BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
Technical Report #86-43		

4. TITLE (and Subtitle)

On Bayes and Empirical Bayes Procedures for Selection Problems

5. TYPE OF REPORT & PERIOD COVERED

Technical

6. PERFORMING ORG. REPORT NUMBER

Technical Report #86-43

7. AUTHOR(s)

Shanti S. Gupta and TaChen Liang

8. CONTRACT OR GRANT NUMBER(s)

N00014-84-C-0167

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Purdue University
Department of Statistics
West Lafayette, IN 47907

10. PROGRAM ELEMENT, PROJECT, TASK, AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Office of Naval Research
Washington, DC

12. REPORT DATE

September 1986

13. NUMBER OF PAGES

30

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)

Asymptotically optimal; Bayes procedure; empirical Bayes procedure; essentially complete; selection and ranking.

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

In this paper, we describe selection and ranking procedures using Bayesian or Empirical Bayes approaches. Section 2 of this paper deals with the problem of selecting the best population or selecting a subset containing the best population through Bayesian approach. An essentially complete class is obtained for a class of reasonable loss functions. A control condition, called P*-condition, is used to filter out poor procedures. In Section 3, we first set up a general formulation of empirical Bayes framework for selection problems. Several empirical Bayes frameworks are discussed based on the underlying statistical models. Two selection

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

problems dealing with binomial and uniform distributions are discussed in detail.

On Bayes and Empirical Bayes Procedures for Selection Problems*

by

Shanti S. Gupta and TaChen Liang
Purdue University

Technical Report #86-43
TR-86-43

Department of Statistics
Purdue University

September 1986

* This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

On Bayes and Empirical Bayes Procedures for Selection Problems*

by

Shanti S. Gupta and TaChen Liang
Purdue University

Abstract

In this paper, we describe selection and ranking procedures using Bayesian or Empirical Bayes approaches. Section 2 of this paper deals with the problem of selecting the best population or selecting a subset containing the best population through Bayesian approach. An essentially complete class is obtained for a class of reasonable loss functions. A control condition, called P*-condition, is used to filter out poor procedures. In Section 3, we first set up a general formulation of empirical Bayes framework for selection problems. Several empirical Bayes frameworks are discussed based on the underlying statistical models. Two selection problems dealing with binomial and uniform distributions are discussed in detail.

AMS 1980 Subject Classification: Primary 62F07; Secondary 62C10

KEY WORDS: Asymptotically optimal; Bayes procedure; empirical Bayes procedure; essentially complete; selection and ranking.

* This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Introduction

A common problem faced by an experimenter is one of comparing several populations (processes, treatments). These may be, for example, different varieties of a grain or different drugs for a specific disease. In other words, we have $k(\geq 2)$ populations and each population is characterized by the value of a parameter of interest, say, θ, which may be, in the example of drugs, an appropriate measure of the effectiveness of a drug. The classical approach to this problem is to test the homogeneity hypothesis $H_0 : \theta_1 = \theta_2 = \cdots = \theta_k$, where $\theta_1, \ldots, \theta_k$ are the values of the parameter for these populations. However, the classical tests of homogeneity are inadequate in the sense that they do not answer a frequently encountered experimenter's question, namely, how to identify the "best" population or how to select the more promising (worthwhile) subset of the populations for further experimentation.

The formulation of a k-sample problem as a multiple decision problem enables the experimenter to answer questions regarding the selection of the best or a subset containing the best population. The formulation of multiple decision procedure in the framework of selection and ranking procedures has been accomplished generally using either the indifference zone approach or the subset selection approach. The former approach was introduced by Bechhofer (1954). Substantial contribution to the early and subsequent developments in the subset selection theory has been made by Gupta (1956, 1965). A discussion of their differences and various modifications that have taken place since then can be found in Gupta and Panchapakesan (1979).

In many situations, an experimenter may have some prior information about the parameters of interest, and he would like to use this information to make an appropriate

decision. In this sense, the classical ranking and selection procedures may seem conservative if the prior information has not been considered. If the information at hand can be quantified into a single prior distribution, one would like to apply a Bayes procedure since it achieves the minimum of Bayes risks among a class of decision procedures. In his recent book, Berger (1985) discusses several approaches to select a prior distribution based on the information at hand. Some contributions to multiple decision problems using Bayesian approach have been made by Bickel and Yahav (1977), Chernoff and Yahav (1977), Deely and Gupta (1968), Goel and Rubin (1977), Gupta and Hsiao (1981), Gupta and Miescke (1984), Gupta and Yang (1985), Berger and Deely (1986), Guttman and Tiao (1964), Miescke (1979) and Roth (1978), among others. Readers are referred to Box and Tiao (1973) and Berger (1985) for general Bayesian inference in statistical analysis.

However, it is usually difficult, perhaps impossible, to quantify the prior information through a single prior. Therefore, it is suggested, (for example, see Robbins (1964)), that the prior information is quantified through a class Γ of subjectively plausible priors. Blum and Rosenblatt (1967) and Berger and Berliner (1986) have used this idea in statistical inference. One of the approaches, through the consideration of a class Γ of subjectively plausible priors, is the so-called Γ-minimax approach. One would like to apply the Γ-minimax procedure which minimizes the supremum of the Bayes risk over the class Γ of priors. Some contributions to multiple decision problems using this criterion have been made by Gupta and Hsiao (1981), Gupta and Huang (1977), Gupta and Kim (1980), Huang and Tseng (1983), Miescke (1981) and Randles and Hollander (1971). Also, Deely (1965) studied some selection problems through empirical Bayes approach assuming that the prior distribution belongs to a class of distributions with some unknown hyperparameters.

The empirical Bayes approach in statistical decision theory is appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem with respect to an unknown (or partially known) prior distribution on the parameter space. One then uses the accumulated observations to improve the decision procedure at each stage. This approach is due to Robbins (1956, 1964, 1983). Empirical Bayes procedures have been derived for multiple decision problems by Deely (1965) for selecting a subset containing the best population. Van Ryzin (1970), Huang (1975), Van Ryzin and Susarla (1977) also studied other multiple decision problems by using the empirical Bayes approach. Recently, Gupta and Hsiao (1983) and Gupta and Leu (1983) have studied empirical Bayes procedures for selecting good populations with respect to a standard or a control. Gupta and Liang (1984, 1986) have studied empirical Bayes procedures for the problem of selecting the best population or selecting populations better than a standard or a control with underlying populations being binomially distributed. Many such empirical Bayes procedures have been shown asymptotically optimal in the sense that the risk for the n-th decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was fully known and the Bayes procedure with respect to this prior distribution was used.

In the present paper, we describe selection and ranking procedures using prior distributions or using the information contained in the past data. Section 2 of this paper deals with the problem of selecting the best population through Bayesian approach. An essentially complete class is obtained for a class of reasonable loss functions. We also discuss Bayes-P* selection procedures which are better than the classical subset selection procedures in terms of the size of selected subset. In Section 3, we first set up a general

formulation of empirical Bayes framework for selection and ranking problems. Several empirical Bayes frameworks are discussed based on the underlying statistical models. Two selection problems dealing with binomial and uniform populations are discussed in detail.

