SOME RECENT DEVELOPMENTS IN MULTIPLE DECISION THEORY: A BRIEF OVERVIEW*

by

Shanti S. Gupta, Purdue University
and
Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-26

October 1979

*This research was supported by the Office of Naval Research under Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

SOME RECENT DEVELOPMENTS IN MULTIPLE DECISION THEORY: A BRIEF OVERVIEW*

by

Shanti S. Gupta, Purdue University
and
Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

Many statistical situations in which it is customary to employ hypothesis testing, really involve a choice between more than two decisions. In such problems, when the hypothesis is rejected, one wants to know in which of a number of possible ways the actual situation (true state of nature) differs from the one postulated by the null hypothesis. By formulating the problem as one involving only two decisions, we not only neglect to differentiate between certain alternative decisions, which may differ considerably in their consequences, but we may also be using an inappropriate acceptance region for the hypothesis.

The traditional approach to hypotheses testing problems is inadequate and unrealistic in the sense that it is not formulated in a way to answer the experimenter's question, namely, how to identify the best population? If fact, the method of least significant differences based on the t-test has been used in the past to detect differences between the true unknown means of

---

*This research was supported by the Office of Naval Research under Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

different varieties and thereby choosing the population which is "best" i.e. the one with the largest mean. But this method is indirect, less efficient and does not provide an overall probability of a correct selection. If the null hypothesis is not rejected, then there is no decision made in the traditional approach. Furthermore, when performing a test one may commit one of two errors: rejecting the hypothesis when it is true (error of the first kind) or accepting it when it is false (error of the second kind). It is desirable to carry out the test in a manner which keeps the probabilities of the two types of error to a minimum. It is customary to assign a bound to the probability of incorrectly rejecting the hypothesis when it is true, and to attempt to minimize the other probability subject to this condition. Unfortunately, when the number of observations is given, both probabilities cannot be controlled simultaneously by the classical approach (see Lehmann (1959)). The decision-theoretic approach provides an effective tool to overcome these difficulties in some reasonable ways. Actually, the cases described above can be formulated as general multiple decision problems. To this end, we start by defining the space $\mathfrak{a}$ of actions of the statistician consisting of a finite number ($k \geq 2$) of points, $\mathfrak{a} = \{a_1, \ldots, a_k\}$. There are two distinct types of multiple decision problems that seem to arise in practice. In one the parameter space Θ is partitioned into k subsets $\Theta_1, \ldots, \Theta_k$, according to the increasing value of a single real-valued function $r(\underline{\theta})$. The action a_i is preferred if $\underline{\theta} \in \Theta_i$. This type of multiple decision problem is called monotone. This approach has been studied by Karlin and Rubin (1956) and Brown, Cohen and Strawderman (1976). For example, if an experimenter is comparing two treatments with means θ_1 and θ_2, he might have available to him only a finite number of actions among which he has preference based on the magnitude of the differences of the means $\theta_2-\theta_1$: A particular case occurs when he may choose from the three alternatives:

(i) prefer treatment 1 over treatment 2,

(ii) prefer treatment 2 over treatment 1,

(iii) no preference (cf. Ferguson 1967).

Another important class of multiple decision problems for selection is where the treatments are classified into a superior category (the selected items) and an inferior one. In general, selection problems have been treated under several different formulations. A basic distinction corresponds to Model I and II cases in the analysis of variance. In Model I, the treatments being classified are considered fixed; only the observations made on each of them are random. In Model II, on the other hand, the treatments themselves are drawn at random from some population and would, therefore, change under replications of the experiment.

Lehmann (1961) (and earlier authors, see Paulson (1949), Bechhofer (1954), and Gupta (1956)) have formulated the selection problems according to the following goals:

Goal 1. We wish to select only a single population (if possible the best one): the variety to be planted, the production method we are going to adopt, etc. As a slight generalization, we may wish to select a fixed number, say two or three.

Goal 2. The number of populations to be selected (or the subset size) is not fixed in advance but is determined by the observations. This arises, for example, when we wish to select all worthwhile treatments, or if we want to be reasonably sure that the selected group contains the best treatment.

Goal 3. The subset size is determined by the observed data subject to an upper bound specified in advance. It may, for example, be desirable to investigate all treatments that appear promising but budget restrictions may limit the research program to the investigation of at most three treatments.

Among the early investigators of multiple decision procedures are Paulson (1949), Bahadur (1950), Bahadur and Robbins (1950). The formulation of such procedures in the framework of selection and ranking procedures has been generally accomplished either using the indifference zone approach or the (random-size) subset selection approach. The former approach was introduced by Bechhofer (1954). Substantial contributions to the early and subsequent developments in the subset selection theory have been made by Gupta starting from his work in 1956.

