ON EMPIRICAL BAYES SELECTION PROCEDURES

by

Shanti S. Gupta
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

and

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, MI 48202

Technical Report # 91-59C

Department of Statistics
Purdue University

October, 1991

*This research was supported in part by NSF Grants DMS–8923071 and DMS–8717799 at Purdue University.

△

ON EMPIRICAL BAYES SELECTION PROCEDURES

by

Shanti S. Gupta
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

and

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, MI 48202

Abstract

This paper surveys the empirical Bayes methodology for ranking and selection problems. Three empirical Bayes approaches are discussed. They are nonparametric empirical Bayes, parametric empirical Bayes and hierarchical empirical Bayes. For each of them, two kinds of empirical Bayes procedures are considered. One is to incorporate information from past data to improve the current decision. The other is to incorporate information from each other so as to simultaneously improve the decision for each of the component problems under study. Certain important models including Poisson, binomial and hypergeometric distributions are investigated. The empirical Bayes methodology is discussed through these examples.

AMS 1980 Subjective Classification: 62F07, 62C12.

Keywords and phrases: Ranking and selection, Bayes, Asymptotic optimality, Nonparametric empirical Bayes, Parametric empirical Bayes, Hierarchical empirical Bayes.

1. Introduction

A common statistical problem faced by an experimenter is one of comparing several populations (treatments). Suppose that there are $k(\geq 2)$ populations $\pi_1, \ldots, \pi_k$, and for each i, π_i is characterized by the value of a parameter of interest, say θ_i. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ denote the ordered values of the parameters $\theta_1, \ldots, \theta_k$. The population associated with $\theta_{[k]}$ is called the best population. For a given standard θ_0, a population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. In many practical situations, an experimenter may be interested in the selection of the best population and/or the selection of all good populations. These problems are known as selection and ranking problems. The formulation of selection and ranking procedures has been accomplished generally using the indifference zone approach (see Bechhofer (1954)) or the subset selection approach (see 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. 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. Some contributions to selection and ranking problems using Bayesian approach have been made by Deely and Gupta (1968, 1988), Bickel and Yahav (1977), Chernoff and Yahav (1977), Goel and Rubin (1977), Gupta and Hsu (1978), Miescke (1979), Gupta and Miescke (1984), Gupta and Yang (1985), and Berger and Deely (1988), among many others.

The empirical Bayes approach in statistical decision theory is typically 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 prior distribution on the parameter space. One then uses information borrowed from other sources to improve the decision procedure for each component. This approach is due to Robbins (1956, 1964). Empirical Bayes procedures have been derived for multiple decision problems by Deely (1965). Recently, Gupta and Hsiao (1983), and Gupta and Liang (1986, 1988, 1989a, b, 1991a, b, c) have investigated empirical Bayes procedures for several selection problems. Many such empirical Bayes procedures have been shown to be asymptotically optimal in the sense that the component Bayes risk will converge to the optimal Bayes risk which would have been obtained if the prior distribution were fully known, and the Bayes procedure with respect to this prior distribution was used.

The present paper is concerned with the selection and ranking problem using the empirical Bayes approach. Two kinds of empirical Bayes procedures will be considered. One is to incorporate information from accumulated past data to improve the current decision. The other is to incorporate information from each other so as to simultaneously improve the decision for each of the component problems under study. The paper is organized in the following way. We briefly introduce the Bayes and empirical Bayes selection problems in Section 2. Through Sections 3–5, we consider certain important selection problems including Poisson, binomial and hypergeometric distributions. The empirical Bayes methodology is discussed through these examples. Certain simulation results are provided to show the small sample performance of the related empirical Bayes procedures.

2. Formulation of Bayes and Empirical Bayes Selection Problems

2.1 Bayes Selection Problems and Procedures

Let $\theta_i \in \Theta \subset \mathbb{R}$ denote the unknown characteristic of interest associated with the population $\pi_i, i = 1, \ldots, k$. Let $X_1, \ldots, X_k$ be random variables representing the k populations $\pi_1, \ldots, \pi_k$, respectively, with X_i having the probability density function $f_i(x|\theta_i)$. 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 $f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{k} f_i(x_i|\theta_i)$, where $\underset{\sim}{x} = (x_1, \ldots, x_k)$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of θ_i's. The population associated with $\theta_{[k]}$ is called a best population. For a given standard θ_0, a population π_i is said to be good if $\theta_i \geq \theta_0$ and bad otherwise. Let $\Omega = \{\underset{\sim}{\theta}|\theta_i \in \Theta, i = 1, \ldots, k\}$ denote the parameter space. Also, it is assumed that the value of the parameter θ_i is a realization of a random variable Θ_i having a prior distribution G_i and $\Theta_1, \ldots, \Theta_k$ are mutually independent. Hence $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k)$ have a joint prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$ on the parameter $\underset{\sim}{\theta}$ over the parameter space Ω.

In many situations, an experimenter is interested in identifying the best population or selecting the more promising subset of the k populations for further experimentation.

For a specified selection goal, an action is a subset of the set $\{1,\ldots,k\}$. When action $S \subset \{1,\ldots,k\}$ is taken, it means that population π_i is included in the selected subset if $i \in S$. Let $\mathcal{A}$ denote the action space. For each $\underset{\sim}{\theta} \in \Omega$ and $S \in \mathcal{A}$, 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] such that $\sum_{S\in\mathcal{A}} d(\underset{\sim}{x}, S) = 1$ for all $\underset{\sim}{x} \in \mathcal{X}$, where $\mathcal{X}$ is the sample space of $\underset{\sim}{X}$. $d(\underset{\sim}{x}, S)$ can be viewed as the probability of taking action S when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed.

Let $\mathcal{D}$ be the class of all decision procedures. For each $d \in \mathcal{D}$, let $r(\underset{\sim}{G}, d)$ denote the associated Bayes risk. Then, $r(\underset{\sim}{G}) \equiv \inf_{d\in\mathcal{D}} r(\underset{\sim}{G}, d)$ is the minimum Bayes risk. An optimal decision procedure, denoted by $d_{\underset{\sim}{G}}$, is obtained if $d_{\underset{\sim}{G}}$ has the property that $r(\underset{\sim}{G}, d_{\underset{\sim}{G}}) = r(\underset{\sim}{G})$. Such a procedure is called Bayes with respect to $\underset{\sim}{G}$. Under some regularity conditions,

$$r(\underset{\sim}{G}, d) = \int_{\mathcal{X}} \sum_{S\in\mathcal{A}} d(\underset{\sim}{x}, S) \left[\int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}) \right] \underset{\sim}{f}(\underset{\sim}{x}) d\underset{\sim}{x}$$

where $\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x})$ is the joint posterior distribution of $\underset{\sim}{\theta}$ given $\underset{\sim}{X} = \underset{\sim}{x}$, $\underset{\sim}{f}(\underset{\sim}{x}) = \overset{k}{\underset{i=1}{\pi}} f_i(x_i)$ and $f_i(x_i) = \int_{\Theta} f_i(x_i|\theta_i) dG_i(\theta_i)$ is the marginal probability density function of X_i.

For each fixed $\underset{\sim}{x} \in \mathcal{X}$, let

$$\begin{aligned} \Delta_{\underset{\sim}{G}}(\underset{\sim}{x}, S) &= \int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}), \\ A(\underset{\sim}{x}) &= \{S \in \mathcal{A} | \Delta_{\underset{\sim}{G}}(\underset{\sim}{x}, S) = \min_{S'\in\mathcal{A}} \Delta_{\underset{\sim}{G}}(\underset{\sim}{x}, S')\}. \end{aligned}$$

Then, the Bayes decision procedure $d_{\underset{\sim}{G}}$ clearly satisfies that $\sum_{S\in A(\underset{\sim}{x})} d_{\underset{\sim}{G}}(\underset{\sim}{x}, S) = 1$.

It should be noted that the Bayes decision procedures vary for different selection problems and goals, and depend on the loss function chosen. Also, the Bayes decision procedure is very sensitive to the prior distribution which is obtained through quantifying prior information into a single prior distribution.

2.2 Empirical Bayes Selection Procedures

In this subsection, we continue with the general setup of the early subsection. However, we assume only the existence of a prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \overset{k}{\underset{i=1}{\pi}} G_i(\theta_i)$ on $\underset{\sim}{\theta}$ over Ω; the form of the prior distributions $G_i, i = 1, \ldots, k$, are either unknown or partially known.

We use the empirical Bayes approach. Two kinds of empirical Bayes procedures will be considered. One is to incorporate information from the accumulated past data to improve the current decision. The other is to incorporate information from each other so as to simultaneously improve the decision for each of the component decision problems.

Incorporating Information from Past Observations

According to the usual empirical Bayes framework, for each $i = 1, \ldots, k$, let X_{ij} denote the random observation taken from π_i at stage j. Let Θ_{ij} denote the random characteristic of π_i at stage j. Given $\Theta_{ij} = \theta_{ij}, X_{ij}$ has the conditional probability density function $f_i(x|\theta_{ij})$. 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 $\underset{\sim}{\Theta}_j, j = 1, \ldots, n$, are mutually independent and have the same prior distribution $\underset{\sim}{G}$, though $\underset{\sim}{\Theta}_j$ are not observable. Also, let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the present random observation.

Consider an empirical Bayes decision procedure $d_n((\underset{\sim}{x}; \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n), S) \equiv d_n(\underset{\sim}{x}, S)$, which is a function of the present observation $\underset{\sim}{x}$ and the past random observations $\underset{\sim}{X}_1, \ldots,$ $\underset{\sim}{X}_n$. Let $r(\underset{\sim}{G}, d_n)$ be the conditional Bayes risk associated with the empirical Bayes procedure d_n, conditional on the past observations $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. That is,

$$r(\underset{\sim}{G}, d_n) = \int_{\mathcal{X}} \sum_{S \in \mathcal{A}} d_n(\underset{\sim}{x}, S) \int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}) \underset{\sim}{f}(\underset{\sim}{x}) d\underset{\sim}{x}.$$

Also, let $E[r(\underset{\sim}{G}, d_n)]$ be the overall Bayes risk of the empirical Bayes procedure d_n. That is,

$$E[r(\underset{\sim}{G}, d_n)] = \int_{\mathcal{X}} \sum_{S \in \mathcal{A}} E[d_n(\underset{\sim}{x}, S)] \int_{\Omega} L(\underset{\sim}{\theta}, S) d\underset{\sim}{G}(\underset{\sim}{\theta}|\underset{\sim}{x}) \underset{\sim}{f}(\underset{\sim}{x}) d\underset{\sim}{x},$$

where the expectation E is taken with respect to $(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)$. Note that $r(\underset{\sim}{G}, d_n) - r(\underset{\sim}{G}) \geq 0$ since $r(\underset{\sim}{G})$ is the minimum Bayes risk among the class of all decision procedures $\mathcal{D}$. Hence $E[r(\underset{\sim}{G}, d_n)] - r(\underset{\sim}{G}) \geq 0$. Either of the two non-negative difference can be used as a measure of optimality of the empirical Bayes procedure d_n. A sequence of empirical Bayes procedures $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution $\underset{\sim}{G}$ if $E[r(\underset{\sim}{G}, d_n)] - r(\underset{\sim}{G}) \to 0$ as $n \to \infty$. The problem concerned here is to construct empirical Bayes procedures possessing the desired asymptotic optimality. Gupta and Liang (1986, 1988, 1989a, b) have investigated several empirical Bayes procedures for certain selection problems under this empirical Bayes framework.