2. Bayesian Approach

2.1. Notations and Formulation of the Selection Problem

Let $\theta_i \epsilon \Theta \subset \mathbb{R}$ denote the unknown characteristic of interest associated with population $\pi_i, i = 1, \ldots, k$. Let $X_1, \ldots, X_k$ be random variables representing the k populations $\pi_i, i = 1, \ldots, k$, respectively, with X_i having the probability density function (or probability frequency function in discrete case) $f_i(x|\theta_i)$. In many cases, X_i is a sufficient statistic for θ_i. It is assumed that given $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$, $\underset{\sim}{X} = (X_1, \ldots, X_k)$ have a joint probability density function $\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{k} f_i(x_i|\theta_i)$, where $\underset{\sim}{x} = (x_1, \ldots, x_k)$. Let $\theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of θ_i's and let $\pi_{[i]}$ denote the unknown population associated with $\theta_{[i]}$. The population $\pi_{[k]}$ will be called the best population. If there are more than one population satisfying this condition, we arbitrarily tag one of them and call it the best one. Also, we let $\Omega = \{\underset{\sim}{\theta} | \theta_i \epsilon \Theta, i = 1, \ldots, k\}$ denote the parameter space; also denote by $G(\cdot)$ a prior distribution on $\underset{\sim}{\theta}$ over Ω.

Let $\mathcal{A}$ be the action space consisting of all the $2^k - 1$ nonempty subset of the set $\{1, \ldots, k\}$. When action S is taken, we mean that population π_i is included in the selected subset if $i \epsilon S$. For each $\underset{\sim}{\theta} \epsilon \Omega$ and $S \epsilon \mathcal{A}$, let $L(\underset{\sim}{\theta}, S)$ denote the loss incurred when $\underset{\sim}{\theta}$ is the true state of nature and the action S is taken. A decision procedure d is defined to be a mapping from $\mathcal{X} \times \mathcal{A}$ into $[0, 1]$, where $\mathcal{X}$ is the sample space of $\underset{\sim}{X} = (X_1, \ldots, X_k)$. That

is, for $\underset{\sim}{x}\epsilon\mathcal{X}$ and $S\epsilon\mathcal{A}$, $d(\underset{\sim}{x}, S)$ is the probability of taking action S when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. Let D* be the set of all decision procedures $d(\underset{\sim}{x}, S)$.

For each $d\epsilon D^*$, let $B(d, G)$ denote the associated Bayes risk. That is,

$$B(d,G) = \int_{\Omega}\int_{\mathcal{X}} \sum_{S\epsilon\mathcal{A}} d(\underset{\sim}{x}, S)L(\underset{\sim}{\theta}, S)\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta})d\underset{\sim}{x}dG(\underset{\sim}{\theta}). \tag{2.1}$$

Then, $B(G) = \inf_{d\epsilon D^*} B(d, G)$ is the minimum Bayes risk.

An optimal decision procedure, denoted by d_G, is obtained if d_G has the property that

$$B(d_G, G) = B(G). \tag{2.2}$$

Such a procedure is called Bayes with respect to G. Under some regularity conditions, we can write (2.1) as

$$B(d,G) = \int_{\mathcal{X}} \sum_{S\epsilon\mathcal{A}} d(\underset{\sim}{x}, S) \int_{\Omega} L(\underset{\sim}{\theta}, S)\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta})d\underset{\sim}{x}. \tag{2.3}$$

Now Let

$$r_G(\underset{\sim}{x}, S) = \int_{\Omega} L(\underset{\sim}{\theta}, S)\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta})dG(\underset{\sim}{\theta}), \tag{2.4}$$

and

$$A_G(\underset{\sim}{x}) = \{S\epsilon\mathcal{A}|r_G(\underset{\sim}{x}, S) = \min_{S'\epsilon\mathcal{A}} r_G(\underset{\sim}{x}, S')\}. \tag{2.5}$$

Then, a sufficient condition for (2.2) is that d_G satisfies

$$\sum_{S\epsilon A_G(\underset{\sim}{x})} d_G(\underset{\sim}{x}, S) = 1. \tag{2.6}$$

2.2. An Essentially Complete Class of Decision Procedures

In this subsection, we consider a class of loss functions possessing the following properties:

Let H denote the group of all permutations of the components of a k-component vector.

Definition 2.1: A loss function L has property T if

(a) $L(\underset{\sim}{\theta}, S) = L(h\underset{\sim}{\theta}, hS)$ for all $\underset{\sim}{\theta}\epsilon\Omega, S\epsilon\mathcal{A}$ and $h\epsilon H$, and

(b) $L(\underset{\sim}{\theta}, S') \leq L(\underset{\sim}{\theta}, S)$ if the following holds for each pair (i,j) with $\theta_i \leq \theta_j$: $i\epsilon S, j\notin S$ and $S' = (S - \{i\}) \cup \{j\}$.

The property (a) assures the invariance under permutation and property (b) assures the monotonicity of the loss function. In many situations, a loss function satisfying these assumptions seems quite natural. Some examples of such a loss function are:

(2.7a) $L(\underset{\sim}{\theta}, S) = a|S| + [\theta_{[k]} - \theta_S]$ (Goel and Rubin (1977));

(2.7b) $L(\underset{\sim}{\theta}, S) = \frac{1}{|S|}\sum_{j\epsilon S}(\theta_{[k]} - \theta_j) + bI_{\{\theta_{[k]} > \theta_S\}}$ (Bickel and Yahav (1977));

(2.7c) $L(\underset{\sim}{\theta}, S) = c(\theta_{[k]} - \theta_S) - \frac{1}{|S|}\sum_{j\epsilon S}\theta_j$ (Chernoff and Yahav (1977));

(2.7d) $L(\underset{\sim}{\theta}, S) = |S| + eI_{\{\theta_{[k]} > \theta_S\}}$ (Gupta and Hsu (1978));

(2.7e) $L(\underset{\sim}{\theta}, S) = \sum_{j\epsilon S}\alpha(|S|)(\theta_{[k]} - \theta_j)$ (Deely and Gupta (1968));

where, $|S|$ denotes the cardinality of the subset S, $\theta_S = \max_{j \epsilon S} \theta_j$, a, b, c and e are positive constants, $\alpha(\cdot)$ is a positive function on the set $\{1, 2, \ldots, k\}$ and I_A denotes the indicator function of set A.

Note that different loss functions have different interpretations. For further discussion, see Gupta and Hsu (1978).

We now let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(k)}$ denote the ordered observations. Here the (i) can be viewed as the (unknown) index of the population associated with the observation $x_{(i)}$. For each $j = 1, \ldots, k$, let $S_j = \{(k), \ldots, (k-j+1)\}$, and the remaining subsets S_j be associated one-to-one with $j = k+1, \ldots, 2^k - 1$, arbitrarily. Also, let $\mathcal{A}_m = \{S \epsilon \mathcal{A} || S| = m\}, m = 1, \ldots, k$, and $D_1^* = \{d \epsilon D^* | \sum_{j=1}^{k} d(\underset{\sim}{x}, S_j) = 1 \text{ for all } \underset{\sim}{x} \epsilon \mathcal{X}\}$.

