GAMMA MINIMAX ROBUSTNESS OF BAYES RULES

by

Roger L. Berger

FSU Statistics Report M463

April, 1978
Department of Statistics
The Florida State University
Tallahassee, Florida 32306
Bayes rules are shown to be robust in multiple decision problems in the sense that they retain an optimality property, \( \Gamma \)-minimaxity, when the original distributions are replaced by families of \( \varepsilon \)-contaminated versions of themselves and the prior is replaced by a family of priors on these \( \varepsilon \)-contaminations. Thus, the Bayes rule is robust against inaccurately specified parameter spaces and, hence, inaccurately specified priors. Bounds are obtained on the amount of contamination which can be present with the Bayes rule remaining \( \Gamma \)-minimax.

Key Words: \( \varepsilon \)-contamination, least favorable prior.
1. INTRODUCTION

The basic formulation of a decision problem usually includes a 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 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 $\epsilon$-contaminated model is useful 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 $\epsilon$ that the distribution is something totally different and unspecified.

Considering the uncertainty inherent in the $\epsilon$-contaminated model, it seems unreasonable that the experimenter could then specify an exact prior on this class of $\epsilon$-contaminated distributions. On the other hand, it may be the case that the experimenter can restrict the prior to lie in some sub-class of all prior distributions. To use this partial prior information, Robbins (1964) has proposed the $\Gamma$-minimax criterion. The $\Gamma$-minimax criterion requires the use of a decision rule which minimizes the maximum of the Bayes risk over the sub-class of priors.

The main result of this paper, found in Section 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 $\epsilon$, it is $\Gamma$-minimax when the sub-class of priors is a class of priors on the family of $\epsilon$-contaminations. In this sense, the Bayes rule is robust against inaccurately specified distributions of the observation (and, hence, inaccurately specified priors). Section 2 contains some basic results on $\Gamma$-minimaxity which are used in Section 3. Section 4 contains bounds on $\epsilon^*$, the amount of contamination which can be present with the Bayes rule remaining $\Gamma$-minimax. Section 5 relates this work to the special case of hypothesis testing studied by Huber (1965).
2. NOTATION AND $\Gamma$-MINIMAXITY

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

This definition is due to Blum and Rosenblatt (1967).

**Definition 2.1.** A decision rule $\delta^*$ is called a $\Gamma$-minimax rule if

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

$\Gamma$-minimaxity provides a reasonable criterion for the choice of a decision rule for those problems in which it is known that the prior is in the set $\Gamma$ but no more specific information is available. The following definition of a "least favorable" prior is natural in the $\Gamma$-minimax situation.

**Definition 2.2.** A prior $\gamma^* \in \Gamma$ is called **least favorable** if, for some $\Gamma$-minimax rule $\delta^*$,

$$B(\gamma^*, \delta^*) = \sup_{\Gamma} B(\gamma, \delta^*).$$
Many authors have found \( \Gamma \)-minimax rules by finding Bayes rules versus least favorable priors. Randles and Hollander (1971) and Gupta and Huang (1975, 1977) used the following result which was proved by George (1969). Because of their different definition of least favorable, Jackson et al. (1970) and DeRouen and Mitchell (1974) were required to verify a stronger condition, namely, equality in (2.1). This result is similar to a standard result on minimaxity (Ferguson (1967), Theorem 1, page 90).

**Theorem 2.1.** If a decision rule \( \delta^* \) is Bayes with respect to a prior \( \gamma^* \in \Gamma \) and, for all \( \gamma \in \Gamma \),

\[(2.1) \quad B(\gamma^*, \delta^*) \geq B(\gamma, \delta^*),\]

then \( \delta^* \) is \( \Gamma \)-minimax and \( \gamma^* \) is least favorable.

Corollary 2.1 is interesting for two reasons. It deals with the specific type of structure which will be used in Section 3. It also elucidates a method of finding \( \Gamma \)-minimax rules which was used by Gupta and Huang (1975, 1977) but whose relationship with Bayes rules and least favorable priors was not explained.

