MINIMAX, ADMISSIBLE, AND GAMMA-MINIMAX MULTIPLE DECISION RULES*

by

Roger L. Berger
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #489

May 1977

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

INTRODUCTION

Multiple decision problems are decision theory problems in which the action space has a finite number of elements. Two different types of multiple decision problems are considered, herein. These two types of problems are subset selection problems and robustness of Bayes rules problems.

Subset selection problems arise because the classical tests of homogeneity are often inadequate in practical situations where the experimenter has to make decisions regarding k ($\geq$ 2) populations, treatments, or processes. The inadequacy is centered in the fact that only two decisions, accept or reject, are available to the experimenter. The experimenter is faced with the problem of what further action is appropriate, if the hypothesis is rejected. This inadequacy may be alleviated by formulating the problems as multiple decision problems aimed at selection or ranking of the k populations. Among the first researchers to formulate the problems in this way were Mosteller (1948), Paulson (1949), and Bahadur (1950).

In the twenty-five years since these early papers, ranking and selection problems have been an active area of statistical research. Gupta (1956, 1965) proposed what has come to be called the subset selection formulation of the problem. In this formulation, the experimenter obtains a subset of the k populations which contains the population of interest with a fixed minimum probability over the whole parameter space. Studies of optimality properties and comparisons between selection rules have been made by Seal (1957), Lehmann (1961), Studden (1967), Deely and Gupta (1968), Deverman (1969), and Schaafsma (1969). The purpose of the first two chapters of this thesis is to study minimaxity and admissibility properties of subset selection rules.

In Chapter I, minimaxity of subset selection rules with respect to the expected subset size and expected number of non-best populations selected is considered. The problem is formulated in a decision theoretic setting in Section 1.1. Section 1.2 contains the formal definition of selection rules. In Section 1.3, the minimax value of a selection problem for a wide class of distributions is obtained. Location and scale parameter problems are considered in Section 1.4. For these problems, it is shown that two rules proposed by Gupta (1965) are minimax, if the distributions have monotone likelihood ratio. Necessary conditions for minimaxity are derived in Section 1.5. A class of rules proposed by Seal (1955) is studied in Section 1.6. Conditions are provided under which rules in this class are not minimax.

In Chapter II, risk is measured in terms of the maximum probability of including any non-best population in the selected subset. This risk is defined in Section 2.1. Translation invariant, scale invariant and just rules are defined and characterized in Section 2.2. In

Section 2.3, a minimaxity result is proven which, when applied in location or scale parameter problems, shows that two rules proposed by Gupta (1965) are minimax and admissible in the set of non-randomized, just, and invariant rules. The behavior of a class of rules proposed by Seal (1955) with respect to this risk is examined in Section 2.4. Section 2.5 addresses a specific problem, first considered by Seal (1955), involving normal populations when the parameters are in a slippage configuration.

Robustness questions can arise in almost any kind of statistical inference. They concern the behavior of a statistical procedure if the underlying assumptions are violated. In many decision problems, the exact parametric form of the distribution of the observations is specified. The question arises, what if this specification is incorrect. The ϵ-contaminated model is common in studies of this problem. Huber (1967, 1972) and Andrews et al. (1972) are examples of studies of this nature and the latter two provide extensive bibliographies of the field. In the ϵ-contaminated model, the form of the distribution is specified only with probability $1-\epsilon$, the probability being ϵ that the distribution is something totally different and unspecified. Considering the uncertainty inherent in the ϵ-contaminated model, it seems unreasonable that the experimenter could then specify an exact prior on the ϵ-contaminations. But if restriction can be made to some sub-class of all prior distributions, this partial prior information should be used. Blum and Rosenblatt (1967) proposed the Γ-minimax criterion for the selection of decision rules in the presence of partial prior information. The Γ-minimax criterion has been studied in a variety of problems by Jackson et al. (1970), Randles and Hollander (1971), Solomon (1972a, 1972b), DeRouen and Mitchell (1974), and Gupta and Huang (1975, 1977).

Chapter III of this thesis considers the Γ-minimaxity of Bayes multiple decision rules. The main result, found in Section 3.3, is that, in a finite parameter space multiple decision problem, the usual Bayes rule, ignoring any contamination, is robust in that, for small ϵ, it is Γ-minimax when the sub-class of priors is a class of priors on the family of ϵ-contaminations. In this sense, the Bayes rule is robust against inaccurately specified distributions. Section 3.2 includes some Γ-minimaxity results which are used in Section 3.3. Bounds on the amount of contamination which can be present with the Bayes rule remaining Γ-minimax are found in Section 3.4. Section 3.5 relates this work to the special case of hypothesis testing studied by Huber (1965).

CHAPTER I

MINIMAXITY OF SUBSET SELECTION RULES WITH RESPECT TO THE EXPECTED SUBSET SIZE

In this chapter, minimaxity of subset selection rules with respect to the expected subset size and expected number of non-best populations selected is considered. The problem is formulated in a decision-theoretic setting in Section 1.1. In Section 1.2, selection rules are formally defined. Theorems 1.3.1 and 1.3.4 of Section 1.3 provide the minimax value of a selection problem for a wide class of distributions. Location and scale parameter problems are considered in Section 1.4. Theorems 1.4.2 and 1.4.4 assert that if the distributions have monotone likelihood ratio, then two rules proposed by Gupta are minimax. Section 1.5 considers necessary conditions for minimaxity. In Section 1.6, a class of selection rules proposed by Seal is studied. Theorem 1.6.1 exhibits an undesirable feature of rules in this class and Theorem 1.6.3 provides conditions under which rules in this class are not minimax.

1.1 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 problems.

The sample space, $\mathcal{X}$, is a subset of k-dimensional Euclidean space, $\mathbb{R}^k$, where $k \geq 2$. Often the sample space will be a k-dimensional product, viz., $\mathcal{X} = A \times A \times \ldots \times A$, where A is a measurable subset of the real line.

The parameter space, Θ, is a subset of k-dimensional Euclidean space. Often Θ will be a k-dimensional product. A distinguishing feature of subset-selection problems is that there is some correspondence between the ith coordinate x_i of the observaticn vector $\underline{x}$ and the ith coordinate θ_i of the parameter vector $\underline{\theta}$. Often the coordinates of the observation are stochastically independent and the distribution of x_i depends only on θ_i. In general, the observation x_i comes from a population (process, treatment, etc.) π_i which has the parameter θ_i associated with it.

The action space $\mathcal{a}$ consists of the 2^k-1 non-empty subsets of $\{1,2,\ldots,k\}$. An action a is the selection of some subset of the k populations. $i \in a$ means that π_i is included in the selected subset.

Based upon the parameter, one of the populations will be classified as best. Usually this will be the population associated with the largest or smallest coordinate of the parameter. If more than one population could be classified as best according to the above criterion, then one of these is arbitrarily tagged as the best. This is done only to insure the continuity of certain important functions of the parameter. The resulting partition of Θ will be denoted by $\{\Theta_i\colon\ i = 1,2,\ldots,k\}$ where

(1.1.1) $$\Theta_i = \{\underline{\theta} \in \Theta\colon\ \pi_i \text{ is the best population}\}.$$

The selection of any subset which contains the best population is called a correct selection, denoted by CS. Let P* by any pre-assigned fixed number such that $1/k < P^* < 1$. It has been traditional in the literature to consider only selection rules R which satisfy the P*-condition, viz.,

$$\inf_{\Theta} P_\theta(CS|R) \geq P^*. \tag{1.1.2}$$

Only rules which satisfy the P*-condition will be considered in this thesis.

Having ensured 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 thesis are the following:

(1.1.3)
i) Selection of any given non-best population
ii) Number of populations selected (S)
iii) Number of non-best populations selected (S').

To complete the decision-theoretic formulation of a subset selection problem, the σ-fields which accompany the sets $\mathcal{X}$, Θ and $\mathcal{A}$ must be specified. Since $\mathcal{A}$ is finite, the discrete σ-field is used. For all applications in this thesis, if $\mathcal{X}$ or Θ is countable, the discrete σ-field is used; if $\mathcal{X}$ or Θ is uncountable, the Borel σ-field is used.

1.2. Definition of Selection Rules

Definition 1.2.1. A measurable function, $\delta: \mathcal{X} \times \mathcal{A} \to [0,1]$, is a selection rule if for each $\underline{x} \in \mathcal{X}$

$$\sum_{a \in \mathcal{A}} \delta(\underline{x},a) = 1. \tag{1.2.1}$$

$\delta(\underline{x},a)$ is the conditional probability of selecting subset a having observed $\underline{x}$.

Definition 1.2.2. The k functions defined by

$$\varphi_i(\underline{x}) = \sum_{\{a:\ i \in a\}} \delta(\underline{x},a) \qquad i = 1,2,\ldots,k \tag{1.2.2}$$

are the individual selection probabilities. $\varphi_i(\underline{x})$ is the conditional probability of including population 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, for an illustration of this fact). But the risk of any rule, for losses defined in terms of the quantities (1.1.3), 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 and henceforth the following definition of a selection rule will be used.

Definition 1.2.3. A selection rule, R, is a measurable function from $\mathcal{X}$ into $\mathbb{R}^k$, $R: \mathcal{X} \to (\varphi_1(\underline{x}),\varphi_2(\underline{x}),\ldots,\varphi_k(\underline{x}))$ where $0 \le \varphi_i(\underline{x}) \le 1$ $i = 1,2,\ldots,k$, $\underline{x} \in \mathcal{X}$.

Definition 1.2.4. A selection rule is called non-randomized if all the φ_i's take on only the values 0 or 1.

Clearly a non-randomized selection rule is completely specified by the k sets, $A_i = \{\underline{x} \in \mathcal{X}: \varphi_i(\underline{x}) = 1\}$. A_i is the set of observations for which population π_i is included in the selected subset. The A_i's may not be disjoint but, since all the selection rules considered herein always select a non-empty subset, it is true that $\bigcup_{i=1}^{k} A_i = \mathcal{X}$.

1.3. Minimax Value for Losses S and S'

In this section, the minimax value for a subset selection problem is computed when loss is measured either in terms of the size of the selected subset, S, or the number of non-best populations selected, S'. It will be shown that for a wide variety of problems, the minimax value is kP* when S is used and (k-1)P* when S' is used.

Both S and S' are random variables which take on positive integer values. Let $E_{\underline{\theta}}(S|R)$ denote the expected value of S when the selection rule R is used and $\underline{\theta}$ is the true parameter value. $E_{\underline{\theta}}(S'|R)$ is defined similarly.

Definition 1.3.1. A selection rule, R*, is minimax with respect to S if

(1.3.1) $$\sup_{\Theta} E_{\underline{\theta}}(S|R^*) = \inf_{R} \sup_{\Theta} E_{\underline{\theta}}(S|R)$$

where the inf is over all selection rules which satisfy the P*-condition (1.1.2). The value on the right hand side of (1.3.1) 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 (1.3.1).

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 minimax problems considered herein.

The following subset of the parameter space will be of interest in finding the minimax values. Let

(1.3.2) $$\Theta_0 = \{\underline{\theta} \in \Theta: \ \underline{\theta} \in \overline{\Theta_i} \text{ for all } i = 1,2,\ldots,k\}$$

where Θ_i was defined in (1.1.1) and $\overline{\Theta_i}$ denotes the closure of Θ_i in the usual topology on $\mathbb{R}^k$.

Theorem 1.3.1. Suppose Θ_0 is non-empty. Suppose there exists $\underline{\theta}_0 \in \Theta_0$ such that $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is upper semicontinuous at $\underline{\theta}_0$ for all R and all $i = 1,2,\ldots,k$. Then the minimax value with respect to S is kP^*.

Proof. It is clear that

(1.3.3) $$E_{\underline{\theta}}(S|R) = \sum_{i=1}^{k} P_{\underline{\theta}}(\text{select } \pi_i|R).$$

The "no data rule" defined by $\varphi_i(\underline{x}) \equiv P^*$ has $P_{\underline{\theta}}(\text{select } \pi_i|\varphi) = P^*$ for all $\underline{\theta}$ and all i. So $E_{\underline{\theta}}(S|\varphi) = kP^*$ for all $\underline{\theta}$ and the minimax value can be no greater than kP^*.

On the other hand, let R denote any rule which satisfies the P*-condition. Let $\{\underline{\theta}_{ji}: j = 1,2,\ldots\}$ be a sequence in Θ_i which converges to $\underline{\theta}_0$. ($\underline{\theta}_0 \in \overline{\Theta_i}$ guarantees the existence of such a sequence). Since

$\underline{\theta}_{ji} \in \Theta_i$, selection of π_i is a correct selection at $\underline{\theta}_{ji}$. Hence by the upper semicontinuity and the P*-condition we have

$$\begin{aligned} P_{\underline{\theta}_0}(\text{select } \pi_i|R) &\geq \overline{\lim_{j\to\infty}}\, P_{\underline{\theta}_{ji}}(\text{select } \pi_i|R) \\ &= \overline{\lim_{j\to\infty}}\, P_{\underline{\theta}_{ji}}(CS|R) \\ &\geq P^*. \end{aligned}$$

This is true for all $i = 1,\ldots,k$ so using it in (1.3.3) yields

$$\text{(1.3.4)} \qquad \sup_{\Theta} E_{\underline{\theta}}(S|R) \geq E_{\underline{\theta}_0}(S|R) \geq kP^*.$$

(1.3.4) is true for any rule R so the minimax value can be no less than kP^*. ||

The hypothesis that Θ_0 is non-empty is satisfied in almost any problem as the following two examples demonstrate. The upper semicontinuity hypothesis appears more formidable but Theorem 1.3.3 will show that, in a wide variety of problems, the functions in question are, in fact, continuous on Θ.

Example 1.3.1. Suppose $\underline{X} = (X_1,X_2,\ldots,X_k)$ has a multinomial distribution with cell probabilities $\underline{p} = (p_1,p_2,\ldots,p_k)$. Suppose one wishes to select the cell associated with the largest or smallest cell probability. In this problem $\Theta = \{(p_1,p_2,\ldots,p_k)\colon p_i \geq 0 \ \forall i \text{ and } \sum_{i=1}^{k} p_i = 1\}$. Θ_0 consists of the single point $(\frac{1}{k}, \frac{1}{k},\ldots,\frac{1}{k})$. If in this problem Θ were restricted so as to exclude a neighborhood of $(\frac{1}{k}, \frac{1}{k},\ldots,\frac{1}{k})$, say the experimenter knew that the largest p_i was at least 2/3, then Θ_0

would be empty and Theorem 1.3.1 would not be applicable.

Example 1.3.2. Suppose $\Theta = I \times I \times \ldots \times I$ (k times) where I is an interval on the real line. This is often the situation when the k populations are independent. Further suppose that best is defined in terms of the population associated with the largest or smallest parameter value. Then $\Theta_0 = \{(\theta, \theta, \ldots, \theta): \theta \in I\}$ is the set of parameter points which have all coordinates equal.

It should be noted that in both Examples 1.3.1 and 1.3.2, the determination of Θ_0 did not depend on which population was tagged as best in the cases where two or more of the parameter values could be tagged as best. This is one indication that the tagging of the population in these cases is truly inconsequential in most problems.

The following theorem may be useful in determining the minimax value in problems that have an empty Θ_0.

Theorem 1.3.2. Suppose there exists a $\underline{\theta}_0$ such that $\underline{\theta}_0 \in \overline{\Theta_i}$ for m different values of i, $2 \leq m \leq k$. Further suppose $P_{\underline{\theta}}(\text{select } \pi_i | R)$ is upper semicontinuous at $\underline{\theta}_0$ for all R and all i = 1,2,...,k. Then the minimax value with respect to S is at least m P*.

Proof. As in Theorem 1.3.1, consider sequences in Θ_i which converge to $\underline{\theta}_0$ for each i for which $\underline{\theta}_0 \in \overline{\Theta_i}$. This yields that for each of these subscripts, i, $P_{\underline{\theta}_0}(\text{select } \pi_i | R) \geq P^*$. Since there are m such subscripts, using (1.3.3) yields $E_{\underline{\theta}_0}(S|R) \geq mP^*$. This being true for all rules, R, implies the minimax value is at least mP*. ||

Theorem 1.3.3 (see Chung (1970), problem 10, page 100) will provide conditions under which the continuity assumptions of Theorems 1.3.1 and 1.3.2 are satisfied.

Theorem 1.3.3. Suppose $\{f_{\underline{\theta}}(\underline{x}) : \underline{\theta} \in \Theta\}$ are densities with respect to a measure μ which satisfy

(i) as $\underline{\theta} \to \underline{\theta}_0$, $f_{\underline{\theta}}(\underline{x}) \to f_{\underline{\theta}_0}(\underline{x})$ a.e. μ

or (ii) as $\underline{\theta} \to \underline{\theta}_0$, $f_{\underline{\theta}}(\underline{x}) \to f_{\underline{\theta}_0}(\underline{x})$ in μ measure.

Then for any bounded measurable function $\psi(\underline{x})$, $E_{\underline{\theta}}\psi(\underline{x})$ is continuous at $\underline{\theta}_0$.

Proof. $(f_{\underline{\theta}_0}-f_{\underline{\theta}})^+ \leq f_{\underline{\theta}_0}$ and $\int f_{\underline{\theta}_0}\, d\mu = 1 < \infty$ so by the dominated convergence theorem

$$\lim_{\underline{\theta}\to\underline{\theta}_0} \int (f_{\underline{\theta}_0}-f_{\underline{\theta}})^+ d\mu = \int \lim_{\underline{\theta}\to\underline{\theta}_0} (f_{\underline{\theta}_0}-f_{\underline{\theta}})^+ d\mu = 0.$$

$$(f_{\underline{\theta}_0}-f_{\underline{\theta}})^- = (f_{\underline{\theta}_0}-f_{\underline{\theta}})^+ - (f_{\underline{\theta}_0}-f_{\underline{\theta}})$$

so $(f_{\underline{\theta}_0}-f_{\underline{\theta}})^-$ is integrable and in fact

$$\int (f_{\underline{\theta}_0}-f_{\underline{\theta}})^- d\mu = \int (f_{\underline{\theta}_0}-f_{\underline{\theta}})^+ d\mu.$$

So

$$\lim_{\underline{\theta}\to\underline{\theta}_0} \int |f_{\underline{\theta}_0}-f_{\underline{\theta}}|\, d\mu = \lim_{\underline{\theta}\to\underline{\theta}_0} \int (f_{\underline{\theta}_0}-f_{\underline{\theta}})^+ d\mu + \lim_{\underline{\theta}\to\underline{\theta}_0} \int (f_{\underline{\theta}_0}-f_{\underline{\theta}})^- d\mu$$

$$= 0.$$

Thus if $\psi(\underline{x}) \leq M < \infty$ a.e. μ,

$$\lim_{\underline{\theta}\to\underline{\theta}_0} |E_{\underline{\theta}_0}\psi(\underline{x}) - E_{\underline{\theta}}\psi(\underline{x})| = \lim_{\underline{\theta}\to\underline{\theta}_0} |\int\psi f_{\underline{\theta}_0} d\mu - \int\psi f_{\underline{\theta}} d\mu|$$

$$\leq \lim_{\underline{\theta}\to\underline{\theta}_0} \int M\cdot|f_{\underline{\theta}_0} - f_{\underline{\theta}}| d\mu = 0,$$

i.e., $E_{\underline{\theta}}\psi(\underline{x})$ is continuous at $\underline{\theta}_0$. ||

Corollary 1.3.1. Suppose that for each $\underline{\theta} \in \Theta$ the random vector $\underline{X}$ has density $f_{\underline{\theta}}(\underline{x})$ with respect to a measure μ and suppose that for each $\underline{\theta}_0 \in \Theta$ condition (i) or (ii) of Theorem 1.3.3 holds. Then for any selection rule, R, and all $i = 1,2,\ldots,k$, $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is a continuous function on Θ.