Theorem 2.1: Suppose that $f_i(x_i|\theta_i) = f(x_i|\theta_i), i = 1, \ldots, k$, where the pdf $f(x|\theta)$ possesses the monotone likelihood ratio (MLR) property, and the prior distribution G is symmetric on Ω. Then,

(a) for each $m = 1, \ldots, k$, $r_G(\underset{\sim}{x}, S_m) \leq r_G(\underset{\sim}{x}, S)$ for all $S \epsilon \mathcal{A}_{k-m+1}$, $\underset{\sim}{x} \epsilon \mathcal{X}$, and

(b) D_1^* is an essentially complete class in D^*;

provided that the loss function has property T.

Proof: The proof for part (a) is analogous to that of Theorem 3.3 of Gupta and Yang (1985). For part (b), let d be any decision procedure in D^*. Consider the decision procedure d^* defined as: for $\underset{\sim}{x} \epsilon \mathcal{X}$,

$$d^*(\underset{\sim}{x}, S_m) = \sum_{S \epsilon \mathcal{A}_{k-m+1}} d(\underset{\sim}{x}, S), \; m = 1, \ldots, k;$$

and

$$d^*(\underset{\sim}{x}, S) = 0, \quad S \neq S_m, \ m = 1, \ldots, k.$$

Then, $d^* \epsilon D_1^*$. Also, by part (a) and (2.3), one can see that

$$B(d^*, G) \leq B(d, G).$$

Hence, the proof of part (b) is completed.

Let $A'_G(\underset{\sim}{x}) = \{S_j | 1 \leq j \leq k, \ r_G(\underset{\sim}{x}, S_j) = \min_{1 \leq i \leq k} r_G(\underset{\sim}{x}, S_i)\}$. Then, under the condition of Theorem 2.1, any Bayes procedure d_G satisfies $\sum_{S_j \epsilon A'_G(\underset{\sim}{x})} d_G(\underset{\sim}{x}, S_j) = 1$ for $\underset{\sim}{x} \epsilon \mathcal{X}$.

Goel and Rubin (1977) choose the loss function (2.7a) and study the behavior of the corresponding Bayes procedure in great detail. Bickel and Yahav (1977) assume that $\theta_{[1]}, \ldots, \theta_{[k]}$ are known and consider the loss function (2.7b). They obtain the best invariant procedure for the normal pdf and then depart from the decision theoretic approach to simplifying this procedure as $k \to \infty$. Chernoff and Yahav (1977) consider the loss function (2.7c) and compare the performance of the Bayes procedure with other procedures in a "normal model" on the basis of Monte Carlo results.

2.3. Bayes Procedures wrt Additive Loss Functions

Deely and Gupta (1968) consider the loss function $L(\underset{\sim}{\theta}, S)$ corresponding to the choice of S given by

$$L(\underset{\sim}{\theta}, S) = \sum_{j \epsilon S} \alpha_{S_j} (\theta_{[k]} - \theta_j). \tag{2.8}$$

Some examples of such a loss function are:

(2.8a) $L(\underset{\sim}{\theta}, S) = \sum_{j \epsilon S} (\theta_{[k]} - \theta_j)$, sum of losses;

(2.8b) $L(\underset{\sim}{\theta}, S) = \frac{1}{|S|} \sum_{j \epsilon S} (\theta_{[k]} - \theta_j)$, average loss;

(2.8c) $L(\underset{\sim}{\theta}, S) = (k + 1 - |S|) \sum_{j \epsilon S} (\theta_{[k]} - \theta_j)$.

Note that all these three loss functions have the property T.

Deely and Gupta (1968) proved that when $\alpha_{Sj} = \alpha > 0$ for all $S \epsilon \mathcal{A}$ and $j \epsilon S$, then the Bayes procedure always selects exactly one population. Miescke (1979) slightly generalized the result of Deely and Gupta (1968). He considered the loss function

$$L(\underset{\sim}{\theta}, S) = \sum_{i \epsilon S} \alpha(|S|) \ell_i(\underset{\sim}{\theta}), \tag{2.9}$$

where $\alpha(\cdot)$ is nonnegative function on the set $\{1, \ldots, k\}$. We cite his result as follows:

Theorem 2.2. Let $m\alpha(m) \geq \alpha(1), m = 1, 2, \ldots, k$. If the ℓ_i's are nonnegative, then there exists a Bayes procedure which always selects exactly one population.

Theorem 2.2 does not hold if the nonnegativity of ℓ_i for all $i = 1, \ldots, k$ is not satisfied. For example, consider the loss function

$$L(\underset{\sim}{\theta}, S) = \sum_{i \epsilon S} (\theta_{[k]} - \theta_i - \varepsilon), \tag{2.10}$$

where $\varepsilon > 0$ is a given constant. This loss function can be used for the problem of selecting populations close to the best. With this loss function, it is possible to select more than one population.

2.4. Bayes-P* Selection Procedures

In this section, we continue with the general setup of Section 2.1.

A selection procedure $\underset{\sim}{\psi} = (\psi_1, \ldots, \psi_k)$ is defined to be a mapping from $\mathcal{X}$ to $[0,1]^k$, where $\psi_i(\underset{\sim}{x}) : \mathcal{X} \to [0,1]$ is the probability that π_i is included in the selected subset when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. A selection procedure $\underset{\sim}{\psi}$ is called nonrandomized if all ψ_i's are 0 or 1; otherwise, it is a randomized procedure. A correct selection (CS) is defined to be the selection of any subset that includes the best population.

Let d by any decision procedure considered in earlier sections. A selection procedure $\underset{\sim}{\psi}^d = (\psi_1^d, \ldots, \psi_k^d)$ associated with d can be obtained by letting

$$\psi_i^d(\underset{\sim}{x}) = \sum_{S \ni i} d(\underset{\sim}{x}, S), \tag{2.11}$$

where the summation being over all the subsets containing i.

In the decision-theoretic approach, a Bayes decision (selection) procedure always provides a decision with minimum risk under a certain loss. However, since, in practice, one always has the difficulty in figuring out what the loss may be and the Bayesian result is quite sensitive to the loss used, in this sense, a Bayes procedure does not mean that its quality is good enough to pass a certain level. For guaranteeing the quality of decision (selection) procedures one would like to have a "quality control" about the class of all possible decision (selection) procedures. That is, any procedure with lower quality will be removed, even though it might be the cheapest one under some losses. Analogous to classical subset selection approach, Gupta and Yang (1985) set up a control condition using the Bayesian approach.

