*The Annals of Statistics*
1979, Vol. 7, No. 6, 1333–1338

# MINIMAX SUBSET SELECTION FOR LOSS MEASURED BY SUBSET SIZE[1]

BY ROGER L. BERGER

*Florida State University*

A subset selection problem is formulated as a multiple decision problem. Then, restricting attention to rules which attain a certain minimum probability of correct selection, the minimax value is computed, under general conditions, for loss measured by subset size and number of non-best populations selected. Applying this to location and scale problems, previously proposed rules are found to be minimax. But for problems involving binomial, multinomial and multivariate noncentrality parameters, such as $\chi^2$ and $F$, previously proposed rules are found to be not minimax.

**1. Introduction.** A subset selection problem is a multiple decision problem which has the goal of determining in which of $k$ partition sets of the parameter space the true parameter lies. Restricting attention to rules which insure a certain minimum probability, $P^*$, of making a correct decision, minimaxity is investigated for loss measured by subset size. The minimax value is found to be $kP^*$ under general conditions involving only the topological structure of the parameter space and the continuity of certain functions of the parameter. These results include problems involving nuisance parameters and (possibly unequal) sample sizes greater than one. Using these results, rules proposed by Gupta (1965) are found to be minimax in location and scale parameter problems. Other rules, proposed for selection in terms of binomial and multinomial probabilities and multivariate noncentrality parameters, are shown to be not minimax.

**2. Multiple decision theory formulation.** A subset selection problem may be formulated as a multiple decision theory problem. The specific choice of the action space sets the subset selection problem apart from other multiple decision theory problems.

$\mathcal{X} \subset R^q$ is the sample space. $\Theta \subset R^r$ is the parameter space. The observation $\mathbf{X} = (X_1, \cdots, X_q)$ is a random vector with cumulative distribution function (cdf) $F(\mathbf{x}; \boldsymbol{\theta})$. It is assumed that there exists a partition of $\Theta$ denoted by $\{\Theta_i \colon i = 1, \cdots, k\}$ ($k \geq 2$). Often this partition is determined by the largest or smallest coordinate of (some subset of) the parameter. If a particular parameter point could be placed in more than one set of this partition, e.g., two coordinates of the

Received September 1977; revised September 1978.

[1]This paper forms a part of the author's doctoral dissertation submitted to Purdue University. This research was supported by the Office of Naval Research Contract N00014-75-C-0455 and the Air Force Office of Scientific Research Contract AFOSR-72-2350E.

*AMS* 1970 *subject classifications.* Primary 62F07; secondary 62C05.

*Key words and phrases.* Minimax subset selection, expected subset size, expected number of non-best populations, multiple decision.

parameter are tied as largest, then the point is arbitrarily put in one of the sets. This is done so the partition is well defined and, in some problems, this insures the continuity of the risk functions. The general goal of a subset selection problem is to determine, based on the observation, which of the $k$ partition sets contains the true parameter. The action space $\mathcal{Q}$ consists of the $2^k - 1$ nonempty subsets of $\{\pi_1, \cdots, \pi_k\}$ where $\pi_i$ is the statement $\boldsymbol{\theta} \in \Theta_i$. So the action $\{\pi_1, \pi_2\}$ corresponds to the decision $\boldsymbol{\theta} \in \Theta_1 \cup \Theta_2$. The $\pi_i$'s correspond to what have been called populations in the earlier subset selection literature. In this terminology, for a given $\boldsymbol{\theta}$, the "best population" is the one true $\pi_i$ and the other $(k - 1)\pi_i$'s are the "nonbest populations." So a statement like "the best population is the one associated with the largest parameter value" means $\Theta_i = \{\boldsymbol{\theta}: \theta_i = \max_{1 \leqslant j \leqslant r} \theta_j\}$ (with the exception that if $\theta_i$ tied with other $\theta_j$'s as the largest, that parameter point may not be in $\Theta_i$). By not assuming equality of $k$, $q$ and $r$, this formulation covers problems involving nuisance parameters and (possibly unequal) sample sizes greater than one.