Proof. Let $\psi_i(\underline{x})$, $i = 1,2,\ldots,k$, denote the individual selection probabilities for R. For any $i = 1,2,\ldots,k$, $P_{\underline{\theta}}(\text{select } \pi_i|R) = E_{\underline{\theta}}\psi_i(\underline{X})$. Since $0 \leq \psi_i(\underline{x}) \leq 1$, Theorem 1.3.3 applies. ||

Example 1.3.3. Suppose $\underline{\theta}$ is a location parameter and $\underline{X}$ has density $f_{\underline{\theta}}(\underline{x}) = f(\underline{x}-\underline{\theta})$ with respect to Lebesgue measure, μ, on R^k. Suppose $f(\underline{x})$ is continuous a.e. μ. Let A be the set of discontinuities of f. Then for a fixed $\underline{\theta}_0 \in \mathbb{R}^k$, the set of $\underline{x}$ for which $f(\underline{x}-\underline{\theta})$ is not continuous at $\underline{\theta}_0$ is $\{\underline{x}: \underline{x} = \underline{y} + \underline{\theta}_0, \underline{y} \in A\} = A + \underline{\theta}_0$ and $\mu(A + \underline{\theta}_0) = \mu(A) = 0$. So (i) of Theorem 1.3.3 is satisifed for every $\underline{\theta}_0 \in \mathbb{R}^k$ and $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is a continuous function of $\underline{\theta}$ on $\mathbb{R}^k$ for any R.

Example 1.3.4. Suppose $\underline{\theta}$ is a scale parameter and $\underline{X}$ has density $f_{\underline{\theta}}(\underline{x}) = f(x_1/\theta_1,\ldots,x_k/\theta_k)/\theta_1\cdot\theta_2\cdot\ldots\cdot\theta_k$ with respect to Lebesgue measure, μ, on $\mathbb{R}^k$. Suppose $f(\underline{x})$ is continuous a.e. μ. Let A be the set of discontinuities of f. Then for a fixed $\underline{\theta}_0 \in (0,\infty)\times\ldots\times(0,\infty)$, the set of $\underline{x}$ for which $f_{\underline{\theta}}(\underline{x})$ is not continuous at $\underline{\theta}_0$ is $\{\underline{x}: x_i = y_i\theta_i, i = 1,\ldots,k, \underline{y} \in A\} = A\cdot\underline{\theta}_0$ and $\mu(A) = 0$ implies $\mu(A\cdot\underline{\theta}_0) = 0$. So (i) of Theorem 1.3.3 is satisfied for every $\underline{\theta}_0$ and $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is a continuous function of $\underline{\theta}$ for any R.

Example 1.3.5. Suppose $\underline{X}$ has a multinomial distribution as in Example 1.3.1. The sample space is $\mathcal{X} = \{(x_1,x_2,\ldots,x_k): x_i \in \{0,1,\ldots,k\}, \Sigma x_i = N\}$. The density with respect to counting measure on $\mathcal{X}$ is given by

$$f_{\underline{p}}(\underline{x}) = \frac{N!}{x_1!\ldots x_k!}\, p_1^{x_1}\cdot p_2^{x_2}\ldots p_k^{x_k}.$$ For every $\underline{x} \in \mathcal{X}$

this is a polynomial in $\underline{p}$ and so is continuous in $\underline{p}$. So again, condition (i) of Theorem 1.3.3 is satisfied.

We will end this section by stating theorems analogous to Theorems 1.3.1 and 1.3.2 which give the minimax value when the loss is in terms of S', the number of non-best populations selected.

Theorem 1.3.4. Suppose Θ_0 is non empty. Suppose there exists $\underline{\theta}_0 \in \Theta_0$ such that $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is upper semicontinuous at $\underline{\theta}_0$ for all R and all $i = 1,2,\ldots,k$. Then the minimax value with respect to S' is $(k-1)P^*$.

Proof. Let $\pi_{(1)},\pi_{(2)},\ldots,\pi_{(k-1)}$ denote the k-1 non-best populations at $\underline{\theta}_0$. Then for any R,

$$E_{\underline{\theta}_0}(S'|R) = \sum_{i=1}^{k-1} P_{\underline{\theta}_0}(\text{select } \pi_{(i)}|R). \tag{1.3.4}$$

Let $\Theta_{(i)}$ be the subset of Θ where $\pi_{(i)}$ is best. By considering a sequence $\underline{\theta}_j \in \Theta_{(i)}$ converging to $\underline{\theta}_0$, we get $P_{\underline{\theta}_0}(\text{select } \pi_{(i)}|R) \geq P^*$ $i = 1,2,\ldots,k-1$. So

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

for any R. But $\varphi_i(\underline{x}) \equiv P^*$ is a rule for which $\sup_{\Theta} E_{\underline{\theta}}(S'|\varphi) = (k-1)P^*$. ||

Theorem 1.3.5. Suppose there exists a $\underline{\theta}_0$ such that $\underline{\theta}_0 \in \Theta_i$ for m different values of i, $2 \leq m \leq k$. Further suppose $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is upper semicontinuous at $\underline{\theta}_0$ for all R and all $i = 1,2,\ldots,k$. Then the minimax value with respect to S' is at least $(m-1)P^*$.

Proof. The proof is the same as that of Theorem 1.3.2 except it uses 1.3.4. Since for one of the m subscripts, $\underline{\theta}_0 \in \Theta_i$, i.e., π_i is best, the bound P* is obtained for only m-1 of the summands in 1.3.4. Hence $E_{\underline{\theta}_0}(S'|R) \geq (m-1)P^*$ for any R. ||

1.4. Minimaxity of Two Classical Rules

In this section, the results of Section 1.3 are used to show that, in location and scale parameter problems, two rules which have been proposed and studied by Gupta (1965) are minimax with respect to both S, the size of the selected subset, and S', the number of non-best populations selected.

1.4.1. The Location and Scale Parameter Problems

Throughout this section it is assumed that $X_1, X_2, \ldots, X_k$ are independent random variables. The c.d.f. of X_i is $F_{\theta_i}(x_i)$ which has density $f_{\theta_i}(x_i)$ with respect to μ, Lebesgue measure on the real line. If $F_{\theta_i}(x_i) = F(x_i - \theta_i)$ then $\underline{\theta}$ is called a location parameter. If $F_{\theta_i}(x_i) = F(x_i/\theta_i), \underline{\theta}$ is called a scale parameter. Here it is to be assumed that the distribution $F(\cdot)$ is known and is the same for all populations. Only the parameter values θ_i are unknown. In a location parameter problem, the parameter space is $\Theta = \mathbb{R}^k$. In a scale parameter problem, $\Theta = (0,\infty)\times(0,\infty)\times\ldots\times(0,\infty)$ (k times). In both cases the best population will be the one associated with the largest parameter value. With the appropriate modifications, results analogous to those which follow could be obtained if the population associated with the smallest parameter value is considered best. In both location and scale problems Θ_0, defined in (1.3.2), is given by $\Theta_0 = \{(\theta_1, \theta_2, \ldots, \theta_k) \in \Theta\colon \theta_1 = \theta_2 = \ldots = \theta_k\}$.

The following two selection rules have been proposed and studied by Gupta (1965).

<u>Definition 1.4.1</u>. For a location parameter problem define the selection rule R_1 by

$$R_1\colon \text{ select } \pi_i \text{ iff } x_i \geq \max_{1 \leq j \leq k} x_j - d \qquad i = 1,2,\ldots,k$$

where d is chosen to be the smallest positive constant such that the P*-condition (1.1.2) is satisfied.

Definition 1.4.2. For a scale parameter problem define the selection rule R_2 by

$$R_2: \text{ select } \pi_i \text{ iff } x_i \geq c \cdot \max_{1 \leq j \leq k} x_j \qquad i = 1,2,\ldots,k$$

where $0 < c < 1$ is the largest constant such that the P*-condition (1.1.2) is satisfied.

The properties of the rules R_1 and R_2, proven by Gupta (1965), which will be of use herein are summarized in the following.

Theorem 1.4.1. Let R denote R_1 in a location parameter problem or R_2 in a scale parameter problem.

a) $$\inf_{\Theta} P_{\underline{\theta}}(CS|R) = \inf_{\Theta_0} P(CS|R) = P^*$$

and $P_{\underline{\theta}}(CS|R) = P^*$ for all $\underline{\theta} \in \Theta_0$.

b) $E_{\underline{\theta}}(S|R) = kP^*$ for all $\underline{\theta} \in \Theta_0$.

c) If the density $f_{\underline{\theta}}(x)$ has monotone likelihood ratio (MLR) then

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

Theorem 1.4.2. a) Suppose $\underline{\theta}$ is a location parameter, i.e., the density of X_i is $f_{\theta_i}(x_i) = f(x_i - \theta_i)$. Suppose $f_\theta(x)$ has MLR and f is continuous a.e. μ. Then R_1 is minimax with respect to S.

b) Suppose $\underline{\theta}$ is a scale parameter, i.e., the density of X_i is $f_{\theta_i}(x_i) = f(x_i/\theta_i)/\theta_i$. Suppose $f_\theta(x)$ has MLR and f is continuous a.e. μ. Then R_2 is minimax with respect to S.

Remark 1.4.1. This generalizes a result of Gupta and Studden (1966). They proved that R_1 (or R_2) is minimax among all permutation invariant rules. Theorem 1.4.2 asserts minimaxity among all rules.

Proof. Let R denote R_1 in a location parameter problem or R_2 in a scale parameter problem. Let μ^k be Lebesgue measure on $\mathbb{R}^k$. Then $f(\underset{\sim}{x}) = \prod_{i=1}^{k} f(x_i)$ is continuous a.e. μ^k. By Example 1.3.3 or Example 1.3.4, the continuity assumption of Theorem 1.3.1 is satisfied. So the minimax value with respect to S is kP*. By Theorem 1.4.1c, $\sup_{\Theta} E_{\underset{\sim}{\theta}}(S|R) = kP^*$ so R is minimax. ||

Theorem 1.4.3. Let R denote R_1 in a location parameter problem or R_2 in a scale parameter problem. Suppose $\sup_{\Theta} E_{\underset{\sim}{\theta}}(S|R) = \sup_{\Theta_0} E_{\underset{\sim}{\theta}}(S|R)$ and the minimax value with respect to S' is (k-1)P*. Then R is minimax with respect to S'.

Proof. Let $\underset{\sim}{\theta}_0 \in \Theta_0$ and $\underset{\sim}{\theta} \in \Theta$. By Theorem 1.4.1b,

$$E_{\underset{\sim}{\theta}_0}(S|R) = \sup_{\Theta_0} E_{\underset{\sim}{\theta}}(S|R) = \sup_{\Theta} E_{\underset{\sim}{\theta}}(S|R).$$

So

$$\begin{aligned} 0 &\leq E_{\underset{\sim}{\theta}_0}(S|R) - E_{\underset{\sim}{\theta}}(S|R) \\ &= E_{\underset{\sim}{\theta}_0}(S'|R) + P_{\underset{\sim}{\theta}_0}(CS|R) - E_{\underset{\sim}{\theta}}(S'|R) - P_{\underset{\sim}{\theta}}(CS|R) \\ &= E_{\underset{\sim}{\theta}_0}(S'|R) - E_{\underset{\sim}{\theta}}(S'|R) + (P_{\underset{\sim}{\theta}_0}(CS|R) - P_{\underset{\sim}{\theta}}(CS|R)) \\ &\leq E_{\underset{\sim}{\theta}_0}(S'|R) - E_{\underset{\sim}{\theta}}(S'|R) \end{aligned}$$

the last inequality being the result of Theorem 1.4.1a. So using Theory 1.4.1a and b yields

$$\sup_{\Theta} E_{\underline{\theta}}(S'|R) = E_{\underline{\theta}_0}(S'|R)$$

$$= E_{\underline{\theta}_0}(S|R) - P_{\underline{\theta}_0}(CS|R)$$

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

Hence R is minimax with respect to S'. ||

Theorem 1.4.4. a) Suppose $\underline{\theta}$ is a location parameter, i.e., the density of X_i is $f_{\theta_i}(x_i) = f(x_i - \theta_i)$. Suppose $f_\theta(x)$ has MLR and f is continuous a.e. μ. Then R_1 is minimax with respect to S'.
b) Suppose $\underline{\theta}$ is a scale parameter, i.e., the density of X_i is $f_{\theta_i}(x_i) = f(x_i/\theta_i)/\theta_i$. Suppose $f_\theta(x)$ has MLR and f is continuous a.e. μ. Then R_2 is minimax with respect to S'.

Proof. Let R denote R_1 in a location parameter problem or R_2 in a scale parameter problem. As in Theorem 1.4.2, Theorem 1.3.4 is applicable so the minimax value is $(k-1)P^*$. By Theorem 1.4.1c, $\sup_{\Theta} E_{\underline{\theta}}(S|R) = \sup_{\Theta_0} E_{\underline{\theta}}(S|R)$. So by Theorem 1.4.3, R is minimax with respect to S'. ||

Example 1.4.1. Using Example 1, page 330 of Lehmann (1959), location parameter densities which satisfy the hypotheses of Theorem 1.4.2 and Theorem 1.4.4 and hence for which R_1 is minimax include normal, exponential, rectangular, logistic and Laplace (double exponential). Furthermore, all of the order statistics from the above distributions have MLR. Leong (1976) proposed using R_1 to select the Laplace population with the largest mean where X_i is the median of n observations from π_i. Since the distribution of X_i has MLR with θ_i as a location parameter, the rule proposed by Leong is minimax.

Example 1.4.2. Using Example 2, page 331 of Lehmann (1959), scale parameter densities which satisfy the hypotheses of Theorem 1.4.2 and Theorem 1.4.4 and hence for which R_2 is minimax include normal, exponential, Laplace and Cauchy.

1.5. Necessary Conditions for Minimaxity

In Section 1.3, the behavior of a selection rule on the set Θ_0 (see (1.3.2)) was important in determining the minimax value of a selection problem. Analysis similar to that in Section 1.3 will yield necessary conditions on the behavior of a selection rule on Θ_0 which must be satisfied by any minimax selection rule. These conditions are principally of use in proving that certain rules, in violating these conditions, are not minimax. Theorem 1.5.1 provides the necessary conditions for minimaxity with respect to S and Theorem 1.5.2 the analogous conditions for S'.

Theorem 1.5.1. Let R be a minimax selection rule with respect to S. Suppose $P_{\underline{\theta}}(\text{select } \pi_i|R)$ is upper semicontinuous at each $\underline{\theta}_0 \in \Theta_0$ for all $i = 1,2,\ldots,k$. Then for all $\underline{\theta}_0 \in \Theta_0$,

a) $P_{\underline{\theta}_0}(\text{select } \pi_i|R) = P^* = \inf_{\Theta} P_{\underline{\theta}}(CS|R)$ for all $i = 1,2,\ldots,k$

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

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

Remark 1.5.1. Condition a) of Theorem 1.5.1 implies condition b) and the first equality in c) as well as a) and the first equality of b) in Theorem 1.5.2. If one wishes to verify these conditions for a given rule, to check if it might be minimax with respect to S or S', only 1.5.1a need be verified.

Proof. Fix $\underline{\theta}_0 \in \Theta_0$. As in the proof of Theorem 1.3.1, it follows that

(1.5.1) $$P_{\underline{\theta}_0}(\text{select } \pi_i | R) \geq P^* \text{ for all } i = 1,2,\ldots,k.$$

By considering the "no data rule", $\psi_i(\underline{x}) \equiv P^*$, it can be seen that the minimax value is no greater than kP^*. So, since R is minimax and (1.5.1) is true,

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

So all the inequalities are equalities and c) is true. In view of (1.5.1) and (1.5.2), a) is true. b) follows from a) since $P_{\underline{\theta}_0}(CS|R) = P_{\underline{\theta}_0}(\text{select } \pi_i|R)$ where $\underline{\theta} \in \Theta_i$. ||

Nagel (1970, Chapters 1 and 2) found that a condition related to 1.5.1b, viz.,

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

was an important property of just selection rules. Conditions 1.5.1 a) and b) have long been recognized (c.f. 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 for which one of the k populations has arbitrarity been tagged as best, e.g., a location or scale parameter problem in which best is defined in terms of the largest or smallest parameter value. Theorem 1.5.1 verifies that, in terms of minimaxity considerations, the intuition is justified.

Theorem 1.5.2. Let R be a minimax selection rule with respect to S'. Suppose $P_{\underline{\theta}}(\text{select } \pi_i | R)$ is upper semicontinuous at each $\underline{\theta}_0 \in \Theta_0$ for all $i = 1,2,\ldots,k$. Then for all $\underline{\theta}_0 \in \Theta_0$,

a) $P_{\underline{\theta}_0}(\text{select } \pi_i | R) = P^* = \inf_{\Theta} P_{\underline{\theta}}(CS|R)$ for all $i = 1,2,\ldots,k$, $i \neq j$, where $\underline{\theta}_0 \in \Theta_j$

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

Remark 1.5.2: For those problems in which the random variables $X_1, X_2,\ldots,X_k$ are exchangeable for all $\underline{\theta} \in \Theta_0$ and the rule R is invariant under permutations (symmetric), the following is true for any $\underline{\theta} \in \Theta_0$:

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

In such a problem, then, 1.5.2a implies 1.5.1a, b and c. So for many problems, the necessary conditions derived in Theorem 1.5.1 for minimaxity with respect to S and those derived in Theorem 1.5.2 for minimaxity with respect to S' are essentially the same.