**Corollary 2.1.** Let \( F = F_0 \cup F_1 \cup \ldots \cup F_k \) where the \( F_i \)'s are disjoint. Suppose

\[(2.2) \quad L(F_0, a) = 0 \quad \text{for all} \quad F_0 \in F_0 \quad \text{and all} \quad a \in A.\]

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

\[(2.3) \quad R(F_i, \delta^*) \leq R(F^*_i, \delta^*) \quad \text{for all} \quad F_i \in F_i, \quad 1 \leq i \leq k.\]
Let

\[(2.4) \quad \Gamma = \{ \gamma : \gamma(F_0) \geq \pi_0, \gamma(F_i) \leq \pi_i, 1 \leq i \leq k, \sum_{i=0}^{k} \gamma(F_i) = 1 \}.\]

Then \( \delta^* \) is \( \Gamma \)-minimax and \( \gamma^* \) is least favorable.

**Remark 2.1.** \( F_0 \) has been called the indifference zone in some problems. If the true distribution lies in \( F_0 \), any action results in zero loss by (2.2). \( F_0 = \phi \) is allowed in which case \( \pi_0 = 0 \) and the inequalities in (2.4) are replaced by equalities. It is clear that the choice of \( F_0^{*} \in F_0 \) is inconsequential. Any such \( F_0 \) will suffice.

**PROOF:** (2.2) implies \( R(F_0, \delta^*) = 0 \) for all \( F_0 \in F_0 \). So, for \( \gamma \in \Gamma \),

\[B(\gamma, \delta^*) = \sum_{i=1}^{k} \int_{F_i} R(F_i, \delta^*) d\gamma(F_i) \leq \sum_{i=1}^{k} \pi_i R(F_i^*, \delta^*) = B(\gamma^*, \delta^*).\]

Theorem 2.1 yields the result.||

---

3. \( \Gamma \)-MINIMAX ROBUSTNESS OF BAYES RULES

For the remainder of this paper, problems which have finite action spaces (i.e., multiple decision problems) and finite parameter spaces will be considered.

The case when the parameter space and the action space both have two elements (i.e., hypothesis testing) was considered by Huber (1965). The following results address problem (iii) in Huber (1965). Conditions are given under which a Bayes rule \( \delta^* \) is \( \Gamma \)-minimax when each of the original distributions is replaced by a
family of \( \varepsilon \)-contaminated versions of itself and the original prior is replaced by a class of priors on these \( \varepsilon \)-contaminations. In this sense, the Bayes rule is robust in that it retains an optimality property, \( \Gamma \)-minimaxity, even if the prior distribution and parameter space were not originally specified exactly correctly. Huber (1973) discusses the problem of inaccurate priors and distributions from another view.

Let \( F_1, \ldots, F_k \) be \( k \) unique probability measures on \( X \). They all have densities with respect to a measure \( \mu \) 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 \( A = \{a_1, \ldots, a_r\} \) be the finite action space. Let \( \pi = \{\pi_1, \ldots, \pi_k\} \) denote a prior on the parameter space \( \{F_1, \ldots, F_k\} \) where \( \pi_i \geq 0, \sum \pi_i = 1 \) and the prior probability of \( F_i \) is \( \pi_i \). Define

\[
S(x, a_j, f) = \sum_{i=1}^{k} \pi_i L(F_i, a_j)f_i(x).
\]

In this problem, the Bayes risk for any decision rule \( \delta \) is given by

\[
B(\pi, \delta) = \int_{\mathcal{X}} \sum_{j=1}^{r} \delta(a_j|x)S(x, a_j, f)\,d\mu(x).
\]

(3.2) is minimized and, hence, a Bayes rule is given by

\[
\delta^*(a_j|x) = \begin{cases} 
1 & \text{if } S(x, a_j, f) < \min_{m \neq j} S(x, a_m, f) \\
\alpha_j(x) & \text{otherwise}
\end{cases}
\]

(3.3)
where \( \alpha_j(x) = 0 \) if \( S(x, a_j, f) > \min_{1 \leq m \leq r} S(x, a_m, f) \) and \( \sum_j \alpha_j(x) = 1 \) for all \( x \).

By an appropriate choice of the \( \alpha_j(x) \), \( \delta^* \) can obviously be chosen to be a non-randomized decision rule, i.e., \( \delta(a_j|x) \in \{0, 1\} \) for all \( a_j \in A \) and all \( x \in X \). It shall be assumed that \( \delta^* \) is non-randomized. Now \( \delta^* \) can be considered a function from \( X \) into \( A \) and the simplified notation, \( \delta^*(x) = a_j \) if \( \delta^*(a_j|x) = 1 \), shall be used, henceforth.

