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

By Roger L. Berger
Florida State University

Abbreviated title: MINIMAX SUBSET SELECTION

Mimeograph Series #507

August 1977

[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.

Minimax Subset Selection for Loss Measured by Subset Size

1. INTRODUCTION. A subset selection problem may be formulated as a multiple decision problem. The distinguishing feature of a subset selection problem is the goal of determining in which of k partition sets of the parameter space the true parameter lies. In subset selection problems, attention is usually restricted to rules which insure a certain minimum probability, P^*, of making a correct decision. Restricting attention to these rules, minimaxity is investigated for loss measured by subset size and number of non-best populations selected. The minimax values are found to be kP^* and $(k-1)P^*$, respectively, 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 when the populations are independent and the densities have monotone likelihood ratio. Other rules, proposed for selection in terms of binomial and multinomial probabilities and multivariate non-centrality parameters, are shown to be not minimax. A class of rules, proposed by Seal (1955) for the location parameter problem, is also investigated. For certain values of k and P^*, rules in this class 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.

The sample space $\mathcal{X}$ is a subset of q-dimensional Euclidean space $\mathbb{R}^q$. The parameter space Θ is a subset of $\mathbb{R}^r$. The observation $\underline{X} = (X_1,\ldots,X_q)$ is a random vector with cumulative distribution function (c.d.f.) $F(\underline{x};\underline{\theta})$. It is assumed that there exists a partition of Θ denoted by $\{\Theta_i: \; i=1,\ldots,k\}$ $(k \geq 2)$. 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 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 $\mathcal{a}$ consists of the 2^k-1 non-empty subsets of $\{\pi_1,\ldots,\pi_k\}$ where π_i is the statement $\underline{\theta}\varepsilon\Theta_i$. So the action $\{\pi_1,\pi_2\}$ corresponds to the decision $\underline{\theta}\varepsilon\Theta_1 \cup \Theta_2$. The π_i's correspond to what have been called populations in the earlier subset selection literature. In this terminology, for a given $\underline{\theta}$, the "best population" is the one true π_i and the other (k-1) π_i's are the "non-best populations." So a statement like,"the best population is the one associated with the largest parameter value," means $\Theta_i = \{\underline{\theta}: \; \theta_i = \max_{1 \leq j \leq r} \theta_j\}$ (with the exception that if θ_i is tied with other θ_j's as the largest, that parameter point may not be in Θ_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. The σ-field associated with the sets $\mathcal{X}$, Θ and $\mathcal{a}$ will be the discrete σ-field if the set is countable and the Borel σ-field if the set is uncountable.

A measurable function, δ: $\mathcal{X} \times \mathcal{A} \to [0,1]$, is called a selection rule provided that, for each $\underline{x} \in \mathcal{X}$, $\sum_{\mathcal{A}} \delta(\underline{x},a) = 1$. $\delta(\underline{x},a)$ is the probability of selecting subset a having observed $\underline{x}$. The k functions defined by $\varphi_i(\underline{x}) = \sum_{\{a:\pi_i \varepsilon a\}} \delta(\underline{x},a)$ are the individual selection probabilities. $\varphi_i(\underline{x})$ is the probability of including π_i in the selected subset having observed $\underline{x}$. A selection rule is not, in general, completely determined by its individual selection probabilities (see Nagel (1970), Example 1.2.1). But the risk of any rule, for losses defined in terms of the quantities (2.3), 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 pre-assigned fixed number such that $1/k < P^* < 1$. It has been traditional in the literature to consider only selection rules which satisfy the P^*-condition, viz.,

$$\inf_{\Theta} P_{\underline{\theta}}(CS|\varphi) \geq P^*. \tag{2.1}$$

This is obviously equivalent to the following k inequalities being satisfied,

$$\inf_{\Theta_i} E_{\underline{\theta}} \varphi_i(\underline{X}) = \inf_{\Theta_i} P_{\underline{\theta}}(\text{select } \pi_i | \varphi) \geq P^*, \ i=1,\ldots,k. \tag{2.2}$$

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

Having insured a high probability of correct selection through the P^*-condition, one would prefer a rule which selects small subsets, that is, a rule which rejects non-best populations effectively. To reflect this, the loss in a subset selection problem might be measured in several ways. The criteria used in this paper are the following, (2.3) i) Number of populations selected (S)

ii) Number of non-best populations selected (S').

So the risk of a selection rule, $R(\underline{\theta},\varphi)$, is given by i) the expected subset size, $E_{\underline{\theta}}(S|\varphi)$, or ii) the expected number of non-best populations selected, $E_{\underline{\theta}}(S'|\varphi)$.