Incorporating Information from Other Components

We now consider the case where it is assumed that the k prior distributions $G_1, \ldots, G_k$ are identical, but there is no past observation available. Under this assumption, the empirical Bayes idea can still be employed. We may incorporate information from each of the k populations to make an appropriate decision for the concerned selection problem. Let d_k be a decision procedure constructed under such consideration (the detailed methods will be discussed later through some examples), and let $r(\underset{\sim}{G}, d_k)$ denote the corresponding Bayes risk. Since $r(\underset{\sim}{G})$ is the minimum Bayes risk, $r(\underset{\sim}{G}, d_k) - r(\underset{\sim}{G}) \geq 0$. An empirical Bayes procedure d_k is said to be asymptotically optimal if $r(\underset{\sim}{G}, d_k) - r(\underset{\sim}{G}) \to 0$ as $k \to \infty$. One may desire to construct empirical Bayes procedures having such asymptotic optimality. Gupta and Liang (1991b, c) have studied several empirical Bayes selection problems using this empirical Bayes approach.

Approaches for Constructing Empirical Bayes Procedures

There are three main approaches for constructing empirical Bayes procedures, according to how much we know about the prior distribution $\underset{\sim}{G}$, namely, nonparametric empirical Bayes, parametric empirical Bayes and hierarchical empirical Bayes, respectively.

For the nonparametric empirical Bayes approach, one assumes that the form of the prior distribution $\underset{\sim}{G}$ is completely unknown. In this situation, one may either use the information obtained from other sources (may be either from the past data or from the other components) to estimate the prior distribution $\underset{\sim}{G}$, then do a Bayesian analysis based on the estimated prior or represent the Bayes procedure in terms of the unknown prior, and then use the data to estimate the behavior of the Bayes decision procedure directly. Gupta and Hsiao (1983) and Gupta and Liang (1986, 1988, 1991a, b) have studied some selection problems using the nonparametric empirical Bayes approach.

For the parametric empirical Bayes approach, it is assumed that the prior distribution $\underset{\sim}{G}$ is a member of some parameter family Γ and is indexed by some unknown parameter(s), say λ. Hence the prior distribution is denoted by $\underset{\sim}{G}_\lambda$. Suppose now an estimate $\hat{\lambda}$ depending on the data can be found and we denote the prior distribution associated with $\hat{\lambda}$ by $\underset{\sim}{G}_{\hat{\lambda}}$. Note that $\underset{\sim}{G}_{\hat{\lambda}}$ is also a member of the family Γ. We use $\underset{\sim}{G}_{\hat{\lambda}}$ to estimate the unknown prior $\underset{\sim}{G}_\lambda$. We then follow the usual Bayesian analysis and derive the Bayes procedure $d_{\underset{\sim}{G}_{\hat{\lambda}}}$

with respect to the estimated prior distribution $G_{\hat{\lambda}}$. Using this line of parametric empirical Bayes approach, Gupta and Liang (1989a, b) have studied empirical Bayes selection procedure for selecting the most probable event in a multinomial distribution and for selecting the best population from among k binomial populations.

For the hierarchical empirical Bayes approach, it is assumed that the prior distribution of component i belongs to some parameter family Γ and is indexed by a parameter (or parameters) λ_i and the λ_i's are assumed to be iid, follow a hierarchical prior distribution. This hierarchical prior distribution may be either known or indexed by an unknown parameter (or parameters). In the latter case, the unknown parameter(s) should be estimated. One then follows a hierarchical Bayesian analysis. A decision procedure derived through this framework is called a hierarchical empirical Bayes procedure.

In the following sections, the empirical Bayes methods will be more detailedly discussed through some examples.

3. Selecting Good Poisson Populations

3.1 Formulation of the Selection Problem

Let $\pi_1, \ldots, \pi_k$ denote k independent populations. For each $i = 1, \ldots, k$, let X_i denote a random observation arising from population π_i, having a Poisson distribution with probability function $f_i(x|\theta_i)$ where

$$f_i(x|\theta_i) = e^{-\theta_i}\theta_i^x/x!,\ x = 0, 1, 2, \ldots; \theta_i > 0.$$

Let $\theta_0 > 0$ be a known standard. Population π_i is said to be good if $\theta_i \geq \theta_0$ and bad otherwise. The goal is to select all good populations and exclude all bad populations.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)|\theta_i > 0, i = 1, \ldots, k\}$ be the parameter space and let $\mathcal{A} = \{\underset{\sim}{a} = (a_1, \ldots, a_k)|a_i = 0, 1; i = 1, \ldots, k\}$ be the action space. When action $\underset{\sim}{a}$ is taken, it means that population π_i is selected as a good population if $a_1 = 1$, and excluded as a bad one if $a_i = 0$. For each $\underset{\sim}{\theta} \in \Omega$, and $\underset{\sim}{a} \in \mathcal{A}$, the loss function $L(\underset{\sim}{\theta}, \underset{\sim}{a})$ is defined to be:

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \sum_{i=1}^{k} a_i(\theta_0 - \theta_i)I(\theta_0 - \theta_i) + \sum_{i=1}^{k}(1 - a_i)(\theta_i - \theta_0)I(\theta_i - \theta_0) \qquad (3.1)$$

where $I(x) = 1(0)$ if $x \geq (<) 0$.

It is assumed that for each i, the parameter θ_i is a realization of a random variable Θ_i which has a prior distribution G_i. It is also assumed that $\Theta_1, \ldots, \Theta_k$ are mutually independent so that $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k)$ has a joint prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$.

Let $\mathcal{X}$ denote the sample space of $\underset{\sim}{X} = (X_1, \ldots, X_k)$. A selection rule $\underset{\sim}{d} = (d_1, \ldots, d_k)$ is defined to be a mapping from $\mathcal{X}$ into $[0,1]^k$, such that $d_i(\underset{\sim}{x})$ is the probability of selecting population π_i as a good population when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. Let $\mathcal{D}$ be the class of all selection rules, and for each $\underset{\sim}{d} \in \mathcal{D}$, let $r(\underset{\sim}{G}, \underset{\sim}{d})$ denote the associated Bayes risk. Then $r(\underset{\sim}{G}) \equiv \inf_{\underset{\sim}{d} \in \mathcal{D}} r(\underset{\sim}{G}, \underset{\sim}{d})$ is the minimum Bayes risk. One can see that for each $\underset{\sim}{d} \in \mathcal{D}$,

$$r(\underset{\sim}{G}, \underset{\sim}{d}) = \sum_{i=1}^{k} r_i(\underset{\sim}{G}, d_i) \tag{3.2}$$

where

$$r_i(\underset{\sim}{G}, d_i) = \sum_{\underset{\sim}{x} \in \mathcal{X}} [\theta_0 - \varphi_i(x_i)] d_i(\underset{\sim}{x}) \prod_{j=1}^{k} f_j(x_j) + C_i, \tag{3.3}$$

where $\varphi_i(x_i) = E[\Theta_i | X_i = x_i] = h_i(x_i + 1)/h_i(x_i)$ is the posterior mean of Θ_i given $X_i = x_i, h_i(x_i) = f_i(x_i)/a(x_i)$, $f_i(x_i) = \int_0^\infty f_i(x_i|\theta) dG_i(\theta) = \int_0^\infty e^{-\theta}\theta^{x_i}/x_i! dG_i(\theta) = a(x_i)h_i(x_i)$ is the marginal probability function of the random variable X_i, and $a(x_i) = (x_i!)^{-1}, h_i(x_i) = \int_0^\infty e^{-\theta}\theta^{x_i} dG_i(\theta)$ and $C_i = \int_{\theta_0}^\infty (\theta - \theta_0) dG_i(\theta)$.

It follows that a Bayes rule, say $\underset{\sim}{d}_{\underset{\sim}{G}} = (d_{\underset{\sim}{G}1}, \ldots, d_{\underset{\sim}{G}k})$, is clearly given by: for each $i = 1, \ldots, k$,

$$d_{\underset{\sim}{G}_i}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_i(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The minimum Bayes risk is: $r(\underset{\sim}{G}) = r(\underset{\sim}{G}, \underset{\sim}{d}_{\underset{\sim}{G}}) = \sum_{i=1}^{k} r_i(\underset{\sim}{G}, d_{\underset{\sim}{G}_i})$.

When the prior distribution $\underset{\sim}{G}$ is unknown, it is not possible to apply the Bayes rule $\underset{\sim}{d}_{\underset{\sim}{G}}$ for the selection problem. In the following, the empirical Bayes approach is employed.

3.2 Incorporating Information from Past Observations

According to the usual empirical Bayes framework, it is assumed that for each $i = 1, \ldots, k$, there are marginally iid past random observations $X_{i1}, \ldots, X_{in}$ with marginal probability function $f_i(x)$ available when a decision is made. Three empirical Bayes selection rules are constructed according to how much we know about the prior distribution $\underset{\sim}{G}$.

3.2.1. A Nonparametric Empirical Bayes Rule

It is assumed that the prior distribution $\underset{\sim}{G}$ is completely unknown. Thus, a nonparametric empirical Bayes approach is employed. It should be noted that $\varphi_i(x_i)$ is increasing in x_i for each $i = 1, \ldots, k$. Therefore the Bayes rule $\underset{\sim}{d}_{\underset{\sim}{G}}$ is a monotone selection rule. Thus, it is desirable that the considered empirical Bayes rule be also monotone.

For each $i = 1, \ldots, k$, and $x = 0, 1, 2, \ldots$, define

$$f_{in}(x) = n^{-1} \sum_{j=1}^{n} I_{\{x\}}(X_{ij})$$
$$h_{in}(x) = f_{in}(x)/a(x).$$

Let $N_{in} = \max_{1 \le j \le n} X_{ij} - 1$ and for each $x = 0, 1, \ldots, N_{in}$, define

$$\varphi_{in}(x) = [h_{in}(x+1) + \delta_n]/[h_{in}(x) + \delta_n],$$

where $\delta_n > 0$ is such that $\delta_n = o(1)$.

