ON THE ASYMPTOTIC OPTIMALITY OF CERTAIN EMPIRICAL BAYES SIMULTANEOUS TESTING PROCEDURES FOR POISSON POPULATIONS *

by

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

and

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

Technical Report # 89-22C

Department of Statistics
Purdue University

July 1989
Revised October 1989
Second Revision June 1990
Third Revision November 1990

---

* This research was supported in part by NSF Grants DMS–8702620, and DMS–8923071 at Purdue University.

★

ON THE ASYMPTOTIC OPTIMALITY OF CERTAIN EMPIRICAL BAYES SIMULTANEOUS TESTING PROCEDURES FOR POISSON POPULATIONS *

by

Shanti S. Gupta
Department of Statistics
Purdue University
West Lafayette, Indiana

and

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, Michigan

ABSTRACT

This paper is concerned with the problem of simultaneous testing for n-component decisions. Under the specific statistical model, the n components share certain similarity. Thus, empirical Bayes approach is employed. We give a general formulation of this empirical Bayes decision problem with a specialization to the problem of selecting good Poisson populations. Three empirical Bayes methods are used to incorporate information from different sources for making a decision for each of the n components. They are: nonparametric empirical Bayes, parametric empirical Bayes and hierarchical empirical Bayes. For each of them, a corresponding empirical Bayes decision rule is proposed. The asymptotic optimality properties and the convergence rates of the three empirical Bayes rules are investigated. It is shown that for each of the three empirical Bayes rules, the rate of convergence is at least of order $O(\exp(-cn + \ln n))$ for some positive constant c, where the value of c varies depending on the empirical Bayes rule used.

Key Words and Phrases: Asymptotic optimality; isotonic regression; nonparametric empirical Bayes; parametric empirical Bayes; hierarchical empirical Bayes; n-component decision problem.

AMS 1980 Subject Classification: 62C12, 62C25, 62F07

* This research was supported in part by NSF Grants DMS-8702620, and DMS-8923071 at Purdue University.

1. Introduction

Let $\pi_1, \ldots, \pi_n$ denote n independent populations. For each $i = 1, \ldots, n$, population π_i is characterized by a parameter θ_i. Let θ_0 denote a standard or a control. The problem of selecting populations with respect to a control has been extensively studied in the literature. Dunnett [3] and Gupta and Sobel [10] have considered problems of selecting a subset containing all populations better than a control using some natural procedures. Lehmann [12] and Spjøtvoll [20] have treated the problem using methods from the theory of testing hypotheses. Randles and Hollander [16], Gupta and Kim [6], Miescke [14] and Gupta and Miescke [7] have derived optimal procedures via minimax or gamma–minimax approaches. The reader is referred to Gupta and Panchapakesan [8, 9] for an overview of this research area. In this paper, we study the problem of selecting good populations from among n populations using the empirical Bayes approach.

For each $i = 1, \ldots, n$, let X_i denote a random observation arising from population π_i with probability density function $f(x|\theta_i)$. The observation X_i may be thought of as the value of a sufficient statistic for the parameter θ_i based on several iid observations taken from π_i. Let θ_0 be a known constant. This θ_0 can be used as a standard level to evaluate each of the n populations. Population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. Our goal is to select all the good populations and exclude all the bad populations.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n) | f(x|\theta_i) \text{ is well–defined}, i = 1, \ldots, n\}$ be the parameter space and let $\mathcal{A} = \{\underset{\sim}{a} = (a_1, \ldots, a_n) | a_i = 0, 1, i = 1, \ldots, n\}$ 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 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:

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \sum_{i=1}^{n} a_i(\theta_0 - \theta_i) I(\theta_0 - \theta_i) + \sum_{i=1}^{n} (1 - a_i)(\theta_i - \theta_0) I(\theta_i - \theta_0) \tag{1.1}$$

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. It is also assumed that the n random variables $\Theta_i, i = 1, \ldots, n$, are independently distributed with a common but unknown prior distribution G. Thus, $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_n)$

has a joint prior distribution $\underset{\sim}{G}(\underset{\sim}{\theta}) = \prod_{i=1}^{n} G(\theta_i)$ over the parameter space Ω. Under the preceding assumptions, $X_1, \ldots, X_n$ are iid with the marginal probability density function $f(x) = \int f(x|\theta)dG(\theta)$.