3. MINIMAX VALUES FOR LOSSES S AND S'. A selection rule $\varphi^*\varepsilon\mathcal{D}_P^*$ is minimax with respect to S if

$$\sup_{\Theta} E_{\underline{\theta}}(S|\varphi^*) = \inf_{\mathcal{D}_P^*} \sup_{\Theta} E_{\underline{\theta}}(S|\varphi). \tag{3.1}$$

The value on the right side of (3.1) is called the minimax value with respect to S of the selection problem. Minimaxity with respect to S' is defined by replacing S with S' in (3.1).

Schaafsma (1969) considered minimaxity in multiple decision problems in a very general setting. But he did not restrict attention to rules which satisfy the P^*-condition. In this unrestricted problem he found that a minimax rule (with respect to S or S') never selects a subset consisting of more than one population. This will certainly not be the case in the restricted minimaxity of (3.1).

The following subset of the parameter space will be of interest in finding the minimax values. Let $\Theta_0 = \{\underline{\theta}\varepsilon\Theta\colon \ \underline{\theta}\varepsilon\overline{\Theta}_i$ for all $i=1,\ldots,k\}$ where $\overline{A}$ denotes the closure of A.

Theorem 3.1. Suppose Θ_0 is non-empty. Suppose there exists $\underline{\theta}_0\varepsilon\Theta_0$ such that $P_{\underline{\theta}}(\text{select } \pi_i|\varphi)$ is upper semicontinuous at $\underline{\theta}_0$ for all $\varphi\varepsilon\mathcal{D}_P^*$ and all $i=1,\ldots,k$. Then the minimax value with respect to S is kP^* and the minimax value with respect to S' is $(k-1)P^*$.

Proof. Let $\pi_{(1)},\ldots,\pi_{(k-1)}$ denote the k-1 non-best populations and $\pi_{(k)}$ the best population at $\underline{\theta}_0$. Then the risks at $\underline{\theta}_0$ are

$$E_{\underline{\theta}_0}(S|\varphi) = \sum_{i=1}^{k} P_{\underline{\theta}_0}(\text{select } \pi_{(i)}|\varphi) \tag{3.2}$$

$$E_{\underline{\theta}_0}(S'|\varphi) = \sum_{i=1}^{k-1} P_{\underline{\theta}_0}(\text{select } \pi_{(i)}|\varphi). \tag{3.3}$$

The "no data rule" defined by $\varphi_i^*(\underline{x}) \equiv P^*$, i=1,...,k, has $P_{\underline{\theta}}$ (select $\pi_{(i)}|\varphi^*$) $= P^*$ for all $\underline{\theta}$ and all i. So $E_{\underline{\theta}}(S|\varphi^*) = kP^*$ and $E_{\underline{\theta}}(S'|\varphi^*) = (k-1)P^*$ for all $\underline{\theta}$ and the minimax values can be no greater than kP^* and $(k-1)P^*$ respectively.

On the other hand, let $\Theta_{(i)}$ be the subset of Θ where $\pi_{(i)}$ is best. Since $\underline{\theta}_0 \varepsilon \overline{\Theta}_{(i)}$, and $P_{\underline{\theta}}$ (select $\pi_{(i)}|\varphi$) is upper semicontinuous at $\underline{\theta}_0$,

$$P_{\underline{\theta}_0}(\text{select } \pi_{(i)}|\varphi) \geq \inf_{\Theta_{(i)}} P_{\underline{\theta}}(\text{select } \pi_{(i)}|\varphi)$$

$$= \inf_{\Theta_{(i)}} P_{\underline{\theta}}(CS|\varphi) \geq P^*$$

for any $\varphi \varepsilon \mathcal{D}_{P^*}$. So

$$\sup_{\Theta} E_{\underline{\theta}}(S|\varphi) \geq E_{\underline{\theta}_0}(S|\varphi) \geq kP^* \tag{3.4}$$

and

$$\sup_{\Theta} E_{\underline{\theta}}(S'|\varphi) \geq E_{\underline{\theta}_0}(S'|\varphi) \geq (k-1)P^*$$

for any $\varphi \varepsilon \mathcal{D}_{P^*}$. Thus the minimax values can be no less than kP^* and $(k-1)P^*$ respectively.||