Since $\varphi_{in}(x)$ may not be increasing in x, a smoothed version of $\varphi_{in}(x)$ is given below. Let $\{\varphi^*_{in}(x)\}_{x=0}^{N_{in}}$ be the isotonic regression of $\{\varphi_{in}(x)\}_{x=0}^{N_{in}}$ with random weights $\{W_{in}(x)\}_{x=0}^{N_{in}}$, where $W_{in}(x) = [h_{in}(x) + \delta_n]a(x+1)$. For $y > N_{in}$, let $\varphi^*_{in}(y) = \varphi^*_{in}(N_{in})$. Therefore, $\varphi^*_{in}(x)$ is nondecreasing in x. We may use $\varphi^*_{in}(x)$ to estimate $\varphi_i(x)$. Based on $\varphi^*_{in}(x)$, $i = 1, \ldots, k$, an empirical Bayes rule $\underset{\sim}{d}^*_n = (d^*_{1n}, \ldots, d^*_{kn})$ is proposed as follows: For each $i = 1, \ldots, k$, and $\underset{\sim}{x} \in \mathcal{X}$,

$$d^*_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi^*_{in}(x_i) \ge \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

3.2.2. A Parametric Empirical Bayes Rule

It is assumed that the prior distribution G_i is a gamma distribution with unknown shape and scale parameters α_i and β_i, respectively, $i = 1, \ldots, k$. That is, G_i has a density function $g_i(\theta|\alpha_i, \beta_i)$, where

$$g_i(\theta|\alpha_i, \beta_i) = \beta_i^{\alpha_i} \theta^{\alpha_i - 1} e^{-\beta_i \theta}/\Gamma(\alpha_i), \ \theta > 0.$$

Then, $X_{i1},\ldots,X_{in}$ are iid with marginal probability function $f_i(x) = \Gamma(x+\alpha_i)\beta_i^{\alpha_i}/[\Gamma(\alpha_i)(1+\beta_i)^{x+\alpha_i}x!]$, $x = 0,1,2,\ldots$. Also, $\varphi_i(x) = \frac{x+\alpha_i}{1+\beta_i}$. Straightforward computations yield that $\mu_{i1} \equiv E[X_{i1}] = \alpha_i/\beta_i$, $\mu_{i2} \equiv E[X_{i1}^2] = \alpha_i(\alpha_i+1)\beta_i^{-2} + \alpha_i\beta_i^{-1}$. Thus, $\beta_i = \mu_{i1}(\mu_{i2} - \mu_{i1} - \mu_{i1}^2)^{-1}$ and $\alpha_i = \mu_{i1}^2(\mu_{i2} - \mu_{i1} - \mu_{i1}^2)^{-1}$. Therefore, $\varphi_i(x) = [x(\mu_{i2} - \mu_{i1} - \mu_{i1}^2) + \mu_{i1}^2](\mu_{i2} - \mu_{i1}^2)^{-1}$.

For each $i = 1,\ldots,k$, let $\mu_{i1n} = n^{-1}\sum_{j=1}^{n} X_{ij}$ and $\mu_{i2n} = n^{-1}\sum_{j=1}^{n} X_{ij}^2$. That is, μ_{i1n} and μ_{i2n} are moment estimators of μ_{i1} and μ_{i2}, respectively. Since it is possible that $\mu_{i2n} - \mu_{i1n} - \mu_{i1n}^2 \equiv \gamma_{in} \le 0$ though $\mu_{i2} - \mu_{i1} - \mu_{i1}^2 > 0$, thus, for each $x = 0,1,\ldots$, define

$$\hat{\varphi}_{in}(x) = \begin{cases} \frac{x\gamma_{in}+\mu_{i1n}^2}{\mu_{i2n}-\mu_{i1n}^2} & \text{if } \gamma_{in} > 0, \\ x & \text{otherwise.} \end{cases} \tag{3.6}$$

Then, an empirical Bayes rule $\underset{\sim}{\hat{d}}_n = (\hat{d}_{1n},\ldots,\hat{d}_{kn})$ is proposed as follows: For each $i = 1,\ldots,k$, and $\underset{\sim}{x} \in \mathcal{X}$,

$$\hat{d}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \hat{\varphi}_{in}(x_i) \ge \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

3.2.3. A Hierarchical Empirical Bayes Rule

Suppose that the prior distribution G_i is a gamma distribution with a known shape parameter α_i and an unknown scale parameter β_i. In this situation, the preceding parametric empirical Bayes approach can be applied here. However, an alternative method, called hierarchical empirical Bayes, is introduced in the following.

Since β_i is a scale parameter, it is assumed that β_i has an improper prior $p(\beta_i) = \frac{1}{\beta_i}$, $\beta_i > 0$. Thus, conditional on β_i, $X_{i1},\ldots,X_{in}$ are iid with the probability function $f_i(x|\beta_i) = \int_0^\infty f_i(x|\theta)g_i(\theta|\alpha_i,\beta_i)d\theta = \frac{\Gamma(x+\alpha_i)\beta_i^{\alpha_i}}{x!\Gamma(\alpha_i)(1+\beta_i)^{x+\alpha_i}}$, $x = 0,1,2,\ldots$. Therefore, $X_{i1},\ldots,X_{in}$ has a joint marginal probability function $\underset{\sim}{f}_i(x_{i1},\ldots,x_{in})$, where

$$\begin{aligned} \underset{\sim}{f}_i(x_{i1},\ldots,x_{in}) &= \int_0^\infty \prod_{j=1}^{n} f_i(x_{ij}|\beta)p(\beta)d\beta \\ &= \prod_{j=1}^{n}\left[\frac{\Gamma(x_{ij}+\alpha_i)}{x_{ij}!\Gamma(\alpha_i)}\right]\Gamma(n\alpha_i)\Gamma(b_i - n\alpha_i)/\Gamma(b_i) \end{aligned}$$

where $b_i = n\alpha_i + \sum_{j=1}^{n} x_{ij}$. Thus, the posterior density function of β_i given $(X_{i1}, \ldots, X_{in}) = (x_{i1}, \ldots, x_{in})$ is

$$p(\beta_i | x_{i1}, \ldots, x_{in}) = \beta_i^{n\alpha_i - 1}(1 + \beta_i)^{-b_i}\Gamma(b_i)/[\Gamma(n\alpha_i)\Gamma(b_i - n\alpha_i)],$$

and the posterior mean of β_i given $(x_{i1}, \ldots, x_{in})$ is

$$\beta_{in} \equiv E[\beta_i | x_{i1}, \ldots, x_{in}] = \begin{cases} \frac{n\alpha_i}{\sum_{j=1}^{n} x_{ij} - 1} & \text{if } \sum_{j=1}^{n} x_{ij} \geq 2, \\ \infty & \text{otherwise.} \end{cases}$$

For each $i = 1, \ldots, k$, and $(x_{i1}, \ldots, x_{in})$, define

$$\overline{\varphi}_{in}(x_i) = \begin{cases} \frac{x_i + \alpha_i}{1 + \beta_{in}} & \text{if } \sum_{j=1}^{n} x_{ij} \geq 2; \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

We then propose an empirical Bayes rule $\underset{\sim}{\overline{d}}_n = (\overline{d}_{1n}, \ldots, \overline{d}_{kn})$ as follows: For each $i = 1, \ldots, k$,

$$\overline{d}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{\varphi}_{in}(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

3.2.4. Asymptotic Optimality

For an empirical Bayes selection rule $\underset{\sim}{d}_n$, let $E[r(\underset{\sim}{G}, \underset{\sim}{d}_n)]$ denote the overall Bayes risk. That is,

$$E[r(\underset{\sim}{G}, \underset{\sim}{d}_n)] = \sum_{i=1}^{k} \left[\sum_{x_i=0}^{\infty} [\theta_0 - \varphi_i(x_i)] E_{in}[d_{in}(x_i)] f_i(x_i) + C_i \right]$$

where the expectation E_{in} is taken with respect to $(X_{i1}, \ldots, X_{in})$. Since $r(\underset{\sim}{G})$ is the minimal Bayes risk $E[r(\underset{\sim}{G}, \underset{\sim}{d}_n)] - r(\underset{\sim}{G}) \geq 0$ for all n.

Following Gupta and Liang (1991b), it is easy to obtain the following result. Let $B_i(\theta_0) = \{x | \varphi_i(x) < \theta_0\}$ and let

$$m_i = \begin{cases} \sup B_i(\theta_0) & \text{if } B_i(\theta_0) \neq \phi; \\ -1 & \text{otherwise.} \end{cases}$$

Theorem 3.1. Let $\underset{\sim}{d}_n$ denote any of the three precedingly constructed empirical Bayes selection rules $\underset{\sim}{d}_n^*, \hat{\underset{\sim}{d}}_n$ and $\underset{\sim}{\overline{d}}_n$. Suppose that $\int_0^\infty \theta dG_i(\theta) < \infty$ and $m_i < \infty$ for all $i =$

$1, \ldots, k$. Then, $E[r(\underset{\sim}{G}, \underset{\sim}{d}_n)] - r(\underset{\sim}{G}) = O(\exp(-cn))$ for some positive constant c, where the value of c varies depending on the empirical Bayes selection rule used.

3.3. Incorporating Information from Other Components

We now consider the case where it is assumed that the k prior distributions $G_1, \ldots, G_k$ are identical, but there is no past observations available. Under this assumption, $X_1, \ldots$, X_k are marginally iid with probability function $f(x) = \int_0^\infty e^{-\theta}\theta^x/x! dG(\theta)$ where $G = G_1 = \ldots = G_k$. Therefore, we can still incorporate information from each other to improve the decisions for each of the k component decision problems. The idea is described again through the nonparametric empirical Bayes, the parametric empirical Bayes and the hierarchical empirical Bayes approaches.

3.3.1. A Nonparametric Empirical Bayes Rule

It is assumed that the prior distribution G is completely unknown. Following the discussion of Subsection 3.2.1, a nonparametric empirical Bayes selection rule is constructed as follows.

For each $i = 1, \ldots, k$, let $N_{ik} = \max\limits_{j \neq i} X_j - 1$, and let $f_{ik}(y) = \frac{1}{k-1} \sum\limits_{\substack{j=1 \\ j \neq i}}^{k} I_{\{y\}}(X_j)$, $h_{ik}(y) = f_{ik}(y)/a(y)$, $y = 0, 1, \ldots$. Also, let $\varphi_{ik}(y) = [h_{ik}(y+1) + \delta_k]/[h_{ik}(y) + \delta_k]$ for each $y = 0, 1, \ldots, N_{ik}$, where $\delta_k > 0$ is such that $\delta_k = o(1)$.

Let $\{\varphi^*_{ik}(y)\}_{y=0}^{N_{ik}}$ be the isotonic regression of $\{\varphi_{ik}(y)\}_{y=0}^{N_{ik}}$ with random weights $\{W_{ik}(y)\}_{y=0}^{N_{ik}}$, where $W_{ik}(y) = [h_{ik}(y) + \delta_k]a(y+1)$. For $y > N_{ik}$, let $\varphi^*_{ik}(y) = \varphi^*_{ik}(N_{ik})$. Now, an empirical Bayes rule $\underset{\sim}{d}^*_k = (d^*_{1k}, \ldots, d^*_{kk})$ is proposed as follows: For each $i = 1, \ldots, k$, and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, define

$$d^*_{ik}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi^*_{ik}(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

3.3.2. A Parametric Empirical Bayes Rule

It is assumed that the prior distribution G is a member of gamma distribution family with probability density function $g(\theta|\alpha, \beta)$, where

$$g(\theta|\alpha, \beta) = \beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}/\Gamma(\alpha), \ \theta > 0$$

and both the parameters α and β are unknown. Following the discussion of Subsection 3.2.2, for each $i = 1, \ldots, k$, let $\mu_{1k}(i) = \frac{1}{k-1}\sum_{\substack{j=1 \\ j\neq i}}^{k} X_j$, and $\mu_{2k}(i) = \frac{1}{k-1}\sum_{\substack{j=1 \\ j\neq i}}^{k} X_j^2$. Let $\tau_{ik} = \mu_{2k}(i) - \mu_{1k}(i) - \mu_{1k}^2(i)$. Define

$$\hat{\varphi}_{ik}(x_i) = \begin{cases} \frac{x_i \tau_{ik} + \mu_{1k}^2(i)}{\mu_{2k}(i) - \mu_{1k}^2(i)} & \text{if } \tau_{ik} > 0; \\ x_i & \text{otherwise.} \end{cases} \tag{3.11}$$

An empirical Bayes rule $\underset{\sim}{\hat{d}}_k = (\hat{d}_{1k}, \ldots, \hat{d}_{kk})$ is proposed as follows: For each $i = 1, \ldots, k$ and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, define

$$\hat{d}_{ik}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \hat{\varphi}_{ik}(x_i) \geq \theta_0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

3.3.3. A Hierarchical Empirical Bayes Rule

It is assumed that the prior distribution G is a gamma distribution with a known shape parameter α and an unknown scale parameter β. Similar to that of Subsection 3.2.3, a hierarchical empirical Bayes rule $\underset{\sim}{\overline{d}}_k = (\overline{d}_{1k}, \ldots, \overline{d}_{kk})$ is constructed as follows.