Proof. Fix $\underset{\sim}{\theta}_0 \in \Theta_0$. Let $\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(k-1)}$ denote the k-1 non-best populations at $\underset{\sim}{\theta}_0$. As in the proof of Theorem 1.3.4, it follows that

$$(1.5.3) \qquad P_{\underset{\sim}{\theta}_0}(\text{select } \pi_{(i)}|R) \geq P^*, \quad i = 1,2,\ldots,k-1.$$

By considering the rule, $\psi_i(\underset{\sim}{x}) \equiv P^*$, (1.5.3) and the minimaxity of R, the inequality

$$(1.5.4) \quad (k-1)P^* = \sup_{\Theta} E_{\underset{\sim}{\theta}}(S'|\psi) \geq \sup_{\Theta} E_{\underset{\sim}{\theta}}(S'|R) \geq E_{\underset{\sim}{\theta}_0}(S'|R)$$

$$= \sum_{i=1}^{k-1} P_{\underset{\sim}{\theta}_0}(\text{select } \pi_{(i)}|R) \geq (k-1)P^*$$

is obtained. a) and b) follow as in Theorem 1.5.1. ||

Now the conditions of Theorems 1.5.1 and 1.5.2 can be used to show that some selection rules which have been proposed are not minimax. Also, a method of constructing rules which do satisfy the conditions of Theorems 1.5.1 and 1.5.2, and thus might be minimax, will be outlined.

Example 1.5.1. Consider the multinomial selection problem introduced in Example 1.3.1. The goal is to select the cell with the largest cell probability. Here Θ_0 is the single point $(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$. Example 1.3.5 shows that the continuity assumptions of Theorems 1.5.1 and 1.5.2 will be satisfied for any selection rule. Gupta and Nagel (1967) proposed using rule R_1 of Section 1.4, viz., select π_i iff $x_i \geq \max_{1 \leq j \leq k} x_j - d$ for this problem. Gupta and Nagel found that for some values of k and P*, the $\inf_{\Theta} P_{\underset{\sim}{\theta}}(CS|R_1)$ did not occur at the point $(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$. So condition 1.5.1b is violated and R_1 can not be minimax with respect to S for these values of k and P*. Since Remark 1.5.2 is applicable in this problem, R_1 is not minimax with respect to S' either.

Example 1.5.2. Consider the binomial selection problem in which $\pi_1, \pi_2, \ldots, \pi_k$ are independent binomial populations with success probabilities $p_1, p_2, \ldots, p_k$ and X_i is the number of successes in n observations from π_i. Gupta and Sobel (1960) proposed using the rule R_1 to select a subset including the population associated with the largest p_i. In this problem $\Theta_0 = \{(p_1, p_2, \ldots, p_k) | p_1 = p_2 = \ldots = p_k = p$ $0 \leq p \leq 1\}$. It was realized by the above authors that $E_{\underline{\theta}}(S|R_1)$ was not constant on Θ_0 as required by 1.5.1c, if R_1 were to be minimax. Indeed, $E_{\underline{\theta}}(S|R_1) \to k$ as $\underline{\theta} = (p, p, \ldots, p) \to (1, 1, \ldots, 1)$ and $E_{\underline{\theta}}(S'|R_1) \to$ k-1 in the same limit. Gupta and Sobel (1960) proposed an arcsin transformation of the data but those results are of an asymptotic $(n \to \infty)$ nature and for any finite n, the behavior of $E_{\underline{\theta}}(S|R_1)$ and $E_{\underline{\theta}}(S'|R_1)$ will be the same as above.

Example 1.5.3. Consider again the binomial selection problem introduced in Example 1.5.2. Gupta and Nagel (1971) proposed a conditional rule for this problem which satisfies 1.5.1a and hence all the other first equalities in Theorems 1.5.1b and c and 1.5.2a and b and so may be minimax with respect to S and S'. The rule proposed by Gupta and Nagel is defined in terms of the individual selection probabilities by

$$(1.5.5)\quad \psi_i(x_1,x_2,\ldots,x_k)=\begin{cases}1 & x_i > c_T\\ \rho_T & x_i = c_T\\ 0 & x_i < c_T\end{cases}$$

where $T = \sum_{i=1}^{k} x_i$, and ρ_T and c_T are constants chosen to satisfy

$$(1.5.6)\quad E_{\underline{\theta}}(\psi_i(\underline{X})|T) = P_{\underline{\theta}}(X_i > c_T|T) + \rho_T P_{\underline{\theta}}(X_i = c_T|T) = P^*$$

for all $\underline{\theta} \in \Theta_0$. The important point is that T is a sufficient statistic for $\underline{\theta} \in \Theta_0$. So the distribution of $\underline{X}$ given T does not depend on $\underline{\theta}$ on Θ_0. This is what makes the determination of the constants c_T and ρ_T, independent of $\underline{\theta}$, possible. 1.5.1a is satisfied since, for $\underline{\theta} \in \Theta_0$,

$$(1.5.7)\quad P_{\underline{\theta}}(\text{select } \pi_i|\psi) = E_{\theta}\psi_i(\underline{X}) = E_{\underline{\theta}}E_{\underline{\theta}}(\psi_i(\underline{X})|T)$$

$$= E_{\underline{\theta}}P^* = P^*.$$

This technique of conditioning on a statistic which is sufficient for $\underline{\theta} \in \Theta_0$ seems to be very useful for constructing selection rules which satisfy the conditions of Theorem 1.5.1 and 1.5.2. Nagel (1970, Sections 2.4 and 2.5) proposed rules similar to (1.5.5) which satisfy the conditions of Theorems 1.5.1 and 1.5.2 for selection problems involving Poisson and negative binomial populations.

<u>Example 1.5.4</u>. The following general problem has been considered by Gupta and Panchapakesan (1972). Suppose $\pi_1,\pi_2,\ldots,\pi_k$ are independent populations with absolutely continuous distributions $F_\theta(x_i)$ where $\theta \in I$ an interval on the real line. The family $\{F_\theta\colon \theta \in I\}$ is assumed to be stochastically increasing in θ. If

$\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ denote the ordered parameters, a class of procedures, investigated by Gupta and Panchapakesan (1972), for selecting a subset containing the population associated with $\theta_{[k]}$ are defined as follows:

(1.5.8) R_h: select π_i iff $h(x_i) \geq \max_{1 \leq j \leq k} x_j$

where h is a real valued function satisfying certain regularity conditions. As in Example (1.3.2) $\Theta_0 = \{(\theta,\theta,\ldots,\theta): \theta \in I\}$. For any $\underline{\theta}_0 \in \Theta_0$ and any $i = 1,2,\ldots,k$,

(1.5.9) $$P_{\underline{\theta}_0}(\text{select } \pi_i | R_h) = \int F_{\theta_0}^{k-1}(h(x))dF_{\theta_0}(x).$$

By Theorems 1.5.1 and 1.5.2, if the procedure R_h is to be minimax with respect to S or S', the expression in (1.5.9) must be constant on Θ_0. But Gupta and Panchapakesan (1972) have shown that in many cases of interest (1.5.9) is an increasing function of θ_0. They have proved the following.

<u>Theorem 1.5.3</u>. For the procedure R_h defined by (1.5.8), the expression (1.5.9) is non-decreasing in θ_0 provided that

(1.5.10) $$f_\theta(x) \frac{\partial}{\partial\theta} F_\theta(h(x)) - h'(x)f_\theta(h(x)) \frac{\partial}{\partial\theta} F_\theta(x) \geq 0$$

for all $\theta \in I$ and all x,

where $h'(x) = (d/dx)h(x)$ and f_θ is the density of F_θ. Further, (1.5.9) is strictly increasing in θ_0 if strict inequality holds in (1.5.10) on a set of positive Lebesgue measure.

Gupta and Studden (1970) have established the strict monotonicity of (1.5.9) for the non-central χ^2 and non-central F distributions when the procedure R_h is R_2 of Section 1.4. This is important in the problem of selection in terms of Mahalanobis distance for multivariate normal distributions.

Gupta and Panchapakesan (1969) have established the strict monotonicity of (1.5.9) in the problem of selection in terms of the largest (or smallest) multiple correlation coefficient when the rule R_h is R_2 (or an analogous rule). Both the conditional and unconditional cases are considered as well as two different statistics, the sample multiple correlation coefficient, R^2, and a transform thereof, $R^{*2} = R^2/(1-R^2)$. In all cases the strict monotonicity of (1.5.9) is established.

So in all of the above problems, the proposed rules are not minimax with respect to S or S'. This was previously reported in some cases. But the interesting point here is that one need not necessarily examine $E_{\underline{\theta}}(S|R)$ to determine that a rule is not minimax with respect to S or S'. Often in investigating the least favorable configuration, i.e., that $\underline{\theta}_0$ for which

(1.5.11) $$P_{\underline{\theta}_0}(CS|R) = \inf_{\Theta} P_{\underline{\theta}}(CS|R),$$

one can reduce the problem to investigating the $\inf_{\Theta_0} P_{\underline{\theta}}(CS|R)$. This, for example, is the case with just rules as defined by Nagel (1970). If one finds that $P_{\underline{\theta}}(CS|R)$ is not constant on Θ_0 and some mild continuity assumptions (e.g., (i) or (ii) of Theorem 1.3.3) are satisfied, then the rule, R, is not minimax with respect to S or S'. Thus, the only analysis required, to show that a proposed rule is not minimax, may be the analysis used to find the least favorable configuration.

1.6. Minimaxity Considerations for Seal's Class

Seal (1955) proposed a class of selection rules for the location parameter problem. In this section, a lower bound is obtained for the $\sup_{\Theta} E_{\underline{\theta}}(S|R)$ and the $\sup_{\Theta} E_{\underline{\theta}}(S'|R)$ for rules in this class. This lower bound can then be used to prove that, in certain cases, the rules in this class are not minimax.

Definition 1.6.1. Let $\mathscr{L}$ denote the class of selection rules which have the following form:

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

where $x_{[1]} \leq x_{[2]} \leq \cdots \leq x_{[k-1]}$ 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.

The rule R_1, introduced in Section 1.4, is in the class $\mathscr{L}$. R_1 corresponds to setting $a_{k-1} = 1$ and $a_j = 0$, $j = 1,2,\ldots,k-2$. Comparisons between $E_{\underline{\theta}}(S|R_1)$ and $E_{\underline{\theta}}(S|R)$ for certain other rules, $R \in \mathscr{L}$, have previously been made by Seal (1957) and Deely and Gupta (1968). Those authors considered specific parameter configurations (e.g., slippage configurations) and specific alternatives to R_1. The results which follow differ from the previous work in that the sup over all parameter configurations and all rules in $\mathscr{L}$ are considered. Nevertheless, the following results tend to confirm the work of the previous authors in indicating that, although a certain rule may have smaller expected size than R_1 for some parameter points, over much of the parameter space R_1 has the smaller expected size.

Throughout this section it will be assumed that $\Theta = \mathbb{R}^k$. The following notation will be used. $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ will denote the ordered coordinates of the parameter point $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ so that the best population is the one associated with $\theta_{[k]}$. Sometimes, a sequence of parameter points $<\underline{\theta}_n>$ will be considered in which case the ordered coordinates of $\underline{\theta}_n = (\theta_{n1}, \theta_{n2}, \ldots, \theta_{nk})$ will be denoted by $\theta_{n[1]} \leq \theta_{n[2]} \leq \cdots \leq \theta_{n[k]}$.

The following theorem will be used primarily to obtain a lower bound on the expected subset size. But, as stated, it also points out an intuitively undesirable property of all rules in the class $\mathscr{L}$, except R_1, namely, it is possible to find a parameter point such that $\theta_{[k]} - \theta_{[k-1]}$ is arbitrarily large but the probability of including the population associated with $\theta_{[k-1]}$ is arbitrarily near one.

Theorem 1.6.1. Suppose the observation $\underline{X} = (X_1, X_2, \ldots, X_k)$ satisfies either (i) or (ii):

(i) $\underline{X}$ has density $f(\underline{x}-\underline{\theta})$ with respect to Lebesgue measure on $\mathbb{R}^k$

(ii) the populations are independent with the c.d.f. of X_i given by $F(x_i - \theta_i)$.

Let R be a rule in $\mathscr{L} \setminus \{R_1\}$ defined by constants $\{a_j, d\colon j=1,2,\ldots,k-1\}$. Let $r = \min\{i\colon a_i > 0\}$. Then there exists a sequence of parameter points $<\underline{\theta}_n>$ and a subset $K \subset \{1,2,\ldots,k\}$ of size k-r-1, such that for all $i \in K$, $\lim_{n\to\infty} \theta_{n[k]} - \theta_{ni} = \infty$ and $\lim_{n\to\infty} P_{\underline{\theta}_n}(\text{select } \pi_i|R) = 1$.

Proof: Note that since R is not R_1, $r \leq k-2$ so the set K will be non-empty. 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 R. Define a sequence of subsets of $\mathcal{X}$ as follows:

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

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

First it will be shown that $A_n \subset S_j$ $j = r+1, r+2, \ldots, k-1$ 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 $j \in \{r+1, r+2, \ldots, k-1\}$. Let $\underline{x} \in A_n$. Then $x_{[k-1]} = x_k$, $\{x_{[k-2]}, x_{[k-3]}, \ldots, x_{[r+1]}\} = \{x_{k-1}, x_{k-2}, \ldots, x_{r+1}\} \setminus \{x_j\}$ and $\{x_{[1]}, x_{[2]}, \ldots, x_{[r]}\} = \{x_1, x_2, \ldots, x_r\}$.

Using these facts and the definition of A_n, the following relationships are obvious:

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

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

So using (1.6.1), (1.6.2) and the fact that $a_m = 0$, $m = 1,\ldots,r-1$, it follows that

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

But $x_j > -d$ by the definition of A_n so

$$x_j > \sum_{m=1}^{k-1} a_m x_{[m]} - d, \text{ i.e. } \underline{x} \in S_j.$$

This was true for any $\underline{x} \in A_n$ so $A_n \subset S_j$.

Define a sequence of parameter points $\underline{\theta}_n = (\theta_{n1}, \theta_{n2}, \ldots, \theta_{nk})$ by

$$(1.6.4) \qquad \theta_{nj} = \begin{cases} \frac{3}{2} n, & j = k \\ \frac{1}{2} n, & j = r+1, r+2, \ldots, k-1 \\ c_n - n, & j = 1, 2, \ldots, r \end{cases}$$

Let $K = \{r+1, r+2, \ldots, k-1\}$. For any $j \in K$

$$\lim_{n\to\infty} \theta_{n[k]} - \theta_{n_j} = \lim_{n\to\infty} (\frac{3}{2} n - \frac{1}{2} n) = \infty.$$

Since $A_n \subset S_j$, $j \in K$, for any $j \in K$

$$P_{\underline{\theta}_n}(\text{select } \pi_j | R) = P_{\underline{\theta}_n}(S_j)$$

$$\geq P_{\underline{\theta}_n}(A_n).$$

In case (i)

$$P_{\underline{\theta}_n}(A_n) = \int \underset{A_n}{\cdots} \int f(\underline{x} - \underline{\theta}_n) dx_1 \ldots dx_n$$

$$= \int_n^{2n} \int_{-d}^{n} \cdots \int_{-d}^{n} \int_{-\infty}^{c_n} \int_{-\infty}^{c_n} f(\underline{x} - \underline{\theta}_n) dx_1 \ldots dx_r dx_{r+1} \ldots dx_{k-1} dx_k$$

$$= \int_{-\frac{n}{2}}^{\frac{n}{2}} \int_{-\frac{n}{2}-d}^{\frac{n}{2}} \cdots \int_{-\frac{n}{2}-d}^{\frac{n}{2}} \int_{-\infty}^{n} \cdots \int_{-\infty}^{n} f(\underline{x}) dx_1 \ldots dx_r dx_{r+1} \ldots dx_{k-1} dx_k.$$

In the last expression, the integrand no longer depends on n and the limits of integration go to ∞ and $-\infty$ as appropriate. So, for $j \in K$,

(1.6.5) $$1 \geq \lim_{n\to\infty} P_{\underline{\theta}_n}(\text{select } \pi_j|R) \geq \lim_{n\to\infty} P_{\underline{\theta}_n}(A_n) = 1$$

and the limit is one in case (i).

In case (ii)

$$P_{\underline{\theta}_n}(A_n) = [F(2n-\theta_{nk})-F(n-\theta_{nk})] \prod_{j=r+1}^{k-1} [F(n-\theta_{nj})-F(-d-\theta_{nj})] \times \prod_{j=1}^{r} F(c_n-\theta_{nj})$$

$$= [F(n/2)-F(-n/2)] \prod_{j=r+1}^{k-1} [F(n/2)-F(-d-n/2)] \prod_{j=1}^{r} F(n).$$

In the last expression, all the factors go to one as n approaches ∞. So, for $j \in K$, (1.6.5) holds and the limit is one in case ii). ||

Theorem 1.6.2. Let R be a rule in $\mathscr{C} \setminus \{R_1\}$ defined by constants $\{a_j, d: j = 1,\ldots,k-1\}$. Let $r = \min\{i: a_i > 0\}$. Suppose (i) or (ii) of Theorem 1.6.1 holds. Then

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

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

Proof. Using all the notation defined in the proof of Theorem 1.6.1 we have

(1.6.6)
$$\sup_{\Theta} E_{\underline{\theta}}(S|R) \geq \lim_{n\to\infty} E_{\underline{\theta}_n}(S|R)$$
$$\geq \lim_{n\to\infty} \sum_{m=r+1}^{k} P_{\underline{\theta}_n}(\text{select } \pi_m|R)$$

Theorem 1.6.1 proved that the first k-r-1 terms approached one in the limit. For every $\underset{\sim}{x} \in A_n$, x_k is the largest coordinate so $A_n \subset S_k$ for all n. Thus $P_{\underline{\theta}_n}(\text{select } \pi_k|R) = P_{\underline{\theta}_n}(S_k|R) \geq P_{\underline{\theta}_n}(A_n|R) = 1$ in the limit. Hence the bound k-r for a).

For b), it can be seen from (1.6.4) that π_k is the best population for all $\underline{\theta}_n$. Thus we use the same reasoning as above, except not including $P_{\underline{\theta}_n}(\text{select } \pi_k|R)$ in the sum (1.6.6) to obtain the bound k-r-1 for part b). ||

Now we are in a position to show that for any rule R in the class $\mathscr{L}$, except R_1, there are values of P* for which R is not minimax.

<u>Theorem 1.6.3</u>. Let R be a rule in $\mathscr{L} \setminus \{R_1\}$ defined by constants $\{a_j, d: j = 1,\ldots,k-1\}$. Let $r = \min\{i: a_i > 0\}$. Suppose (i) or (ii) of Theorem 1.6.1 holds. Then

a) if $P^* < (k-r)/k$, R is not minimax with respect to S

b) if $P^* < (k-r-1)/(k-1)$, R is not minimax with respect to S'.

<u>Proof</u>. For part a, the "no data rule", $\psi_i(\underset{\sim}{x}) \equiv P^*$, has $\sup_{\Theta} E_{\underline{\theta}}(S|\psi) = kP^* < k-r \leq \sup_{\Theta} E_{\underline{\theta}}(S|R)$, the last inequality following from Theorem 1.6.2a.

Part b) follows analogously. ||

<u>Corollary 1.6.1</u>. Suppose (i) or (ii) of Theorem 1.6.1 holds. Then

a) if $P^* < 2/k$ no rule in the class $\mathscr{L}$, with the possible exception of R_1, is minimax with respect to S.

b) if $P^* < 1/(k-1)$, no rule in the class $\mathscr{G}$, with the possible exception of R_1, is minimax with respect to S'.