Remark 3.1. The hypothesis that Θ_0 is non-empty is usually satisfied. If $\Theta = I \times I \times \ldots \times I$ (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 $\Theta_0 = \{\underline{\theta} = (\theta,\theta,\ldots,\theta): \theta \varepsilon I\}$. If $\underline{X}$ has a multinomial distribution, $\Theta = \{(\theta_1,\ldots,\theta_k): \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1\}$. If the best population is the coordinate associated with the largest or smallest coordinate of the parameter, then Θ_0 is the single point $(1/k,\ldots,1/k)$. It should be noted that in both of these examples, the determination of Θ_0 did not depend on which population was tagged as best in those cases where two or more of the coordinates were equal and largest (or smallest). It may be argued that in problems like the above, any action is acceptable to the experimenter if $\underline{\theta}\varepsilon\Theta_0$. In this case, one would set $R(\underline{\theta},\varphi) = 0$ for $\underline{\theta}\varepsilon\Theta_0$. But, even allowing this, Theorem 3.1 remains true, in the usual (see Remark 3.2) case where $P_{\underline{\theta}}$ (select $\pi_i|\varphi$) is continuous in $\underline{\theta}$, for (3.4) can be replaced by

$$\sup_{\Theta} E_{\underline{\theta}}(S|\varphi) \geq \lim_{\underline{\theta}\to\underline{\theta}_0} E_{\theta}(S|\varphi) \geq kP^*$$

and similarly for S'.

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 $\underline{X}$ has a density $f(\underline{x};\underline{\theta})$ with respect to a sigma finite measure μ and if $f(\underline{x};\underline{\theta})$ is continuous at $\underline{\theta}_0$ (as a function of $\underline{\theta}$) for almost all (μ) $\underline{x}$, then $E_{\underline{\theta}}\psi(\underline{X})$ is continuous at $\underline{\theta}_0$ for any bounded ψ. Since $P_{\underline{\theta}}$ (select $\pi_i|\varphi$) = $E_{\underline{\theta}}\varphi_i(\underline{X})$ and $0 \leq \varphi_i \leq 1$, this shows that $P_{\underline{\theta}}$ (select $\pi_i|\varphi$) will be a continuous function of $\underline{\theta}$ on Θ for any φ in any problem with densities which are (almost everywhere) continuous in the parameter.

Theorem 3.1 indicates a relationship between minimaxity with respect to S and S'. Theorem 3.2 shows that minimaxity with respect to S' is more easily achieved than minimaxity with respect to S.

Theorem 3.2. Under the assumptions of Theorem 3.1, if $\varphi \in \mathcal{D}_P^*$ is minimax with respect to S, then φ is minimax with respect to S'.

Proof.

$$\sup_{\Theta} E_{\underline{\theta}}(S'|\varphi) = \sup_{\Theta}\{E_{\underline{\theta}}(S|\varphi)-P_{\underline{\theta}}(CS|\varphi)\}$$
$$\leq \sup_{\Theta}\{E_{\underline{\theta}}(S|\varphi)-P^*\}$$
$$= kP^*-P^* = (k-1)P^*. ||$$

Theorems 3.1 and 3.2 can be used to show that in location and scale parameter problems, two rules proposed by Gupta (1965) are minimax. In the following, we will 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 parameter value is best.

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

(3.5) $\quad R_1$: select π_i if $x_i \geq \max_{1\leq j\leq k} x_j - d \qquad i=1,\ldots,k$

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

(3.6) $\quad R_2$: select π_i if $x_i \geq c\cdot \max_{1\leq j\leq k} x_j \qquad i=1,\ldots,k$

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

Theorem 3.3. Suppose $X_1,\ldots,X_k$ are independent. Suppose $\underline{\theta}$ 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, μ, on the real line $((0,\infty))$. Suppose $f_\theta(x)$ has monotone likelihood ratio. Then $R_1(R_2)$ is minimax with respect to S and S'.

Proof. Gupta (1965) proved that under the assumptions of independence and monotone likelihood ratio,

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

The continuity assumption of Theorem 3.1 is satisfied for any location (scale) parameter density with respect to Lebesgue measure (see Royden (1968) problem 17, chapter 4). The result follows from Theorems 3.1 and 3.2.||

Theorem 3.3 generalizes a result of Gupta and Studden (1966). They proved that $R_1(R_2)$ is minimax among all permutation invariant rules in $\mathcal{D}_P^*$. Theorem 3.3 proves minimaxity among all rules in $\mathcal{D}_P^*$.

4. NECESSARY CONDITIONS FOR MINIMAXITY. Any minimax selection rule must satisfy certain equalities on the set Θ_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 with respect to S and Theorem 4.2 the analogous conditions for S'.