The \( \varepsilon \)-contaminated neighborhoods of the distributions \( F_i \) are defined as follows. Let \( 0 \leq \varepsilon \leq 1 \).

\[
(3.4) \quad F_{i\varepsilon} = \{G: G = (1-\varepsilon)F_i + \varepsilon H, H \text{ any probability measure on } X\}.
\]

Since the \( F_i \)'s are all distinct, for all sufficiently small \( \varepsilon \)'s, the \( F_{i\varepsilon} \)'s are all disjoint. Let \( \varepsilon_0 \) be a positive constant such that \( \varepsilon_0 < \varepsilon_{00} \) where

\[
(3.5) \quad \varepsilon_{00} = \sup\{\varepsilon > 0: \text{ all } F_{i\varepsilon} \text{ are disjoint}\}.
\]

An exact expression for \( \varepsilon_{00} \) is given in Section 4. Finally, the class of priors on these \( \varepsilon \)-contaminations which assign probability \( \pi_i \) to \( F_{i\varepsilon} \) is defined by

\[
(3.6) \quad \Gamma_\varepsilon = \{\gamma: \gamma(F_{i\varepsilon}) = \pi_i, 1 \leq i \leq k\}.
\]

It is assumed that the loss for this expanded parameter space consisting of \( u F_{i\varepsilon} \) is given by

\[
(3.7) \quad L(F_i^*, a_j) = L(F_i, a_j) \text{ for all } F_i^* \in F_{i\varepsilon} \text{ and all } a_j \in A
\]

where \( L(F_i, a_j) \) is the loss in the original problem. That is the loss is the same for any \( \varepsilon \)-contaminated version of \( F_i \) as for \( F_i \) itself.
The $\Gamma$-minimax robustness result can now be stated.

Theorem 3.1. Let $\delta^*$ be the Bayes rule defined by (3.3).

Let

$$(3.8) \quad D = \{x: S(x, \delta^*(x), f) < \min_{A \setminus \{\delta^*(x)\}} S(x, a, f)\}$$

be the set of observations for which the Bayes decision is unique. Let

$$(3.9) \quad A_i = \{x: \ell(F_i, \delta^*(x)) = \sup_X \ell(F_i, \delta^*(x))\} \cap D, \ 1 \leq i \leq k.$$ 

Suppose there exist disjoint sets $B_i$ such that $B_i \subset A_i$ and $\mu(B_i) > 0,$ $1 \leq i \leq k$. Then there exists $\epsilon^* > 0$ such that $\delta^*$ is $\Gamma_{\epsilon^*}$-minimax. Furthermore, a least favorable prior which assigns probability $\pi_i$ to $G_i = (1 - \epsilon^*)F_i + \epsilon^*H_i$, where $H_i$ is a linear combination of the $F_j$'s with support $B_i$, exists.

Remark 3.1. Since $A$ is finite, $L(F_i, \delta^*(x))$ takes on only a finite number of values as a function of $x$. So the sup in (3.9) is attained for some $x$.

Remark 3.2. The hypothesis that $\mu(B_i) > 0$ implies that

$$\sup_X L(F_i, \delta^*(x)) = \sup_S L(F_i, \delta^*(x))$$

where $S$ is the support of $F_1 + \ldots + F_k$ (recall $\mu$ is absolutely continuous with respect to $F_1 + \ldots + F_k$). This puts no important restriction on $\delta^*$. Since the behavior of $\delta^*$ on $X \setminus S$ does not affect the fact that $\delta^*$ is Bayes, $\delta^*(x)$ for $x \in X \setminus S$ can always be chosen so that the above equality is true.
Proof. \( k \) distributions of the form specified will be constructed such that, if they are used as the \( F_i^* \)'s in Corollary 2.1, \( \delta^* \) is Bayes with respect to \( \gamma^* \) and (2.3) is satisfied for all \( F_i^* \in F_i^{\text{e}*} \). Then Corollary 2.1 will yield the result.

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

If \( \pi_1 L(F_i^*, \delta^*(x)) = 0 \) on \( A_i \), set \( \epsilon_i = 1 \) and let

\[
(3.10) \quad h_i(x) = (f_1(x) + \ldots + f_k(x))/(F_1 + \ldots + F_k)(B_i)
\]
on \( B_i \) and zero elsewhere. Note, the denominator is positive since \( \mu \) is absolutely continuous with respect to \( F_1 + \ldots + F_k \). So \( h_i \) is obviously a density.