For given $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, let

$$\beta_k = \begin{cases} k\alpha \Big/ \left(\sum_{j=1}^{k} x_j - 1 \right) & \text{if } \sum_{j=1}^{k} x_j \geq 2; \\ \infty & \text{otherwise.} \end{cases}$$

For each $i = 1, \ldots, k$, and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$, define

$$\overline{\varphi}_{ik}(x_i) = \begin{cases} (x_i + \alpha)/(1 + \beta_k) & \text{if } \sum_{j=1}^{k} x_j \geq 2; \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

Define, for each $i = 1, \ldots, k$, and $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$,

$$\overline{d}_{ik}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{\varphi}_{ik}(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

3.3.4. Asymptotic Optimality

Let $\underset{\sim}{d}_k$ denote any of the three precedingly constructed empirical Bayes selection rules. The associated overall Bayes risk $r(\underset{\sim}{G}, \underset{\sim}{d}_k)$ is:

$$r(\underset{\sim}{G}, \underset{\sim}{d}_k) = \sum_{i=1}^{k} r_i(\underset{\sim}{G}, d_{ik}),$$

where

$$r_i(\underset{\sim}{G}, d_{ik}) = E_{ik}E_i[(\theta_0 - \varphi_i(X_i))d_{ik}(\underset{\sim}{X})] + C$$

where the expectation E_i is taken with respect to X_i and the expectation E_{ik} is taken with respect to $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$. Also, here $C = \int_{\theta_0}^{\infty}(\theta - \theta_0)dG(\theta)$.

Since $r(\underset{\sim}{G})$ is the minimal Bayes risk, $r(\underset{\sim}{G}, \underset{\sim}{d}_k) - r(\underset{\sim}{G}) \geq 0$ for all k. For the present problem, Gupta and Liang (1991b) obtained the following strong asymptotic optimality.

Let $B(\theta_0) = \{x | \varphi(x) < \theta_0\}$ where $\varphi(x) = \varphi_1(x) = \ldots = \varphi_k(x)$ since $G_1 = \ldots = G_k$ and let

$$m = \begin{cases} \sup B(\theta_0) & \text{if } B(\theta_0) \neq \phi, \\ -1 & \text{otherwise.} \end{cases}$$

Theorem 3.2 Let $\underset{\sim}{d}_k$ denote any of the three precedingly constructed empirical Bayes selection rules $\underset{\sim}{d}_k^*$, $\hat{\underset{\sim}{d}}_k$ and $\overline{\underset{\sim}{d}}_k$. Suppose that $\int_0^{\infty} \theta dG(\theta) < \infty$ and $m < \infty$. Then, $r(\underset{\sim}{G}, \underset{\sim}{d}_k) - r(\underset{\sim}{G}) = O(\exp(-ck + \ln k))$ for some positive constant c, where the value of c varies depending on the empirical Bayes rule used.

3.4 Small Sample Performance: Simulation Study

In this study, we only consider the case where one may incorporate information from each other among the k populations. A Monte Carlo study were designed to investigate the performance of the three empirical Bayes procedures. We let the prior distribution G be a gamma distribution with $\alpha = 1$ and $\beta = 1$. With this specified prior distribution, the minimum Bayes risk for each of the k component decision problems is $r_i(\underset{\sim}{G}, d_{\underset{\sim}{G}_i}) = e^{-\theta_0} - 4^{-\theta_0}$, where θ_0 is the known standard. Therefore, the total minimum Bayes risk $r(\underset{\sim}{G}) = \sum_{i=1}^{k} r_i(\underset{\sim}{G}, d_{\underset{\sim}{G}_i}) = k(e^{-\theta_0} - 4^{-\theta_0})$.

Let $\underset{\sim}{d}_k = (d_{1k}, \ldots, d_{kk})$ be any of the three proposed empirical Bayes procedures. Since $r(\underset{\sim}{G}, \underset{\sim}{d}_k) - r(\underset{\sim}{G}) = \sum_{i=1}^{k} [r_i(\underset{\sim}{G}, d_{ik}) - r_i(\underset{\sim}{G}, d_{\underset{\sim}{G}_i})] = k(r_1(\underset{\sim}{G}, d_{1k}) - r_1(\underset{\sim}{G}, d_{\underset{\sim}{G}_1})]$, in the following, we have simulated the difference $r_1(G, d_{1k}) - r_1(\underset{\sim}{G}, d_{\underset{\sim}{G}_1})$ by $D(d_{1k}) \equiv R_1(\underset{\sim}{G}, d_{1k}) - r_1(\underset{\sim}{G}, d_{\underset{\sim}{G}_1})$, which is the difference between the conditional Bayes risk of d_{1k} conditional on $(X_2, \ldots, X_k)$ and the minimum Bayes risk. We have then use $k[R_1(\underset{\sim}{G}, d_{1k}) - r_1(\underset{\sim}{G}, d_{\underset{\sim}{G}_1})]$ as an estimator of the difference $r(\underset{\sim}{G}, d_{\underset{\sim}{G}}) - r(\underset{\sim}{G})$.

The simulation scheme is described as follows:

(1) For a fixed k, generate independent random values $X_1, \ldots, X_k$ according to the probability function $f(x)$.

(2) Based on the values $X_1, \ldots, X_k$, construct the empirical Bayes procedure d_{1k} and compute the conditional difference $D(d_{1k}) \equiv R_1(\underset{\sim}{G}, d_{1k}) - r_1(\underset{\sim}{G}, d_{\underset{\sim}{G}_1})$.

(3) The process was repeated 1000 times. The average of the differences based on the 1000 repetitions, which is denoted by $\overline{D}(d_{1k})$, is used as an estimator of the difference $r_1(\underset{\sim}{G}, d_{1k}) - r_1(\underset{\sim}{G}, d_{\underset{\sim}{G}_1})$. Then $k\overline{D}(d_{1k})$ is used as an estimate of the total difference $r(\underset{\sim}{G}, \underset{\sim}{d}_k) - r(\underset{\sim}{G})$.

Table 1 list some simulation results on the performance of the three empirical Bayes procedures $\underset{\sim}{d}_k^*, \hat{\underset{\sim}{d}}_k$ and $\bar{\underset{\sim}{d}}_k$ for the case where $\theta_0 = 1.5$. The notation $SE(\overline{D}(d_{1k}))$ is used to denote the estimated standard errors of the corresponding estimate $\overline{D}(d_{1k})$.

The simulation results indicate that for the empirical Bayes procedure $\bar{\underset{\sim}{d}}_k, k\overline{D}(\bar{d}_{1k})$ tends to zero very fast, and that $k\overline{D}(\bar{d}_{1k}) = 0$ for all $k \geq 100$. Also, for the empirical Bayes procedure $\hat{\underset{\sim}{d}}_k, k\overline{D}(\hat{d}_{1k})$ roughly increases in k for $k \leq 40$, then decreases in k and $k\overline{D}(\hat{d}_{1k}) = 0$ for $k \geq 280$. However, the behavior of the nonparametric empirical Bayes procedure $k\overline{D}(d_{1k}^*)$ was not the same as we might expect. Though $\overline{D}(d_{1k}^*)$ roughly decreases in k, its convergence speed is a little slow so that $k\overline{D}(d_{1k}^*)$ seems to be increasing in k.

In general, $\bar{\underset{\sim}{d}}_k$ performs better than the other two, since $k\overline{D}(\bar{d}_{1k}) \leq k\overline{D}(\hat{d}_{1k}) \leq k\overline{D}(d_{1k}^*)$ for all k listed in the table. This result is reasonable since we have the most information regarding the prior distribution $\underset{\sim}{G}$ when the hierarchical empirical Bayes procedure $\bar{\underset{\sim}{d}}_k$ is applied and we have no information regarding the prior distribution $\underset{\sim}{G}$ when the nonparametric empirical Bays procedure $\underset{\sim}{d}_k^*$ is employed.

Table 1. Small Sample Performance of $\underset{\sim}{d}_k^*, \underset{\sim}{\hat{d}}_k$ and $\underset{\sim}{\bar{d}}_k$.

K	$\overline{D}(d_{1k}^*)$	$SE(\overline{D}(d_{1k}^*))$	$K\overline{D}(d_{1k}^*)$	$\overline{D}(\hat{d}_{1k})$	$SE(\overline{D}(\hat{d}_{1k}))$	$K\overline{D}(\hat{d}_{1k})$	$\overline{D}(\bar{d}_{1k})$	$SE(\overline{D}(\bar{d}_{1k}))$	$K\overline{D}(\bar{d}_{1k})$	$r(\underset{\sim}{G})$
10	0.02419	0.00109	0.24190	0.01803	0.00118	0.18030	0.00666	0.00044	0.06656	0.98130
20	0.01877	0.00088	0.37541	0.01609	0.00108	0.32187	0.00266	0.00029	0.05313	1.96260
30	0.02493	0.00111	0.74800	0.01305	0.00096	0.39154	0.00134	0.00020	0.04031	2.94390
40	0.02787	0.00119	1.11474	0.01167	0.00091	0.46678	0.00072	0.00015	0.02875	3.92521
50	0.02680	0.00116	1.33987	0.00925	0.00099	0.46245	0.00059	000013	0.02969	4.90651
60	0.02553	0.00110	1.53189	0.00581	0.00062	0.34874	0.00019	0.00007	0.01125	5.88781
70	0.02603	0.00106	1.82214	0.00440	0.00050	0.30784	0.	0.	0.	6.86911
80	0.02508	0.00102	2.00718	0.00318	0.00041	0.25453	0.	0.	0.	7.85041
90	0.02407	0.00096	2.16587	0.00236	0.00036	0.21270	0.00003	0.00003	0.00281	8.83171
100	0.02404	0.00101	2.40478	0.00174	0.00030	0.17378	0.	0.	0.	9.81301
120	0.02306	0.00093	2.76767	0.00109	0.00024	0.13124				11.77562
140	0.02149	0.00092	3.00801	0.00078	0.00020	0.10876				13.73822
160	0.01961	0.00082	3.13834	0.00061	0.00017	0.09750				15.70083
180	0.01897	0.00080	3.41499	0.00041	0.00014	0.07453				17.66343
200	0.01810	0.00076	3.61943	0.00032	0.00013	0.06406				19.62603
220	0.01738	0.00072	3.82250	0.00013	0.00006	0.02750				21.58864
240	0.01726	0.00072	4.14188	0.00006	0.00004	0.01500				23.55124
260	0.01634	0.00070	4.24938	0.00003	0.00003	0.00813				25.51384
280	0.01589	0.00067	4.44938	0.	0.	0.				27.47644
300	0.01530	0.00066	4.59140	0.	0.	0.				29.43905

4. Selecting the Best Binomial Population Compared with a Control

It is assumed that given θ_i, X_i follows a $B(M, \theta_i)$ distribution with probability function $f_i(x|\theta_i) = \binom{M}{x}\theta_i^x(1-\theta_i)^{M-x}$, and θ_i has a beta prior distribution G_i with pdf $g_i(\theta_i)$ given by

$$g_i(\theta_i) = \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i\mu_i)\Gamma(\alpha_i(1-\mu_i))}\theta_i^{\alpha_i\mu_i-1}(1-\theta_i)^{\alpha_i(1-\mu_i)-1}, \tag{4.1}$$

where $0 < \mu_i < 1, \alpha_i > 0$, and both α_i and μ_i are unknown. We call this statistical model as a binomial-beta model.