<u>Theorem 4.1</u>. Let φ be a minimax rule with respect to S. Suppose $P_{\underline{\theta}}(\text{select } \pi_i|\varphi)$ is upper semicontinuous for all i=1,...,k at $\underline{\theta}_0 \varepsilon \Theta_0$. Then

a) $$P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi) = P^* = \inf_{\Theta} P_{\underline{\theta}}(CS|\varphi) \text{ for all } i=1,\ldots,k$$

b) $$P_{\underline{\theta}_0}(CS|\varphi) = P^* = \inf_{\Theta} P_{\underline{\theta}}(CS|\varphi)$$

c) $$E_{\underline{\theta}_0}(S|\varphi) = kP^* = \sup_{\Theta} E_{\underline{\theta}}(S|\varphi).$$

<u>Remark 4.1</u>. Condition (a) of Theorem 4.1 implies condition (b) and the first equality in (c) as well as (a) and the first equality in (b) of Theorem 4.2. If one wishes to verify these conditions for a given rule to check if it might be minimax, only 4.1(a) need be verified.

<u>Proof</u>. As in the proof of Theorem 3.1, it follows that

(4.1) $$P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi) \geq P^* \quad \text{for all } i=1,\ldots,k.$$

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

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

All the inequalities are equalities and (a) and (c) are true. (b) follows from (a) since $P_{\underline{\theta}_0}(CS|\varphi) = P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi)$ were $\underline{\theta}_0 \varepsilon \Theta_i$.||

Nagel (1970) found that a condition related to 4.1(b), viz.,

$$\inf_{\Theta} P_{\underline{\theta}}(CS|\varphi) = \inf_{\Theta_0} P_{\underline{\theta}}(CS|\varphi),$$

was an important property of just selection rules. Conditions 4.1(a) and (b) have long been recognized (cf. Gupta and Studden (1966)) as intuitively appealing properties of selection rules. This is especially true for those problems in which Θ_0 consists of those parameter points at which one of the k populations has arbitrarily been tagged as best, e.g., a location or scale parameter problem in which best has been defined in terms of the largest or smallest parameter. Theorem 4.1 verifies that, in terms of minimaxity considerations, the intuition is justified.

<u>Theorem 4.2</u>. Let φ be a minimax rule with respect to S'. Suppose $P_{\underline{\theta}}(\text{select } \pi_i|\varphi)$ is upper semicontinuous for all $i=1,\ldots,k$ at $\underline{\theta}_0 \varepsilon \Theta_0$. Then

a) $P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi) = P^* = \inf_{\Theta} P_{\underline{\theta}}(CS|\varphi)$ for all $i=1,\ldots,k$, $i \neq j$, where $\underline{\theta}_0 \varepsilon \Theta_j$

b) $E_{\underline{\theta}_0}(S'|\varphi) = (k-1)P^* = \sup_{\Theta} E_{\underline{\theta}}(S'|\varphi)$.

<u>Proof</u>. Similar to Theorem 4.1.||

Remark 4.2. For those problems in which the random variables $X_1,\ldots,X_k$ are exchangeable for $\underline{\theta}\varepsilon\Theta_0$ and the rule φ is invariant under permutations (symmetric), the following is true for any $\underline{\theta}\varepsilon\Theta_0$:

$$P_{\underline{\theta}}\ (\text{select } \pi_1|\varphi) = P_{\underline{\theta}}\ (\text{select } \pi_2|\varphi) = \ldots = P_{\underline{\theta}}\ (\text{select } \pi_k|\varphi).$$

In such a problem, then, 4.2(a) implies 4.1(a) and (b). So for these problems, the necessary conditions in Theorem 4.1 and those in Theorem 4.2 are essentially the same.

Remark 4.3. Santner (1975) gives conditions under which $\sup_{\Theta} E_{\underline{\theta}}(S|\varphi) = \sup_{\Theta_0} E_{\underline{\theta}}(S|\varphi)$ $(\sup_{\Theta} E_{\underline{\theta}}(S'|\varphi) = \sup_{\Theta_0} E_{\underline{\theta}}(S'|\varphi))$ which from Theorem 4.1(c) (4.2(b)) is a necessary condition for minimaxity.

In the following three examples, rules which have been proposed by other authors for various problems will be examined. In all cases, the best population is the one associated with the largest parameter value. In all cases the continuity assumptions of Theorems 4.1 and 4.2 are satisfied for all $\underline{\theta}_0\varepsilon\Theta$ since the densities (or probability mass functions) are continuous functions of the parameter (see Remark 3.2).