Let

(2.12) $$p_i(\underset{\sim}{x}) = P(\pi_i \text{ is the best } | \underset{\sim}{X} = \underset{\sim}{x}) = P(\theta_i \text{ is the largest } | \underset{\sim}{X} = \underset{\sim}{x})$$

be the posterior probability that population π_i is the best population when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. Then, for selection procedure $\underset{\sim}{\psi}$, the posterior probability of a correct selection given $\underset{\sim}{X} = \underset{\sim}{x}$ is

(2.13) $$P(CS|\underset{\sim}{\psi}, \underset{\sim}{X} = \underset{\sim}{x}) = \sum_{i=1}^{k} \psi_i(\underset{\sim}{x}) p_i(\underset{\sim}{x}).$$

Definition 2.2. Given a number P^*, $k^{-1} < P^* < 1$, and a prior G on Ω, we say a selection procedure $\underset{\sim}{\psi}$ satisfies the PP*-condition (posterior P*-condition) if

(a) $\psi_i(\underset{\sim}{x}) = 1$ at least for some i, $1 \leq i \leq k$, and

(b) $P(CS|\underset{\sim}{\psi}, \underset{\sim}{X} = \underset{\sim}{x}) \geq P^*$ for all $\underset{\sim}{x} \epsilon \mathcal{X}$.

Note that $\sum_{i=1}^{k} p_i(\underset{\sim}{x}) = 1$ for all $\underset{\sim}{x} \epsilon \mathcal{X}$; hence this kind of selection procedures always exist. We let $\mathcal{C} = \mathcal{C}(P^*)(\mathcal{C}^* = \mathcal{C}^*(P^*))$ be the class of all nonrandomized (randomized) selection procedures satisfying the PP*-condition.

Let $p_{[1]}(\underset{\sim}{x}) \leq \ldots \leq p_{[k]}(\underset{\sim}{x})$ be the ordered $p_i(\underset{\sim}{x})$'s and let $\pi_{(i)}$ be the population associated with $p_{[i]}(\underset{\sim}{x}), i = 1, \ldots, k$. Then a selection procedure $\underset{\sim}{\psi}$ can be completely specified by $\{\psi_{(1)}, \ldots, \psi_{(k)}\}$, where

(2.14) $$\psi_{(i)}(\underset{\sim}{x}) = P(\pi_{(i)} \text{ is selected } | \underset{\sim}{\psi}, \underset{\sim}{X} = \underset{\sim}{x}\}, i = 1, \ldots, k.$$

Gupta and Yang (1985) proposed two selection procedures; one is nonrandomized, say $\underset{\sim}{\psi}^G$ and the other is randomized, say $\underset{\sim}{\psi}^{G^*}$. They are defined as below.

Definition 2.3. Given a number $P^*, k^{-1} < P^* < 1$, and an observation $\underset{\sim}{X} = \underset{\sim}{x}$, let $j = max\{m | \sum_{i=m}^{k} p_{[i]}(\underset{\sim}{x}) \geq P^*\}$.

(a) The nonrandomized selection procedure $\underset{\sim}{\psi}^G$ is defined by $\{\psi^G_{(1)}, \ldots, \psi^G_{(k)}\}$, where

$$\psi^G_{(i)}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i \geq j; \\ 0 & \text{otherwise.} \end{cases}$$

(b) The randomized selection procedure $\underset{\sim}{\psi}^{G^*}$ is defined by $\{\psi^{G^*}_{(1)}, \ldots, \psi^{G^*}_{(k)}\}$, where $\psi^{G^*}_{(k)}(\underset{\sim}{x}) = 1$, and for $1 \leq i \leq k-1$,

$$\psi^{G^*}_{(i)}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i > j; \\ \lambda & \text{if } i = j; \\ 0 & \text{if } i < j; \end{cases}$$

the constant λ is determined so that

$$\lambda p_{[j]}(\underset{\sim}{x}) + \sum_{m=j+1}^{k} p_{[m]}(\underset{\sim}{x}) = P^*.$$

It is clear that $\underset{\sim}{\psi}^G \epsilon C$ and $\underset{\sim}{\psi}^{G^*} \epsilon C^*$. In the following, some optimalities of these two selection procedures are investigated.

Definition 2.4. A selection procedure $\underset{\sim}{\psi}$ is called ordered if for every $\underset{\sim}{x} \epsilon \mathcal{X}, x_i \leq x_j$ implies $\psi_i(\underset{\sim}{x}) \leq \psi_j(\underset{\sim}{x})$. It is called monotone or just if for every $i = 1, \ldots, k$, and $\underset{\sim}{x}, \underset{\sim}{y} \epsilon \mathcal{X}$, $\psi_i(\underset{\sim}{x}) \leq \psi_i(\underset{\sim}{y})$ whenever $x_i \leq y_i, x_j \geq y_j$ for any $j \neq i$.

Some sufficient condition for $\underset{\sim}{\psi}^G (\underset{\sim}{\psi}^{G^*})$ to be ordered and monotone are given below:

Theorem 2.3. (Gupta and Yang (1985)). Let $G(\underset{\sim}{\theta}|\underset{\sim}{x})$ be the posterior cdf of $\underset{\sim}{\theta}$, given $\underset{\sim}{X} = \underset{\sim}{x}$. Let $G(\underset{\sim}{\theta}|\underset{\sim}{x})$ be absolutely continuous and have the generalized stochastic increasing property, that is:

(1) $G(\underset{\sim}{\theta}|\underset{\sim}{x}) = \prod_{i=1}^{k} G_i(\theta_i|\underset{\sim}{x}), G_i(\cdot|\underset{\sim}{x}) =$ posterior cdf of θ_i.

(2) $G_i(t|\underset{\sim}{x}) \geq G_j(t|\underset{\sim}{x})$ for any t, whenever $x_i \leq x_j$.

Then, both $\underset{\sim}{\psi}^G$ and $\underset{\sim}{\psi}^{G^*}$ are ordered and monotone.

Gupta and Yang (1985) also investigated some optimal behavior of these two procedures through the decision-theoretic approach over a class of loss functions.

Definition 2.5. A loss function L has property T' if

(a) L has property T, and

(b) $L(\underset{\sim}{\theta}, S) \leq L(\underset{\sim}{\theta}, S')$ if $S \subset S'$.

Theorem 2.4. (Gupta and Yang (1985)). Under the assumption of Theorem 2.3, the selection procedure $\underset{\sim}{\psi}^G(\underset{\sim}{\psi}^{G^*})$ is Bayes in $C(C^*)$ provided that the loss function has property T'.

Gupta and Yang (1985) investigated the computation of $p_i(\underset{\sim}{x})$ for the "normal model" by using normal and non-informative priors. Berger and Deely (1986) consider another selection problem, and give a more detailed discussion about the computation of $p_i(\underset{\sim}{x})$ under several different priors.

3. Empirical Bayes Approach

In this section, we continue with the general setup of Section 2. However, we assume only the existence of prior distribution G on Ω, and the form of G is unknown or partially known. In Section 3.1, we consider decision procedures for general loss functions. In Sections 3.2 and 3.3, only selection procedures are concerned.