If \( \pi_1 L(F_i^*, \delta^*(x)) > 0 \) on \( A_i \), let

\[
(3.11) \quad h_i(x) = \frac{1 - \epsilon_i}{\epsilon_i \pi_1 L(F_i^*, \delta^*(x))} \left( \min_{A \setminus \{\delta^*(x)\}} S(x, a, f) - S(x, \delta^*(x), f) \right)
\]
on \( B_i \) and zero elsewhere. Since \( B_i \subset D \) and \( \pi_1 L(F_i^*, \delta^*(x)) > 0 \) on \( B_i \), the integral (w.r.t. \( \mu \)) 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 zero and infinity as \( \epsilon_i \) varies between one and zero, \( \epsilon_i \) can be chosen (and will be positive) so that \( h_i \) is a density (i.e., integrates to one).

Let

\[
(3.12) \quad \epsilon^* = \min(\epsilon_i : 0 \leq i \leq k)
\]
with \( \epsilon_0 \) defined in (3.5). 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 are of the form specified in the theorem.

To see that \( \delta^* \) is Bayes with respect to \( \gamma^* \), the inequality in (3.3) must be verified for each \( x \), viz.,

\[
S(x, \delta^*(x), g) \leq \min_{A \setminus \{\delta^*(x)\}} S(x, a, g)
\]

holds. For \( x \notin \bigcup_i B_i \), \( g_i(x) = (1-\epsilon^*)f_i(x) \) for all \( i \) so in (3.13) the \((1-\epsilon^*)\) cancels from both sides and the inequality reduces to that of (3.3). Thus the same decision, \( \delta^*(x) \), is Bayes. For \( x \in B_m \),

\[
S(x, \delta^*(x), g) = (1-\epsilon^*)S(x, \delta^*(x), f) + \epsilon^*\pi_m L(F_m, \delta^*(x)) h_m(x).
\]

If \( \pi_m L(F_m, \delta^*(x)) = 0 \) on \( A_m \), (3.14) equals

\[
(1-\epsilon^*)S(x, \delta^*(x), f) < (1-\epsilon^*) \min_{A \setminus \{\delta^*(x)\}} S(x, a, f)
\]

\[
\leq \min_{A \setminus \{\delta^*(x)\}} S(x, a, g).
\]

The first inequality is true because \( x \notin B_m \subset D \) and the second is true because

\[
(1-\epsilon^*)f_i(x) \leq g_i(x) \text{ for all } x \in X. \text{ If } \pi_m L(F_m, \delta^*(x)) > 0 \text{ on } A_m, \text{ (3.14) equals}
\]

\[
(1-\epsilon^*)S(x, \delta^*(x), f) + \epsilon^* \left( \frac{1-\epsilon_m}{\epsilon_m} \right) \left( \min_{A \setminus \{\delta^*(x)\}} S(x, a, f) - S(x, \delta^*(x), f) \right)
\]

\[
\leq (1-\epsilon^*)S(x, \delta^*(x), f) + (1-\epsilon^*) \left( \min_{A \setminus \{\delta^*(x)\}} S(x, a, f) - S(x, \delta^*(x), f) \right)
\]

\[
= (1-\epsilon^*) \min_{A \setminus \{\delta^*(x)\}} S(x, a, f) \leq \min_{A \setminus \{\delta^*(x)\}} S(x, a, g).
\]
The first inequality is true since, by (3.12), $\varepsilon^* \leq \varepsilon_m$. So (3.13) is verified for all $x$ and $\delta^*$ is Bayes with respect to $\gamma^*$.