Example 4.1. Consider the multinomial selection problem in which the cell associated with the largest cell probability is best. Here $\Theta_0 = (1/k,\ldots,1/k)$. Gupta and Nagel (1967) proposed using the rule R_1 (see (3.5)) for this problem. They found that for some values of k and P^*, the $\inf_{\Theta} P_{\underline{\theta}}(CS|R_1)$ did not occur at $(1/k,\ldots,1/k)$. So 4.1(b) and 4.2(a) are violated and R_1 is not minimax with respect to S or S' for these values of k and P^*.

Example 4.2. Consider the binomial selection problem in which $X_1,\ldots,X_k$ are independent binomial random variables with success probabilities $\theta_1,\ldots,\theta_k$. $\Theta_0 = \{(\theta_1,\ldots,\theta_k) | \theta_1 = \ldots = \theta_k = \theta,\ 0 \leq \theta \leq 1\}$. Gupta and Sobel (1960) proposed using the rule R_1 (see (3.5)) to select a subset including the population associated with the largest θ_i. These authors realized that $E_{\underline{\theta}}(S|R_1)$ was not constant on Θ_0, as required by 4.1(c) if R_1 were to be minimax. Indeed, $E_{\underline{\theta}}(S|R_1) \to k$ as $\underline{\theta} \to (1,\ldots,1)$ and $E_{\underline{\theta}}(S'|R_1) \to (k-1)$. The same is true for the arcsin transformation proposed by these authors.

Example 4.3. The following general problem has been considered by Gupta and Panchapakesan (1972). Suppose $\pi_1,\ldots,\pi_k$ are independent populations with absolutely continuous distributions $F_{\theta_i}(x_i)$ where $\theta_i \varepsilon I$ (an interval on the real line). The family $\{F_\theta\colon\ \theta\varepsilon I\}$ is assumed to be stochastically increasing in θ. Gupta and Panchapakesan investigated a class of procedures for selecting a subset containing the population associated with the largest θ_i defined by:

(4.2) $\qquad R_h\colon$ select π_i iff $h(x_i) \geq \max\limits_{1\leq j\leq k} x_j$

where h is a real valued function satisfying certain regularity conditions. $\Theta_0 = \{(\theta,\ldots,\theta)\colon\ \theta\varepsilon I\}$. For any $\underline{\theta} = (\theta,\ldots,\theta)\varepsilon\Theta_0$ and $i=1,\ldots,k$,

(4.3) $\qquad P_{\underline{\theta}}(\text{select } \pi_i | R_h) = \int F_\theta^{k-1}(h(x))\,dF_\theta(x).$

By Theorems 4.1 and 4.2, if the procedure R_h is to be minimax with respect to S or S', (4.3) must be constant on Θ_0. But Gupta and Panchapakesan have found conditions under which (4.3) will be strictly increasing in θ. Gupta and Studden (1970) have established the strict monotonicity of (4.3) for the non-central χ^2 and non-central F distributions when R_h is R_2 (see (3.6)). This is of interest in the problem of selection in terms of Mahalanobis distance for multivariate normal distributions. Gupta and Panchapakesan (1969) have established the strict monotonicity of (4.3) in the problem of selection in terms of the largest (or smallest) multiple correlation coefficient when R_h is R_2 (or an analogous rule). Both the conditional and unconditional cases were considered as well as two different statistics, R^2, the sample multiple correlation coefficient, and $R^{*2} = R^2/(1-R^2)$. In violating Theorems 4.1(a) and 4.2(a), all of the above rules are not minimax with respect to S or S'.

Remark 4.3. The fact that the above rules are not minimax was previously reported in some cases. But the interesting point is that one need not always examine $E_{\underline{\theta}}(S|\varphi)$ or $E_{\underline{\theta}}(S'|\varphi)$ to determine that a rule is not minimax. Often in investigating the least favorable configuration, i.e., that $\underline{\theta}_0$ for which $P_{\underline{\theta}_0}(CS|\varphi) = \inf_{\Theta} P_{\underline{\theta}}(CS|\varphi)$, one can reduce the problem to investigating $\inf_{\Theta_0} P_{\underline{\theta}}(CS|\varphi)$. This, for example, is the case for just rules as defined by Nagel (1970) and Gupta and Nagel (1971). If one finds that $P_{\underline{\theta}}(CS|\varphi)$ is not constant on Θ_0 (and some mild continuity assumptions are satisfied), then R is not minimax. Thus, the only analysis required, to show that a proposal rule is not minimax, may be the analysis used to find the least favorable configuration.