3.1. Formulation and Summary of the Empirical Bayes Selection Problems

For each $i, i = 1, \ldots, k$, let X_{ij} denote the random observation from π_i at stage j. Let Θ_{ij} denote the random characteristic of π_i at stage j. Conditional on $\Theta_{ij} = \theta_{ij}, X_{ij}|\theta_{ij}$ has pdf (or pf in discrete case) $f_i(x|\theta_{ij})$. Let $\underset{\sim}{X}_j = (X_{1j}, \ldots, X_{kj})$ and $\underset{\sim}{\theta}_j = (\theta_{1j}, \ldots, \theta_{kj})$. Suppose that independent observations $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are available and $\underset{\sim}{\theta}_j, 1 \leq j \leq n$, have the same distribution G for all j, though not observable. We also let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present random observation.

Consider an empirical Bayes decision procedure d_n. Let $B(d_n, G)$ be the Bayes risk associated with the decision procedure d_n. Then

$$B(d_n, G) = \int_{\Omega} E \int_{\mathcal{X}} \sum_{S \in \mathcal{A}} d_n((\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n), S) L(\underset{\sim}{\theta}, S) f(\underset{\sim}{x}|\underset{\sim}{\theta}) d\underset{\sim}{x} dG(\underset{\sim}{\theta}),$$

where $d_n((\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n), S) (\equiv d_n(\underset{\sim}{x}, S))$ is the probability of selecting the subset S when $(\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ is observed, and the expectation E is taken with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. Note that $B(d_n, G) - B(G) \geq 0$, since $B(G)$ is the minimum Bayes risk. This nonnegative difference is always used as a measure of optimality of the decision procedure d_n.

Definition 3.1. A sequence of decision procedures $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution G if $B(d_n, G) \to B(G)$ as $n \to \infty$.

Let $L(\underset{\sim}{\theta}) = \max_{S \in \mathcal{A}} |L(\underset{\sim}{\theta}, S)|$ and assume that $\int L(\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) < \infty$. Following Robbins (1964), one can see that a sufficient condition for the sequence $\{d_n\}$ to be asymptotically optimal is that $d_n(\underset{\sim}{x}, S) \xrightarrow{P} d_G(\underset{\sim}{x}, S)$ for all $\underset{\sim}{x} \in \mathcal{X}$ and $S \in \mathcal{A}$, where "$\xrightarrow{P}$" means convergence in probability (with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$).

Let G_n be a distribution function on the parameter space Ω. Suppose G_n is a function of $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ such that $P\{\lim_{n\to\infty} G_n(\underset{\sim}{\theta}) = G(\underset{\sim}{\theta})$ for every continuous point $\underset{\sim}{\theta}$ of $G\} = 1$, where the probability is with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. Let the loss function $L(\underset{\sim}{\theta}, S)$ and the density $\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta})$ be such that $L(\underset{\sim}{\theta}, S)\underset{\sim}{f}(\underset{\sim}{x}|\underset{\sim}{\theta})$ is bounded and continuous in $\underset{\sim}{\theta}$ for every $S \epsilon \mathcal{A}$. Then $\{d_{G_n}\}$ is asymptotically optimal with respect to G if $\int_\Omega L(\underset{\sim}{\theta}) dG(\underset{\sim}{\theta}) < \infty$, where d_{G_n} is a Bayes procedure with respect to the distribution G_n.

To find G_n, we may assume G to be a member of some parametric family Γ with unknown hyperparameters, say $\underset{\sim}{\lambda} = (\lambda_1, \ldots, \lambda_k)$. Suppose now an estimator $\underset{\sim}{\lambda}_n = (\lambda_{1n}, \ldots, \lambda_{kn})$ depending on the previous observations $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$ can be found such that G_n converges to G with probability one. Note that G_n is also a member in Γ. We then follow the typically Bayesian analysis and derive the Bayes procedure d_{G_n} with respect to the estimated prior distribution G_n. Then, according to the result of Deely (1965), the sequence of empirical Bayes procedures $\{d_{G_n}\}$ is asymptotically optimal. This approach is referred as parametric empirical Bayes. Deely (1965) has derived the empirical Bayes procedures through the parametric empirical Bayes approach in several special cases among which are (a) normal-normal, (b) normal- uniform, (c) binomial-beta, and (d) Poisson-gamma.

In another approach, called nonparametric empirical Bayes, one just assumes that $\underset{\sim}{\theta}_j, j = 1, 2, \ldots$, are independently and identically distributed; however, the form of the prior distribution G on Ω is completely unknown. In this situation, one may either estimate the prior distribution and then proceed to a typical Bayesian analysis or represent the Bayes procedure in terms of the unknown prior and then use the data to estimate the Bayes

procedure directly. The estimation of the prior distribution through the nonparametric empirical Bayes approach has been studied (see Simar (1976) for Poisson distribution and Jewell (1982) for exponential distribution). For the second approach, see Van Ryzin (1970), Van Ryzin and Susarla (1977), Gupta an Hsiao (1983), Gupta and Leu (1983), and Gupta and Liang (1984, 1986), among others.

In the following sections, we consider some selection problems with underlying populations having binomial or uniform distributions. We will use the approach of first looking at the form of the Bayes procedure and then estimating the Bayes procedure directly.

3.2. Empirical Bayes Procedures Related to Binomial Populations

In this section, two selection problems related to binomial populations are discussed: selecting the best among k binomial populations and selecting populations better than a standard or a control. For each i, the observations X_i can be viewed as the number of successes among N independent trials taken from π_i, and the parameter θ_i as the probability of a success for each trial in π_i. Then $X_i|\theta_i$ has probability function $f_i(x|\theta_i) = \binom{N}{x}\theta_i^x(1-\theta_i)^{N-x}, x = 0, 1, \ldots, N$. We let $G_i(\cdot)$ denote the prior distribution of θ_i and assume that $G(\underline{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$.

3.2.1. Selecting the Best Binomial Population

Gupta and Liang (1986) considered the loss function

$$L(\underset{\sim}{\theta}, \{i\}) = \theta_{[k]} - \theta_i \tag{3.1}$$

for the problem of selecting the largest binomial parameter $\theta_{[k]}$ among k binomial populations.

Let $f_i(x) = \int_0^1 f_i(x|\theta) dG_i(\theta), W_i(x) = \int_0^1 \theta f_i(x|\theta) dG_i(\theta)$ and $\varphi_i(x) = W_i(x)/f_i(x)$. Then, from (3.1), following straightforward computation, a randomized Bayes selection procedure, say $\underset{\sim}{\psi}^B = (\psi_1^B, \ldots, \psi_k^B)$, is given below:

$$\psi_i^B(\underset{\sim}{x}) = \begin{cases} |S(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon S(\underset{\sim}{x}); \\ 0 & \text{otherwise;} \end{cases} \tag{3.2}$$

where

$$S(\underset{\sim}{x}) = \{i | \varphi_i(x_i) = \max_{1 \le j \le k} \varphi_j(x_j)\}. \tag{3.3}$$

Here, $\psi_i^B(\underset{\sim}{x})$ is the probability of selecting π_i as the best population given $\underset{\sim}{X} = \underset{\sim}{x}$.