A measurable function $\delta$: $\mathcal{X} \times \mathcal{Q} \rightarrow [0, 1]$, is called a *selection rule* provided that, for each $\mathbf{x} \in \mathcal{X}$, $\Sigma_{\mathcal{Q}} \delta(\mathbf{x}, a) = 1$. $\delta(\mathbf{x}, a)$ is the probability of selecting subset $a$ having observed $\mathbf{x}$. The $k$ functions defined by $\psi_i(\mathbf{x}) = \Sigma_{\{a:\ \pi_i \in a\}} \delta(\mathbf{x}, a)$ are the *individual selection probabilities*. $\psi_i(\mathbf{x})$ is the probability of including $\pi_i$ in the selected subset having observed $\mathbf{x}$. The risk of any rule, for the losses considered herein, can be computed in terms of the individual selection probabilities. For this reason, any two rules which have the same individual selection probabilities shall be considered equivalent.

The selection of any subset which contains the best population is called a correct selection, denoted by CS. Let $P^*$ be any preassigned fixed number such that $1/k < P^* < 1$. As is traditional, the only selection rules to be considered are those which satisfy the *$P^*$-condition*, viz., $\inf_\Theta P_{\boldsymbol{\theta}}(\mathrm{CS}|\psi) \geqslant P^*$. This is obviously equivalent to the following $k$ inequalities being satisfied,

$$\inf_{\Theta_i} E_{\boldsymbol{\theta}} \psi_i(\mathbf{X}) = \inf_{\Theta_i} P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi) \geqslant P^*, \qquad i = 1, \cdots, k.$$

The set of all selection rules which satisfy the $P^*$-condition is denoted by $\mathfrak{D}_{P^*}$.

Having insured a high probability of correct selection through the $P^*$-condition, one would prefer a rule which selects small subsets. To reflect this, the loss used in this paper is the number of populations selected, $S$. So the risk of a selection rule, $R(\boldsymbol{\theta}, \psi)$, is given by the expected subset size, $E_{\boldsymbol{\theta}}(S|\psi)$. Other reasonable loss functions are discussed by Bickel and Yahav (1977), Chernoff and Yahav (1977) and Goel and Rubin (1977).

## 3. Minimax value for the loss $S$.

THEOREM 3.1. *Let $\Theta_0 = \{\boldsymbol{\theta} \in \Theta: \boldsymbol{\theta} \in \overline{\Theta}_i$ for all $i = 1, \cdots, k\}$ where $\overline{\mathcal{Q}}$ denotes the closure of $\mathcal{Q}$. Suppose there exists $\boldsymbol{\theta}_0 \in \Theta_0$ such that $P_{\boldsymbol{\theta}}($select $\pi_i|\psi)$ is upper semicontinuous at $\boldsymbol{\theta}_0$ for all $\psi \in \mathfrak{D}_{P^*}$ and all $i = 1, \cdots, k$. Then $\inf_{\mathfrak{D}_{P^*}} \sup_\Theta E_{\boldsymbol{\theta}}(S|\psi) = kP^*$.*

PROOF. The risk at $\boldsymbol{\theta}_0$ is

$$E_{\boldsymbol{\theta}_0}(S|\psi) = \Sigma_{i=1}^{k} P_{\boldsymbol{\theta}_0}(\text{select } \pi_i|\psi). \tag{3.1}$$

The "no data rule" defined by $\psi_i^*(\mathbf{x}) \equiv P^*$, $i = 1, \cdots, k$, has $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi^*) = P^*$ for all $\boldsymbol{\theta}$ and all $i$. So $E_{\boldsymbol{\theta}}(S|\psi^*) = kP^*$ for all $\boldsymbol{\theta}$ and the minimax value is no greater than $kP^*$.