Finally, inequality (2.3) must be verified. Let $F'_1 \in F_{\epsilon^*}$, say

$$F'_1 = (1-\epsilon^*)F_1 + \epsilon^*H'.$$

$$R(F'_1, \delta^*) = \int L(F'_1, \delta^*(x))dF'_1(x)$$

$$= (1-\epsilon^*)\int L(F_1, \delta^*(x))dF_1(x) + \epsilon^*\int L(F_1, \delta^*(x))dH^*(x)$$

$$\leq (1-\epsilon^*)\int L(F_1, \delta^*(x))dF'_1(x) + \epsilon^*\sup_{\delta^*} L(F_1, \delta^*(x))$$

$$= R(G_1, \delta^*)$$

The last equality is true because $H_1$ gives probability one to $B_1 \subset A_1$, the set where $L(F_1, \delta^*(x)) = \sup_{\delta^*} L(F_1, \delta^*(x))$. So (2.3) is verified.||

It can be seen from the proof of Theorem 3.1 that the important features of the $H_1$'s are i) $H_1$ has its support on $A_1$ and ii) the probability is distributed on $A_1$ in such a way that $\delta^*$ is Bayes with respect to $\gamma^*$. Thus, there are many least favorable priors. This one illustrates that the least favorable distributions need not be particularly pathological. Rather, the least favorable distributions may be simply linear combinations of the other distributions in the original problem. The explicit construction of Theorem 3.1 will be used in Section 4 to obtain bounds on $\epsilon^*$, i.e., bounds on the amount of contamination which can be present with $\delta^*$ still remaining $\Gamma_{\epsilon^*}$-minimax.
4. BOUNDS FOR $\epsilon^*$

In Section 3, the Bayes rule $\delta^*$ was shown to be $\Gamma_{\epsilon^*}$-minimax for some $\epsilon^* > 0$. In this section, bounds for $\epsilon^*$ are obtained. Recall that (3.12) gives $\epsilon^* = \min \{\epsilon_i : 0 \leq i \leq k\}$. The bounds obtained in this section are bounds on $\epsilon_0$ and bounds on $\min \{\epsilon_i : 1 \leq i \leq k\}$. The bounds are all sharp in that they are attained for certain problems.

Berger (1977) has proved a theorem which gives the exact value of $\epsilon_{00} = \sup \\{\epsilon > 0 : \text{all } F_{i\epsilon} \text{ are disjoint}\}$. Theorem 2 of Berger (1977) states that $F_{i\epsilon}$ and $F_{j\epsilon}$ are disjoint if and only if

$$\epsilon < 1 - \left[ F_i (f_j (X)/f_i (X) \leq 1) + F_j (f_j (X)/f_i (X) > 1) \right]^{-1} = \epsilon_{ij}. \quad \text{So}$$

$$\epsilon_{00} = \min \{\epsilon_{ij} : 1 \leq i < j \leq k\}. \quad \text{Given the } F_i \text{'s and corresponding densities } f_i, \text{ the } \epsilon_{ij} \text{ can be computed. } \epsilon_0 \text{ is any positive number less than } \epsilon_{00}.$$  

Now bounds for $\min \{\epsilon_i : 1 \leq i \leq k\}$ will be considered. Let

$$c_i = \pi_i \sup \limits_{x} L(F_i, \delta^*(x)), \quad 1 \leq i \leq k. \quad (4.1)$$

It will be assumed that $c_i > 0, \ 1 \leq i \leq k$. So (3.11), rather than (3.10), of Theorem 3.1 is being considered. Recall that (3.10) is the trivial case and the corresponding $\epsilon_i$ is one. For any measurable set $B$, let

$$I(B) = \int \limits_{B \setminus \{\delta^*(x)\}} \min \limits_{A_i} S(x, a, f) - S(x, \delta^*(x), f) d\mu(x). \quad (4.2)$$

Recall that on the sets $A_i$ of (3.9), 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 satisfying $0 \leq c \leq I(A_i)$.  

The sets $B_i$, $1 \leq i \leq k$, will be said to "satisfy the inclusion conditions" if the $B_i$ are all disjoint, $u(B_i) > 0$, and $B_i \subseteq A_i$, $1 \leq i \leq k$. These were the conditions required of the $B_i$ in Theorem 3.1. Setting the integral of (3.11) equal to one and solving for $\epsilon_i$ (as a function of $B_i$) yields

\begin{equation}
\epsilon_i(B_i) = I(B_i)/(c_i + I(B_i)).
\end{equation}

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

Finally, the interest is not in bounds on $\min\{\epsilon_i : 1 \leq i \leq k\}$, per se. This can always be made small by choosing one of the $B_i$ small. The $B_i$ are to be chosen to make this quantity as large as possible, since, large values of $\epsilon^*$ correspond to wide classes of priors with respect to which $\delta^*$ is $\Gamma$-minimax. Thus, bounds are desired for

\begin{equation}
\epsilon^* = \sup_{B} \min\{\epsilon_i(B_i) : 1 \leq i \leq k\}
\end{equation}