For each $i = 1, \ldots, n$, let $\mathcal{X}_i$ be the sample space of X_i, and let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$. Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ and let $\underset{\sim}{x} = (x_1, \ldots, x_n)$ be the observed value of $\underset{\sim}{X}$. A selection rule $\underset{\sim}{d} = (d_1, \ldots, d_n)$ is defined to be a mapping from $\mathcal{X}$ into $[0,1]^n$ such that $d_i(\underset{\sim}{x})$ is the probability of selecting π_i as a good population given $\underset{\sim}{X} = \underset{\sim}{x}$. Let D be the class of all selection rules, and let $r(G, \underset{\sim}{d})$ denote the Bayes risk associated with each $\underset{\sim}{d} \in D$. Then, $r(G) = \inf_{\underset{\sim}{d} \in D} r(G, \underset{\sim}{d})$ is the minimum Bayes risk.

The Bayes risk associated with any rule $\underset{\sim}{d} \in D$ can be rewritten as

$$r(G, \underset{\sim}{d}) = \sum_{i=1}^{n} r_i(G, d_i) \tag{1.2}$$

where

$$r_i(G, d_i) = \int_{\mathcal{X}} [\theta_0 - \varphi_i(x_i)] d_i(\underset{\sim}{x}) \prod_{j=1}^{n} f(x_j) d\underset{\sim}{x} + C \tag{1.3}$$

where $\varphi_i(x_i) = E[\Theta_i | X_i = x_i] = \int \theta f(x_i|\theta)dG(\theta)/f(x_i)$, the posterior mean of Θ_i given $X_i = x_i$, and $C = \int_{\theta_0}^{\infty} (\theta - \theta_0)dG(\theta)$.

Since the value C is independent of the selection rule $\underset{\sim}{d}$, from (1.3), a Bayes rule, say $\underset{\sim}{d}_B = (d_{1B}, \ldots, d_{nB})$ is clearly given by

$$d_{iB}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_i(x_i) \geq \theta_0, \\ 0 & \text{otherwise,} \end{cases} \tag{1.4}$$

and the minimum Bayes risk is: $r(G) = \sum_{i=1}^{n} r_i(G, d_{iB})$.

Since the prior distribution G is unknown, it is not possible to apply the Bayes rule $\underset{\sim}{d}_B$ for the selection problem at hand. However, the selection problem under study can be viewed as that in which we are dealing with a Bayes decision problem having n components with a common unknown prior distribution. Thus, the empirical Bayes approach of Robbins [17, 18] can be employed here. We use all the observations obtained from the n populations to form a decision for each of the n–component problems.

Let $\varphi_{in}(x_i|\underset{\sim}{x}(i))$ be an estimator of $\varphi_i(x_i)$ based on $(x_1,\ldots,x_n)$ where $\underset{\sim}{x}(i) = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$. We then define a selection rule $\underset{\sim}{d}_n = (d_{1n},\ldots,d_{nn})$ as follows:

$$d_{in}(x_i|\underset{\sim}{x}(i)) \equiv d_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_{in}(x_i|\underset{\sim}{x}(i)) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{1.5}$$

The associated Bayes risk of the selection rule $\underset{\sim}{d}_n$ is:

$$r(G,\underset{\sim}{d}_n) = \sum_{i=1}^{n} r_i(G,d_{in}) \tag{1.6}$$

where

$$r_i(G,d_{in}) = E_i R_i(G,d_{in}) \tag{1.7}$$

and

$$R_i(G,d_{in}) = \int_{\mathcal{X}_i} [\theta_0 - \varphi_i(x_i)] d_{in}(x_i|\underset{\sim}{X}(i)) f(x_i) + C. \tag{1.8}$$

In (1.7), the expectation E_i is taken with respect to $\underset{\sim}{X}(i) \equiv (X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$. Recall that $r_i(G,d_{iB})$ is the minimum Bayes risk for the i–th component problem. Thus, $r_i(G,d_{in}) - r_i(G,d_{iB}) \geq 0$ and therefore, $r(G,\underset{\sim}{d}_n) - r(G) \geq 0$. For the empirical Bayes selection rule $\underset{\sim}{d}_n$ to be useful, we always desire that the average nonnegative difference $(r(G,\underset{\sim}{d}_n) - r(G))/n$ or the total nonnegative difference $r(G,\underset{\sim}{d}_n) - r(G)$ be small.

Definition 1.1

(a) A decision rule $\underset{\sim}{d}_n$ is said to be weakly asymptotically optimal relative to the (unknown) prior G if $(r(G,\underset{\sim}{d}_n) - r(G))/n \to 0$ as $n \to \infty$.

(b) A decision rule $\underset{\sim}{d}_n$ is said to be strongly asymptotically optimal relative to the (unknown) prior G if $r(G,\underset{\sim}{d}_n) - r(G) \to 0$ as $n \to \infty$.