Population π_i is called the best if $\theta_i = \max_{i \le j \le k} \theta_j$, and said to be good if $\theta_i \ge \theta_0$ where $0 < \theta_0 < 1$ and bad otherwise. The parameter θ_0 is assumed to be known and may be viewed as a required standard. The selection goal is to select a population which should be the best among the k binomial populations $\pi_1, \ldots, \pi_k$ and be good compared with the standard. If no population is good, we select none.

Let $\mathcal{A} = \{i|i = 0, 1, \ldots, k\}$ be the action space. When action $i \ne 0$ is taken, it means that population π_i is selected as the best among the k binomial populations and believed to be good compared with the standard θ_0. When action $i = 0$ is taken, it means that all populations are excluded as bad populations. For the parameter $\underset{\sim}{\theta}$ and action i, the loss function $L(\underset{\sim}{\theta}, i)$ is defined as:

$$L(\underset{\sim}{\theta}, i) = \max(\theta_{[k]}, \theta_0) - \theta_i. \tag{4.2}$$

Let $\mathcal{X}$ be the sample space of $\underset{\sim}{X} = (X_1, \ldots, X_k)$. A selection rule $\underset{\sim}{d} = (d_0, d_1, \ldots, d_k)$ is defined to be a mapping from $\mathcal{X}$ into $[0,1]^{k+1}$ such that $0 \le d_i(\underset{\sim}{x}) \le 1, i = 0, 1, \ldots, k$, and $\sum_{i=0}^{k} d_i(\underset{\sim}{x}) = 1$. $d_i(\underset{\sim}{x})$ can be viewed as the probability of taking action i when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. We denote the associated Bayes risk of $\underset{\sim}{d}$ by $r(\underset{\sim}{G}, \underset{\sim}{d})$. Under the precedingly described statistical model,

$$\begin{aligned} r(\underset{\sim}{G}, \underset{\sim}{d}) &= \int_\Omega \prod_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=0}^{k} L(\underset{\sim}{\theta}, i) d_i(\underset{\sim}{x}) \prod_{j=1}^{k} f_j(x_j|\theta_j) d\underset{\sim}{G}(\underset{\sim}{\theta}) \\ &= C - \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=0}^{k} d_i(\underset{\sim}{x}) \varphi_i(\underset{\sim}{x}) \underset{\sim}{f}(\underset{\sim}{x}), \end{aligned} \tag{4.3}$$

where $C = \int_\Omega \max(\theta_{[k]}, \theta_0) d\underset{\sim}{G}(\underset{\sim}{\theta}), \varphi_i(\underset{\sim}{x}) = E[\Theta_i|\underset{\sim}{X} = \underset{\sim}{x}] = \frac{x_i + \alpha_i\mu_i}{M + \alpha_i}$ is the posterior mean of Θ_i gives $\underset{\sim}{X} = \underset{\sim}{x}$ for each $i = 1, \ldots, k$, and $\varphi_0(\underset{\sim}{x}) = \theta_0$.

For each $\underset{\sim}{x} \in \mathcal{X}$, let

$$A(\underset{\sim}{x}) = \{i | \varphi_i(\underset{\sim}{x}) = \max_{1 \le j \le k} \varphi_j(\underset{\sim}{x}) \text{ and } \varphi_i(\underset{\sim}{x}) \ge \theta_0, i = 1, \ldots, k\}. \tag{4.4}$$

Then, a randomized Bayes rule, say $\underset{\sim}{d}_{\underset{\sim}{G}} = (d_{\underset{\sim}{G}_0}, \ldots, d_{\underset{\sim}{G}_k})$ can be obtained as follows:

For each $\underset{\sim}{x} \in \mathcal{X}$, if $A(\underset{\sim}{x}) \neq \phi$, define, for each $i = 0, 1, \ldots, k$,

$$d_{\underset{\sim}{G}_i}(\underset{\sim}{x}) = \begin{cases} |A(\underset{\sim}{x})|^{-1} & \text{if } i \in A(\underset{\sim}{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{4.5}$$

and if $A(\underset{\sim}{x}) = \phi$, define, $d_{\underset{\sim}{G}_0}(\underset{\sim}{x}) = 1$ and $d_{\underset{\sim}{G}_i}(\underset{\sim}{x}) = 0, i = 1, \ldots, k$, where $|A(\underset{\sim}{x})|$ is the cardinality of the set $A(\underset{\sim}{x})$.

The minimum Bayes risk is $r(\underset{\sim}{G}) = r(\underset{\sim}{G}, \underset{\sim}{d}_{\underset{\sim}{G}})$.

In view of (4.4) and (4.5), it is clear that the Bayes selection rule $\underset{\sim}{d}_{\underset{\sim}{G}}$ is sensitive to the values of α_i and μ_i. When the values of α_i and μ_i are unknown, the Bayes rule $\underset{\sim}{d}_{\underset{\sim}{G}}$ cannot be applied. It is assumed certain past observations from each of the k populations are available. Hence, a parametric empirical Bayes approach is employed.

A Parametric Empirical Bayes Selection Rule

Following the usual empirical Bayes framework, it is assumed that for each $i = 1, \ldots, k$, there are marginally iid past random observations $X_{i1}, \ldots, X_{in}$ with marginal probability function $f_i(x) = \int_0^1 f_i(x|\theta) g_i(\theta) d\theta$ available when a decision is made. It is also assumed that the random observations obtained from the k populations are mutually independent.

Under the binomial-beta statistical model, for each $i = 1, \ldots, k$,

$$\begin{aligned} &E[X_i/M] = \mu_i, \\ &E[(X_i/M)^2] = \mu_i/M + (\alpha_i\mu_i + 1)\mu_i(M-1)/(M(\alpha_i+1)) \equiv \mu_{i2}. \end{aligned} \tag{4.6}$$

From (4.6), through straightforward computation, the parameter α_i can be written as $\alpha_i = B_i/A_i$, where

$$\begin{cases} B_i = \mu_i - \mu_{i2} \\ A_i = \mu_{i2} - \mu_i/M + \mu_i^2/M - \mu_i^2. \end{cases} \tag{4.7}$$

Note that $B_i > 0$ and hence $A_i > 0$ since $\alpha_i > 0$. Hence μ_i and μ_{i2} satisfy that $\mu_i/M - \mu_i^2/M + \mu_i^2 < \mu_{i2} < \mu_i$. From (4.7), α_i can be viewed as a function of μ_i and μ_{i2}

for $\mu_i \in (0,1)$ and $\mu_{i2} \in (\mu_i/M - \mu_i^2/M + \mu_i^2, \mu_i)$. For each fixed μ_i, α_i is decreasing in μ_{i2} and $\lim_{\mu_{i2}\to\mu_i} \alpha_i = 0$, $\lim_{\mu_{i2}\to a_i} \alpha_i = \infty$, where $a_i = \mu_i/M - \mu_i^2/M + \mu_i^2$.

Let μ_{in} and μ_{i2n} be the moment estimators of μ_i and μ_{i2}, respectively. That is, $\mu_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}/M$, and $\mu_{i2n} = \frac{1}{n}\sum_{j=1}^{n}(X_{ij}/M)^2$. Also, let $A_{in} = \mu_{i2n} - \mu_{in}/M + \mu_{in}^2/M - \mu_{in}^2$, $B_{in} = \mu_{in} - \mu_{i2n}$. Since it is possible that $A_{in} \leq 0$, we consider the following estimator α_{in} for α_i, where

$$\alpha_{in} = \begin{cases} B_{in}/A_{in} & \text{if } A_{in} > 0, \\ \infty & \text{otherwise.} \end{cases} \tag{4.8}$$

We then propose an estimator $\varphi_{in}(\underset{\sim}{x})$ for the posterior mean $\varphi_i(\underset{\sim}{x}) = \frac{x_i+\alpha_i\mu_i}{M+\alpha_i}$, where

$$\varphi_{in}(\underset{\sim}{x}) = \begin{cases} (x_i + \alpha_{in}\mu_{in})/(M+\alpha_{in}) & \text{if } \alpha_{in} < \infty, \\ \mu_{in} & \text{if } \alpha_{in} = \infty. \end{cases} \tag{4.9}$$

Now, an empirical Bayes selection rule $\underset{\sim}{d}_n^* = (d_{0n}^*, \ldots, d_{kn}^*)$ for the selection problem under study is proposed as follows:

For each $\underset{\sim}{x} \in \mathcal{X}$, let

$$A_n^*(\underset{\sim}{x}) = \{i | \varphi_{in}(\underset{\sim}{x}) = \max_{1\leq j\leq k} \varphi_{jn}(\underset{\sim}{x}) \text{ and } \varphi_{in}(\underset{\sim}{x}) \geq \theta_0, i = 1, \ldots, k\}. \tag{4.10}$$

If $A_n^*(\underset{\sim}{x}) \neq \phi$, define, for each $i = 0, 1, \ldots, k$,

$$d_{in}^*(\underset{\sim}{x}) = \begin{cases} |A_n^*(\underset{\sim}{x})|^{-1} & \text{if } i \in A_n^*(\underset{\sim}{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{4.11}$$

and if $A_n^*(\underset{\sim}{x}) = \phi$, define $d_{on}^*(\underset{\sim}{x}) = 1$ and $d_{in}^*(\underset{\sim}{x}) = 0, i = 1, \ldots, k$.