Bechhofer (1954) considered the problem of ranking k normal means. In order to explain the basic formulation, consider the problem of selecting the population with the largest mean from k normal populations with unknown means μ_i, i = 1,2,...,k, and a common known variance σ^2. Let $\bar{X}_i$, i = 1,...,k, denote the means of independent samples of size n from these populations. The "natural" procedure (which has been shown to have optimum properties by Hall (1959) and Eaton (1967)) is to select the population that yields the largest $\bar{X}_i$'s. The experimenter would, of course, need a guarantee that this procedure will pick the population having the largest μ_i with a probability not less than a specified level P*. For the problem to be meaningful P* lies between $\frac{1}{k}$ and 1. Since we do not know the true configuration of the μ_i, we look for the least favorable configuration (LFC) for which the probability of a correct selection (PCS) is minimized. This LFC is given by $\mu_1 = \ldots = \mu_k$; the corresponding PCS = 1/k and hence the probability guarantee cannot be met whatever be the sample size n. A natural modification is to insist on the minimum probability guarantee whenever the best population is sufficiently superior to the next best. In other words, the experimenter specifies a positive constant Δ^* and requires that the PCS is at least P* whenever $\mu_{[k]} - \mu_{[k-1]} \geq \Delta^*$, where $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$ denote the ordered means. Now the minimization of PCS is over the part Ω_{Δ^*} of the parameter space in which $\mu_{[k]} - \mu_{[k-1]} \geq \Delta^*$. The complement of Ω_{Δ^*} is called the indifference zone for obvious reason. The LFC in Ω_{Δ^*} is given by $\mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \Delta^*$. The problem is then to determine the minimum sample size required in order to have PCS $\geq$ P* for all $\underline{\mu}$ such that $\mu_{[k]} - \mu_{[k-1]} \geq \Delta^*$.

Bechhofer's formulation is more general than what is described above. His general ranking problem includes, for example, selection of the t best populations.

In the subset selection approach, the goal is to select a nonempty subset of the populations so as to include the best population. Here the size of the selected subset is random and is determined by the observations themselves. In the case of normal populations with unknown means $\mu_1, \ldots, \mu_k$, and a common known variance σ^2, the rule proposed by Gupta (1956) selects the population that yields $\bar{X}_i$ iff $\bar{X}_i \geq \max_{1 \leq j \leq k} \bar{X}_j - \frac{d\sigma}{\sqrt{n}}$, where $d = d(k, P^*) > 0$ is determined so that the PCS is at least P*. Here a correct selection is the selection of any subset that includes the population with the largest μ_i. Thus, the LFC is with regard to the whole parameter space Ω. Under this formulation, for given k and P* we determine the constant d. The rule explicitly involves n. In general, the rule will involve a constant which depends on k, P* and n. The performance of a subset selection procedure is studied by evaluating the expected subset size and its supremum over the parameter space Ω.

To study useful good procedures, we look for some optimal rules in a small class of decision rules. Using the general decision theory of Wald (1950), we can construct optimal selection procedures according to several different approaches. The indifference zone and subset selection approaches can be combined to guarantee the infimum of the probability of a correct selection over a preference zone and minimize the supremum of the expected size of selected subset over the parameter space (cf. Gupta and Huang (1976)). In many cases, we do not know whether the true parametric configuration belongs to the preference zones. If the best and the second best are not very much different it is reasonable to select a subset. We would like to keep the subset size under control. This can be achieved by increasing the sample size which also increases the probability of a correct selection. Sometimes, we have some partial information about the parameter space; in this case, the so-called Γ-minimax selection rules have been studied by Gupta and Huang (1977).

In general, we are interested in introducing a wide class of decision rules, the so-called subset selection procedures for the k populations $\pi_1,\ldots,\pi_k$. We are given one observation x_i from each population π_i, $i = 1,2,\ldots,k$. The vector of observations is $\underline{x} = (x_1,\ldots,x_k)$. Each π_i is characterized by a parameter θ_i. Let $\mathcal{G}$ denote the action space consisting of the 2^k subsets of the set $\{1,2,\ldots,k\}$. A measurable function d defined on $\mathcal{X} \times \mathcal{G}$ is called a selection procedure provided that for each fixed k-vector of observations $\underline{x} \in \mathcal{X}$, $d(\underline{x},a) \geq 0$ and $\Sigma_{\mathcal{G}} d(\underline{x},a) = 1$. Thus if $\underline{X} = \underline{x}$ is an observed vector of observations, $d(\underline{x},a)$ is the probability of selecting the subset $a \in \mathcal{G}$. It can be seen that

$\delta_i(\underline{x}) = \sum_{a \ni i} d(\underline{x},a)$ (summation over those subsets a containing i) denotes the probability of selecting the ith population. The functions $\delta_1, \delta_2, \ldots, \delta_k$ will be referred to as the individual selection probabilities. Note that the selection procedure d is completely specified by the individual selection probabilities whenever the latter take on only the values zero and one.