Note that $\varphi_i(x)$ is the Bayes estimator of the parameter θ_i under the squared error loss given $X_i = x$. One can see that $\varphi_i(x)$ is increasing in x for $i = 1, \ldots, k$ and hence $\underset{\sim}{\psi}^B$ is a monotone selection procedure.

A. Formulation of the Empirical Bayes Framework

Due to the surprising quirk that $\varphi_i(x)$ can not be consistently estimated in the usual empirical Bayes sense (see Robbins (1964) and Samuel (1963)), an idea of Robbins in setting up the empirical Bayes framework for binomial populations is used below.

For each $i, i = 1, \ldots, k$, at stage j, consider $N+1$ independent trials from π_i. Let X_{ij} and Y_{ij}, respectively, stand for the number of successes in the first N trials and the last trial. Let $\underset{\sim}{Z}_j = ((X_{ij}, Y_{ij}), \ldots, (X_{kj}, Y_{kj}))$ denote the observations at the jth stage, $j = 1, \ldots, n$. We also let $\underset{\sim}{X}_{n+1} = \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present observations.

By the monotonicity of the estimators $\varphi_i(x), 1 \leq i \leq k$, in terms of Bayes risks, one can see that all monotone procedures form an essentially complete class in the set of all selection procedures. In view of this fact, it is reasonable to require that the appropriate empirical Bayes procedures possess the above mentioned monotone property. For this purpose, we first need to have some monotone empirical Bayes estimators for $\varphi_i(x), 1 \leq i \leq k$. Gupta and Liang (1986), by using isotonic regression method, proposed two monotone empirical Bayes estimators for $\varphi_i(x)$.

B. The Proposed Monotone Empirical Bayes Selection Procedures

For each $x = 0, 1, \ldots, N$, and $n = 1, 2, \ldots$, define

$$f_{in}(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_{ij}) + n^{-1}; \tag{3.4}$$

$$W_{in}(x) = \frac{1}{n} \sum_{j=1}^{n} Y_{ij} I_{\{x\}}(X_{ij}) + n^{-1}; \tag{3.5}$$

Also, let $V_{ij} = X_{ij} + Y_{ij}, j = 1, 2, \ldots$ Define

$$\tilde{W}_{in}(x) = \{[\frac{x+1}{n(N+1)} \sum_{j=1}^{n} I_{\{x+1\}}(V_{ij})] \wedge [\frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_{ij})]\} + n^{-1}, \tag{3.6}$$

where $a \wedge b = \min\{a, b\}$. Let

$$\varphi_{in}(x) = W_{in}(x)/f_{in}(x); \tag{3.7}$$

(3.8) $$\tilde{\varphi}_{in}(x) = \tilde{W}_{in}(x)/f_{in}(x);$$

and for each $0 \leq x \leq N$, define

(3.9) $$\varphi^*_{in}(x) = \max_{0\leq s\leq x} \min_{s\leq t\leq N} \{\sum_{y=s}^{t} \varphi_{in}(y)/(t-s+1)\};$$

(3.10) $$\tilde{\varphi}^*_{in}(x) = \max_{0\leq s\leq x} \min_{s\leq t\leq N} \{\sum_{y=s}^{t} \tilde{\varphi}_{in}(y)/(t-s+1)\}.$$

By (3.9) and (3.10), one can see that both $\varphi^*_{in}(x)$ and $\tilde{\varphi}^*_{in}(x)$ are increasing in x. Gupta and Liang (1986) proposed $\varphi^*_{in}(x)$ (or $\tilde{\varphi}^*_{in}(x)$) as an estimator of $\varphi_i(x)$. They also proposed two empirical Bayes selection procedures, say $\underset{\sim}{\psi}^*_n = (\psi^*_{1n}, \ldots, \psi^*_{kn})$, and $\underset{\sim}{\tilde{\psi}}_n = (\tilde{\psi}_{1n}, \ldots, \tilde{\psi}_{kn})$, which are given below, respectively:

(3.11) $$\psi^*_{in}(\underset{\sim}{x}) = \begin{cases} |S^*_n(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon S^*_n(\underset{\sim}{x}); \\ 0 & \text{otherwise,} \end{cases}$$

where

(3.12) $$S^*_n(\underset{\sim}{x}) = \{i | \varphi^*_{in}(x_i) = \max_{1\leq j\leq k} \varphi^*_{jn}(x_j)\};$$

and

(3.13) $$\tilde{\psi}_{in}(\underset{\sim}{x}) = \begin{cases} |\tilde{S}_n(\underset{\sim}{x})|^{-1} & \text{if } i \epsilon \tilde{S}_n(\underset{\sim}{x}); \\ 0 & \text{otherwise;} \end{cases}$$

where

(3.14) $$\tilde{S}_n(\underset{\sim}{x}) = \{i | \tilde{\varphi}^*_{in}(x_i) = \max_{1\leq j\leq k} \tilde{\varphi}^*_{jn}(x_j)\}.$$

Due to the increasing property of the estimators $\varphi^*_{in}(x), \tilde{\varphi}^*_{in}(x),\ 1 \leq i \leq k$, one can see that $\underset{\sim}{\psi}^*_n$ and $\underset{\sim}{\tilde{\psi}}_n$ are both monotone selection procedures.

Asymptotic Optimality of $\{\underset{\sim}{\psi}_n^*\}$ and $\{\tilde{\underset{\sim}{\psi}}_n\}$.

Without ambiguity, we still use $B(\underset{\sim}{\psi}, G)$ to denote the Bayes risk associated with the selection procedure $\underset{\sim}{\psi}$ when G is the true prior distribution.

Gupta and Liang (1986) proved that the two sequences of selection procedures $\{\underset{\sim}{\psi}_n^*\}$ and $\{\tilde{\underset{\sim}{\psi}}_n\}$ have the following asymptotically optimal property:

$$B(\underset{\sim}{\psi}_n^*, G) - B(\underset{\sim}{\psi}^B, G) \leq 0(exp(-c_1 n)),$$

and

$$B(\tilde{\underset{\sim}{\psi}}_n, G) - B(\underset{\sim}{\psi}^B, G) \leq 0(exp(-c_2 n)),$$

for some positive constants c_1 and c_2.

3.2.2. Selecting Populations Better Than A Control

Let $\theta_0 \epsilon (0,1)$ denote a control parameter. Population π_i is said to be good if $\theta_i \geq \theta_0$ and bad if $\theta_i < \theta_0$. Gupta and Liang (1984) considered the loss function

$$L(\underset{\sim}{\theta}, S) = \sum_{i \epsilon S} (\theta_0 - \theta_i) I_{(0,\theta_0)}(\theta_i) + \sum_{i \notin S} (\theta_i - \theta_0) I_{(\theta_0,1)}(\theta_i), \tag{3.15}$$

for the problem of selecting (excluding) all good (bad) populations. For the loss function (3.15), the first summation is the loss due to selecting some bad populations, and the second summation is the loss due to not selecting some good populations. The value of the control parameter θ_0 is either known or unknown. When θ_0 is unknown, a sample from the control population, say π_0, is needed. To be consistent with the notation used in earlier sections, we assume θ_0 is known. We note that Gupta and Liang (1984) have studied the case when θ_0 is unknown.