The overall expected Bayes risk of the empirical Bayes selection rule $\underset{\sim}{d}_n^*$ is

$$E[r(\underset{\sim}{G}, \underset{\sim}{d}_n^*)] = C - \sum_{\underset{\sim}{x}\in\mathcal{X}} \left\{ \sum_{i=0}^{k} E[d_{in}^*(\underset{\sim}{x})]\varphi_i(\underset{\sim}{x}) \right\} \underset{\sim}{f}(\underset{\sim}{x}). \tag{4.12}$$

where $C = \int_\Omega \max(\theta_{[k]}, \theta_0) d\underset{\sim}{G}(\underset{\sim}{\theta})$ and the expectation E is taken with respect to $X_{ij}, j = 1, \ldots, n, i = 1, \ldots, k$. Note that $E[r(\underset{\sim}{G}, \underset{\sim}{d}_n^*)] - r(\underset{\sim}{G}) \geq 0$ since $r(\underset{\sim}{G})$ is the minimum Bayes risk.

Following a proof analogous to that of Gupta and Liang (1989a), for the empirical Bayes selection rule $\underset{\sim}{d}_n^*$, we have the following asymptotic optimality.

Theorem 4.1 $E[r(\underset{\sim}{G}, \underset{\sim}{d}_n^*)] - r(\underset{\sim}{G}) = O(\exp(-cn))$ for some positive constant c.

5. Selection Rules for Sampling Inspection

The hypergeometric distribution arises in sampling without replacement from a finite population. Consider a finite population, say, a batch of M items, which is inspected for defective. One takes a sample of size m without replacement from the population. Let X denote the random number of defectives in the sample. Also, let d denote the number of defectives in the population. Then, the random variable X has a probability function

$$f(x|d) = \binom{d}{x}\binom{M-d}{m-x}/\binom{M}{m}, \tag{5.1}$$

where $\max(0, d+m-M) \le x \le \min(m, d)$. Such a finite population with the probability model (5.1) is denoted by $\pi(M, m, d)$.

Consider the problem of acceptance sampling for k independent hypergeometric populations, say $\pi_i = \pi(M_i, m_i, d_i)$, $i = 1, \ldots, k$. Let d_{i0} be a positive integer such that $0 < d_{i0} < M_i$, $i = 1, \ldots, k$. d_{i0} is a given number used as a standard to evaluate the quality of the population π_i. Population π_i is said to be good and acceptable if $d_i < d_{i0}$, and bad otherwise. Our goal is to select all good populations and to exclude all bad populations.

5.1. Formulation of the Selection Problem

Let X_i denote a random variable arising from the population $\pi_i = \pi(M_i, m_i, d_i)$. Conditional on d_i, X_i has a hypergeometric distribution with probability function

$$f_i(x|d_i) = \binom{d_i}{x}\binom{M_i-d_i}{m_i-x}/\binom{M_i}{m_i}.$$

It is assumed that $X_1, \ldots, X_k$ are mutually independent so that $(X_1, \ldots, X_k)$ has a joint probability function $f(\underset{\sim}{x}|\underset{\sim}{d}) = \prod_{i=1}^{k} f_i(x_i|d_i)$, $\underset{\sim}{x} = (x_1, \ldots, x_k), \underset{\sim}{d} = (d_1, \ldots, d_k)$. It is also assumed that for each i, d_i follows a $B(M, \theta_i)$ distribution and that $d_1, \ldots, d_k$ are mutually independent. Note that θ_i is the probability that any item in π_i will be defective. It follows that given θ_i, X_i has a marginal probability function

$$f_i(x|\theta_i) = \sum_{d=x}^{M_i-m_i+x} f_i(x|d)g_i(d|\theta_i) \tag{5.2}$$

$$= \binom{m_i}{x} \theta_i^x (1-\theta_i)^{m_i - x}, \; x = 0, 1, \ldots, m_i,$$

where

$$g_i(d_i|\theta_i) = \binom{M_i}{d_i} \theta_i^{d_i} (1-\theta_i)^{M_i - d_i}, \; d_i = 0, 1, \ldots, M_i.$$

Also, $X_1, \ldots, X_k$ are marginally mutually independent.

Let $\Omega = \{\underset{\sim}{d} = (d_1, \ldots, d_k) | 0 \le d_i \le M_i, \; i = 1, \ldots, k\}$ denote the parameter space and let $\mathcal{A} = \{\underset{\sim}{a} = (a_1, \ldots, a_k) | a_i = 0, 1; i = 1, \ldots, k\}$ be the action space. When action $\underset{\sim}{a}$ is taken, it means that population π_i is selected as a good population if $a_i = 1$, and excluded as a bad one if $a_i = 0$. For the parameter $\underset{\sim}{d}$ and action $\underset{\sim}{a}$, the loss function $L(\underset{\sim}{d}, \underset{\sim}{a})$ is defined as follows:

$$L(\underset{\sim}{d}, \underset{\sim}{a}) = \sum_{i=1}^{k} (1 - a_i) L_{i0}(d_i) + \sum_{i=1}^{k} a_i L_{i1}(d_i) \tag{5.3}$$

where $L_{i0}(d_i)$ and $L_{i1}(d_i)$ are bounded functions and satisfy

$$L_{i0}(d_i) \begin{cases} = 0 & \text{if } d_i \ge d_{i0} \\ > 0 & \text{and nonincreasing in } d_i \text{ for } d_i < d_{i0}; \end{cases}$$

$$L_{i1}(d_i) \begin{cases} = 0 & \text{if } d_i \le d_{i0} \\ > 0 & \text{and nondecreasing in } d_i \text{ for } d_i > d_{i0}. \end{cases}$$

Let $\mathcal{X}$ denote the sample space of $(X_1, \ldots, X_k)$. A selection rule $\underset{\sim}{\delta}_k = (\delta_{k1}, \ldots, \delta_{kk})$ is defined to be a mapping from the sample space $\mathcal{X}$ into $[0,1]^k$, such that for each $\underset{\sim}{x} \in \mathcal{X}$, $\delta_{ki}(\underset{\sim}{x})$ is the probability of selecting π_i as a good population, $i = 1, \ldots, k$.

Gupta and Liang (1991c) consider the following two cases:

Case 1. It is assumed that $\theta_1 = \ldots = \theta_k = \theta$, and the value of the common parameter θ is fixed, but unknown.

Case 2. It is assumed that for each $i = 1, \ldots, k$, the parameter θ_i is a realization of a random variable Θ_i, and $\Theta_1, \ldots \Theta_k$ are iid, having a beta prior distribution Beta$(\alpha\mu, \alpha(1-\mu))$, with probability function $h(\theta|\alpha, \mu)$,

$$h(\theta|\alpha, \mu) = \frac{\Gamma(\alpha)}{\Gamma(\alpha\mu)\Gamma(\alpha(1-\mu))} \theta^{\alpha\mu - 1} (1-\theta)^{\alpha(1-\mu)-1}, \; 0 < \theta < 1,$$

where $0 < \mu < 1$, $\alpha > 0$, and the values of both the parameters α and μ are fixed but unknown.

It is assumed that there is no past data available for the current decision problem. In such an instance, Gupta and Liang (1991c) investigate certain empirical Bayes selection rules by incorporating information from each of the k populations.

5.2. A Parametric Empirical Bayes Selection Rule for Case 1

In order to construct the empirical Bayes selection rules, as a first step, we derive the Bayes rule.

A Bayes Selection Rule Relative to θ

Consider the unknown hyperparameter θ as fixed. For each $i = 1, \ldots, k$, define

$$C_i(\theta) = \sum_{d_i=0}^{d_{i0}-1} L_{i0}(d_i) g_i(d_i|\theta), \tag{5.4}$$

$$H_i(x_i, \theta) = \sum_{d_i=d_{i0}+1}^{M_i-m_i+x_i} L_{i1}(d_i) g_i(d_i|x_i, \theta) - \sum_{d_i=x_i}^{d_{i0}-1} L_{i0}(d_i) g_i(d_i|x_i, \theta),$$

where $g_i(d_i|x_i, \theta) = f_i(x_i|d_i) g_i(d_i|\theta)/f_i(x_i|\theta)$ is the posterior probability function of d_i given $X_i = x_i$, and $\sum_{d=s}^{t} \equiv 0$ if $t < s$. Let $r_{ki}(\theta, \delta_{ki})$ denote the i–th component Bayes risk, $i = 1, \ldots, k$, and let $r_k(\theta, \underset{\sim}{\delta}_k)$ be the overall Bayes risk of the selection rule $\underset{\sim}{\delta}_k = (\delta_{k1}, \ldots, \delta_{kk})$. Then,

$$\begin{cases} r_{ki}(\theta, \delta_{ki}) = \sum_{\mathcal{X}} \delta_{ki}(\underset{\sim}{x}) H_i(x_i, \theta) f(\underset{\sim}{x}|\theta) + C_i(\theta), \\ r_k(\theta, \underset{\sim}{\delta}_k) = \sum_{i=1}^{k} r_{ki}(\theta, \delta_{ki}), \end{cases} \tag{5.5}$$

where $f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{k} f_i(x_i|\theta)$.

Therefore, relative to the fixed parameter θ, a Bayes rule, denoted by $\underset{\sim}{\delta}_k^\theta = (\delta_{k1}^\theta, \ldots, \delta_{kk}^\theta)$ can be obtained as:

For each $\underset{\sim}{x} \in \mathcal{X}$, $i = 1, \ldots, k$,

$$\delta_{ki}^\theta(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } H_i(x_i, \theta) < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

The minimum Bayes risk is $r_k(\theta, \underset{\sim}{\delta}_k^\theta) = \sum_{i=1}^{k} r_{ki}(\theta, \delta_{ki}^\theta)$.

A Parametric Empirical Bayes Selection Rule

Under the statistical model of Case 1, $X_1, \ldots, X_k$ are mutually independent with $X_i \sim B(m_i, \theta), i = 1, \ldots, k$. Hence, $X_1 + \ldots + X_k \sim B(\sum_{i=1}^{k} m_i, \theta)$. Therefore $\hat{\theta} = (X_1 + \ldots + X_k)/\sum_{i=1}^{k} m_i$ is an unbiased estimator of and sufficient statistic for the parameter θ. Gupta and Liang (1991c) use $H_i(x_i, \hat{\theta})$ to estimate $H_i(x_i, \theta)$, and propose a parametric empirical Bayes selection rule $\underset{\sim}{\delta}_k^* = (\delta_{k1}^*, \ldots, \delta_{kk}^*)$ given as follows: For each $i = 1, \ldots, k$, and $\underset{\sim}{x} \in \mathcal{X}$,

$$\delta_{ki}^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } H_i(x_i, \hat{\theta}) < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

Asymptotic Optimality of $\underset{\sim}{\delta}_k^*$

Gupta and Liang (1991c) establish the following asymptotic optimality of the parametric empirical Bayes selection rule $\underset{\sim}{\delta}_k^*$.

Theorem 5.1. Suppose that $m_* \le m_i \le M_i \le M^*$ for all $i = 1, \ldots, k$, and $L_{ij}(d_i) \le L^*$ for all $j = 0, 1$ and $i = 1, \ldots, k$, where the bound values L^*, m_* and M^* are independent of k for all k. Then, under the statistical model of Case 1, for each $\theta \in (0, 1)$,

$$r_k(\theta, \underset{\sim}{\delta}_k^*) - r_k(\theta, \underset{\sim}{\delta}_k^\theta) = O(\exp(-c(\theta)k + \ln k)),$$

where $c(\theta) > 0$ depends on $\theta \in (0, 1)$.