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

Theorem 4.1.

\begin{equation}
\epsilon^* \leq I(uA_i)/\left(\sum_{i=1}^k c_i + I(uA_i)\right).
\end{equation}

If there exist $B_j$, $1 \leq j \leq k$, which satisfy the inclusion conditions and also satisfy

\begin{equation}
I(B_j) = c_j I(uA_i)/\sum_{i=1}^k c_i, \quad 1 \leq j \leq k,
\end{equation}

then this bound is attained.
Proof. If $B_j$ satisfies (4.6), then (4.3) yields

$$
(4.7) \quad \varepsilon_j(B_j) = I(uA_i) / \sum_{i} c_i + I(uA_i)).
$$

Let $B_1, \ldots, B_k$ satisfy the inclusion conditions. Since the $B_i$'s are disjoint and satisfy $uB_i \subseteq uA_i$, $\sum_i I(B_i) \leq I(uA_i)$. So at least one $B_j$ satisfies $I(B_j) \leq c_j I(uA_i) / \sum_i c_i$. (4.7) yields $\varepsilon(B_j) \leq I(uA_i) / \sum_i c_i + I(uA_i))$ and $\min \varepsilon(B_j): 1 \leq j \leq k$ is less than or equal to the same value. The $B_i \ldots, B_k$ were arbitrary, so (4.5) is true. The second assertion follows from (4.7).

A lower bound for $\varepsilon^*$ is more complicated to write down. It involves how the $A_i$ overlap in a particular problem. To write down a lower bound, this notation will be used. Let

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

There are at most $2^k - 1$ sets in $E$, the sets in $E$ are all disjoint and $E$ is non-empty since $uE = (uA_i) \setminus N$ where $\mu(N) = 0$. For $E \in E$, define

$$
\phi(E) = \{i_1, \ldots, i_m: E \subseteq A_i \cap A_j, 1 \leq j \leq m\}. \phi(E)$ has between one and $k$ elements for every $E$.

Theorem 4.2. Let

$$
M_j = \sum_{E: j \in \phi(E)} (I(E) / \sum_{m \in \phi(E)} c_m), 1 \leq j \leq k.
$$

Then $\varepsilon^* \geq \min \{M_j / (1 + M_j): 1 \leq j \leq k\}$. 

Proof. Denote the elements in \( E \) by \( E_1, \ldots, E_n \). Partition each \( E_i \) into the number of elements in \( \phi(E_i) \) disjoint measurable subsets, \( \{E_{ij} : j \in \phi(E_i)\} \), satisfying \( I(E_{ij}) = c_j I(E_i)/\sum_{m \in \phi(E_i)} c_m \). Let \( B_j = \bigcup_{i \in \phi(E_i)} E_{ij}, 1 \leq j \leq k \).

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

\[
I(B_j) = \sum_{i : j \in \phi(E_i)} I(E_{ij}) = \sum_{i : j \in \phi(E_i)} (c_j I(E_i)/\sum_{m \in \phi(E_i)} c_m).
\]

(4.3) yields

\[
\epsilon_j(B_j) = M_j/(1+M_j)
\]
since this choice of \( B_1, \ldots, B_k \) satisfy the inclusion conditions, the result follows.

Example 4.1. An example in which the upper bound of Theorem 4.1 is attained is the following. Let \( k = 3 \). Let \( F_1, F_2, \) and \( F_3 \) be normal distributions with means \(-1, 0, \) and \( 1\) respectively and common variance 1. Let \( A = \{a_1, a_2, a_3\} \) where \( a_i \) corresponds to classifying the observation as coming from \( F_i \). Suppose the prior is \( \pi_1 = \pi_3 = .3 \) and \( \pi_2 = .4 \). Assume 0-1 loss so \( L(F_i, a_j) = 0 \) if \( i = j \) and 1 if \( i \neq j \). Then the Bayes rule is of the form

(4.8) \[\delta^*(x) = \begin{cases} a_1 & x < x_1 \\ a_2 & x_1 \leq x \leq x_2 \\ a_3 & x_2 < x \end{cases}\]

where \( a_i \) this prior and loss \(-x_1 = x_2 = .80\). Using normal tables, it can be computed that \( I(uA_i) = .2510 \). So the upper bound of (4.5) is \( .205 \). Let \( B_1 = (0, \infty) \), \( B_2 = (-\infty, -c) \cup (c, \infty) \) and \( B_3 = (-c, 0) \). Choosing \( c = 1.80 \) results
in (4.6) being satisfied and these $B_i$ satisfy the inclusion conditions.