Proof. Any rule, excluding R_1, in $\mathscr{G}$ has $r \leq k-2$. So, the smallest upper bound given in Theorem 1.6.3a is $2/k$. Hence part a) is true.

Part b) follows analogously. ||

CHAPTER II

MINIMAXITY AND ADMISSIBILITY OF SELECTION RULES IN TERMS OF THE PROBABILITY OF INCLUDING NON-BEST POPULATIONS IN THE SELECTED SUBSET

In this chapter, risk is measured in terms of the maximum probability of including any non-best population in the selected subset. This risk is defined in Section 2.1. In Section 2.2, translation invariant, scale invariant and just rules are defined and some characterizations of them are given. Theorem 2.3.1 is a minimaxity result which, when applied in the location and scale parameter problems, shows that the rules R_1 and R_2 are minimax and admissible in the class of non-randomized, just and invariant rules. Section 2.4 examines the behavior of the class $\mathcal{S}$ of selection rules (Definition 1.6.1) with respect to this risk. Section 2.5 addresses a specific problem, first considered by Seal (1955), regarding the probabilities of accepting and rejecting best and non-best populations when the parameters are in a slippage configuration.

2.1. Risk Measured by Probabilities of Including Non-best Populations

In Chapter 1, risk was measured by the expected size of the selected subset. In attempting to keep this quantity small, it is conceivable that, at a particular parameter point, the probability of including a particular non-best population could be relatively large while the sum of all such probabilities could be relatively small when compared to k-1. In this chapter, risk is measured in terms of the probability of including each non-best population in the selected subset and rules are examined which attempt to keep all such probabilities small. Specifically, in this chapter we define the risk of a selection rule φ by

$$M(\underline{\theta},\varphi) = \max_{\{i:\underline{\theta}\notin\Theta_i\}} P_{\underline{\theta}}(\text{select } \pi_i|\varphi) \tag{2.1.1}$$

where Θ_i is the subset of the parameter space where π_i is the best population. Since at each parameter point exactly one population is best, the maximum is over k-1 quantities and is simply the maximum of the probabilities of selecting each of the non-best populations.

2.2. Just and Invariant Procedures and an Ordering of Distributions

In this section, certain classes of selection rules which will be of importance in later sections are defined. An ordering property of distribution functions is also introduced. These concepts are clarified somewhat by means of some characterization lemmas and are illustrated by means of some location and scale parameter examples.

Definition 2.2.1. A selection rule is *just* if for every $i = 1,2,\ldots,k$, $\varphi_i(x_1,x_2,\ldots,x_k)$ is a non-decreasing function of x_i and a non-increasing function of x_j, $j \neq i$.

The concept of justness is appealing if the best population is the one associated with the largest parameter value and the increase of a parameter value causes the distribution of the observation to be stochastically larger. In such a case, justness means that the probability of selecting a given population does not decrease if the observation becomes more favorable with respect to that population. Location and scale parameters are common examples of this monotonic behavior. The importance of just rules was recognized by Studden (1967, Lemma 3.1 (iv)) and was defined and investigated in more generality by Nagel (1970). See also Gupta and Nagel (1971).

Definition 2.2.2. A selection rule is called translation invariant if for every $\underline{x} \in \mathbb{R}^k$, for every $c \in \mathbb{R}$ and for every $i = 1,2,..,k$, $\varphi_i(x_1+c,x_2+c,\ldots,x_k+c) = \varphi_i(x_1,x_2,\ldots,x_k)$.

Definition 2.2.3. A selection rule is called scale invariant if for every $\underline{x} \in \mathbb{R}^k$, for every $c \in (0,\infty)$ and for every $i = 1,2,\ldots,k$, $\varphi_i(cx_1,cx_2,\ldots,cx_k) = \varphi_i(x_1,x_2,\ldots,x_k)$.

Restriction to translation invariant and scale invariant rules is appealing in location and scale parameter problems, respectively. The rationale behind this restriction is similar to that in hypotheses testing, namely, that if the parameter vector is translated, the population which is best remains unchanged.

Lemmas 2.2.1 and 2.2.2 provide useful characterization of selection rules which are both just and translation or scale invariant.

Lemma 2.2.1. A selection rule, $\varphi(\underline{x}) = (\varphi_1(\underline{x}), \varphi_2(\underline{x}),\ldots,\varphi_k(\underline{x}))$, is just and translation invariant if and only if the following two conditions hold:

(i) for every $i = 1,2,\ldots,k$, φ_i is a function only of the set of differences $\{x_j-x_i;\ j = 1,2,\ldots,k,\ j \neq i\}$,

(ii) if $\underline{x}$ and $\underline{y}$ satisfy $x_j-x_i \leq y_j-y_i$ for every $j \neq i$, then $\varphi_i(\underline{x}) \geq \varphi_i(\underline{y})$.

Proof. φ is translation invariant if and only if (i) holds because the differences are a maximal invariant for the translation group (see Lehmann (1959) p. 216). Suppose φ is just and translation invariant. Let $\underline{x}$ and $\underline{y}$ be as in (ii). Then using first invariance and then justness yields

$$\begin{aligned}\varphi_i(\underline{x}) &= \varphi_i(x_1-x_i+y_i,\ x_2-x_i+y_i,\ldots,x_k-x_i+y_i)\\ &= \varphi_i(x_1-x_i+y_i,\ x_2-x_i+y_i,\ldots,y_i,\ldots,x_k-x_i+y_i)\\ &\geq \varphi_i(y_1,y_2,\ldots,y_i,\ldots,y_k) = \varphi_i(\underline{y})\end{aligned}$$

so (ii) is true.

Now suppose (ii) is true. Fix $\underline{x} \in \mathbb{R}^k$, $\epsilon \geq 0$ and $i \neq j$. Then $x_j + \epsilon - x_i \geq x_j-x_i$ and all other differences are equal so by (ii), $\varphi_i(x_1,x_2,\ldots,x_k) \geq \varphi_i(x_1,x_2,\ldots,x_j+\epsilon,\ldots,x_k)$, i.e., φ_i is non-increasing in x_j, $j \neq i$. Also, $x_j-(x_i+\epsilon) \leq x_j-x_i$ for every $j \neq i$ so by (ii) $\varphi_i(x_1,x_2,\ldots,x_i+\epsilon,\ldots,x_k) \geq \varphi_i(x_1,x_2,\ldots,x_i,\ldots,x_k)$, i.e., φ_i is non-decreasing in x_i. Hence φ is just. ||

Lemma 2.2.2. Suppose $x_i > 0$, $y_i > 0$, $i = 1,2,\ldots,k$. A selection rule $\varphi(\underline{x}) = (\varphi_1(\underline{x}), \varphi_2(\underline{x}),\ldots,\varphi_k(\underline{x}))$ is just and scale invariant if and only if the following two conditions hold:

(i) for every $i = 1,2,\ldots,k$, φ_i is a function only of the set of quotients $\{x_j/x_i: j = 1,2,\ldots,k,\ j \neq i\}$

(ii) if $\underset{\sim}{x}$ and $\underset{\sim}{y}$ satisfy $x_j/x_i \leq y_j/y_i$ for every $j \neq i$, then $\varphi_i(\underset{\sim}{x}) \geq \varphi_i(\underset{\sim}{y})$.

Proof. Replace the differences in Lemma 2.2.1 with the quotients. ||

The following ordering property of distributions was introduced by Lehmann (1952) and further investigated by Lehmann (1955). The usefulness of this concept will become apparent, forthwith.

Definition 2.2.4. A subset $A \subset \mathbb{R}^k$ is monotone if $\underset{\sim}{x} \in A$ and $\underset{\sim}{y}$ satisfies $y_i \leq x_i$, $i = 1,2,\ldots,k$, implies $\underset{\sim}{y} \in A$.

Definition 2.2.5. A family of probability distributions on $\mathbb{R}^k$, $\{F_{\underset{\sim}{\theta}}: \underset{\sim}{\theta} \in \Theta \subset \mathbb{R}^k\}$, is said to have the stochastic increasing property (SIP) if $\underset{\sim}{\theta} \in \Theta$, $\underset{\sim}{\theta}' \in \Theta$, and $\theta_i \leq \theta_i'$ for every $i = 1,2,\ldots,k$, implies

$$P_{\underset{\sim}{\theta}}(A) = \int_A dF_{\underset{\sim}{\theta}} \geq \int_A dF_{\underset{\sim}{\theta}'} = P_{\underset{\sim}{\theta}'}(A)$$

for all monotone sets A.

Lehmann (1952) suggested the following method of proving that a family has the SIP. For $\underset{\sim}{\theta}$ and $\underset{\sim}{\theta}'$ as in Definition 2.2.5, prove the existence of random variables $Z_1,Z_2,\ldots,Z_r$ and functions f_i and g_i such that $X_i = f_i(Z_1,Z_2,\ldots,Z_r)$, $Y_i = g_i(Z_1,Z_2,\ldots,Z_r)$, $X_i \leq Y_i$ for every $i = 1,2,\ldots,k$, and the c.d.f.'s of $(X_1,X_2,\ldots,X_k)$ and $(Y_1,Y_2,\ldots,Y_k)$ are $F_{\underset{\sim}{\theta}}$ and $F_{\underset{\sim}{\theta}'}$ respectively.

In Lemmas 2.2.1 and 2.2.2 it was seen that if consideration is restricted to translation invariant or scale invariant procedures, the distribution of differences or quotients of the observations would be of interest. The following two lemmas show that these random vectors have the SIP in location and scale problems.

Lemma 2.2.3. Suppose $\underline{\theta} \in \mathbb{R}^k$ is a location parameter in the distribution of $\underline{X} = (X_1,\ldots,X_k)$. Then the distribution of $\underline{X}^* = (X_1-X_i,\ldots,X_{i-1}-X_i,X_{i+1}-X_i,\ldots,X_k-X_i)$ depends on $\underline{\theta}$ only through the parameter $\underline{\theta}^* = (\theta_1-\theta_i,\ldots,\theta_{i-1}-\theta_i,\theta_{i+1}-\theta_i,\ldots,\theta_k-\theta_i)$ and the family of distributions of $\underline{X}^*$ has the SIP in terms of $\underline{\theta}^*$.

Proof. Let $F(\underline{x}-\underline{\theta})$ be the cdf of $\underline{X}$. Let $\underline{Y} = (Y_1,\ldots,Y_k)$ be a random vector with c.d.f. $F(\underline{y})$. Let G be the c.d.f. of $(Y_1-Y_i,\ldots,Y_{i-1}-Y_i,$ $Y_{i+1}-Y_i,\ldots,Y_k-Y_i)$ and $G_{\underline{\theta}}$ the c.d.f. of $\underline{X}^*$. Then $(Y_1+\theta_1,\ldots,Y_k+\theta_k)$ has the same distribution as $\underline{X}$ so, for any constants $c_1,c_2,\ldots,c_{k-1}$,

$$\begin{aligned} G_{\underline{\theta}}(c_1,\ldots,c_{k-1}) &= P_{\underline{\theta}}(X_1-X_i \le c_1,\ldots,X_k-X_i \le c_{k-1}) \\ &= P(Y_1+\theta_1-Y_i-\theta_i \le c_1,\ldots,Y_k+\theta_k-Y_i-\theta_i \le c_{k-1}) \\ &= G(c_1-(\theta_1-\theta_i),\ldots,c_{k-1}-(\theta_k-\theta_i)). \end{aligned}$$

So the distribution of $\underline{X}^*$ depends on $\underline{\theta}$ only through $\underline{\theta}^*$ and in fact $\underline{\theta}^*$ is a location parameter for $\underline{X}^*$. It is easily seen (see Lehmann (1955)) that any location parameter family has the SIP. ||

Lemma 2.2.4. Suppose $\underline{\theta} \in (0,\infty)^k$ is a scale parameter in the distribution of $\underline{X} = (X_1,\ldots,X_k)$, $X_i > 0$ for all i. Then the distribution of $\underline{X}^* = (X_1/X_i,\ldots,X_{i-1}/X_i,X_{i+1}/X_i,\ldots,X_k/X_i)$ depends on $\underline{\theta}$ on through the parameter $\underline{\theta}^* = (\theta_1/\theta_i,\ldots,\theta_{i-1}/\theta_i,\theta_{i+1}/\theta_i,\ldots,\theta_k/\theta_i)$ and the family of distributions of $\underline{X}^*$ has the SIP in terms of $\underline{\theta}^*$.

Proof. The proof is the same as that of Lemma 2.2.3 with differences replaced by quotients and "location" by "scale". ||

2.3. Minimaxity and Admissibility of Two Classical Rules

In this section it will be shown that the rules R_1 and R_2 are minimax and admissible with respect to M for the location and scale parameter problems, respectively, if consideration is restricted to those rules which are non-randomized, just and translation (respectively, scale) invariant. The discussion will be in terms of the location parameter problem with the understanding that the same discussion is true in the scale parameter problem if differences are replaced by quotients as in Section 2.2.

In Section 1.2, it is explained that a non-randomized selection rule is completely determined by k sets $A_1,\ldots,A_k$ where A_i is the set of observations for which population π_i is included in the selected subset. By Lemma 2.2.1, a non-randomized rule is just and translation invariant if and only if $\underline{x} \in A_i$ or $\underline{x} \in A_i^c$ (A^c denotes the complement of A) can be determined from only the differences $\{x_j - x_i : j \neq i\}$ and A_i is monotone in these differences. In determining a rule which is minimax with respect to M, the quantity to be minimized is

$$\sup_{\Theta} M(\underline{\theta},\varphi) = \sup_{\Theta} \max_{\{i:\underline{\theta}\notin\Theta_i\}} P_{\underline{\theta}}(\text{select } \pi_i|\varphi)$$

$$= \max_{1 \le i \le k} \sup_{\Theta_i^c} P_{\underline{\theta}}(\text{select } \pi_i | \varphi)$$

$$= \max_{1 \le i \le k} \sup_{\Theta_i^c} P_{\underline{\theta}}(A_i).$$

This can be minimized by minimizing each of the terms $\sup_{\Theta_i^c} P_{\underline{\theta}}(A_i)$ separately with the only restriction being $\bigcup_{i=1}^{k} A_i = \mathcal{X}$ so that at least one population is always selected. Finally, the P*-condition (1.1.2) for a non-randomized rule is equivalent to

$$\inf_{\Theta_i} P_{\underline{\theta}}(A_i) \ge P^* \quad i = 1,2,\ldots,k. \tag{2.3.1}$$

Thus the following lemma has been proven.

<u>Lemma 2.3.1</u>. Let $A_1,\ldots,A_k$ be sets which satisfy (i) - (iv). Then the non-randomized selection rule defined by $A_1,\ldots,A_k$ is minimax with respect to M in the class of selection rules which satisfy the P*-condition and are non-randomized, just and translation invariant.

(i) $\bigcup_{i=1}^{k} A_i = \mathcal{X}$

(ii) $\inf_{\Theta_i} P_{\underline{\theta}}(A_i) \ge P^*$

(iii) A_i is a function only of the differences $\{x_j - x_i\colon j=1,\ldots,k,\ j \ne i\}$ and is monotone in these differences.

(iv) $\sup_{\Theta_i^c} P_{\underline{\theta}}(A_i) = \inf_{\mathcal{A}_i} \sup_{\Theta_i^c} P_{\underline{\theta}}(A)$

where $\mathcal{A}_i = \{A \subset \mathcal{X}\colon$ A satisfies (ii) and (iii) for the subscript i$\}$.

It should be pointed out that $\mathcal{G}_i$ is the set of all just, translation invariant selection regions for π_i which satisfy the P*-condition.

The form of a region, S, which, subject to (ii) and (iii), satisfies (iv) is given by Theorem 2.3.1 which is an extension of Lehmann's (1952) Theorem 4.1, page 545.

Theorem 2.3.1. Let the joint distribution of $(Y_1,\ldots,Y_k)$ be $F_{\underline{\gamma}}(y_1,\ldots,y_k)$ where the parameter space is the finite or infinite open rectangle $\underline{\gamma}_i < \gamma_i < \bar{\gamma}_i$ and the sample space is the finite or infinite open rectangle $\underline{y}_i < y_i < \bar{y}_i$, independent of the $\underline{\gamma}$. Suppose $P_{\underline{\gamma}}(S)$ is a continuous function of $\underline{\gamma}$ for any monotone set S. Suppose the family $\{F_{\underline{\gamma}}\}$ has the SIP, that the marginal distribution of Y_i depends only on γ_i and that Y_i converges in probability to $\underline{y}_i$ as $\gamma_i \to \underline{\gamma}_i$. Let $\underline{\gamma}^* = (\gamma_1^*,\ldots,\gamma_k^*)$ be a fixed parameter point and define

$$\Gamma = \{\underline{\gamma}: \gamma_i \leq \gamma_i^*,\ i = 1,\ldots,k\}. \tag{2.3.2}$$

Let $\mathcal{S}$ be the collection of all monotone sets which satisfy

$$\inf_{\Gamma} P_{\underline{\gamma}}(S) \geq P^*. \tag{2.3.3}$$

Then a region $S^* \in \mathcal{S}$ which satisfies

$$\sup_{\Gamma^c} P_{\underline{\gamma}}(S^*) = \inf_{\mathcal{S}} \sup_{\Gamma^c} P_{\underline{\gamma}}(S) \tag{2.3.4}$$

is given by

$$S^* = \{\underline{y}: y_i \leq a_i,\ i = 1,\ldots,k\}, \tag{2.3.5}$$

where the constants a_i are determined by

$$(2.3.6) \qquad P_{\gamma^*}(S^*) = P^*$$

and

$$(2.3.7) \qquad P_{\gamma_1^*}(Y_1 \le a_1) = P_{\gamma_2^*}(Y_2 \le a_2) = \ldots = P_{\gamma_k^*}(Y_k \le a_k).$$

Furthermore, if for every i, the distribution of Y_i given γ_i^* has the entire interval $(\underline{y}_i, \bar{y}_i)$ as its support, the region S* is the essentially unique element of $\mathcal{S}$ which is minimax, i.e., satisfies (2.3.4).

Proof. For any set of constants $y_j > \underline{y}_j$ and any i = 1,2,...,k

$$(2.3.8) \qquad \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j \ne i}} P_{\gamma}(Y_1 \le y_1, \ldots, Y_k \le y_k) = P_{\gamma_i}(Y_i \le y_i)$$

because

$$P(Y_1 \le y_1, \ldots, Y_k \le y_k) = P(Y_i \le y_i) - P(Y_i \le y_i, Y_j > y_j \text{ for at least one } j \ne i)$$

$$\ge P(Y_i \le y_i) - \sum_{j \ne i} P(Y_j > y_j)$$

and every term $P(Y_j > y_j)$ converges to zero in the limit of (2.3.8) because of the convergence in probability. The $\le$ inequality is immediate.