For the loss function (3.15), a nonrandomized Bayes selection procedure $\underset{\sim}{\alpha}^B = (\alpha_1^B, \ldots, \alpha_k^B)$ is given by

$$\alpha_i^B(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_i(x_i) \geq \theta_0; \\ 0 & \text{otherwise,} \end{cases} \tag{3.16}$$

where $\alpha_i^B(\underset{\sim}{x})$ is the probability of selecting π_i as a good population given $\underset{\sim}{X} = \underset{\sim}{x}$.

Note that $\underset{\sim}{\alpha}^B$ is also a monotone selection procedure. Hence, based on the estimators $\varphi_{in}^*(x)$ and $\tilde{\varphi}_{in}^*(x)$, two intuitive empirical Bayes procedures, say $\underset{\sim}{\alpha}_n^* = (\alpha_{1n}^*, \ldots, \alpha_{kn}^*)$ and $\tilde{\underset{\sim}{\alpha}}_n = (\tilde{\alpha}_{1n}, \ldots, \tilde{\alpha}_{kn})$ can be obtained where

$$\alpha_{in}^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_{in}^*(x_i) \geq \theta_0; \\ 0 & \text{otherwise;} \end{cases} \tag{3.17}$$

and

$$\tilde{\alpha}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \tilde{\varphi}_{in}^*(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

Similarly, one can show that these two sequences of selection procedures $\{\underset{\sim}{\alpha}_n^*\}$ and $\{\tilde{\underset{\sim}{\alpha}}_n\}$ have the following asymptotically optimal property:

$$B(\underset{\sim}{\alpha}_n^*, G) - B(\underset{\sim}{\alpha}^B, G) \leq 0(exp(-c_3 n)),$$

and

$$B(\tilde{\underset{\sim}{\alpha}}_n, G) - B(\underset{\sim}{\alpha}^B, G) \leq 0(exp(-c_4 n))$$

for some positive constants c_3 and c_4.

3.3. Empirical Bayes Procedures Related to Uniform Populations

In this section, we assume that the random variables X_i, $1 \leq i \leq k$, have uniform distributions $U(0, \theta_i), \theta_i > 0$ and unknown. The parameter space is $\Omega = \{\underset{\sim}{\theta} | \theta_i > 0,\ 1 \leq i \leq k\}$. It is also assumed that the prior distribution G on Ω has the form of $G(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$, where $G_i(\cdot)$ is a distribution on $(0, \infty)$, $i = 1, \ldots, k$.

Let $\theta_0 > 0$ be a known control parameter. Gupta and Hsiao (1983) considered the loss function

$$L(\underset{\sim}{\theta}, S) = L_1 \sum_{i \notin S} (\theta_i - \theta_0) I_{(\theta_0, \infty)}(\theta_i) + L_2 \sum_{i \in S} (\theta_0 - \theta_i) I_{(0, \theta_0)}(\theta_i), \tag{3.19}$$

where L_i, $i = 1, 2$, are positive and known, for the problem of selecting populations better than a standard θ_0.

Let $m_i(x)$ be the marginal pdf of X_i and $M_i(x)$ be the marginal distribution of X_i. Then, we have

$$m_i(x) = \int_x^\infty \frac{1}{\theta} dG_i(\theta) \quad for\ x > 0, \tag{3.20}$$

and

$$M_i(x) = \int_0^x \int_t^\infty \frac{1}{\theta} dG_i(\theta) dt = x m_i(x) + G_i(x). \tag{3.21}$$

Note that the marginal pdf $m_i(x)$ is continuous and decreasing in x.

By direct computation, a Bayes procedure $\underset{\sim}{\psi}^B = (\psi_1^B, \ldots, \psi_k^B)$ for this selection problem is given by

$$\psi_i^B(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } (x_i \geq \theta_0) \text{ or } (x_i < \theta_0 \text{ and } \Delta_{iG}(x_i) \geq 0); \\ 0 & \text{otherwise;} \end{cases} \tag{3.22}$$

where

$$\Delta_{iG}(x_i) = L_2 m_i(x_i)(x_i - \theta_0) + L_2[M_i(\theta_0) - M_i(x_i)] + L_1[1 - M_i(\theta_0)]. \tag{3.23}$$

By the decreasing property of the pdfs $m_i(x)$, $1 \leq i \leq k$, one can see that $\Delta_{iG}(x)$, $1 \leq i \leq k$, are increasing in x for $x < \theta_0$; and hence, the Bayes procedure $\underset{\sim}{\psi}^B$ has the monotone property.

Empirical Bayes Procedures

To form an empirical Bayes procedure, we first need to have some estimators, say $m_{in}(x)$ and $M_{in}(x)$, for $m_i(x)$ and $M_i(x)$, respectively. Due to the decreasing property of $m_i(x)$, we require that the estimators $m_{in}(x), n = 1, 2, \ldots$, possess the same property. Once an estimator m_{in} is obtained, we let

$$M_{in}(x) = \int_0^x m_{in}(y)dy, \tag{3.24}$$

and

$$\Delta_{in}(x) = L_2 m_{in}(x)(x - \theta_0) + L_2[M_{in}(\theta_0) - M_{in}(x)] + L_1[1 - M_{in}(\theta_0)]. \tag{3.25}$$

Then, an empirical Bayes procedure $\underset{\sim}{\psi}_n = (\psi_{1n}, \ldots, \psi_{kn})$ can be given as follows:

$$\psi_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } (x_i \geq \theta_0) \text{ or } (x_i < \theta_0 \text{ and } \Delta_{in}(x_i) \geq 0); \\ 0 & \text{otherwise.} \end{cases} \tag{3.26}$$

This empirical Bayes procedure $\underset{\sim}{\psi}_n$ is a monotone procedure if $m_{in}(x)$, $1 \leq i \leq k$, are decreasing in x. We use the method of Grenander (1956) to obtain such an estimator having the decreasing property.