5. MINIMAXITY CONSIDERATIONS FOR SEAL'S CLASS. Seal (1955) proposed a class of rules for the location parameter problem. The rules were proposed for the independent normal populations problem but might reasonably be used in any location problem. In this section, a lower bound is obtained for $\sup_{\Theta} E_{\underline{\theta}}(S|\varphi)$ and $\sup_{\Theta} E_{\underline{\theta}}(S'|\varphi)$ for rules in this class. This lower bound is then used to prove that, in certain cases, the rules in this class are not minimax.

Definition 5.1. Let $\mathscr{C}$ denote the class of selection rules having the form:

$$\varphi_{\underline{a}}: \text{ select } \pi_i \text{ iff } x_i \geq \sum_{j=1}^{k-1} a_j x_{[j]} - d$$

where $x_{[1]} \leq \ldots \leq x_{[k-1]}$ are the ordered observations excluding x_i, a_j are non-negative constants with $\sum_{j=1}^{k-1} a_j = 1$, and d is the smallest positive constant for which the P^*-condition is satisfied.

R_1 (see (3.5)) is in $\mathscr{C}$ and corresponds to setting $a_{k-1}=1$, $a_j=0, j=1,\ldots,k-2$. Comparisons between $E_{\underline{\theta}}(S|R_1)$ and $E_{\underline{\theta}}(S|\varphi)$ for certain other rules, $\varphi \varepsilon \mathscr{C}$, have previously been made by Seal (1957) and Deely and Gupta (1968). These authors considered specific parameter configurations (e.g., slippage configurations) and specific alternatives to R_1. In the following results, the sup over all parameter configurations and all rules in $\mathscr{C}$ are considered. But, as have the previous authors' works, these results shed some favorable light on R_1.

Throughout this section, it will be assumed that $\Theta = \mathbb{R}^k$. The c.d.f. of $\underline{X}$ is $F(\underline{x}-\underline{\theta})$. The following notation will be used. $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ will denote the ordered coordinates of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ so that the best population is the (unknown) one associated with $\theta_{[k]}$. Sometimes, a sequence of parameter points $<\underline{\theta}_n>$ will be considered in which case $\theta_{n[1]} \leq \ldots \leq \theta_{n[k]}$ will denote the ordered coordinates of $\underline{\theta}_n = (\theta_{n1},\ldots,\theta_{nk})$.

Theorem 5.1 will be used to obtain a lower bound on the expected subset size. As stated, it also points out an intuitively undesirable property of all rules in $\mathscr{C}$, except R_1, namely, there exist parameter points such that $\theta_{[k]}-\theta_{[k-1]}$ is arbitrarily large but the probability of including the population associated with $\theta_{[k-1]}$ in the selected subset is arbitrarily near one.

Theorem 5.1. Let $\varphi_{\underline{a}} \varepsilon \mathscr{C}\setminus\{R_1\}$. Let $r=\min\{i: a_i > 0\}$. Then there exists a sequence of parameter points $<\underline{\theta}_n>$ and a subset $K \subset \{1,\ldots,k\}$ of size k-r-1 such that for $i \varepsilon K$, $\lim_{n\to\infty} \theta_{n[k]}-\theta_{ni}=\infty$ and $\lim_{n\to\infty} P_{\underline{\theta}_n}(\text{select } \pi_i|\varphi_{\underline{a}})=1$.

Remark 6.1. For $\varphi_{\underline{a}} = R_1$, k-r-1=0 so the theorem is vacuously true for R_1 also. For any $\varphi_{\underline{a}} \neq R_1$, $r \leq k-2$ so K will be non-empty.

Proof. Let $S_i = \{\underline{x}:\ x_i \geq \sum_{j=1}^{k-1} a_j x_{[j]} - d\}$ be the selection region for π_i using $\varphi_{\underline{a}}$. Define a sequence of subsets of $\mathcal{X}$ by

$$(5.1)\qquad A_n = \{\underline{x}:\ 2n \geq x_k > n,\ n \geq x_j > -d,\ j=r+1,r+2,\ldots,k-1,\quad c_n \geq x_i,\ i=1,2,\ldots,r\}$$

where $c_n = (-n-2n\cdot a_{k-1})/a_r$.