In studying the performance of subset selection procedures, the measures of loss most often used have been related to incorrect selection $ICS(\underline{\theta},a)$ where $\underline{\theta} = (\theta_1, \ldots, \theta_k)$, and the number of elements $|a|$ in the selected subset a. More recently, Goel and Rubin (1977) studied the subset selection problem from a Bayesian point of view using loss functions that are linear combinations of $\theta_{[k]} - \max_{j \in a} \theta_j$ and $|a|$. Bickel and Yahav (1977) studied the behavior of Bayes procedures as $k \to \infty$ using loss functions that are linear combinations of $ICS(\underline{\theta},a)$ and $\theta_{[k]} - \sum_{j \in a} \theta_j/|a|$. Chernoff and Yahav (1977), employing Monte Carlo techniques, compared the integrated risks with respect to exchangeable normal priors of Bayes procedures, and Bechhofer type procedures using loss functions that are linear combinations of $\theta_{[k]} - \max_{j \in a} \theta_j$ and $\theta_{[k]} - \sum_{j \in a} \theta_j/|a|$. Gupta and Hsu (1978) used Monte Carlo study which parallels that of Chernoff and Yahav in that exchangeable normal priors are used but differs in that the loss functions considered are linear combinations of $ICS(\underline{\theta},a)$ and $|a|$. Several classical methods have been compared by Gupta and Hsu (1978). The four loss combinations that have been used are presented in Figure 1.

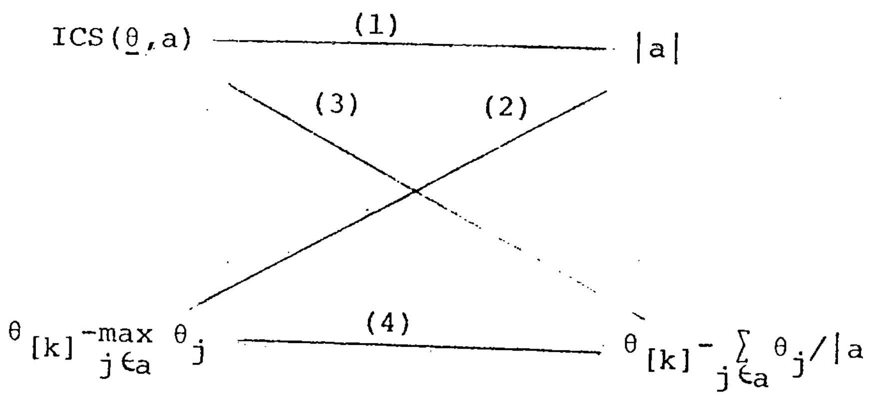

Figure 1

Note that the different combinations have different interpretations. The combinations (1) and (2) correspond to situations where the subset selection procedure is used as a screening procedure. For example, in developing a new drug, a pharmaceutical company may start with a number of ingredients known to have beneficial effects (and side effects) from previous experiments, and then obtain a collection of potentially good formulae for combining these ingredients in different proportions. After the first stage of testing, one wants to reject those formulae that are evidently non-best and retain those that still seem potentially best for further study. Eventually, if the development is successful, only one formula will be marked. Corresponding to this situation then, loss functions that depend only on the best selected and the size selected are reasonable. On the other hand, the component $\theta_{[k]} - \sum_{j \in a} \theta_j/|a|$ in the combinations (3) and (4) correspond to situation where all those selected will be used. This is the case, for example, when one purchases stocks for long term investment. One would purchase stocks of more than one company to guard against the possibility of gross errors, and all the stocks purchased contribute to the gain or loss.

Investigation of these problems more in the framework of general multiple decision theory is a topic of current research interest. The field of selection and ranking problems has grown steadily over the years as also evidenced by the publication of the following books: Bechhofer, Kiefer and Sobel (1968), Gibbons, Olkin and Sobel (1977), Gupta and Panchapakesan (1979) and Gupta and Huang (1979).