On the other hand, since $\boldsymbol{\theta}_0 \in \overline{\Theta}_i$ and $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi)$ is upper semicontinuous at $\boldsymbol{\theta}_0$,

$$P_{\boldsymbol{\theta}_0}(\text{select } \pi_i|\psi) \geqslant \inf_{\Theta_i} P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi) = \inf_{\Theta_i} P_{\boldsymbol{\theta}}(\text{CS}|\psi) \geqslant P^*$$

for any $\psi \in \mathfrak{D}_{P^*}$. So

$$\sup_{\Theta} E_{\boldsymbol{\theta}}(S|\psi) \geqslant E_{\boldsymbol{\theta}_0}(S|\psi) \geqslant kP^* \tag{3.2}$$

for any $\psi \in \mathfrak{D}_{P^*}$. Thus the minimax value is no less than $kP^*$. □

REMARK 3.1. If $\Theta = I \times I \times \cdots \times I$ ($k$ times) where $I$ is an interval on the real line and if the best is defined in terms of the largest or smallest coordinate of the parameter, then $\Theta_0 = \{\boldsymbol{\theta} = (\theta, \theta, \cdots, \theta): \theta \in I\}$. If $\mathbf{X}$ has a multinomial distribution, $\Theta = \{(\theta_1, \cdots, \theta_k): \theta_i \geqslant 0, \Sigma_{i=1}^{k}\theta_i = 1\}$ and $\Theta_0$ is the single point $(1/k, \cdots, 1/k)$. It may be argued that in problems like these any action is acceptable to the experimenter if $\boldsymbol{\theta} \in \Theta_0$. In this case, one could set the loss to zero, not $S$, for $\boldsymbol{\theta} \in \Theta_0$. This would yield $R(\boldsymbol{\theta}, \psi) = 0$ for $\boldsymbol{\theta} \in \Theta_0$. But, even allowing this, Theorem 3.1 remains true, in the usual case (see Remark 3.2) where $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi)$ is continuous in $\boldsymbol{\theta}$, for (3.2) can be replaced by

$$\sup_{\Theta} E_{\boldsymbol{\theta}}(S|\psi) \geqslant \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} E_{\boldsymbol{\theta}}(S|\psi) \geqslant kP^*.$$

REMARK 3.2. The upper semicontinuity assumption of Theorem 3.1 is much less formidable than it appears. For example, Chung (1970) (problem 10, page 100) can be generalized to state that if $\mathbf{X}$ has a density $f(\mathbf{x}; \boldsymbol{\theta})$ with respect to a sigma finite measure $\mu$ and if $f(\mathbf{x}; \boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_0$ (as a function of $\boldsymbol{\theta}$) for almost all $(\mu)\mathbf{x}$, then $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi)$ is continuous at $\boldsymbol{\theta}_0$.

REMARK 3.3. Another loss, closely related to $S$, which has been considered in subset selection problems is $S'$, the number of nonbest populations selected. Under the assumptions of Theorem 3.1, the minimax value for $S'$ is $(k-1)P^*$. Under the assumptions of Theorem 3.1, if $\psi \in \mathfrak{D}_{P^*}$ is minimax with respect to $S$, then $\psi$ is minimax with respect to $S'$ since

$$\sup_{\Theta} E_{\boldsymbol{\theta}}(S'|\psi) = \sup_{\Theta}\{E_{\boldsymbol{\theta}}(S|\psi) - P_{\boldsymbol{\theta}}(\text{CS}|\psi)\} \leqslant \sup_{\Theta}\{E_{\boldsymbol{\theta}}(S|\psi) - P^*\}$$

$$= kP^* - P^* = (k-1)P^*.$$

Theorem 3.1 will be used to show that in location and scale parameter problems, two rules proposed by Gupta (1965) are minimax. Consider the case in which the population associated with the largest parameter value is best. With the appropriate modifications, analogous results could be obtained if the population associated with the smallest value is best.

Gupta (1965) proposed and studied the following two rules. For a location parameter problem, define the rule $R_1$ by

(3.3) $$R_1\text{: select } \pi_i \text{ iff } x_i \geqslant \max_{1 \leqslant j \leqslant k} x_j - d \qquad i = 1, \cdots, k$$

where $d > 0$ is the smallest constant such that the $P^*$-condition is satisfied. For a scale parameter problem, define the rule $R_2$ by

(3.4) $$R_2\text{: select } \pi_i \text{ iff } x_i \geqslant c \cdot \max_{1 \leqslant j \leqslant k} x_j \qquad i = 1, \cdots, k$$

where $0 < c < 1$ is the largest constant such that the $P^*$-condition is satisfied.