Let $K = \{r+1,\ldots,k-1\}$. First it will be shown that $A_n \subseteq S_j$ for all $j \varepsilon K$, for all large n. Since $a_{k-1} \geq 0$ and $a_r > 0$, $c_n \leq -n/a_r < -d$ for all large n. Fix such an n and $j \varepsilon K$. Let $\underline{x} \varepsilon A_n$. Then $x_{[k-1]} = x_k$, $\{x_{[r+1]},\ldots x_{[k-2]}\} = \{x_{r+1},\ldots,x_{k-1}\} \setminus \{x_j\}$ (this set is empty if r=k-2) and $\{x_{[1]},\ldots,x_{[r]}\} = \{x_1,\ldots,x_r\}$. Using these facts and (5.1), (5.2) and (5.3) are obvious.

$$(5.2)\qquad a_{k-1}\, x_{[k-1]} + a_r\, x_{[r]} \leq a_{k-1}\cdot 2n + a_r c_n = -n$$

$$(5.3)\qquad \sum_{m=r+1}^{k-2} a_m x_{[m]} \leq \max\{x_{[r+1]},\ldots,x_{[k-2]}\} \leq n$$

Using (5.2), (5.3) and the fact that $a_m = 0$ m=1,...,r-1 it follows that

$$(5.4)\qquad \sum_{m=1}^{k-1} a_m x_{[m]} - d = \sum_{m=r}^{k-1} a_m x_{[m]} - d \leq -n+n-d = -d.$$

But $x_j > -d$ by (5.1) so $\underline{x} \varepsilon S_j$. This is true for any $\underline{x} \varepsilon A_n$ so $A_n \subset S_j$ for all $j \varepsilon K$.

Define a sequence of parameter points $\underline{\theta}_n = (\theta_{n1},\ldots,\theta_{nk})$ by

$$(5.5)\qquad \theta_{nj} = \begin{cases} 3n/2 & j = k \\ n/2 & j = r+1,\ldots,k-1 \\ c_n - n & j = 1,\ldots,r \end{cases}$$

For any $j \varepsilon K$, $\lim_{n\to\infty} \theta_{n[k]} - \theta_{nj} = \lim_{n\to\infty} (3n/2-n/2) = \infty$.

$$
\begin{aligned}
P_{\underline{\theta}_n}(A_n) &= P_{\underline{0}}(A_n - \underline{\theta}_n) \\
&= P(n/2 \geq Y_k > -n/2,\ n/2 \geq Y_j > -n/2-d, \\
&\qquad j = r+1,\ldots,k-1,\ n \geq Y_i > -\infty,\ i=1,\ldots,r)
\end{aligned} \tag{5.6}
$$

where $\underline{Y} = (Y_1,\ldots,Y_k)$ has c.d.f. $F(\underline{y})$. (5.6) converges to 1 as $n\to\infty$ since all the limits converge to ∞ or $-\infty$ as appropriate. Since $A_n \subset S_j$ for all $j \in K$,

$$
\lim_{n\to\infty} P_{\underline{\theta}_n}(\text{select } \pi_j | \varphi_{\underline{a}}) = \lim_{n\to\infty} P_{\underline{\theta}_n}(S_j) \geq \lim_{n\to\infty} P_{\underline{\theta}_n}(A_n) = 1 \tag{5.7}
$$

for all $j \in K$.||

Theorem 5.2. Let $\varphi_{\underline{a}} \in \mathscr{G}$. Let $r=\min\{i: a_i > 0\}$. Then

a) $\sup_{\Theta} E_{\underline{\theta}}(S|\varphi_{\underline{a}}) \geq k-r$

b) $\sup_{\Theta} E_{\underline{\theta}}(S'|\varphi_{\underline{a}}) \geq k-r-1.$

Proof. If $\varphi_{\underline{a}} = R_1$, $k-r=1$ and $k-r=0$ so (a) and (b) are obviously true.

For any $\varphi_{\underline{a}} \in \mathscr{G}\setminus\{R_1\}$, using the notation defined in the proof of Theorem 5.1 we have

$$
\begin{aligned}
\sup_{\Theta} E_{\underline{\theta}}(S|\varphi_{\underline{a}}) &\geq \lim_{n\to\infty} E_{\underline{\theta}_n}(S|\varphi_{\underline{a}}) \\
&\geq \lim_{n\to\infty} \sum_{m=r+1}^{k} P_{\underline{\theta}_n}(\text{select } \pi_m|\varphi_{\underline{a}})
\end{aligned} \tag{5.8}
$$

Theorem 5.1 proved the first k-r-1 terms converge to one in the limit. For every $\underline{x} \in A_n$, x_k is the largest coordinate so $A_n \subset S_k$ for every n. Thus (5.7) holds with $j=k$. Hence the bound k-r for (a).