5.3. **A Hierarchical Empirical Bayes Selection Rule for Case 2**

A Hierarchical Bayes Selection Rule wrt Beta $(\alpha\mu, \alpha(1-\mu))$ Prior

For each $i = 1, \ldots, k$, let

$$f_i(x_i|\alpha, \mu) = \int_0^1 f_i(x_i|\theta)h(\theta|\alpha, \mu)d\theta$$

$$f_i(x_i, d_i|\alpha, \mu) = \int_0^1 f_i(x_i|d_i) g_i(d_i|\theta) h(\theta|\alpha, \mu) d\theta$$
$$g_i(d_i|x_i, \alpha, \mu) = f_i(x_i, d_i|\alpha, \mu)/f_i(x_i|\alpha, \mu)$$
$$C_i = \int_0^1 C_i(\theta) h(\theta|\alpha, \mu) d\theta$$
$$Q_i(x_i, \alpha, \mu) = \sum_{d_i=d_{i0}+1}^{M_i-m_i+x_i} L_{i1}(d_i) g_i(d_i|x_i, \alpha, \mu) - \sum_{d_i=x_i}^{d_{i0}-1} L_{i0}(d_i) g_i(d_i|x_i, \alpha, \mu)$$
$$f(\underset{\sim}{x}|\alpha, \mu) = \prod_{i=1}^{k} f_i(x_i|\alpha, \mu).$$

For fixed values of the parameters α and μ, we denote the i–th component Bayes risk and the overall Bayes risk of the selection rule $\underset{\sim}{\delta}_k = (\delta_{k1}, \ldots, \delta_{kk})$ by $R_{ki}(\alpha, \mu, \delta_{ki})$ and $R_k(\alpha, \mu, \underset{\sim}{\delta}_k)$, respectively. Then, under the corresponding statistical model,

$$\begin{cases} R_{ki}(\alpha, \mu, \delta_{ki}) = \sum_{\mathcal{X}} \delta_{ki}(\underset{\sim}{x}) Q_i(x_i, \alpha, \mu) f(\underset{\sim}{x}|\alpha, \mu) + C_i, \\ R_k(\alpha, \mu, \underset{\sim}{\delta}_k) = \sum_{i=1}^{k} R_{ki}(\alpha, \mu, \delta_{ki}). \end{cases} \tag{5.8}$$

Under this hierarchical statistical model, a Bayes selection rule, called the hierarchical Bayes selection rule, is $\underset{\sim}{\delta}_k^{\mathrm{HB}} = (\delta_{k1}^{\mathrm{HB}}, \ldots, \delta_{kk}^{\mathrm{HB}})$, where for each $\underset{\sim}{x} \in \mathcal{X}$,

$$\delta_{ki}^{\mathrm{HB}}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } Q_i(x_i, \alpha, \mu) < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

The minimum Bayes risk is $R_k(\alpha, \mu, \underset{\sim}{\delta}_k^{\mathrm{HB}}) = \sum_{i=1}^{k} R_{ki}(\alpha, \mu, \delta_{ki}^{\mathrm{HB}})$.

A Hierarchical Empirical Bayes Selection Rule

Under the statistical model of Case 2, for each $i = 1, \ldots, k$,

$$E[X_i] = m_i \mu$$
$$E[X_i^2] = \frac{m_i \alpha (1-\mu)\mu}{\alpha + 1} + \frac{m_i^2 (\alpha\mu + 1)\mu}{\alpha + 1} \equiv \mu_{i2}$$

where μ_{i2} is decreasing in α and tends to $m_i\mu(1-\mu)+m_i^2\mu^2$ as α tends to infinity. Therefore

$$E\left[\sum_{i=1}^{k} X_i\right] = \mu \sum_{i=1}^{k} m_i \text{ and}$$

$$E\left[\sum_{i=1}^{k} X_i^2\right] = \frac{\alpha\mu(1-\mu)\sum_{i=1}^{k} m_i}{\alpha+1} + \frac{(\alpha\mu+1)\mu\sum_{i=1}^{k} m_i^2}{\alpha+1}.$$

Hence

$$\alpha = \frac{\mu\sum_{i=1}^{k} m_i^2 - \sum_{i=1}^{k}\mu_{i2}}{\sum_{i=1}^{k}\mu_{i2} - \mu^2\sum_{i=1}^{k} m_i^2 - \mu(1-\mu)\sum_{i=1}^{k} m_i}. \tag{5.10}$$

We may use $\hat{\mu} = \sum_{i=1}^{k} X_i / \sum_{i=1}^{k} m_i$ to estimate μ and $\sum_{i=1}^{k} X_i^2$ to estimate $\sum_{i=1}^{k}\mu_{i2}$. Note that $\mu_{i2} - m_i^2\mu^2 - m_i\mu(1-\mu) > 0$ for each $i = 1, \ldots, k$, and hence $\sum_{i=1}^{k}\mu_{i2} - \mu^2\sum_{i=1}^{k} m_i^2 - \mu(1-\mu)\sum_{i=1}^{k} m_i > 0$. However, it is possible that

$$D(X_1, \ldots, X_k) \equiv \sum_{i=1}^{k} X_i^2 - \hat{\mu}^2\sum_{i=1}^{k} m_i^2 - \hat{\mu}(1-\hat{\mu})\sum_{i=1}^{k} m_i \leq 0.$$

Motivated by the form of (5.10) and the decreasing property of μ_{i2} with respect to α, one may define

$$\hat{\alpha} = \begin{cases} \frac{\hat{\mu}\sum_{i=1}^{k} m_i^2 - \sum_{i=1}^{k} X_i^2}{D(X_1,\ldots,X_k)} & \text{if } D(X_1, \ldots, X_k) > 0, \\ \infty & \text{otherwise.} \end{cases} \tag{5.11}$$

Note that when x_i and μ are kept fixed, $Q_i(x_i, \infty, \mu) \equiv \lim_{\alpha\to\infty} Q_i(x_i, \alpha, \mu)$ exists. Gupta and Liang (1991c) use $Q_i(x_i, \hat{\alpha}, \hat{\mu})$ to estimate $Q_i(x_i, \alpha, \mu)$ and propose a hierarchical empirical Bayes selection rule $\hat{\underset{\sim}{\delta}}_k = (\hat{\delta}_{k1}, \ldots, \hat{\delta}_{kk})$ defined as follows: For each $i = 1, \ldots, k$ and $\underset{\sim}{x} \in \mathcal{X}$,

$$\hat{\delta}_{ki}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } Q_i(x_i, \hat{\alpha}, \hat{\mu}) < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

Asymptotic Optimality of $\hat{\underset{\sim}{\delta}}_k$

Gupta and Liang (1991c) establish the following asymptotic optimality of the hierarchical empirical Bayes selection rule $\hat{\underset{\sim}{\delta}}_k$.

Theorem 5.2. Suppose that $1 < m_* \leq m_i \leq M_i \leq M^*$ for all $i = 1, \ldots, k$, and $L_{ij}(d_i) \leq L^*$ for all $j = 0, 1$; $i = 1, \ldots, k$, where the values of the bounds L^*, m_* and M^* are independent of k. Then, under the statistical model of Case 2, for each pair of the values (α, μ), $0 < \alpha < \infty$, $0 < \mu < 1$, the empirical Bayes selection rule $\hat{\underset{\sim}{\delta}}_k$ is asymptotically optimal in the sense that

$$R_k(\alpha, \mu, \hat{\underset{\sim}{\delta}}_k) - R_k(\alpha, \mu, \underset{\sim}{\delta}_k^{\text{HB}}) = O(\exp(-\tau k + \ln k)),$$

where $\tau > 0$ depends on the values of (α, μ).

5.4 Small Sample Performance: Simulation Studies

In the Monte Carlo studies, we have assumed that

$$M_1 = \ldots = M_k = M \quad m_1 = \ldots = m_k = m \quad d_{10} = \ldots = d_{k0} = d_0. \tag{5.13}$$

Under the preceding assumption and the statistical model of Case 1, for the Bayes selection rule $\underset{\sim}{\delta}_k^\theta = (\delta_{k1}^\theta, \ldots, \delta_{kk}^\theta), r_{k1}(\theta, \delta_{k1}^\theta) = \ldots = r_{kk}(\theta, \delta_{kk}^\theta)$ and $r_k(\theta, \underset{\sim}{\delta}_k^\theta) = k r_{kk}(\theta, \delta_{kk}^\theta)$. Also, for the parametric empirical Bayes selection rule $\underset{\sim}{\delta}_k^* = (\delta_{k1}^*, \ldots, \delta_{kk}^*)$, it can be seen that $r_{k1}(\theta, \delta_{k1}^*) = \ldots = r_{kk}(\theta, \delta_{kk}^*)$ and therefore $r_k(\theta, \underset{\sim}{\delta}_k^*) = k r_{kk}(\theta, \delta_{kk}^*)$.

Similarly, under the assumption (5.13) and the statistical model of Case 2, for the hierarchical Bayes selection rule $\underset{\sim}{\delta}_k^{HB} = (\delta_{k1}^{HB}, \ldots, \delta_{kk}^{HB}), R_{k1}(\alpha, \mu, \delta_{k1}^{HB}) = \ldots = R_{kk}(\alpha, \mu, \delta_{kk}^{HB})$ and $R_k(\alpha, \mu, \underset{\sim}{\delta}_k^{HB}) = k R_{kk}(\alpha, \mu, \delta_{kk}^{HB})$. Also, for the hierarchical empirical Bayes selection rule $\hat{\underset{\sim}{\delta}}_k = (\hat{\delta}_{k1}, \ldots, \hat{\delta}_{kk}), R_{k1}(\alpha, \mu, \hat{\delta}_{k1}) = \ldots = R_{kk}(\alpha, \mu, \hat{\delta}_{kk})$ and $R_k(\alpha, \mu, \hat{\underset{\sim}{\delta}}_k) = k R_{kk}(\alpha, \mu, \hat{\delta}_{kk})$.

Therefore, in the following, we simulated the differences $r_{kk}(\theta, \delta_{kk}^*) - r_{kk}(\theta, \delta_{kk}^\theta)$ and $R_{kk}(\alpha, \mu, \hat{\delta}_{kk}) - R_{kk}(\alpha, \mu, \delta_{kk}^{HB})$, and used $k[r_{kk}(\theta, \delta_{kk}^*) - r_{kk}(\theta, \delta_{kk}^\theta)]$ to estimate $r_k(\theta, \underset{\sim}{\delta}_k^*) - r_k(\theta, \underset{\sim}{\delta}_k^\theta)$ and used $k[R_{kk}(\alpha, \mu, \hat{\delta}_{kk}) - R_{kk}(\alpha, \mu, \delta_{kk}^{HB})]$ to estimate $R_k(\alpha, \mu, \hat{\underset{\sim}{\delta}}_k) - R_k(\alpha, \mu,$

$\underset{\sim}{\delta}_k^{HB}$), respectively. In the simulation studies, the following linear loss functions $L_{i1}(d_i)$ and $L_{i0}(d_i)$ are used, where

$$\begin{cases} L_{i1}(d_i) = (d_i - d_{i0})I(d_i > d_{i0}), \\ L_{i0}(d_i) = (d_{i0} - d_i)I(d_{i0} > d_i). \end{cases}$$

The simulation scheme used is described as follows:

Case 1. The Parameter θ Being Fixed

(1) For any fixed value of the parameter θ and a given value of m, generate $k-1$ independent random numbers $X_1 \ldots X_{k-1}$ from a $B(m, \theta)$ distribution.