Let $X^n_{i(1)} \le X^n_{i(2)} \le \ldots \le X^n_{i(n)}$ be the ordered observations of the first n observations taken from π_i. Let F_{in} be the empirical distribution based on $X_{i1}, \ldots, X_{in}$. For each $j,\ 1 \le j \le n$, let

$$\beta_{ij} = \min_{s \le j-1} \max_{t \ge j} \frac{F_{in}(X^n_{i(t)}) - F_{in}(X^n_{i(s)})}{X^n_{i(t)} - X^n_{i(s)}}, \tag{3.27}$$

when $X^n_{i(0)} \equiv 0$, and define

$$m_{in}(x) = \begin{cases} 0 & \text{for } x \le 0; \\ \beta_{ij} & \text{for } X^n_{i(j-1)} < x \le X^n_{i(j)}; \\ 0 & \text{for } x > X^n_{i(n)}. \end{cases} \tag{3.28}$$

From (3.27) and (3.28), one can see that the estimator $m_{in}(x)$ is decreasing in x. Thus, the empirical Bayes procedures $\underset{\sim}{\psi}_n$ defined by (3.24 $\sim$ 3.28) is a monotone procedure. It is known that both estimators $M_{in}(x)$ and $m_{in}(x)$ have strong consistency property. Hence, $\Delta_{in}(x)$ is a strongly consistent estimator of $\Delta_{iG}(x)$. Then by Theorem 2.1 of Gupta and Hsiao (1983), the sequence of empirical Bayes procedures $\{\underset{\sim}{\psi}_n\}$ is asymptotically optimal provided $\int_0^\infty \theta dG_i(\theta) < \infty$ for each $i = 1, \ldots, k$.

3.4. Remark on the Monotonicity of Empirical Bayes Selection Procedures

The monotonicity of selection procedures is an important property in many selection problems. Under some regularity conditions, Miescke (1979) showed that every Bayes procedure is monotone. Hence, the class of all monotone selection procedures form an essentially complete class among the class of all selection procedures. In other words, a non-monotone selection procedure is always inadmissible in terms of the Bayes risk.

Generally, the monotonicity of a Bayes selection procedure is due to the monotonicity of the posterior expectation of loss functions (or functions related to loss functions).

Therefore, in the empirical Bayes selection problems, one of the most important things is to construct monotone estimators for each related monotone function.

The techniques to construct monotone empirical Bayes estimators have been studied by van Houwelingen (1976, 1977) for continuous one-parameter exponential family and also for a class of discrete distributions with monotone likelihood ratio property. Stijnen (1982, 1985) and van Houwelingen and Stijnen (1983) have studied the same problem for the continuous one-parameter exponential family. Those techniques can (only) be applied to selection problems with underlying distributions being in the above mentioned family of distributions. Further studies are needed to investigate the asymptotic behavior for each related empirical Bayes selection procedure. The present authors are planning to do some work along these lines.

References

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16–39.

Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Berger, J. and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *Ann. Statist.* **14**, 461–486.

Berger, J. and Deely, J. J. (1986). A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology. Technical Report #86-8, Department of Statistics, Purdue University, West Lafayette, Indiana.

Bickel, P. J. and Yahav, J. A. (1977). On selecting a subset of good populations. *Statistical Decision Theory and Related Topics–II* (Eds. S. S. Gupta and D. S. Moore), Academic, New York, 37–55.

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

Box, G. E. P. and Tiao, G. C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, Massachusetts.

Chernoff, H. and Yahav, J. A. (1977). A subset selection problem employing a new criterion. *Statistical Decision Theory and Related Topics–II* (Eds. S. S. Gupta and D. S. Moore), Academic, New York, 93–119.

Deely, J. J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo. Ser. No. 45), Department of Statistics, Purdue University, West Lafayette, Indiana.

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

Goel, P. K. and Rubin, H. (1977). On selecting a subset containing the best population – A Bayesian approach. *Ann. Statist.* **5**, 969–983.

Grenander, U. (1956). On the theory of mortality measurement. Part II. *Skand. Akt.* **39**, 125–153.

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

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

Gupta, S. S. and Hsiao, P. (1981). On Γ-minimax, minimax, and Bayes procedures for selecting populations close to a control. *Sankhyā B* **43**, 291–318.

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plan. Infer.* **8**, 87–101.

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

Gupta, S. S. and Huang, D. Y. (1977). On some Γ-minimax selection and multiple comparison procedures. *Statistical Decision Theory and Related Topics – II* (Eds. S. S. Gupta and D. S. Moore), 139–155.

Gupta, S. S. and Kim, W. C. (1980). Γ-minimax and minimax rules for comparison of treatments with a control. *Recent Developments in Statistical Inference and Data Analysis* (Ed. K. Matusita), North Holland Publishing Co., Amsterdam, The Netherlands, 55–71.

Gupta, S. S. and Leu, L. Y. (1983). On Bayes and empirical Bayes rules for selecting good populations. Technical Report #83–37, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Liang, T. (1984). Empirical Bayes rules for selecting good binomial populations. To appear in *The Proceedings of the Symposium on Adaptive Statistical Procedures and Related Topics*.

Gupta, S. S. and Liang, T. (1986). Empirical Bayes rules for selecting the best binomial population. Technical Report #86-13, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Miescke, K. (1984). On two-stage Bayes selection procedures. *Sankhyā B46*, 123–134.

Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures. Wiley, New York.

Gupta, S. S. and Yang, H. M. (1985). Bayes-P* subset selection procedures for the best population. *J. Statist. Plan. Infer.* **12**, 213–233.

Guttman, I. and Tiao, G. C. (1964). A Bayesian approach to some best population problems. *Ann. Math. Statist.* **35**, 825–835.

Huang, D. Y. and Tseng, S. T. (1983). Γ-optimal decision procedures for selecting the best population in randomized complete block design. *Commun. Statist.-Theor. Methods* **12**, 341–354.

Huang, W. T. (1975). Bayes approach to a problem of partitioning k normal populations. *Bull. Inst. Math. Acad. Sinica* **3**, 87–97.

Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10**, 479–484.

Miescke, K. (1979). Bayesian subset selection for additive and linear loss functions. *Commun. Statist.-Theor. Meth.* **A8(12)**, 1205–1226.

Miescke, K. (1981). Γ-minimax selection procedures in simultaneous testing problems. *Ann. Statist.* **9**, 215–219.

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

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Probab.* University of California Press, 155–163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1–19.

Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11**, 713–723.

Roth, A. J. (1978). A new procedure for selecting a subset containing the best normal population. *J. Amer. Statist. Assoc.* **73**, 613–617.

Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypothesis. *Ann. Math. Statist.* **34**, 1370–1385.

Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4**, 1200–1209.

Stijnen, Th. (1982). A monotone empirical Bayes estimator and test for the one-parameter continuous exponential family based on spacings. *Scand. J. Statist.* **9**, 153–158.

Stijnen, Th. (1985). On the asymptotic behaviour of empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **13**, 403–412.

van Houwelingen, J. C. (1976). Monotone empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **4** 981–989.

van Houwelingen, J. C. (1977). Monotonizing empirical Bayes estimators for a class of discrete distributions with monotone likelihood ratio. *Statist. Neerl.* **31**, 95–104.

van Houwelingen, J. C. and Stijner, Th. (1983). Monotone empirical Bayes estimators for the continuous one-parameter exponential family. *Statist. Neerl.* **37**, 29–43.

Van Ryzin, J. (1970). On some nonparametric empirical Bayes multiple decision problems. *Proc. First Internat. Symp. Nonparametric Techniques in Statistic Inference* (Ed. M. L. Puri), 585–603.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. *Ann. Statist.* **5**, 172–181.