THEOREM 3.2. *Suppose $X_1, \cdots, X_k$ are independent. Suppose $\boldsymbol{\theta}$ is a location (scale) parameter and $X_i$ has density $f_{\theta_i}(x_i) = f(x_i - \theta_i)(f(x_i/\theta_i)/\theta_i)$ with respect to Lebesgue measure, $\mu$, on the real line ((0, $\infty$)). Suppose $f_\theta(x)$ has monotone likelihood ratio. Then $R_1(R_2)$ is minimax.*

PROOF. Gupta (1965) proved that under these assumptions,

$$\sup_\Theta E_{\boldsymbol{\theta}}(S|R_1(R_2)) = \sup_{\Theta_0} E_{\boldsymbol{\theta}}(S|R_1(R_2)) = kP^*.$$

The continuity assumption of Theorem 3.1 is satisfied (see Royden (1968) problem 17, chapter 4). The result follows from Theorem 3.1. □

Theorem 3.2 generalizes a result of Gupta and Studden (1966). They proved that $R_1(R_2)$ is minimax among all permutation invariant rules in $\mathfrak{D}_{P^*}$.

REMARK 3.4. A natural question is which minimax rules are also admissible. For example, the no data rule, $\psi^*$, of Theorem 3.1 is minimax but inadmissible since it is dominated by $R_1$ (under the assumptions of Theorem 3.2, for example). Berger and Gupta (1977) showed that $R_1$ is admissible in a class of invariant rules when loss is measured by the maximum probability of an incorrect selection. But the question of admissibility in the class $\mathfrak{D}_{P^*}$ for loss measured by subset size remains an open question.

**4. Necessary conditions for minimaxity.** Any minimax selection rule must satisfy certain equalities on the set $\Theta_0$. These necessary conditions are principally of use in proving that certain rules, in violating these conditions, are not minimax. Theorem 4.1 provides the necessary conditions for minimaxity.

THEOREM 4.1. *Let $\psi$ be a minimax rule. Suppose $P_{\boldsymbol{\theta}}$(select $\pi_i|\psi$) is upper semicontinuous for all $i = 1, \cdots, k$ at $\boldsymbol{\theta}_0 \in \Theta_0$. Then*

(a) $P_{\boldsymbol{\theta}_0}$(*select* $\pi_i|\psi) = P^* = \inf_\Theta P_{\boldsymbol{\theta}}(\text{CS}|\psi)$ *for all* $i = 1, \cdots, k$;

(b) $P_{\boldsymbol{\theta}_0}(\text{CS}|\psi) = P^* = \inf_\Theta P_{\boldsymbol{\theta}}(\text{CS}|\psi)$;

(c) $E_{\boldsymbol{\theta}_0}(S|\psi) = kP^* = \sup_\theta E_{\boldsymbol{\theta}}(S|\psi)$.

PROOF. As in the proof of Theorem 3.1, it follows that

(4.1) $$P_{\boldsymbol{\theta}_0}(\text{select } \pi_i|\psi) \geqslant P^* \qquad \text{for all} \quad i = 1, \cdots, k.$$

By considering the "no data rule" $\psi_i^*(\mathbf{x}) \equiv P^*$, it follows that the minimax value is no greater than $kP^*$ so, since $\psi$ is minimax and (4.1) is true,

$$kP^* \geq \sup_{\Theta} E_{\boldsymbol{\theta}}(S|\psi) \geq E_{\boldsymbol{\theta}_0}(S|\psi) = \Sigma_{i=1}^k P_{\boldsymbol{\theta}_0}(\text{select } \pi_i|\psi) \geq kP^*.$$

All the inequalities are equalities and (a) and (c) are true. (b) follows from (a) since $P_{\boldsymbol{\theta}_0}(\mathrm{CS}|\psi) = P_{\boldsymbol{\theta}_0}(\text{select } \pi_i|\psi)$ where $\boldsymbol{\theta}_0 \in \Theta_i$. □

REMARK 4.1. Gupta and Nagel (1971) found that a condition related to condition (b) of Theorem 4.1, viz., $\inf_{\Theta} P_{\boldsymbol{\theta}}(\mathrm{CS}|\psi) = \inf_{\Theta_0} P_{\boldsymbol{\theta}}(\mathrm{CS}|\psi)$, was an important property of just selection rules. Conditions (a) and (b) of Theorem 4.1 have long been recognized (cf. Gupta and Studden (1966)) as intuitively appealing properties of selection rules. Santner (1975) gives conditions under which $\sup_{\Theta} E_{\boldsymbol{\theta}}(S|\psi) = \sup_{\Theta_0} E_{\boldsymbol{\theta}}(S|\psi)$ which from condition (c) of Theorem 4.1 is a necessary condition for minimaxity.