(2) Let x_k be an observed value from a $B(m, \theta)$ distribution. Use $X_1 \ldots X_{k-1}$ and x_k to estimate θ and construct the parametric empirical Bayes selection rule δ_{kk}^*. Then, compute the conditional regret Bayes risk of δ_{kk}^* (conditional on $X_1 \ldots X_{k-1}$) by $D_k^\theta(X_1 \ldots X_{k-1}) = r_{kk}(\theta, \delta_{kk}^* | X_1 \ldots X_{k-1}) - r_{kk}(\theta, \delta_{kk}^\theta)$.

(3) The above process was repeated 500 times. The average of the regret Bayes risk based on the 500 repetitions denoted by $\overline{D}_k^\theta$ was used as an estimate of the regret Bayes risk $r_{kk}(\theta, \delta_{kk}^*) - r_{kk}(\theta, \delta_{kk}^\theta)$. Then, we used $k\overline{D}_k^\theta$ as an estimate of the total regret Bayes risk $r_k(\theta, \underset{\sim}{\delta}_k^*) - r_k(\theta, \underset{\sim}{\delta}_k^\theta)$. A summary of the simulated results is given in Table 2.

Case 2. The Parameter θ with a Prior Distribution Beta $(\alpha\mu, \alpha(1-\mu))$

(1) For given values of α and μ, we generated $k-1$ random variables from a distribution having the probability function $f(x|\alpha, \mu)$ where $f(x|\alpha, \mu) = \int_0^1 f(x|\theta)h(\theta|\alpha, \mu)d\theta$.

Then we followed steps (2) and (3) analogous to steps (2) and (3) of Case 1, by just replacing the Bayes risk by the corresponding Bayes risks of the hierarchical empirical Bayes selection rule $\hat{\delta}_{kk}$ and the hierarchical Bayes selection rule δ_{kk}^{HB}, respectively. We denote the average of the conditional regret risks based on 500 repetitions by $\overline{D}_k^{\alpha\mu}$. $\overline{D}_k^{\alpha\mu}$ was used to estimate the regret Bayes risk $R_{kk}(\alpha, \mu, \hat{\delta}_{kk}) - R_{kk}(\alpha, \mu, \delta_{kk}^{HB})$. Then, we used $k\overline{D}_k^{\alpha\mu}$ to estimate the total regret Bayes risk $R_k(\alpha, \mu, \hat{\underset{\sim}{\delta}}_k) - R_k(\alpha, \mu, \underset{\sim}{\delta}_k^{HB})$.

The results of these simulations are reported in Table 3 and Table 4, respectively.

The simulated results indicate that in each of the two models, the total regret risks converges to zero as the value of k become large. Of course, the rates of convergence vary according to the models. In Case 1 model, for small values of k, $k\overline{D}_k^\theta$ is small and $k\overline{D}_k^\theta$ tends to zero gradually. While in Case 2 model, though for small values of k, the $k\overline{D}_k^{\alpha\mu}$ values are larger, however, the rate of convergence of $k\overline{D}_k^{\alpha\mu}$ to zero is very fast.

Table 2. The Small Sample Performance of $\underset{\sim}{\delta}_k^*$

$M = 100,\ m = 20,\ d_0 = 6$, and $\theta = 0.02$.

k	$\overline{D}_k^\theta$	$k\overline{D}_k^\theta$	$SE(\overline{D}_k^\theta)$
10	2.9878×10^{-3}	29.8785×10^{-3}	0.6169×10^{-3}
20	0.5254×10^{-3}	10.5073×10^{-3}	0.0822×10^{-3}
30	0.1811×10^{-3}	5.4324×10^{-3}	0.0369×10^{-3}
40	0.1097×10^{-3}	4.3891×10^{-3}	0.0190×10^{-3}
50	0.0719×10^{-3}	3.5932×10^{-3}	0.0047×10^{-3}
60	0.0774×10^{-3}	4.6418×10^{-3}	0.0190×10^{-3}
70	0.0539×10^{-3}	3.7705×10^{-3}	0.0043×10^{-3}
80	0.0386×10^{-3}	3.0882×10^{-3}	0.0038×10^{-3}
90	0.0359×10^{-3}	3.2319×10^{-3}	0.0037×10^{-3}
100	0.0310×10^{-3}	3.0972×10^{-3}	0.0035×10^{-3}
120	0.0269×10^{-3}	3.2319×10^{-3}	0.0033×10^{-3}
140	0.0162×10^{-3}	2.2623×10^{-3}	0.0026×10^{-3}
160	0.0148×10^{-3}	2.3700×10^{-3}	0.0025×10^{-3}
180	0.0076×10^{-3}	1.3755×10^{-3}	0.0018×10^{-3}
200	0.0108×10^{-3}	2.1546×10^{-3}	0.0021×10^{-3}
250	0.0022×10^{-3}	0.5611×10^{-3}	0.0010×10^{-3}
300	0.0031×10^{-3}	0.9426×10^{-3}	0.0012×10^{-3}
350	0.0004×10^{-3}	0.1571×10^{-3}	0.0004×10^{-3}
400	0.0004×10^{-3}	0.1795×10^{-3}	0.0004×10^{-3}

Table 3. The Small Sample Performance of $\underset{\sim}{\delta}_k^{\text{HB}}$

$M = 100,\ m = 20,\ d_0 = 6,\ \alpha = 10$ and $\mu = 0.02$.

k	$\overline{D}_k^{\alpha\mu}$	$k\overline{D}_k^{\alpha\mu}$	$SE(\overline{D}_k^{\alpha\mu})$
10	36.7345×10^{-3}	367.3446×10^{-3}	1.8780×10^{-3}
20	30.8095×10^{-3}	616.1903×10^{-3}	1.8231×10^{-3}
30	26.0696×10^{-3}	782.0877×10^{-3}	1.7493×10^{-3}
40	17.4362×10^{-3}	697.4465×10^{-3}	1.5324×10^{-3}
50	17.0976×10^{-3}	854.8797×10^{-3}	1.5213×10^{-3}
60	11.6805×10^{-3}	700.8327×10^{-3}	1.3069×10^{-3}
70	9.3106×10^{-3}	651.7408×10^{-3}	1.1856×10^{-3}
80	9.8963×10^{-3}	791.7044×10^{-3}	1.2592×10^{-3}
90	6.7713×10^{-3}	609.4202×10^{-3}	1.0280×10^{-3}
100	8.6335×10^{-3}	863.3451×10^{-3}	1.1468×10^{-3}
120	3.8935×10^{-3}	467.2222×10^{-3}	0.7938×10^{-3}
140	3.5550×10^{-3}	497.6932×10^{-3}	0.7600×10^{-3}
160	2.2009×10^{-3}	352.1094×10^{-3}	0.6030×10^{-3}
180	0.8464×10^{-3}	152.3550×10^{-3}	0.3770×10^{-3}
200	0.5079×10^{-3}	101.5700×10^{-3}	0.2926×10^{-3}
250	0.3386×10^{-3}	84.6417×10^{-3}	0.2392×10^{-3}
300	0.1693×10^{-3}	50.7850×10^{-3}	0.1693×10^{-3}
350	0.	0.	0.
400	0.	0.	0.

Table 4. The Small Sample Performance of $\underset{\sim}{\delta}_k^{\mathrm{HB}}$

$M = 100,\ m = 20,\ d_0 = 6,\ \alpha = 1$ and $\mu = 0.02$.

k	$\overline{D}_k^{\alpha\mu}$	$k\overline{D}_k^{\alpha\mu}$	$SE(\overline{D}_k^{\alpha\mu})$
10	4.4973×10^{-3}	44.9733×10^{-3}	0.5255×10^{-3}
20	4.6379×10^{-3}	92.7573×10^{-3}	0.5324×10^{-3}
30	3.8649×10^{-3}	115.9467×10^{-3}	0.4921×10^{-3}
40	3.3027×10^{-3}	132.1090×10^{-3}	0.4590×10^{-3}
50	2.1784×10^{-3}	108.9196×10^{-3}	0.3793×10^{-3}
60	1.9676×10^{-3}	118.0547×10^{-3}	0.3616×10^{-3}
70	1.1946×10^{-3}	83.6221×10^{-3}	0.2851×10^{-3}
80	0.5622×10^{-3}	44.9732×10^{-3}	0.1974×10^{-3}
90	0.7730×10^{-3}	69.5680×10^{-3}	0.2307×10^{-3}
100	0.3514×10^{-3}	35.1353×10^{-3}	0.1565×10^{-3}
120	0.2108×10^{-3}	25.2975×10^{-3}	0.1215×10^{-3}
140	0.2811×10^{-3}	39.3516×10^{-3}	0.1401×10^{-3}
160	0.0703×10^{-3}	11.2433×10^{-3}	0.0703×10^{-3}
180	0.	0.	0.
200	0.	0.	0.
250	0.	0.	0.
300	0.	0.	0.
350	0.	0.	0.
400	0.	0.	0.

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.O. and Deely, J.J. (1988). A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology. *J. Amer. Statist. Assoc.* **83**, 364–373.

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 Press, New York, 37–55.

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 Press, 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* **A30**, 37–50.

Deely, J.J. and Gupta, S.S. (1988). Hierarchical Bayesian selection procedures for the best binomial population. Technical Report #88–21, Department of Statistics, Purdue University, West Lafayette, Indiana.

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

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, North Carolina.

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

Gupta, S.S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J.*

Statist. Plann. Inference **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 Liang, T. (1986). Empirical Bayes rules for selecting good binomial populations. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes–Monograph Series, Vol. 8, 110–128.

Gupta, S.S. and Liang, T. (1988). Empirical Bayes rules for selecting the best binomial populations. *Statistical Decision Theory and Related Topics–IV* (Eds. S.S. Gupta and J.O. Berger), Springer–Verlag, Vol. I, 213–224.

Gupta, S.S. and Liang, T. (1989a). Selecting the best binomial population: parametric empirical Bayes approach. *J. Statist. Plann. Inference* **23**, 21–31.

Gupta, S.S. and Liang, T. (1989b). Parametric empirical Bayes rules for selecting the most probable multinomial event. Chapter 18 in *Contribution to Probability and Statistics: Essays in Honor of Ingram Olkin*, Springer–Verlag, 318–328.

Gupta, S.S. and Liang, T. (1991a). On empirical Bayes selection rules for negative binomial populations. To appear in *The Proceedings of the Indo–US Workshop on Bayesian Analysis in Statistics and Econometrics.*

Gupta, S.S. and Liang, T. (1991b). On the asymptotic optimality of certain empirical Bayes simultaneous testing procedures for Poisson populations. *Statistics & Decisions*, Vol. **9**, 263–283.

Gupta, S.S. and Liang, T. (1991c). On empirical Bayes selection rules for sampling inspection. Technical Report #91-35C, Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S.S. and Miescke, K.J. (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 populations. *J. Statist. Plann. Inference* **12**, 213–233.

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

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, University of California Press, 157–163.

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