$\varphi \in \mathcal{C} \Rightarrow g^* \varphi g^{-1} \in \mathcal{C} \quad \text{for all } g \in G.$$

The following are examples of some orderings satisfying (i)–(iii).

EXAMPLE 1. *Minimax.* $\varphi \leq \varphi'$ if $\sup R(\theta, \varphi') \leq \sup R(\theta, \varphi)$

In this case, Lemma 2 reduces to Theorem 8.6.4 of Blackwell and Girshick (1954).

EXAMPLE 2. *Minimax regret.* For a finite decision problem this ordering is given by $\varphi \leq \varphi'$ if

$$\sup [R(\theta, \varphi') - \min_i L_i(\theta)] \leq \sup [R(\theta, \varphi) - \min_i L_i(\theta)]$$

EXAMPLE 3. *Average risk.* $\varphi \leq \varphi'$ if for some specified θ_0

$$\sum R(\bar{g}_i \theta_0, \varphi')/N \leq \sum R(\bar{g}_i \theta_0, \varphi)/N.$$

Lemma 2, in conjunction with the Bahadur-Goodman theorem proves, for example, that the procedure $\varphi^{(0)}$ of that theorem maximizes the minimum probability of a correct decision whenever the best population is sufficiently much better than the second best (and hence establishes an optimum property of the procedures given in [4], [5], [6] and [11]). This result was proved by the method of least favorable distributions by Hall (1959). It also follows that $\varphi^{(0)}$ minimizes, for example, $\sup [\max \theta_i - \sum \theta_j E_{\varphi_j}(X)]$, a measure of regret considered in [1] and [3].

REFERENCES

- [1] BAHADUR, RAGHU RAJ (1950). On a problem in the theory of k populations. *Ann. Math. Statist.* **20** 362-375.
- [2] BAHADUR, RAGHU RAJ and GOODMAN, LEO A. (1952). Impartial decision rules and sufficient statistics. *Ann. Math. Statist.* **23** 553-562.
- [3] BAHADUR, RAGHU RAJ and ROBBINS, HERBERT (1950). The problem of the greater mean. *Ann. Math. Statist.* **21** 469-487.
- [4] BECHHOFFER, ROBERT E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16-39.
- [5] BECHHOFFER, ROBERT E., ELENAGHRABI, SALAH and MORSE, NORMAN (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. *Ann. Math. Statist.* **30** 102-119.
- [6] BECHHOFFER, ROBERT E. and SOBEL, MILTON (1954). A single-sample multiple decision procedure for ranking variances of normal populations. *Ann. Math. Statist.* **25** 273-289.
- [7] BLACKWELL, DAVID and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.
- [8] HALL, WM. JACKSON (1959). The most-economical character of some Bechhofer and Sobel decision rules. *Ann. Math. Statist.* **30** 964-969.
- [9] HALL, W. J., WIJSMAN, R. A., and GHOSH, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575-614.
- [10] LEHMANN, E. L. (1959). Optimum invariant tests. *Ann. Math. Statist.* **30** 881-884.
- [11] SOBEL, MILTON and HUYETT, MARYLIN J. (1957). Selecting the best one of several binomial populations. *Bell System Tech. J.* **36** 537-576.