For any $S \in \mathcal{S}$, the SIP implies that $\lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j \ne i}} P_{\gamma}(S)$ exists and the limit will be denoted by $\beta_i(S|\gamma_i)$. The SIP and continuity of $P_{\gamma}(S)$ also imply that

$$(2.3.9) \qquad \sup_{\Gamma^c} P_{\gamma}(S) = \max_{1 \le i \le k} \beta_i(S|\gamma_i^*).$$

Since for the region S^* given by (2.3.5), (2.3.7) and (2.3.8) imply that $\beta_1(S^*|\gamma_1^*) = \beta_2(S^*|\gamma_2^*) = \ldots = \beta_k(S^*|\gamma_k^*)$, if the theorem were false, an $S \in \mathcal{S}$ could be found which simultaneously decreases all k quantities. But this can not happen. For let $S \in \mathcal{S}$. Let $\underline{y} \in S \cap S^{*c}$. (Such a $\underline{y}$ exists unless S is essentially the same as S^* because of (2.3.3) and (2.3.6).) For some $i = 1,2,\ldots,k$, $y_i > a_i$ since $\underline{y} \in S^{*c}$

$$P(S^* \cap S^c) \leq P(\bigcup_{j\neq i} \{Y_i \leq a_i, Y_j > y_j\}) \tag{2.3.10}$$

$$\leq \sum_{j\neq i} P(Y_i \leq a_i, Y_j > y_j)$$

$$\leq \sum_{j\neq i} P(Y_j > y_j).$$

As $\gamma_j \to \underline{\gamma}_j$, all the terms $P_{\gamma_j}(Y_j > y_j) \to 0$.

$$\beta_i(S|\gamma_i^*) = \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j\neq i}} P(S|\gamma_1,\ldots,\gamma_i^*,\ldots,\gamma_k)$$

$$= \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j\neq i}} P(S^*) + \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j\neq i}} P(S \cap S^{*c}) - \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j\neq i}} P(S^* \cap S^c) \tag{2.3.11}$$

$$= \beta_i(S^*|\gamma_i^*) + \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j\neq i}} P(S \cap S^{*c}) - \lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j\neq i}} P(S^* \cap S^c).$$

From (2.3.10) the last limit is zero, so $\beta_i(S|\gamma_i^*) \geq \beta_i(S^*|\gamma_i^*)$ and the first part of the theorem is proven.

Furthermore,

$$P(S \cap S^{*c}) \geq P(Y_1 \leq y_1,\ldots,Y_i \leq y_i,\ldots,Y_k \leq y_k)$$

$$- P(Y_1 \leq y_1,\ldots,Y_i \leq a_i,\ldots,Y_k \leq y_k).$$

As $\gamma_j \to \underline{\gamma}_j$, $j \neq i$, by (2.3.8) the right hand side converges to

$$P_{\gamma_i^*}(Y_i \leq y_i) - P_{\gamma_i^*}(Y_i \leq a_i). \tag{2.3.12}$$

So if the support of the distribution of Y_i given γ_i^* is the entire interval $(\underline{y}_i, \bar{y}_i)$, (2.3.12) is greater than zero and by (2.3.11), $\beta_i(S|\gamma_i^*) > \beta_i(S^*|\gamma_i^*)$. Hence by (2.3.9)

$$\sup_{\Gamma^c} P_{\underline{\gamma}}(S) > \sup_{\Gamma^c} P_{\underline{\gamma}}(S^*) \tag{2.3.13}$$

and S^* is the essentially unique element of $\mathcal{S}$ which is minimax.||

The form of the region S^* in Theorem 2.3.1 becomes particularly simple if the joint distribution of $(Y_1,\ldots,Y_k)$ is symmetric (i.e., the random variables are exchangeable) given $\underline{\gamma}^*$. Then (2.3.7) implies $a_1 = \ldots = a_k = a$, where a is determined by (2.3.6) and the minimax region is

$$S^* = \{\underline{y}: \max_{1 \leq i \leq k} y_i \leq a\}. \tag{2.3.14}$$

Finally, to apply Theorem 2.3.1 to the selection problem, the following lemma will be used.

<u>Lemma 2.3.2</u>. Suppose the random variables $\underline{X} = (X_1,\ldots,X_k)$ are exchangeable. Then the k-1 random variables $X_1 - X_i, \ldots, X_{i-1} - X_i, X_{i+1} - X_i, \ldots, X_k - X_i$ are exchangeable.

<u>Proof</u>. Let $c_1,\ldots,c_{k-1}$ be any fixed constants. Let
$A = \{\underline{x}: x_1 - x_i \leq c_1, \ldots, x_{i-1} - x_i \leq c_{i-1}, x_{i+1} - x_i \leq c_i, \ldots, x_k - x_i \leq c_{k-1}\}$.

Any permutation, σ, of the k-1 differences corresponds to a permutation, σ', of $X_1,\ldots,X_k$ which leaves X_i fixed. So

$$P(X_1-X_i \le c_1,\ldots,X_k-X_i \le c_{k-1}) = P(\underline{X} \in A)$$

(2.3.15) $$= P((X_{\sigma'^{-1}(1)},\ldots,X_i,\ldots,X_{\sigma'^{-1}(k)}) \in A)$$

$$= P(X_{\sigma^{-1}(1)}-X_i \le c_1,\ldots,X_{\sigma^{-1}(k)}-X_i \le c_{k-1}). \quad ||$$

Theorem 2.3.2. Let $\underline{X} = (X_1,\ldots,X_k)$ have a density $f(\underline{x}-\underline{\theta})$, $\underline{\theta} \in \mathbb{R}^k$, with respect to Lebesgue measure μ on $\mathbb{R}^k$. Suppose f is continuous a.e. μ, the support of f is $\mathbb{R}^k$, and f is symmetric (i.e., the random variables are exchangeable if $\theta_1 = \ldots = \theta_k$). Suppose the best population is the one associated with the largest parameter. Then R_1 is minimax with respect to M in the class of non-randomized, just and translation invariant rules which satisfy the P*-condition. Furthermore R_1 is the unique minimax rule in this class so R_1 is admissible in this class.

Proof. Let

$$A_i = \{\underline{x}: x_i \ge \max_{1\le j\le k} x_j - d\} \qquad i = 1,\ldots,k.$$

The sets $A_1,\ldots,A_k$ define the selection rule R_1. (i) - (iii) of Lemma 2.3.1 are obviously satisfied and (iv) must be verified.

Fix i = 1,2,...,k. To apply Theorem 2.3.1, let $Y_1 = X_1-X_i,\ldots,Y_{k-1} = X_k-X_i$ (omitting X_i-X_i) and $\gamma_1 = \theta_1-\theta_i,\ldots,\gamma_{k-1} = \theta_k-\theta_i$ (omitting $\theta_i-\theta_i$). Here the sample space and parameter space are $\mathbb{R}^{k-1}$. By Theorem 1.3.3 and Example 1.3.3, $P_{\underline{\theta}}(A)$ is a continuous function of

$\underline{\theta}$ for any measurable A, thus for any set monotone in $\underline{Y}$, $P_{\underline{\gamma}}(S)$ is a continuous function of $\underline{\gamma}$. Lemma 2.2.3 establishes the SIP of $\{F_{\underline{\gamma}}(\underline{y})\colon \underline{\gamma} \in \mathbb{R}^{k-1}\}$. Since $\underline{\theta}$ is a location parameter, the marginal distribution of Y_j depends only on γ_j and in fact γ_j is a location parameter in this distribution so the convergence in probability assumption of Theorem 2.3.1 is true.

Let $\underline{\gamma}^* = (0,\ldots,0)$ so that the set Γ in Theorem 2.3.1 is equivalent to

$$\overline{\Theta_i} = \{\underline{\theta}\colon \theta_j \leq \theta_i,\ j = 1,\ldots,k,\ j \neq i\}.$$

Because of the continuity of $P_{\underline{\theta}}(A)$ and $P_{\underline{\gamma}}(S)$ in terms of $\underline{\theta}$ and $\underline{\gamma}$, the fact that Γ is $\overline{\Theta_i}$ rather than Θ_i is unimportant since the sup's and inf's are all the same taken over a set or its closure. (2.3.3) simply ensures the P* condition on Θ_i.

Because f is symmetric, by Lemma 2.3.2, the distribution of $\underline{Y}$ given $\underline{\gamma}^*$ is also symmetric so the remark following Theorem 2.3.1 is relevant and an A_i satisfying (iv) of Lemma 2.3.1 is the S* of Theorem 2.3.1 given by

$$S^* = \{\underline{y}\colon y_j \leq d, \qquad j = 1,2,\ldots,k-1\}$$

$$= \{\underline{x}\colon x_j - x_i \leq d,\ j \neq i\} = A_i.$$

Since the support of f is $\mathbb{R}^k$, the support of the distribution of Y_j given γ_j^* is $\mathbb{R}$. So the uniqueness follows. Any unique minimax rule is admissible. ||

As previously mentioned, an analogous result is true in the scale parameter problem for the rule R_2. For completeness this is stated as

Theorem 2.3.3. Let $\underline{X} = (X_1,\ldots,X_k)$ have a density $f(x_1/\theta_1,\ldots,x_k/\theta_k)/\theta_1\cdot\theta_2\cdot\ldots\cdot\theta_k$, $\underline{\theta} \in (0,\infty)^k$, with respect to Lebesgue measure μ on $\mathbb{R}^k$. Suppose f is continuous a.e. μ, the support of f is $(0,\infty)^k$, and f is symmetric (i.e., the random variables are exchangeable if $\theta_1 = \ldots = \theta_k$). Suppose the best population is the one associated with the largest parameter. Then R_2 is minimax with respect to M in the class of non-randomized, just and scale invariant rules which satisfy the P*-condition. Furthermore, R_2 is the unique minimax rule in the class so R_2 is admissible in this class.

If the best population is the one associated with the smallest parameter, similar results may be obtained. The rule R_3:

$$\text{select } \pi_i \text{ if } x_i \leq \min_{1\leq j\leq k} x_j + d$$

is admissible in the location parameter problem. The rule R_4:

$$\text{select } \pi_i \text{ if } x_i \leq c\cdot\min_{1\leq j\leq k} x_j$$

is admissible in the scale parameter problem.

It is interesting to compare the results obtained in this chapter, for risk measured in terms of the maximum probability of accepting any non-best population, with those obtained in Chapter 1, for risk measured in terms of expected subset size. In the expected subset size problems, the least favorable parameter configuration was that in which all parameters were equal. For the risk considered in this chapter the least favorable configuration is in the limit as two parameters remain fixed and equal and all of the remaining converge to $-\infty$ (or zero in the scale case). So both rules R_1 and R_2 seem to do well at these two very different parameter configurations. One could hope that this indicates that these rules are good throughout the parameter space.

The restriction to non-randomized rules made in this section can not, in general, be dropped. In fact, using the notation as in the proof of Theorem 2.3.2, it is true that

$$
\begin{aligned}
\sup_{\Theta} M(\underline{\theta}, R_1) &= P_{Y_1^*}(Y_1 \leq a) \\
&> P_{\underline{Y}^*}(Y_1 \leq a, \ldots, Y_{k-1} \leq a) \\
&= P^* \\
&= \sup_{\Theta} M(\underline{\theta}, \varphi^*)
\end{aligned}
\tag{2.3.16}
$$

where $\varphi^* \equiv P^*$ is the "no data rule". So R_1 is not minimax if randomized rules are allowed.

2.4. Seal's Class and the Probability of Accepting any Non-best Population.

In the previous section, the rule R_1 was shown to be minimax and admissible with respect to M. In this section, the behavior of the other rules in the class $\mathscr{L}$(Definition 1.6.1) is briefly investigated.

Theorem 1.6.1 showed that for any rule R in $\mathscr{G}$, excluding R_1, and in a wide variety of location parameter problems, there exists a sequence of parameter points $<\underline{\theta}_n>$ such that $\underline{\theta}_n \notin \Theta_{k-1}$ for any n but $P_{\underline{\theta}_n}(\text{select } \pi_{k-1}|R) \to 1$ as $n \to \infty$. So

$$\sup_{\Theta} M(\underline{\theta},R) \geq \lim_{n\to\infty} P_{\underline{\theta}_n}(\text{select } \pi_{k-1}|R) = 1 \tag{2.4.1}$$

which is the worst possible upper bound.

In Theorem 1.6.3, it was necessary to place conditions on P* and k in order to assert that the rules in $\mathscr{G}$ were not minimax with respect to S or S'. But that is not necessary for M. For, using the notation as in the proof of Theorem 2.3.2, it is true that for any R in $\mathscr{G}$, excluding R_1,

$$\begin{aligned} \sup_{\Theta} M(\underline{\theta},R_1) &= P_{\gamma_1^*}(Y_1 \leq a) \\ &< 1 = \sup_{\Theta} M(\underline{\theta},R). \end{aligned} \tag{2.4.2}$$

The fact is that the risk M is aimed exactly at controlling this type of behavior. As explained in Section 2.1, this risk attempts to keep the probability of accepting any non-best population small. When considering expected subset size, it may not be too serious if the probability of selecting one non-best population is large if all of the other such probabilities are small. In this case, the sum remains small. But, for M, this situation causes the risk to be large.

2.5. An "Optimal" Rule for the Slippage Configuration in the Normal Means Problem.

Seal (1955) proposed the following problem involving the minimization of the probability of accepting a non-best population.

Suppose $\pi_1,\ldots,\pi_k$ are independent normal populations with common known variances σ^2. And suppose the unknown mean vector was in the slippage configuration, i.e., $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$. The problem as proposed by Seal was to find that rule in the class $\mathscr{C}$ (Definition 1.6.1) which

i) maximizes the probability of retaining in the selected group the population with the unequal mean if this mean is larger than the common mean of the other k-1 populations; and

ii) maximizes the probability of not retaining the population with the unequal mean if this mean is smaller than the common mean of the other k-1 populations.

Seal showed "approximately" (Seal's term) that the above goal was achieved by the rule R_5:

(2.5.1) $\quad$ select π_i if $x_i \geq \frac{1}{k-1} \sum_{j\neq i} x_j - d.$

In this section, this result is proven explicitly and the result is extended to a wider class of rules than $\mathscr{C}$.

Lemma 2.5.1. Among all rules which satisfy

(2.5.2) $\quad P_{\underline{\theta}_0}(\text{select } \pi_i|R) = P^*$, for all $\underline{\theta}_0 = (\theta_0,\ldots,\theta_0)$,

R_5 maximizes $P_{\underline{\theta}}(\text{select } \pi_i | R)$ for $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$, $\theta \in \mathbb{R}$, $\delta > 0$, where the unequal mean is the ith.

Proof. Fix $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$. Let $\underline{\theta}^* = (\theta^*,\ldots,\theta^*)$ where $\theta^* = \theta+\delta/k$. By the Neyman-Pearson Lemma (page 83 of Lehmann (1959)), the individual selection probability φ_i which maximizes

$$P_{\underline{\theta}}(\text{select } \pi_i) = E_{\underline{\theta}}\ \varphi_i(\underline{X})$$

subject to

$$P^* = P_{\underline{\theta}^*}(\text{select } \pi_i) = E_{\underline{\theta}^*}\ \varphi_i(\underline{X})$$

is given by

$$(2.5.3) \quad \varphi_i(\underline{x}) = \begin{cases} 1 & \text{if } \dfrac{f_{\underline{\theta}}(\underline{x})}{f_{\underline{\theta}^*}(\underline{x})} \geq c \\ 0 & \qquad\qquad < \end{cases}$$

where $f_{\underline{\theta}}(\underline{x}) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma}(x_i-\theta_i)^2}$. Some algebra yields

$$(2.5.4) \quad \varphi_i(\underline{x}) = \begin{cases} 1 & \text{if } x_i \geq \dfrac{1}{k-1} \sum_{j\neq i} x_j - \dfrac{k\sigma^2 \ln c}{(k-1)\delta} + \dfrac{\delta}{2} \\ 0 & \end{cases}$$

where c is chosen to satisfy $P_{\underline{\theta}^*}(\text{select } \pi_i) = P^*$. (2.5.4) is the rule R_5 and R_5 satisfies (2.5.2) so R_5 is the maximizing rule. ||

Note that in the proof of Lemma 2.5.1, if any other $\underline{\theta}^*$ were used, a rule which did not satisfy (2.5.2) would be obtained in

(2.5.3). Also note that the equality in (2.5.2) can not be weakened to a $\geq$. The inequality in the Neyman-Pearson Lemma gives $\leq$. Obviously the rule $\varphi_i \equiv 1$ satisfies $P_{\underline{\theta}_0}(\text{select } \pi_i) = 1 \geq P^*$ while $P_{\underline{\theta}}(\text{select } \pi_i) = 1$ is maximized.

Lemma 2.5.2. Among all rules which satisfy

$$(2.5.5) \qquad P_{\underline{\theta}_0}(\text{select } \pi_i) \geq P^*, \quad \text{for all } \underline{\theta}_0 = (\theta_0,\ldots,\theta_0),$$

R_5 minimizes $P_{\underline{\theta}}(\text{select } \pi_i)$ for $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$, $\theta \in \mathbb{R}$, $\delta < 0$, where the unequal mean is the ith.

Proof. Using the Neyman-Pearson Lemma as in Lemma 2.5.1,

$$\varphi_i(\underline{x}) = \begin{cases} 1 & \text{if } \dfrac{f_{\underline{\theta}}(\underline{x})}{f_{\underline{\theta}^*}(\underline{x})} \leq c \\ 0 & \qquad\qquad > \end{cases}$$

is obtained in place of (2.5.3) since we are now minimizing rather than maximizing. But, since $\delta < 0$, the inequality gets reversed and in the end (2.5.4) is obtained as before. R_5 satisfies (2.5.5) so R_5 is the minimizing rule. Also, this time the inequality in the Neyman-Pearson Lemma is in the right direction so the inequality in (2.5.5) is permissible. ||

Theorem 2.5.1. Among all rules which satisfy the P*-condition, R_5 minimizes $P_{\underline{\theta}}(\text{select } \pi_i)$ for $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$, $\theta \in \mathbb{R}$, $\delta < 0$, where the unequal mean is the ith.

Proof. This is immediate from Lemma 2.5.2 since any rule satisfying the P*-condition satisfies (2.5.5). ||

Theorem 2.5.2. Let $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$ where the unequal mean is the ith. Among all just and translation invariant rules which satisfy

$$\text{(2.5.6)} \qquad \inf_{\Theta_i} P_{\underline{\theta}}(CS) = P^*,$$

R_5 minimizes $P_{\underline{\theta}}(\text{select } \pi_i)$ if $\delta<0$ and maximizes $P_{\underline{\theta}}(\text{select } \pi_i)$ if $\delta > 0$.