REMARK 4.2. If $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi)$ is a continuous function of $\boldsymbol{\theta}$, condition (a) of Theorem 4.1 requires that $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi)$ be constant on $\Theta_0$ if $\psi$ is to be minimax. Rules which do not satisfy this condition and so are not minimax include the following: the rule $R_1$ (see (3.3)) proposed for the binomial selection problem by Gupta and Sobel (1960); the rule $R_2$ (see (3.4)) proposed for the noncentral $\chi^2$ and noncentral $F$ selection problem by Gupta and Studden (1970); and the rule $R_2$ proposed for the multiple correlation coefficient selection problem by Gupta and Panchapakesan (1969). Gupta and Panchapakesan (1972) studied a general class of rules and gave conditions under which $P_{\boldsymbol{\theta}}(\text{select } \pi_i|\psi)$ will not be constant on $\Theta_0$ for rules in their class.

REMARK 4.3. Gupta and Nagel (1967) proposed using rule $R_1$ in the multinomial selection problem. They found that for some values of $k$ and $P^*$ the $\inf_{\Theta} P_{\boldsymbol{\theta}}(\mathrm{CS}|R_1)$ did not occur at $(1/k, \cdots, 1/k)$. So $R_1$ is not minimax for these values of $k$ and $P^*$.

**Acknowledgment.** My thanks go to my thesis advisor Professor Shanti S. Gupta. His advise and encouragement were most helpful throughout the preparation of this paper.

## REFERENCES

[1] BERGER, R. L. and GUPTA, S. S. (1977). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. Mimeo Series No. 495, Dept. Statist., Purdue Univ.

[2] BICKEL, P. J. and YAHAV, J. A. (1977). On selecting a set of good populations. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. S. Moore, eds.), 37–55 Academic Press.

[3] CHERNOFF, H. and YAHAV, J. (1977). A subset selection problem employing a new criterion. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. S. Moore, eds.), 93–119 Academic Press.

[4] CHUNG, K. L. (1974). *A Course in Probability Theory*. Academic Press.

[5] GOEL, P. K. and RUBIN, H. (1977). On selecting a subset containing the best population—a Bayesian approach. *Ann. Statist.* **5** 969–983.

[6] GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7** 225–245.

[7] GUPTA, S. S. and NAGEL, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. *Sankhyā Ser. B* **29** 1–34.

[8] GUPTA, S. S. and NAGEL, K. (1971). On some contributions to multiple decision theory. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. Yackel, eds.), 79–102 Academic Press.

[9] GUPTA, S. S. and PANCHAPAKESAN, S. (1969). Some selection and ranking procedures for multivariate populations. In *Multivariate Analysis II* (P. R. Krishnaiah, ed.), 475–505 Academic Press.

[10] GUPTA, S. S. and PANCHAPAKESAN, S. (1972). On a class of subset selection procedures. *Ann. Math. Statist.* **43** 814–822.

[11] GUPTA, S. S. and SOBEL, M. (1960). Selecting a subset containing the best of several binomial populations. In *Contributions to Probability and Statistics* (I. Olkin, et al., eds.), 224–248 Stanford Univ. Press.

[12] GUPTA, S. S. and STUDDEN, W. J. (1966). Some aspects of selection and ranking procedures with applications. Mimeo Series No. 81, Dept. Statist., Purdue Univ.

[13] GUPTA, S. S. and STUDDEN, W. J. (1970). On some selection and ranking procedures with applications to multivariate populations. In *Essays in Probability and Statistics* (R. C. Bose, et al., eds.) Univ. North Carolina Press, 327–338.

[14] ROYDEN, H. L. (1968). *Real Analysis*. MacMillan.

[15] SANTNER, T. J. (1975). A restricted subset selection approach to ranking and selection problems. *Ann. Statist.* **3** 334–349.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306