A related problem (see Goal 3) in which the number of populations selected does not exceed a given number m (< k), was solved by Gupta and Santner (1973) and Santner (1975). More recently, Kiefer (1975, 1976, 1977) has made several significant contributions in the very interesting area of multiple decision problems based on conditional inference.

References

[1] Bahadur, R. R. (1950). On a problem in the theory of k populations. Ann. Math. Statist. 21, 362-375.

[2] Bahadur, R. R. and Robbins, H. (1950). The problem of the greater mean. Ann. Math. Statist. 21, 469-87; Correction: 22 (1951), 310.

[3] 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.

[4] Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. The University of Chicago Press, Chicago.

[5] Bickel, P. J. and Yahav, J. A. (1977). On selecting a set of good populations. Statistical Decision Theory and Related Topics II, (ed. S. S. Gupta and D. S. Moore), 37-55. Academic Press.

[6] Brown, L. D., Cohen, A. and Strawderman, W. E. (1976). A complete class theorem for strict monotone likelihood ratio with applications. Ann. Statist. 4, 712-722.

[7] Chernoff, H. and Yahav, J. A. (1977). A subset selection problem employing a new criterion. Statistical Decision Theory and Related Topics II, (ed. S. S. Gupta and D. S. Moore), 93-119. Academic Press.

[8] Eaton, M. L. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist. 38, 124-137.

[9] Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

[10] Gibbons, J. D., Olkin, I. and Sobel, M. (1977). Selecting and Ordering Populations. John Wiley & Sons, New York.

[11] Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population - a Bayesian approach. Ann. Statist. 5, 969-983.

[12] Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. thesis, Univ. of North Carolina.

[13] Gupta, S. S. and Huang, D. Y. (1977). On some Γ-minimax selection and multiple comparison procedures. Statistical Decision Theory and Related Topics (Ed. Gupta, S. S. and Moore, D. S.), pp. 139-155.

[14] Gupta, S. S. and Huang, D. Y. (1976). A note on optimal subset selection procedure. Dept. of Statist., Mimeo Ser. #470, Purdue Univ., W. Lafayette, IN 47907.

[15] Gupta, S. S. and Huang, D. Y. (1979). Multiple Statistical Decision Theory. To appear in Lecture Note Series, Springer Verlag, New York.

[16] Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Methodology of Selection and Ranking. To appear in John Wiley & Sons.

[17] Gupta, S. S. and Santner, T. J. (1973). On selection and ranking procedures - A restricted subset selection rule. Bull. Inst. Internat. Statist. Vol. 45, Book 1, pp. 478-486.

[18] Gupta, S. S. and Hsu, J. C. (1978). On the performance of some subset selection procedures. Commun. Statist.-Simula. Computa., B7(6), 561-591.

[19] Hall, W. J. (1959). The most economical character of Bechhofer and Sobel decision rules. Ann. Math. Statist. 30, 964-969.

[20] Karlin, S. and Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. Ann. Math. Statist. 27, 272-299.

[21] Kiefer, J. (1975). Conditional confidence approach in multidecision problems. Proceedings of the Fourth Dayton Multivariate Conference, ed. P. R. Krishnaiah, Amsterdam: North Holland Publishing Co., 143-158.

[22] Kiefer, J. (1976). Admissibility of conditional confidence procedures. Ann. Statist. 4, 836-865.

[23] Kiefer, J. (1977). Conditional confidence statements and confidence estimators. JASA, 72, 789-808.

[24] Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley, New York.

[25] Lehmann, E. L. (1961). Some model I problems of selection. Ann. Math. Statist. 32, 990-1012.

[26] Paulson, E. (1949). A multiple decision procedure for certain problems in analysis of variance. Ann. Math. Statist. 20, 95-98.

[27] Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. Ann. Statist. 3, 334-349.

[28] Wald, A. (1950). Statistical Decision Functions, New York: John Wiley & Sons.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #79-26	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SOME RECENT DEVELOPMENTS IN MULTIPLE DECISION THEORY: A BRIEF OVERVIEW		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #79-26
7. AUTHOR(s) Huang, D. Y. and Gupta, S. S.		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE October 1979
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for pubic release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple decision problems, selection and ranking, multiple comparisons, conditional confidence procedure, Bayes, indifference zone approach, subset selection approach.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper surveys some recent developments in multiple decision (selection and ranking) theory. The spirit and philosophy underlying this approach to comparing k populations are discussed briefly. Criteria to construct optimal multiple decision procedures, which are based on some 'modifications' of decision theory, are discussed. The combination of the indifference zone and the subset selection approach into a general framework is described.