Proof. Let φ be any just, translation invariant rule which satisfies (2.5.6). By Lemma 2.2.1, $\varphi_i(\underline{x})$ is a function of only $\{x_j-x_i,\ j=1,\ldots,k,\ j \neq i\}$ and is non-increasing in these differences. Since $\underline{\theta}$ is a location parameter, by Lemma 2.2.3, the distribution of the differences depends only upon and has the SIP in terms of the parameter $\underline{\theta}^* = (\theta_1-\theta_i,\ldots,\theta_k-\theta_i)$. Lehmann (1955) has shown that this implies $E_{\underline{\theta}}\varphi_i(\underline{X})$ is non-increasing in terms of the differences $\theta_j-\theta_i$, $j \neq i$. Let

$$\Gamma = \{\underline{\theta}\colon \theta_j-\theta_i \leq 0,\ j = 1,\ldots,k,\ j \neq i\} = \overline{\Theta_i}.$$

Fix $\underline{\theta}_0 = (\theta_0,\ldots,\theta_0)$. Then for any $\underline{\theta} \in \Gamma$, $\theta_j-\theta_i \leq \theta_0-\theta_0 = 0$, for all j, so

$$\begin{aligned} P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi) &= E_{\underline{\theta}_0}\varphi_i(\underline{X}) \\ &\leq E_{\underline{\theta}}\varphi_i(\underline{X}) \\ &= P_{\underline{\theta}}(\text{select } \pi_i|\varphi). \end{aligned}$$

Hence

$$P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi) \leq \inf_{\Theta_i} P_{\underline{\theta}}(\text{select } \pi_i|\varphi)$$

$$= \inf_{\Theta_i} P_{\underline{\theta}}(CS|\varphi) = P^*.$$

But the continuity (in $\underline{\theta}$) of $P_{\underline{\theta}}(\text{select } \pi_i|\varphi)$ implies

$$P_{\underline{\theta}_0}(\text{select } \pi_i|\varphi) \geq \inf_{\Theta_i} P_{\underline{\theta}}(\text{select } \pi_i|\varphi) = P^*.$$

So every just, translation invariant rule which satisfies (2.5.6) satisfies (2.5.2) and (2.5.5). Thus the result follows from Lemmas 2.5.1 and 2.5.2. ||

Seal's result is a special case of Theorem 2.5.2 since every rule in $\mathscr{L}$ is just and translation invariant and satisfies (2.5.6) for all $i = 1,2,\ldots,k$.

As stated, the class of rules considered in Theorem 2.5.2 includes non-symmetric rules. But as in Remark 1.5.2, if a rule R is symmetric, then for any $\underline{\theta}_0 = (\theta_0,\ldots,\theta_0)$,

$$P_{\underline{\theta}_0}(\text{select } \pi_1|R) = \ldots = P_{\underline{\theta}_0}(\text{select } \pi_k|R).$$

So if $\inf_{\Theta} P_{\underline{\theta}}(CS|R) = P^*$, where R is a just, translation invariant, symmetric rule, then

$$\inf_{\Theta_j} P_{\underline{\theta}}(CS|R) = \inf_{\Theta_j} P_{\underline{\theta}}(\text{select } \pi_j|R)$$

$$= P_{\underline{\theta}_0}(\text{select } \pi_j|R)$$

$$= P^*$$

for all j. So the following is proven.

Corollary 2.5.1. Let $\underline{\theta} = (\theta,\ldots,\theta,\theta+\delta,\theta,\ldots,\theta)$ where the unequal mean is the ith. Among all just, translation invariant, and symmetric rules which satisfy

$$\inf_{\Theta} P_{\underline{\theta}}(CS) = P^*,$$

(note equality), R_5 minimizes $P_{\underline{\theta}}(\text{select } \pi_i)$ if $\delta < 0$ and maximizes $P_{\underline{\theta}}(\text{select } \pi_i)$ if $\delta > 0$.

The results of this section put a very favorable light on the rule R_5 if the parameter is in a slippage configuration. But it should be remembered that in Theorem 1.6.2, R_5 is a worst possible case in that $r = 1$ so

$$\sup_{\Theta} E_{\underline{\theta}}(S|R_5) \geq k-1$$

and

$$\sup_{\Theta} E_{\underline{\theta}}(S'|R_5) \geq k-2.$$

Also, in Section 2.4 it was seen that

$$\sup_{\Theta} M(\underline{\theta},R_5) = 1$$

for M defined by (2.1.1), again the worst possible situation. So if the experimenter has no prior knowledge of true parameter configuration, R_5 should perhaps be avoided, considering its very poor behavior for some parameter values.

CHAPTER III

ROBUSTNESS OF BAYES RULES IN MULTIPLE DECISION PROBLEMS

3.1 Introduction

The basic formulation of a decision problem usually includes the specification of the parametric form of the distribution of the observations. If a Bayesian formulation is employed, a probability distribution on the parameter space, the prior, is then specified. But, although the experimenter may have an idea about the form of the distribution of the observations, exact specification of the parametric form may be difficult or impossible. The ϵ-contaminated model is common in studies of this problem (see, e.g., Andrews et al. (1972)). In this model the form of the distribution is specified only with probability $1-\epsilon$, the probability being ϵ that the distribution is something totally different and unspecified. For example, Huber (1965) found robust hypothesis tests in the ϵ-contaminated model.

Considering the uncertainty inherent in an ϵ-contaminated model, it seems unreasonable that the experimenter could then specify an exact prior distribution on these ϵ-contaminated distributions. On the other hand, it may be the case that the experimenter can restrict the prior to be in some sub-class of all prior distributions and it would seem desirable to use this partial prior information in the decision problem. Robbins (1964) has suggested that attention should be given to this case when the prior is restricted to be in a sub-class of all priors. Blum and Rosenblatt (1967) proposed the Γ-minimax criterion for selection of decision rules in the presence of this partial prior information. The Γ-minimax criterion requires the use of a decision rule which minimizes the maximum of the Bayes risk over the sub-class. The Γ-minimax criterion has been studied in a variety of problems by Jackson et al. (1970), Randles and Hollander (1971), Solomon (1972a, 1972b), De Rouen and Mitchell (1974), and Gupta and Huang (1975, 1977). Bayesian criticism of the Γ-minimax criterion has been offered by Watson (1974). None of these authors, however, dealt with the ϵ-contaminated model.

The main result of this chapter, found in Section 3.3, is that, in a finite parameter space multiple decision (i.e., finite action space) problem, the usual Bayes rule, ignoring any contamination, is robust in that, for small ϵ, it is Γ-minimax when the sub-class of priors is a class of priors on the family of ϵ-contaminations. In this sense, the Bayes rule is robust against inaccurately specified distributions of the observation (and, hence, inaccurately specified priors). Section 3.2 includes some basic results on Γ-minimaxity which are used in Section 3.3. Section 3.4 gives bounds on ϵ^*, the amount of contamination which can be present with the Bayes rule remaining Γ-minimax. Section 3.5 relates this work to the special case of hypothesis testing studied by Huber (1965).

3.2 Formulation and Γ-minimaxity

The elements of a decision problem will be denoted in the following manner. $\mathcal{X} \subset \mathbb{R}^n$ is the sample space of the random vector X. $\mathfrak{F} = \{F: F \in \mathfrak{F}\}$ is a set of distributions on $\mathcal{X}$. $\mathcal{G} = \{a: a \in \mathcal{G}\}$ is the action space. $L(F,a): \mathfrak{F} \times \mathcal{G} \to [0,\infty)$ is the loss function. A decision rule $\delta(a|x)$, is for each $x \in \mathcal{X}$, a probability measure on $\mathcal{G}$. The set of all decision rules is $\mathfrak{D}$. $R(F,\delta)$ will denote the risk of a decision rule δ at the point F. A probability measure γ on $\mathfrak{F}$ is called a prior and the Bayes risk of a decision rule δ with respect to a prior γ is denoted by $B(\gamma,\delta)$. $\Gamma = \{\gamma: \gamma \in \Gamma\}$ will denote a set of priors on $\mathfrak{F}$. The σ-fields associated with the various sets will usually not be of importance with the exception that the σ-field associated with $\mathfrak{F}$ must contain all of the single points {F} so that priors which put all their mass on a finite number of distributions are valid.

The following definition is due to Blum and Rosenblatt (1967).

Definition 3.2.1. A decision rule δ^* is called a Γ-minimax decision rule if

$$\sup_{\Gamma} B(\gamma,\delta^*) = \inf_{\mathfrak{D}} \sup_{\Gamma} B(\gamma,\delta).$$

If Γ consists of one prior, a Γ-minimax decision rule is Bayes with respect to that prior. If Γ consists of all priors on $\mathfrak{F}$, a Γ-minimax decision rule is minimax in the usual sense. The concept of Γ-minimaxity is useful in those cases when the prior distribution can be only partially specified in that it is known a priori that the prior is in the set Γ but no more specific information is available.

The following definition of a "least favorable" prior is useful in the Γ-minimax situation. This is not the same definition as used by Jackson et al. (1970).

<u>Definition 3.2.2.</u> A prior $\gamma^* \in \Gamma$ is called <u>least favorable</u> if, for some Γ-minimax rule δ^*,

$$B(\gamma^*,\delta^*) = \sup_{\Gamma} B(\gamma,\delta^*).$$

Many authors have found Γ-minimax rules by finding Bayes rules versus "least favorable" priors. Randles and Hollander (1971) and Gupta and Huang (1975, 1977) used the following result although it was never stated in this generality. Because of their different definition of "least favorable", Jackson et al. (1970) and De Rouen and Mitchell (1974) were required to verify a stronger condition, namely, equality in (3.2.1). The similarity between this result and a standard result on minimaxity (Theorem 1, page 90, Ferguson (1967)) is interesting.

<u>Theorem 3.2.1</u>. If a decision rule δ^* is Bayes with respect to a prior $\gamma^* \in \Gamma$ and, for all $\gamma \in \Gamma$,

$$B(\gamma,\delta^*) \leq B(\gamma^*,\delta^*) \tag{3.2.1}$$

then δ^* is Γ-minimax and γ^* is least favorable.

<u>Proof</u>. The following inequalities show that δ^* is Γ-minimax. Then (3.2.1) shows γ^* is least favorable.

$$\sup_{\Gamma} B(\gamma,\delta^*) = B(\gamma^*,\delta^*)$$

$$= \inf_{\mathcal{D}} B(\gamma^*,\delta)$$

$$\leq \inf_{\mathcal{D}} \sup_{\Gamma} B(\gamma,\delta)$$

$$\leq \sup_{\Gamma} B(\gamma,\delta^*). \quad ||$$

Corollary 3.2.1 is of interest for two reasons. It deals with the specific type of structure which will be used in Section 3.3. But, also, it elucidates a method of finding Γ-minimax rules which has been used by Gupta and Huang (1975, 1977) but whose relationship with Bayes rules and least favorable priors had not been explained.

<u>Corollary 3.2.1</u>. Let $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1 \cup \ldots \cup \mathcal{F}_k$ where the unions are disjoint. Suppose

(3.2.2) $$L(F_0,a) = 0 \quad \text{for all } F_0 \in \mathcal{F}_0 \text{ and all } a \in \mathcal{A}.$$

Let $\pi_0,\pi_1,\ldots,\pi_k$ be non-negative constants with $\pi_0 +\ldots+ \pi_k = 1$. Suppose there exist $F_i^* \in \mathcal{F}_i$, $i = 0,1,\ldots,k$, such that the Bayes rule δ^* for the prior $\gamma^*(\{F_i^*\}) = \pi_i$, $i = 0,1,\ldots,k$, has the property that

(3.2.3) $$R(F_i,\delta^*) \leq R(F_i^*,\delta^*) \quad \text{for all } F_i \in \mathcal{F}_i,\ i = 1,2,\ldots,k.$$

Let

(3.2.4) $$\Gamma = \{\gamma\colon \gamma(\mathcal{F}_0) \geq \pi_0,\ \gamma(\mathcal{F}_i) \leq \pi_i,\ i = 1,2,\ldots,k,\ \sum_{i=0}^{k} \gamma(\mathcal{F}_i)=1\}.$$

Then δ^* is Γ-minimax and γ^* is least favorable.

Remark 3.2.1. $\mathfrak{F}_0$ has been called the "indifference zone" in some problems. If the true distribution lies in $\mathfrak{F}_0$, any action results in zero loss by (3.2.2). $\mathfrak{F}_0 = \phi$ is allowed in which case π_0 and the inequalities in (3.2.4) are replaced by equalities. Also, the choice of $F_0^* \in \mathfrak{F}_0$ is inconsequential. Any such F_0 will suffice.

Proof. (3.2.2) implies $R(F_0,\delta^*) = 0$ for all $F_0 \in \mathfrak{F}_0$. So, for any $\gamma \in \Gamma$,

$$\begin{aligned} B(\gamma,\delta^*) &= \sum_{i=1}^{k} \int_{\mathfrak{F}_i} R(F_i,\delta^*)d\gamma(F_i) \\ &\leq \sum_{i=1}^{k} \pi_i \; R(F_i^*,\delta^*) \\ &= B(\gamma^*,\delta^*). \end{aligned}$$

Theorem 3.2.1 yields the result. ||

3.3 Robustness of Bayes Rules

In this section, problems which have finite action spaces (i.e., multiple decision problems) and finite parameter spaces are considered. The case when the parameter space and the action space both have two elements (i.e., hypothesis testing) was considered by Huber (1965). This work is inspired by problem (iii) in Huber (1965). Conditions are given under which the Bayes rule δ^* is robust in that it is Γ-minimax when each of the original distributions is replaced by a family of ϵ-contaminated versions of itself and the original prior is replaced by a class of priors on these

ϵ-contaminations. So the Bayes rule is robust in that it retains an optimality property, Γ-minimaxity, even if the prior distribution and parameter space were not originally specified exactly correctly.

The method of proof is to construct a least favorable prior against which δ^* is Bayes. The proof is constructive so that, for a given problem, a least favorable prior can be exhibited and bounds for ϵ^*, the amount of contamination allowable, can be computed.

Let $F_1,\ldots,F_k$ be k unique cumulative distribution functions (c.d.f.) on $\mathcal{X}$. They all have densities with respect to a measure μ which is also absolutely continuous with respect to $F_1 +\ldots+ F_k$ (e.g., $\mu = F_1 +\ldots+ F_k$). The densities will be denoted by $f_1(x),\ldots,f_k(x)$. Let $\mathcal{A} = \{a_1,\ldots,a_r\}$ be the finite action space. For brevity the loss will be denoted by $L(F_i,a_j) = L(i,a_j)$. Let $\underline{\pi} = (\pi_1,\ldots,\pi_k)$ denote a prior on the parameter space $\{F_1,\ldots,F_k\}$ where $\pi_i \geq 0$, $\Sigma\pi_i = 1$ and the prior probability of F_i is π_i.

In this problem, the Bayes risk for any decision rule δ is given by

$$(3.3.1) \qquad B(\underline{\pi},\delta) = \int_{\mathcal{X}} \sum_{j=1}^{r} \delta(a_j|x) \sum_{i=1}^{k} \pi_i \, L(i,a_j) f_i(x) d\mu(x).$$

(3.3.1) is minimized and, hence, the Bayes rule is given by

$$(3.3.2) \quad \delta^*(a_j|x) = \begin{cases} 1 & \text{if } \sum_i \pi_i L(i,a_j) f_i(x) < \min_{m\neq j} \sum_i \pi_i L(i,a_m) f_i(x) \\ \alpha_i(x) & \qquad\qquad\qquad\qquad = \\ 0 & \qquad\qquad\qquad\qquad > \end{cases}$$

where $\sum_i \alpha_i(x) = 1$ for all x. By an appropriate choice of the $\alpha_i(x)$,

δ^* can obviously be chosen to be a non-randomized decision rule, i.e., $\delta(a_j|x) \in \{0,1\}$ for all $a_j \in \mathcal{G}$ and all $x \in \mathcal{X}$. It shall be assumed that δ^* is non-randomized. So now δ^* can be considered a function from $\mathcal{X}$ into $\mathcal{G}$ and the simplified notation, $\delta^*(x) = a_j$ if $\delta(a_j|x) = 1$ shall be used, henceforth. Let

$$(3.3.3) \qquad D = \{x: \sum_i \pi_i L(i,\delta^*(x))f_i(x) < \min_{\mathcal{G}\setminus\{\delta^*(x)\}} \sum_i \pi_i L(i,a)f_i(x)\}$$

be the set of observations where the Bayes decision is unique.

The ϵ-contaminated neighborhoods of the distributions F_i are defined as follows. Let $0 \leq \epsilon \leq 1$.

$$\mathfrak{F}_{i\epsilon} = \{G(x): G(x) = (1-\epsilon)F_i(x)+\epsilon H(x), \text{ H any c.d.f. on } \mathcal{X}\}.$$

It can easily be shown that

$$\mathfrak{F}_{i\epsilon} = \{G(x): G(x) = (1-\epsilon')F_i(x) + \epsilon' H(x), \text{ H any c.d.f. on } \mathcal{X}, \; 0 \leq \epsilon' \leq \epsilon\}.$$

So $\mathfrak{F}_{i\epsilon}$ consists of all distributions which are less than or equal to ϵ contaminations of F_i. Since the F_i's are all distinct, for all sufficiently small positive ϵ's, the $\mathfrak{F}_{i\epsilon}$'s are all disjoint. Let ϵ_0 be a positive constant such that $\epsilon_0 < \epsilon_{00}$ where

$$(3.3.4) \qquad \epsilon_{00} = \sup\{\epsilon > 0: \text{ all } \mathfrak{F}_{i\epsilon} \text{ are disjoint}\}.$$

Finally a class of priors on these ϵ-contaminations is defined by

$$\Gamma_\epsilon = \{\gamma: \gamma(\mathfrak{F}_{i\epsilon}) = \pi_i, \quad i = 1,2,\ldots,k\}.$$

Now it is assumed that the loss for this new set of distributions is given by

$$L(F_i',a_j) = L(i,a_j) \quad \text{for all } F_i' \in \mathfrak{F}_{i\epsilon} \text{ and all } a_j \in \mathcal{G}$$

where $L(i,a_j)$ is the loss in the original problem. That is the loss is the same for any ϵ-contaminated version of F_i as for F_i itself.

The robustness result can now be stated.

Theorem 3.3.1. Let δ^* be the Bayes rule defined by (3.3.2). Let

(3.3.5) $$A_i = \{x: L(i,\delta^*(x)) = \sup_{\mathcal{X}} L(i,\delta^*(x))\} \cap D, \; i = 1,2,\ldots,k,$$

where D is defined in (3.3.3). Suppose there exist disjoint sets $B_i \subset A_i$ such that $\mu(B_i) > 0$, $i = 1,\ldots,k$. Then there exists $\epsilon^* > 0$ such that δ^* is Γ_{ϵ^*}-minimax. Furthermore, a least favorable prior, which puts mass π_i on

(3.3.6) $$G_i(x) = (1-\epsilon^*)F_i(x) + \epsilon^* H_i(x)$$

where H_i is a linear combination of the F_j's with support B_i, exists.

Remark 3.3.1. Since $\mathcal{G}$ is finite, $L(i,\delta^*(x))$ takes on only a finite number of values as a function of x. So the sup in (3.3.5) is attained for some x.

Proof. k distributions of the form (3.3.6) will be constructed such that, if they are used as the F_i^*'s in Corollary 3.2.1, δ^* is Bayes against γ^* and (3.2.3) is satisfied for all $F_i' \in \mathfrak{F}_{i\epsilon^*}$. Then Corollary 3.2.1 will yield the result.