From (5.5), π_k is the best population for all $\underline{\theta}_n$. Thus using the same reasoning as above, excluding the term $P_{\underline{\theta}_n}(\text{select } \pi_k|\varphi_{\underline{a}})$ in (5.8), yields the bound k-r-1 for (b).||

Corollary 5.1. Let $\varphi_{\underline{a}} \in \mathscr{G}\setminus\{R_1\}$. Let $r = \min\{i: a_i > 0\}$. Then

(a) if $P^* < (k-r)/k$, $\varphi_{\underline{a}}$ is not minimax with respect to S

(b) if $P^* < (k-r-1)/(k-1)$, $\varphi_{\underline{a}}$ is not minimax with respect to S'.

Proof. The "no data rule", $\varphi_i^*(\underline{x}) \equiv P^*$, has $\sup_{\Theta} E_{\underline{\theta}}(S|\varphi^*) = kP^* < k-r \leq \sup_{\Theta} E_{\underline{\theta}}(S|\varphi_{\underline{a}})$. Hence (a) is true. (b) is analogous.||

Corollary 5.2. (a) If $P^* < 2/k$, no rule in $\mathscr{L}\setminus\{R_1\}$ is minimax with respect to S. (b) If $P^* < 1/(k-1)$, no rule in $\mathscr{L}\setminus\{R_1\}$ is minimax with respect to S'.

Proof. Any rule in $\mathscr{L}\setminus\{R_1\}$ has $r \leq k-2$. So the smallest upper bound in Corollary 5.1(a) is $2/k$. Hence (a) is true. (b) is analogous.||

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

(1) Chung, K. L. (1974). A Course in Probability Theory. Academic Press.

(2) Deely, J. J. and Gupta, S. S. (1968). On the properties of subset selection procedures. Sankhyā Ser. A 30, 37-50.

(3) Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics 7, 225-245.

(4) 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.

(5) Gupta, S. S. and Nagel, K. (1971). On some contributions to multiple decision theory. Statistical Decision Theory and Related Topics (eds. Gupta, S. S. and Yackel, J.) Academic Press, 79-102.

(6) Gupta, S. S. and Panchapakesan, S. (1969). Some selection and ranking procedures for multivariate populations. Multivariate Analysis II (ed. Krishnaiah, P.R.) Academic Press, 475-505.

(7) Gupta, S. S. and Panchapakesan, S. (1972). On a class of subset selection procedures. Ann. Math. Statist. 43, 814-822.

(8) Gupta, S. S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to Probability and Statistics (eds. Olkin, I. et al.) Stanford University Press, 224-248.

(9) Gupta, S. S. and Studden, W. J. (1966). Some aspects of selection and ranking procedures with applications. Mimeo Series No. 81, Dept. of Statistics, Purdue University.

(10) Gupta, S. S. and Studden, W. J. (1970). On some selection and ranking procedures with applications to multivariate populations. Essays in Probability and Statistics (eds. Bose, R.C. et al.) University of North Carolina Press, 327-338.

(11) Nagel, K. (1970). On subset selection rules with certain optimality properties. Ph.D. thesis, Dept. of Statistics, Purdue University.

(12) Royden, H. L. (1968). Real Analysis. MacMillan.

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

(14) Schaafsma, W. (1969). Minimax risk and unbiasedness for multiple decision problems of type I. Ann. Math. Statist. 40, 1684-1720.

(15) Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 26, 387-398.

(16) Seal, K. C. (1957). An optimum decision rule for ranking means of normal populations. Calcutta Statist. Assoc. Bull. 7, 131-150.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Memeograph Series #507	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Minimax Subset Selection For Loss Measured By Subset Size	5. TYPE OF REPORT & PERIOD COVERED Technical
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series #507
7. AUTHOR(s) Roger L. Berger	8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455 AFOSR-72-2350 E
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Purdue University W. Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-243 61102F 9769-05
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, D.C. and Air Force Office of Scientific Research Bolling AFB, DC	12. REPORT DATE August 1977
	13. NUMBER OF PAGES 17
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 public 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)

Minimax subset selection, expected subset size, expected number of non-best populations, multiple decision.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

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 non-centrality parameters, such as χ^2 and F, previously proposed

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. rules are found to be not minimax. A class of rules which has been proposed for the location problem is investigated and rules in this class are found to be not minimax in some cases.