$\varepsilon_{00} = .276 > .205$ so $\delta^*$ is $\Gamma_{.205}$-minimax. The fact that $\delta^*$ is $\Gamma$-minimax for up to 20% contamination reflects favorably on $\delta^*$. The lower bound of Theorem 4.2 is .199 so the two bounds are very close in this problem.

Example 4.2. An example in which the lower bound of Theorem 4.2 is attained is the following

Consider the same classification problem as in Example 4.1 with the only change being in the loss. Rather than 0-1 loss, assume the loss matrix is

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

For this loss and prior, $\delta^*$ is of the form (4.8) with $x_1 = -.95$ and $x_2 = 1.55$.

The lower bound is .026 and it can be shown that this is the best which can be done with the construction of Theorem 3.1. The choice of $B_i$'s which attain this satisfy $B_1 \cup B_2 = (1.55, \infty)$ and $B_3 = (-\infty, -.95)$. So in this problem, $\delta^*$ is $\Gamma_{.026}$-minimax. The upper bound can be computed to be .055. As in Example 4.1, the two bounds do not differ greatly.

Remark 4.1. It should be remembered that the bounds obtained in this section are only for the construction used in Theorem 3.1. So $\delta^*$ may in fact be $\Gamma_\varepsilon$-minimax for $\varepsilon$ greater than indicated by these bounds.
5. HYPOTHESIS TESTING

Hypothesis testing is an important special case of the problem which has been discussed. In a hypothesis testing problem, there are two distributions with densities $f_1(x)$ and $f_2(x)$ and $A = \{a_1, a_2\}$. The loss has the form $L(F_1, a_1) = L(F_2, a_2) = 0$ and $L(F_i, a_j) = L_i > 0$ for $i \neq j$. A version of the Bayes rule is given by

$$
\delta^*(x) = \begin{cases} 
  a_1 & \text{if } r(x) \leq b \\
  a_2 & \text{if } r(x) > b
\end{cases}
$$

where $r(x) = f_2(x)/f_1(x)$ and $b = \pi_1 L_1/\pi_2 L_2$. The sets $A_1$ of (3.9) are $A_1 = \{x: r(x) > b\}$ and $A_2 = \{x: r(x) < b\}$. Since $A_1 \cap A_2 = \emptyset$, $B_1$ and $B_2$ can be any non-empty subsets of $A_1$ and $A_2$. But if the $B_i$'s are of the form $B_1 = \{x: r(x) > c^+ \geq b\}$ and $B_2 = \{x: r(x) < c^- \leq b\}$, where $c^-$ and $c^+$ are constants chosen so that $\epsilon_1 = \epsilon_2 = \epsilon^*$, the densities of the least favorable distributions constructed in Theorem 3.1 are

$$
g_1(x) = \begin{cases} 
  (1-\epsilon^*)f_1(x) & \text{if } r(x) \leq c^- \\
  (1-\epsilon^*)bf_2(x) & \text{if } r(x) > c^-
\end{cases}
$$

and

$$
g_2(x) = \begin{cases} 
  (1-\epsilon^*)f_2(x) & \text{if } r(x) \geq c^-\\
  (1-\epsilon^*)bf_1(x) & \text{if } r(x) < c^-
\end{cases}
$$

These are similar to but not the same as the least favorable pair used by Huber (1965). The probability ratio is
\[
\frac{g_2(x)}{g_1(x)} = \begin{cases} 
  b & \text{if } r(x) \geq c^- \\
  r(x) & \text{if } c^- > r(x) > c^- \\
  b & \text{if } c^- \geq r(x)
\end{cases}
\]

If \( c^- \) and \( c^- \) are chosen so that \( \epsilon_1 \neq \epsilon_2 \), the probability ratio will not have this truncated form.

**BIBLIOGRAPHY**