First, k densities, $h_1, h_2, \ldots, h_k$, which have as their supports $B_1, \ldots, B_k$, respectively, and k positive constants, $\epsilon_1, \epsilon_2, \ldots, \epsilon_k$, are defined.

If $\pi_i L(i, \delta^*(x)) = 0$ on A_i, set $\epsilon_i = 1$ and let

$$h_i(x) = (f_1(x) + \ldots + f_k(x))/(F_1 + \ldots + F_k)(B_i) \tag{3.3.7}$$

on B_i and zero elsewhere. This is obviously a density.

If $\pi_i L(i, \delta^*(x)) > 0$ on A_i, let

$$h_i(x) = \frac{1-\epsilon_i}{\epsilon_i \pi_i L(i, \delta^*(x))} \min_{G \setminus \{\delta^*(x)\}} \sum_{j=1}^{k} \pi_j f_j(x)[L(j,a) - L(j, \delta^*(x))] \tag{3.3.8}$$

on B_i and zero elsewhere. Since $B_i \subset D$ and $\pi_i L(i, \delta^*(x)) > 0$ on B_i, the integral of the right hand side over the set B_i without the $(1-\epsilon_i)/\epsilon_i$ term is a finite positive constant. Since $(1-\epsilon_i)/\epsilon_i$ varies between 0 and ∞ as ϵ_i varies between 1 and 0, ϵ_i can be chosen (and will be positive) so that $h_i(x)$ is a density (i.e., integrates to one).

Now let

$$\epsilon^* = \min_{0 \le i \le k} \{\epsilon_i\} \tag{3.3.9}$$

with ϵ_0 defined in (3.3.4). The claim is that the k distributions, G_i, with densities $g_i(x) = (1-\epsilon^*) f_i(x) + \epsilon^* h_i(x)$, can be used as the least favorable set. They certainly satisfy the restrictions on H_i stated after (3.3.6).

To see that δ^* is Bayes with respect to γ^*, it must be verified that for each x, the inequality in (3.3.2), viz.,

$$(3.3.10) \qquad \sum_{i=1}^{k} \pi_i L(i,\delta^*(x))g_i(x) \le \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_{i=1}^{k} \pi_i L(i,a)g_i(x)$$

holds. For $x \notin \bigcup_i B_i$, $g_i(x) = (1-\epsilon^*)f_i(x)$ for all i so in (3.3.10) the $(1-\epsilon^*)$ cancels from both sides and the inequality reduces to that of (3.3.2). Thus the same decision, $\delta^*(x)$, is Bayes. For $x \in B_m$,

$$(3.3.11) \quad \sum_i \pi_i L(i,\delta^*(x))g_i(x) = (1-\epsilon^*)\sum_i \pi_i L(i,\delta^*(x))f_i(x) + \epsilon^*\pi_m L(m,\delta^*(x))h_m(x).$$

If $\pi_m L(m,\delta^*(x)) = 0$ on A_m, (3.3.11) equals

$$(1-\epsilon^*)\sum_i \pi_i L(i,\delta^*(x))f_i(x) \le (1-\epsilon^*) \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_i \pi_i L(i,a)f_i(x)$$

$$\le \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_i \pi_i L(i,a)g_i(x).$$

The first inequality is true because δ^* is Bayes in the original problem and the second is true because $(1-\epsilon^*)f_i(x) \le g_i(x)$ for all $x \in \mathcal{X}$. If $\pi_m L(m,\delta^*(x)) > 0$ on A_m, (3.3.11) equals

$$(1-\epsilon^*)\sum_i \pi_i L(i,\delta^*(x))f_i(x) + \epsilon^*\left(\frac{1-\epsilon_m}{\epsilon_m}\right) \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_i \pi_i f_i(x)[L(i,a) - L(i,\delta^*(x))]$$

$$\le (1-\epsilon^*)\sum_i \pi_i L(i,\delta^*(x))f_i(x) + (1-\epsilon^*) \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_i \pi_i f_i(x)[L(i,a) - L(i,\delta^*(x))]$$

$$= (1-\epsilon^*) \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_i \pi_i f_i(x)L(i,a)$$

$$\le \min_{\mathcal{a}\setminus\{\delta^*(x)\}} \sum_i \pi_i g_i(x)L(i,a).$$

The first inequality is true since, by (3.3.9), $\epsilon^* \leq \epsilon_m$. Thus δ^* is Bayes with respect to γ^*.

Finally, inequality (3.2.3) must be verified. Let $F_i' \in \mathcal{F}_{i\epsilon^*}$. So $F_i' = (1-\epsilon^*)F_i + \epsilon^* H'$.

$$
\begin{aligned}
R(F_i',\delta^*) &= \int L(i,\delta^*(x))dF_i'(x) \\
&= (1-\epsilon^*)\int L(i,\delta^*(x))dF_i(x) + \epsilon^*\int L(i,\delta^*(x))dH'(x) \\
&\leq (1-\epsilon^*)\int L(i,\delta^*(x))dF_i(x) + \epsilon^* \sup_{\mathcal{X}} L(i,\delta^*(x)) \\
&= R(G_i,\delta^*).
\end{aligned}
$$

The last equality is true because H_i puts all its mass on $B_i \subset A_i$, the set where $L(i,\delta^*(x)) = \sup_{\mathcal{X}} L(i,\delta^*(x))$. So (3.2.3) is verified.||

As is obvious from the proof of Theorem 3.3.1, all that was important about the H_i's was that i) H_i had had its support in A_i and ii) the mass was put on A_i in such a way that δ^* was Bayes with respect to γ^*. Thus, there are many least favorable priors. This particular one was chosen to illustrate that the least favorable distributions need not be particularly pathological nor have particularly "heavy tails". But, rather, the least favorable distributions may be simply linear combinations of the other distributions in the original problem. The explicit construction of Theorem 3.3.1 will be used in Section 3.4 to obtain bounds for ϵ^*, i.e., bounds on the amount of contamination which can be present with δ^* still remaining Γ_{ϵ^*}-minimax.

3.4 Bounds for ϵ^*

In Section 3.3, it was found that the Bayes rule δ^* is Γ_{ϵ^*}-minimax for some $\epsilon^* > 0$. In this section bounds are obtained for ϵ^*. Equation (3.3.9) showed that ϵ^* could be chosen to equal $\min_{0\le i\le k}\{\epsilon_i\}$. The bounds obtained in this section are bounds on ϵ_0 and bounds on $\min_{1\le i\le k}\{\epsilon_i\}$. The bounds are all sharp in that they are attained for certain problems.

First, we discuss bounds for ϵ_0, or more precisely, for

$$\epsilon_{00} = \sup\{\epsilon > 0: \ \mathfrak{F}_{i\epsilon} \text{ are all disjoint}\}.$$

Lemma 3.4.1: Let F_1 and F_2 be two distinct c.d.f.'s on $\mathcal{X}$. Let $A = \sup_{\mathcal{X}}|F_1(x)-F_2(x)|$. If $\epsilon_0 < A/(1+A)$, then $\mathfrak{F}_{1\epsilon_0} \cap \mathfrak{F}_{2\epsilon_0}$ is empty.

Proof. Fix ϵ satisfying $0 < \epsilon < A - \epsilon_0(1+A)$. Let $x_0 \in \mathcal{X}$ be such that $|F_1(x_0)-F_2(x_0)| > A-\epsilon > 0$, say $F_1(x_0)-F_2(x_0) > A-\epsilon$. Then

$$\begin{aligned}(1-\epsilon_0)F_1(x_0)-[(1-\epsilon_0)F_2(x_0) + \epsilon_0] &> (1 - \tfrac{A}{1+A})[F_1(x_0)-F_2(x_0)] - \epsilon_0 \\ &> (\tfrac{1}{1+A})(A-\epsilon) - \epsilon_0 \\ &> \tfrac{1}{1+A}(\epsilon_0(1+A)) - \epsilon_0 = 0.\end{aligned}$$

So, for any $G_1 \in \mathfrak{F}_{1\epsilon_0}$ and $G_2 \in \mathfrak{F}_{2\epsilon_0}$,

$$\begin{aligned}G_1(x_0) &= (1-\epsilon_0)F_1(x_0) + \epsilon_0 H_1(x_0) \\ &\ge (1-\epsilon_0)F_1(x_0) \\ &> (1-\epsilon_0)F_2(x_0) + \epsilon_0\end{aligned}$$

$$\geq (1-\epsilon_0)F_2(x_0) + \epsilon_0 H_2(x_0)$$
$$= G_2(x_0).$$

So no G can be in both $\mathfrak{F}_{1\epsilon_0}$ and $\mathfrak{F}_{2\epsilon_0}$ for such a G would have to satisfy $G(x_0) > G(x_0)$. ||

Theorem 3.4.1: Let $A' = \min_{\substack{1\leq i,j\leq k \\ i\neq j}} \sup_{\mathcal{X}} |F_i(x)-F_j(x)|$. Then

$$A'/(1+A') \leq \epsilon_{00} \leq \frac{1}{2}.$$

Proof. $G(x) = \frac{1}{2} F_1(x) + \frac{1}{2} F_2(x)$ is in both $\mathfrak{F}_{1\frac{1}{2}}$ and $\mathfrak{F}_{2\frac{1}{2}}$. So $\epsilon_{00} \leq \frac{1}{2}$.

Fix $i \neq j$. Since $A'/(1+A')$ is an increasing function of A', if $\epsilon_0 < A'/(1+A')$ then

$$\epsilon_0 < A'/(1+A') \leq A/(1+A)$$

where $A = \sup_{\mathcal{X}} |F_i(x)-F_j(x)|$. By Lemma 3.4.1, $\mathfrak{F}_{i\epsilon_0} \cap \mathfrak{F}_{j\epsilon_0} = \phi$. So $\epsilon_{00} \geq A'/(1+A')$. ||

The following two examples show that the bounds of Theorem 3.4.1 are attained in some problems.

Example 3.4.1: Suppose $k = 2$. $\mathcal{X} = \mathbb{R}$. Suppose the support of $F_1 \subset (-\infty,a]$ and the support of $F_2 \subset (a,\infty)$. Since $F_1(a) = 1$ and $F_2(a) = 0$, $\sup_{\mathcal{X}}|F_1(x)-F_2(x)| = 1$. By Theorem 3.4.1, since $A'/(1+A') = 1/2$, $\epsilon_{00} = \frac{1}{2}$. Thus in this example both the lower and upper bounds are attained. A more interesting example of attainment of the lower bound, when it is not 1/2, is in Example 3.4.2.

Example 3.4.2: Let $k = 2$. $\mathcal{X} = \mathbb{R}$. Suppose F_1 and F_2 satisfy i) $F_1(x) \geq F_2(x)$ for all $x \in \mathcal{X}$ and ii) there exists x_0 such that $F_1(x_0)-F_2(x_0) = A > 0$ and $F_1(x)-F_2(x)$ is non-decreasing on $(-\infty,x_0]$ and non-increasing on $[x_0,\infty)$. This will be true, for example if $F(\cdot)$ has a symmetric, unimodal density with respect to Lebesgue measure and $F_1(x) = F(x-\theta_1)$, $F_2(x) = F(x-\theta_2)$ for some $\theta_2 > \theta_1$. In this case, $x_0 = (\theta_2-\theta_1)/2$. For this problem, $\sup_{\mathcal{X}}|F_1(x)-F_2(x)| = A$. It will be shown that if $\epsilon_0 = A/(1+A)$ then $\mathfrak{F}_{1\epsilon_0} \cap \mathfrak{F}_{2\epsilon_0} \neq \phi$. Thus the lower bound of Theorem 3.4.1 will be attained. Let

$$H(x) = \begin{cases} (F_1(x)-F_2(x))/A & x \leq x_0 \\ 1 & x > x_0 \end{cases}.$$

By ii) H is non-negative and increases from 0 to 1 on $(-\infty,x_0]$, i.e. H is a c.d.f. The claim is that $G(x) = (1-\epsilon_0)F_2(x) + \epsilon_0 H(x)$ is in $\mathfrak{F}_{1\epsilon_0} \cap \mathfrak{F}_{2\epsilon_0}$. $G \in \mathfrak{F}_{2\epsilon_0}$ is obvious. Since $\epsilon_0/A = 1-\epsilon_0$, $G(x) = (1-\epsilon_0)F_1(x)$ on $(-\infty,x_0]$. And on (x_0,∞),

$$\begin{aligned} G(x) &= (1-\epsilon_0)F_2(x) + \epsilon_0 \\ &= (1-\epsilon_0)F_1(x) + (1-\epsilon_0)(F_2(x)-F_1(x)) + \epsilon_0 \\ &= (1-\epsilon_0)F_1(x) + \epsilon_0[1-(1-\epsilon_0)(F_1(x)-F_2(x))/\epsilon_0] \\ &= (1-\epsilon_0)F_1(x) + \epsilon_0[1-(F_1(x)-F_2(x))/A] \end{aligned}$$

The function

$$H'(x) = \begin{cases} 0 & x \leq x_0 \\ 1-(F_1(x)-F_2(x))/A & x > x_0 \end{cases}$$

is, because of ii), increasing from 0 to 1 on (x_0,∞), i.e., $H'(x)$ is a c.d.f. So, since $G(x) = (1-\epsilon_0)F_1(x) + \epsilon_0 H'(x)$, $G \in \mathfrak{F}_{1\epsilon_0}$.

Now bounds for $\min_{1 \le i \le k} \{\epsilon_i\}$ will be considered. The following notation will be used. Let

$$c_i = \pi_i \sup_{\mathcal{X}} L(i,\delta^*(x)).$$

We shall assume that $c_i > 0$, $i = 1,\ldots,k$. So (3.3.8), rather than (3.3.7), of Theorem 3.3.1 is being considered. It will be recalled that (3.3.7) was the trivial case and the corresponding ϵ_i was one. For any measurable set B, let

$$(3.4.1) \qquad I(B) = \int_B \min_{\mathcal{A} \setminus \{\delta^*(x)\}} \sum_{j=1}^{k} \pi_j f_j(x)[L(j,a)-L(j,\delta^*(x))]d\mu(x).$$

Recall that on the sets A_i of (3.3.5), this integrand is positive. It will also be assumed that the F_i are continuous so that for any A_i, it is possible to choose $B \subset A_i$ such that $I(B) = c$, where c is any number $0 \le c \le I(A_i)$.

We shall say that the sets B_i, $i = 1,\ldots,k$, "satisfy the inclusion conditions" if the B_i are all disjoint, $\mu(B_i) > 0$, and $B_i \subset A_i$, $i = 1,\ldots,k$. These were the conditions required of the B_i in Theorem 3.3.1. Setting the integral of (3.3.8) equal to one and solving for ϵ_i yields

$$(3.4.2) \qquad \epsilon_i(B_i) = I(B_i)/(c_i + I(B_i)).$$

So ϵ_i is a strictly increasing function of $I(B_i)$.

Finally, the interest is not in bounds on $\min_{1 \le i \le k} \{\epsilon_i\}$, per se. This can always be made small by choosing one of the B_i's small. We want to choose the B_i's so as to make this quantity as large as possible, since, large values of ϵ^* correspond to wide classes of priors with respect to which δ^* is Γ-minimax. Thus the quantity for which bounds are desired is

$$\epsilon' = \sup_{\mathcal{B}} \min_{1 \le i \le k} \{\epsilon_i(B_i)\}$$

where $\mathcal{B} = \{(B_1,\ldots,B_k)$: B_i satisfy the inclusion conditions$\}$.

Theorem 3.4.2.

$$\epsilon' \le I(\bigcup_i A_i)/(\sum_i c_i + I(\bigcup_i A_i)). \tag{3.4.3}$$

If there exist B_i, $i = 1,\ldots,k$, which satisfy the inclusion conditions and also satisfy

$$I(B_j) = c_j \, I(\bigcup_i A_i)/\sum_i c_i \qquad j = 1,\ldots,k, \tag{3.4.4}$$

then this bound is attained.

Proof. The values in (3.4.4) are the solution of the k equations

$$I(B_j)/c_j = I(B_k)/c_k \qquad j = 1,\ldots,k-1$$

$$\sum_i I(B_i) = I(\bigcup_i A_i). \tag{3.4.5}$$

Substituting (3.4.4) into (3.4.2) yields

(3.4.6) $\quad \epsilon_j(B_j) = I(\bigcup_i A_i)/(\sum_i c_i + I(\bigcup_i A_i)) \qquad j = 1,\ldots,k$

So for such a choice of B_j, ϵ_j does not depend on j and the $\min_{1\le j\le k} \{\epsilon_j\}$ is the value given in (3.4.3) proving the second assertion. Now let $B_1',\ldots,B_k'$ satisfy the inclusion conditions. Since the B_i''s are disjoint and satisfy $\bigcup_i B_i' \subset \bigcup_i A$,

$$\sum_i I(B_i') \le I(\bigcup_i A_i)$$

So by (3.4.5), at least one B_j' satisfies

$$I(B_j') \le c_j \; I(\bigcup_i A_i)/\sum_i c_i$$

By (3.4.6), this implies

$$\epsilon_j(B_j') \le I(\bigcup_i A_i)/(\sum_i c_i + I(\bigcup_i A_i))$$

and $\min_{1\le j\le k} \{\epsilon_j(B_j')\}$ is less than or equal to the same value. Since $B_1',\ldots,B_k'$ were arbitrary, (3.4.3) follows. ||

A lower bound for ϵ' is more complicated to write down. It involves how the A_i overlap in a particular problem. But one can be obtained which is computable and is attained in some problems. The following notation is used. Let

$$\mathcal{E} = \{E = C_1 \cap \ldots \cap C_k : C_i = A_i \text{ or } C_i = A_i^c \quad i = 1,\ldots,k, \text{ and } \mu(E) > 0\} \setminus \{A_1^c \cap \ldots \cap A_k^c\}.$$

There are at most 2^k-1 sets in $\mathcal{E}$, the sets in $\mathcal{E}$ are all disjoint and $\mathcal{E}$ is non-empty since $\bigcup_{E\in\mathcal{E}} E = (\bigcup_i A_i)\setminus N$ where $\mu(N) = 0$. For $E \in \mathcal{E}$, let

$$\varphi(E) = \{i_1,\ldots,i_m : E \subset A_{i_j},\ j = 1,\ldots,m\}.$$

$\varphi(E)$ has between one and k elements for every E.

<u>Theorem 3.4.2.</u> Let

$$A_j' = \sum_{\{E:\ j\in\varphi(E)\}} (I(E)/\sum_{m\in\varphi(E)} c_m) \quad j = 1,\ldots,k.$$

Then

$$\epsilon' \geq (\min_{1\leq j\leq k} A_j')/(1 + \min_{1\leq j\leq k} A_j'). \tag{3.4.7}$$

<u>Proof</u>. Let $\mathcal{E} = \{E_1,\ldots,E_n\}$. Partition each E_i into the number of elements in $\varphi(E_i)$ disjoint measurable subsets, $\{E_{ij}: j \in \varphi(E_i)\}$, satisfying

$$I(E_{ij}) = c_j\ I(E_i)/\sum_{m\in\varphi(E_i)} c_m.$$

Now let $B_j = \bigcup_{\{i:j\in\varphi(E_i)\}} E_{ij} \qquad j = 1,\ldots,k.$

The B_j satisfy the inclusion conditions since $j \in \varphi(E_i)$ implies $E_{ij} \subset E_i \subset A_j$ and all the E_{ij} are disjoint. Because of the disjointness,

$$I(B_j) = \sum_{\{i:j\in\varphi(E_i)\}} I(E_{ij})$$

$$= \sum_{\{i: j \in \varphi(E_i)\}} (c_j \, I(E_i) / \sum_{m \in \varphi(E_i)} c_m).$$

Using (3.4.2) yields

$$\epsilon_j(B_j) = A_j'/(1+A_j').$$

So

$$\min_{1 \le j \le k} \{\epsilon_j(B_j')\} = \min_{1 \le j \le k} A_j'/(1+A_j')$$

$$= \min_{1 \le j \le k} A_j'/(1 + \min_{1 \le j \le k} A_j').$$

Since this choice of $(B_1,\ldots,B_k) \subset \mathcal{B}$, (3.4.7) follows. ||

Example 3.4.3: An example in which the upper bound of Theorem 3.4.2 is attained is the following. Let $k = 3$. Let F_1, F_2, and F_3 be normal distributions with means -1, 0, 1 respectively and common variance 1. Let this be a classification problem with $\mathcal{G} = \{a_1, a_2, a_3\}$ where a_i corresponds to classifying the observation as coming from F_i. Suppose the prior is $\pi_1 = \pi_3 = .3$, $\pi_2 = .4$. Assume 0-1 loss so $L(i,a_j) = 0$ if $i = j$ and 1 if $i \neq j$. Then the Bayes rule is of the form

$$\text{(3.4.8)} \qquad \delta^*(x) = \begin{cases} a_1 & x < x_1 \\ a_2 & x_1 \le x \le x_2 \\ a_3 & x_2 < x \end{cases}$$

where for this prior and loss, $x_1 = -.80$ and $x_2 = .80$.
$A_1 = \{x: \delta^*(x) = a_2 \text{ or } a_3\} = [-.80,\infty)$, $A_2 = \{x: \delta^*(x) = a_1 \text{ or } a_3\} = (-\infty,-.80) \cup (.80,\infty)$ and $A_3 = \{x: \delta^*(x) = a_1 \text{ or } a_2\} = (-\infty,.80)$.

It can be computed, using normal tables, that $I(-\infty,-.80) = .0889 = I(.80,\infty)$ and $I(-.80,.80) = .0732$ so $I(\bigcup_i A_i) = .2510$. Thus the upper bound of (3.4.3) is $\epsilon' \leq .2510/1.2510 = .205$. An obvious way to define the B_i's is $B_1 = (0,c)$, $B_2 = (-\infty,-c) \cup (c,\infty)$, and $B_3 = (-c,0)$ where c is chosen so that (see (3.4.4)) $I(B_1) = I(B_3) = .3 \cdot I(\bigcup_i A_i)$ and $I(B_2) = .4 \cdot I(\bigcup_i A_i)$. Solving $I(0,c) = .3 \cdot I(\bigcup_i A_i)$ for c yields $c = 1.80$. Since $c > x_2$, $B_i \subset A_i$ for all i, so this B_i satisfy the inclusion conditions. The lower bound of Theorem 3.4.1 is $(F_2(.5) - F_3(.5))/(1+F_2(.5) - F_3(.5)) = .276 > .205$. Hence δ^* is $\Gamma_{.205}$-minimax. The fact that δ^* is Γ-minimax for up to 20% contamination seems to reflect favorably on δ^*. It is also interesting to note that if the lower bound of Theorem 3.4.3 is computed for this problem, $A_1' = A_3' = .249$ and $A_2' = .254$ so the lower bound is $.249/1.249 = .199$. Hence the range of possible values for ϵ' given by Theorems 3.4.2 and 3.4.3 is very small for this problem viz., $.199 \leq \epsilon' \leq .205$.

Example 3.4.4. An example in which the lower bound of Theorem 3.4.3 is attained is the following. Consider the same classification problem as in Example 3.4.3 with the only change being in the loss. Rather than 0-1 loss, assume the loss matrix is

$$(L(i,a_j)) = (L_{ij}) = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix} .$$

The Bayes rule has the same form, (3.4.8), where, for this loss and prior, it can be computed that $x_1 = -.95$ and $x_2 = 1.55$. For this loss $A_1 = A_2 = \{x: \delta^*(x) = a_3\} = (1.55,\infty)$ and $A_3 = \{x: \delta^*(x) = a_1\} = (-\infty,-.95)$. The set $\mathcal{E}$ contains only two sets, viz., $E_1 = A_1 \cap A_2 \cap A_3^c = A_1$ and $E_2 = A_1^c \cap A_2^c \cap A_3 = A_3$. Computations yield $I(A_1) = .0370$ and $I(A_3) = .0798$. The question is, what choice of B_1, B_2, and B_3 will maximize the $\min\{\epsilon_1,\epsilon_2,\epsilon_3\}$. Obviously, ϵ_3 is maximized by choosing $B_3 = A_3$ which, from (3.4.2), yields

$$\epsilon_3 = I(A_3)/(c_3 + I(A_3)) = .117. \tag{3.4.9}$$

The $\min\{\epsilon_1,\epsilon_2\}$ is maximized by choosing B_1 and B_2 so that $\epsilon_1 = \epsilon_2$. For any choice with inequality, say $\epsilon_1 > \epsilon_2$, the $\min\{\epsilon_1,\epsilon_2\}$ can be increased by making B_1 slightly smaller and B_2 correspondingly larger. Also, it is required that $B_1 \cup B_2 \subset A_1$ but the $\min\{\epsilon_1,\epsilon_2\}$ is maximized by choosing $B_1 \cup B_2 = A_1$ since larger B_i's yield larger ϵ_i's. Finally, from (3.4.2) it is seen that $\epsilon_1 = \epsilon_2$ if and only if $I(B_i)/c_1 = I(B_2)/c_2$ and, substituting $I(B_2) = I(A_1) - I(B_1)$, this yields $I(B_1) = c_1 \cdot I(A_i)/(c_1+c_2)$ and $I(B_2) = c_2 \cdot I(A_1)/(c_1+c_2)$. From (3.4.2) this best choice of B_1 and B_2 yields

$$\epsilon_1 = \epsilon_2 = I(A_1)/(c_1+c_2+I(A_1)) = .026. \tag{3.4.10}$$

But, in Theorem 3.4.3, $A_1' = A_2' = I(A_1)/(c_1+c_2)$ and $A_3' = I(A_3)/c_3$. So the theorem asserts that $\epsilon' \geq \min\{I(A_1)/(c_1+c_2+I(A_1)), I(A_3)/(c_3+I(A_3))\}$. However (3.4.9) and (3.4.10) show that in this problem $\epsilon' = \min(\epsilon_1,\epsilon_3)$ is equal to this lower bound. The lower bound for ϵ_0 given by Theorem 3.4.1 is again .276, as in Example 3.4.3, so finally it can be stated that δ^* is $\Gamma_{.026}$-minimax.

The upper bound of Theorem 3.4.2 can be computed to be .055. As in Example 3.4.3, the upper and lower bounds do not differ greatly.

3.5. Hypothesis Testing

In this section, the important special case of hypothesis testing is considered via two examples. In the first example it is found that for appropriately chosen B_i's, the least favorable distributions of Theorem 3.3.1 are the same as those found by Huber (1965). The second example exhibits the importance of considering ϵ_0 when determining ϵ^*.

Example 3.5.1. In a hypothesis testing problem, there are two distributions with densities $f_1(x)$ and $f_2(x)$ and $G = \{a_1, a_2\}$. The loss has the form $L(1,a_1) = L(2,a_2) = 0$ and $L(i,a_j) = L_i > 0$, $i \neq j$. A version of the Bayes rule is given by

$$\delta^*(x) = \begin{cases} a_1 & \text{if } f_2(x)/f_1(x) \leq \pi_1 L_1/\pi_2 L_2 \\ a_2 & \qquad\qquad\qquad\quad > \end{cases}.$$

The sets A_i of (3.3.5) are given by

$$A_1 = \{x: \delta^*(x) = a_2\} \cap D = \{x: f_2(x)/f_1(x) > \pi_1 L_1/\pi_2 L_2\}$$

and

$$A_2 = \{x: \delta^*(x) = a_1\} \cap D = \{x: f_2(x)/f_1(x) < \pi_1 L_1/\pi_2 L_2\}.$$

Since $A_1 \cap A_2 = \phi$, B_1 and B_2 can be any non-empty subsets of A_1 and A_2. But suppose they are chosen to be of the form

$$B_1 = \{x: f_2(x)/f_1(x) > c'' \geq \pi_1 L_1/\pi_2 L_2\}$$

$$B_2 = \{x: f_2(x)/f_1(x) < c' \leq \pi_1 L_1/\pi_2 L_2\}$$

where c" and c' are constants. Then, if c' and c" are such that $\epsilon_1(B_1) = \epsilon_2(B_2) = \epsilon^*$, the densities of the least favorable distributions of Theorem 3.3.1 are given by

$$g_1(x) = \begin{cases} (1-\epsilon^*)f_1(x) & \text{if } \gamma(x) \leq c'' \\ (1-\epsilon^*)\dfrac{\pi_2 L_2}{\pi_1 L_1} f_2(x) & \text{if } \gamma(x) > c'' \end{cases}$$

$$g_2(x) = \begin{cases} (1-\epsilon^*)f_2(x) & \text{if } \gamma(x) \geq c' \\ (1-\epsilon^*)\dfrac{\pi_1 L_1}{\pi_2 L_2} f_1(x) & \text{if } \gamma(x) < c' \end{cases}$$

where $\gamma(x) = f_2(x)/f_1(x)$. These least favorable distributions are of the same form as found by Huber (1965) and lead to a censored probability ratio $g_2(x)/g_1(x)$ of the form,

$$\frac{g_2(x)}{g_1(x)} = \begin{cases} d_0 & \text{if } \gamma(x) \geq c'' \\ \gamma(x) & \text{if } c'' > \gamma(x) > c' \\ d_1 & \text{if } c' \geq \gamma(x) \end{cases} \tag{3.5.1}$$

where d_0 and d_1 are constants. But, if c' and c" are such that $\epsilon_1(B_1) \neq \epsilon_2(B_2)$, the density ratio has the form

$$\frac{g_2(x)}{g_1(x)} = \begin{cases} (b_1+b_2/\gamma(x))^{-1} & \text{if } \gamma(x) \geq c'' \\ b_3 \cdot \gamma(x) & \text{if } c'' > \gamma(x) > c' \\ b_4+b_5 \cdot \gamma(x) & \text{if } c' \geq \gamma(x) \end{cases}$$

where $b_1,\ldots,b_5$ are constants.

Example 3.5.2. As a continuation of Example 3.5.1, a problem in which the censored probability ratio of (3.5.1) is obtained and which exhibits the importance of including ϵ_0 in (3.3.9) is the following. Let $\mathcal{X} = \mathbb{R}$ and $f_i(x) = f(x-\theta_i)$ for constants θ_1 and θ_2 where $f(x)$ is a density with respect to Lebesgue measure which is symmetric about zero. Suppose $\pi_1 L_1 = \pi_2 L_2$. Then for $c' \geq 1$, fixed, if

$$B_1 = \{x\colon \gamma(x) > c'\} \text{ and } B_2 = \{x\colon \gamma(x) < 1/c'\},$$

because of the symmetry, $P(B_1|F_1) = P(B_2|F_2)$ and $P(B_2|F_1) = P(B_1|F_2)$ are obtained. So, from (3.3.8), the equality

$$\begin{aligned} \epsilon_1/(1-\epsilon_1) &= P(B_1|F_2) - P(B_1|F_1) \\ &= P(B_2|F_1) - P(B_2|F_2) \\ &= \epsilon_2/(1-\epsilon_2) \end{aligned}$$

is obtained. Hence

$$\epsilon_1 = \epsilon_2 = (P(B_2|F_1) - P(B_2|F_2))/(1 + P(B_2|F_1) - P(B_2|F_2)).$$

Thus, if ϵ_0 is ignored and ϵ^* is set equal to ϵ_1, the censored probability ratio of (3.5.1) is obtained. But this does not ensure that $\mathfrak{F}_{1\epsilon^*} \cap \mathfrak{F}_{2\epsilon^*}$ is empty. Indeed, if $c' = 1$, the logical value to choose in order to maximize ϵ^*, the two densities g_1 and g_2 are equal, viz.,

$$g_1(x) = g_2(x) = \begin{cases} (1-\epsilon^*)f_1(x) & \text{if } \gamma(x) \leq 1 \\ (1-\epsilon^*)f_2(x) & \text{if } \gamma(x) > 1 \end{cases} .$$

Thus, $\mathfrak{F}_{1\epsilon^*} \cap \mathfrak{F}_{2\epsilon^*}$ is not empty. ϵ_0 in (3.3.9) can not, in general, be ignored.

BIBLIOGRAPHY

1. Andrews, D. F., et al. (1972). Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton, New Jersey.

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

3. Blum, J. R., and Rosenblatt, J. (1967). On partial a priori information in statistical inference. Ann. Math. Statist., 38, 1671-1678.

4. Chung, K. L. (1974). A Course in Probability Theory. Academic Press, New York.

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

6. De Rouen, T. A., and Mitchell, T. J. (1974). A G_1-minimax estimator for a linear combination of binomial probabilities. JASA, 69, 231-233.

7. Deverman, J. N. (1969). A general selection procedure relative to the t best populations. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

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

9. Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo. Series No. 150, Inst. of Statist., University of North Carolina, Chapel Hill, North Carolina.

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

11. Gupta, S. S., and Huang, D. Y. (1975). A note on Γ-minimax classification procedures. Proceedings of the 40th Session of the International Statistical Institute, Warsaw, Poland, 330-335.

12. Gupta, S. S., and Haung, D. Y. (1977). On some Γ-minimax selection and multiple comparison procedures. Statistical Decision Theory and Related Topics II (ed. Gupta, S. S., and Moore, D. S.), 139-156, Academic Press, New York.

13. 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.

14. Gupta, S. S., and Nagel, K. (1971). On some contribution to multiple decision theory. Statistical Decision Theory and Related Topics (ed. Gupta, S. S., and Yackel, J.), 79-102, Academic Press, New York.

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

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

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

18. Gupta, S. S., and Studden, W. J. (1966). Some aspects of selection and ranking procedures with applications. Mimeo Series No. 81, Department of Statistics, Purdue University, West Lafayette, Indiana.

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

20. Huber, P. J. (1965). A robust version of the probability ratio test. Ann. Math. Statist., 36, 1753-1758.

21. Huber, P. J. (1972). Robust statistics: A review. Ann. Math. Statist., 43, 1041-1067.

22. Jackson, D. A., O'Donovan, T. M., Zimmer, W. J., and Deely, J. J. (1970). G_2-minimax estimators in the exponential family. Biometrika, 57, 439-443.

23. Lehmann, E. L. (1952). Testing multiparameter hypotheses. Ann. Math. Statist., 23, 541-552.

24. Lehmann, E. L. (1955). Ordered families of distributions. Ann. Math. Statist., 26, 399-419.

25. Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, New York.

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

27. Leong, Y. K. (1976). Some results on subset selection procedures. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

28. Mosteller, F. (1948). A k-sample slippage test for an extreme population. Ann. Math. Statist., 19, 58-65.

29. Nagel, K. (1970). On subset selection rules with certain optimality properties. Ph.D. Thesis, Department of Statistics, Purdue University, West Lafayette, Indiana.

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

31. Randles, R. H., and Hollander, M. (1971). Γ-minimax selection procedures in treatment versus control problems. Ann. Math. Statist., 42, 330-341.

32. Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist., 35, 1-20.

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

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

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

36. Solomon, D. L. (1972a). Λ-minimax estimation of a multivariate location parameter. JASA, 67, 641-646.

37. Solomon, D. L. (1972b). Λ-minimax estimation of a scale parameter. JASA, 67, 647-649.

38. Studden, W. J. (1967). On selecting a subset of k populations containing the best. Ann. Math. Statist., 38, 1072-1078.

39. Watson, S. R. (1974). On Bayesian inference with incompletely specified prior distributions. Biometrika, 61, 193-196.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Mimeograph Series #489
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Minimax, Admissible and Gamma-minimax Multiple Decision Rules
5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER: Mimeo. Series #489
7. AUTHOR(s): Roger L. Berger
8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Purdue University, West Lafayette, IN 47907
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-243
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC
12. REPORT DATE: May 1977
13. NUMBER OF PAGES: 87
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)

Subset selection rules, minimaxity, expected subset size, probability of accepting non-best populations, admissibility, Γ-minimaxity robustness of Bayes multiple decision rules, ε-contamination.

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

Multiple decision problems are decision theory problems in which the action space has a finite number of elements. Two different types of multiple decision problems are considered, herein. These two types of problems are subset selection problems and robustness of Bayes rules problems.

Subset selection problems arise because the classical tests of homogeneity are often inadequate in practical situations where the experimenter

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

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) has to make decisions regarding k(> 2) populations, treatments, or processes. This inadequacy may be alleviated by formulating the problems as multiple decision problems aimed at ranking or selection of the k populations. A formulation was proposed in which the population of interest is selected with a fixed minimum probability P* over the entire parameter space. This formulation is called the subset selection formulation.

Chapter I of this paper considers minimaxity of subset selection rules when the risk is the expected size of the selected subset or the expected number of non-best populations selected. The minimax value of the selection problem is computed for a wide class of distributions. It is found that two classical rules are minimax in location and scale parameter problems if the populations are independent and the distributions have monotone likelihood ratio. Necessary conditions for minimaxity are obtained and are used to show that other proposed rules are not minimax. The minimaxity of rules in a proposed class of rules is also investigated. Chaper II of this paper consideres minimaxity and admissibility of selection rules when the risk is the maximum probability of selecting any non-best population. It is found that, if the restriciton is made to non-randomized, just, and translation invariant rules, a classical rule is minimax and admissible in the location parameter problem. The analagous result holds in the scale parameter problem for scale invariant rules. Other rules in a proposed class are found to behave poorly with respect to this risk. But one of these rules is found to have a certain optimality property if the parameters are in a slippage configuration.

A different type of multiple decision problem is considered in Chapter III. The robustness of Bayes rules when the parameter space is finite is considered. The usual Bayes rules are found to be robust in that they are Γ-minimax when the original distributions are replaced by ∈-contaminated versions of themselves. To derive this result, some general results involving the relationship of Bayes rules and Γ-minimax rules are obtained. Bounds are obtained on the amount of contamination which may be present with the Bayes rule remaining Γ-minimax.