Clearly, for a selection rule $\underset{\sim}{d}_n$, the strong asymptotic optimality implies the weak asymptotic optimality. In certain compound decision problems the weak asymptotic optimality of compound decision rules has been studied in the literature by many authors, notably Vardeman [21, 22], Gilliland and Hannan [4], and Gilliland, Hannan and Huang [5], though the formulation of their compound decision problems are different from the one we consider here. However, very surprisingly, it seems that the strong asymptotic optimality has not been investigated so far.

In this paper, we consider the problem of selecting good Poisson populations. According to how much we know about the prior distribution G, three empirical Bayes methods are used to incorporate information from different sources for making a decision for each of the n components. They are: nonparametric empirical Bayes, parametric empirical Bayes and hierarchical empirical Bayes. For each of them, a corresponding empirical Bayes selection rule is proposed. The strong asymptotic optimality of the selection rules is also established. It is shown that for each of the three empirical Bayes selection rules, the rate of convergence is at least of order $O(exp(-cn + \ell n\ n))$ for some positive constant c, where the value of c varies depending on the empirical Bayes rule used. This result indicates the advantage of incorporating all the information from different sources for making a decision for each of the n component problems.

2. Selecting Good Poisson Populations

It is assumed that for each $i = 1, \ldots, n$, the random observation X_i arises from a Poisson population with mean θ_i. That is, $f(x_i|\theta_i) = e^{-\theta_i}\theta_i^{x_i}/(x_i!)$, $x_i = 0, 1, 2, \ldots$. Then, $f(x_i) = \int_0^\infty e^{-\theta}\theta^{x_i}/(x_i!)dG(\theta) = a(x_i)h(x_i)$, where $a(x_i) = 1/x_i!$ and $h(x_i) = \int_0^\infty e^{-\theta}\theta^{x_i}dG(\theta)$, and $\varphi_i(x_i) = h(x_i+1)/h(x_i) \equiv \varphi(x_i)$. Let $\theta_0 > 0$ be the known standard level. The Bayes rule $\underset{\sim}{d}_B = (d_{1B}, \ldots, d_{nB})$ for this problem is:

$$d_{iB}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases}$$

Since the prior distribution G is unknown, it is not possible to apply the Bayes rule $\underset{\sim}{d}_B$ here. Therefore, in the following, empirical Bayes rules are constructed according to how much information we have about the prior distribution G.

2.1. A Nonparametric Empirical Bayes Rule

First, it is assumed that the prior distribution G is completely unknown. Thus, the nonparametric empirical Bayes approach is employed. Note that the Bayes rule $\underset{\sim}{d}_B$ is a monotone rule. That is, for each $i = 1, \ldots, n, d_{iB}(\underset{\sim}{x})$ is nondecreasing in x_i when all the other variables are kept fixed. This follows from the increasing property of $\varphi_i(x_i)$

which can be verified by noting that $f(x|\theta_i)$ has the monotone likelihood ratio. Thus, it is desirable that the considered empirical Bayes rules be monotone.

For each $i = 1, \ldots, n$, let $N_{in} = \max_{j \neq i} X_j - 1$. For each $y = 0, 1, \ldots, N_{in} + 1$, let

$$f_{in}(y) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} I_{\{y\}}(X_j), \tag{2.1}$$

$$h_{in}(y) = f_{in}(y)/a(y). \tag{2.2}$$

Since it is possible that $h_{in}(y)$ may be equal to 0, we define

$$\varphi_{in}(y) = [h_{in}(y+1) + \delta_n]/[h_{in}(y) + \delta_n], \tag{2.3}$$

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

By the forms of $\varphi_i(x_i)$ and $\varphi_{in}(x_i)$, it seems natural to use $\varphi_{in}(x_i)$ as an estimator of $\varphi_i(x_i)$ and one may obtain an empirical Bayes rule as follows: Select π_i as a good population if $\varphi_{in}(x_i) \geq \theta_0$, and exclude π_i as a bad one otherwise. However, this selection rule is not monotone since $\varphi_{in}(y)$ may not possess the increasing property. Thus, we consider a smoothed version of $\varphi_{in}(y)$. Let $\{\varphi_{in}^*(y)\}_{y=0}^{N_{in}}$ be the isotonic regression of $\{\varphi_{in}(y)\}_{y=0}^{N_{in}}$ with random weights $\{W_{in}(y)\}_{y=0}^{N_{in}}$, where $W_{in}(y) = [h_{in}(y) + \delta_n]a(y+1)$. For $y > N_{in}$, define $\varphi_{in}^*(y) = \varphi_{in}^*(N_{in})$. Therefore, $\varphi_{in}^*(y)$ is nondecreasing in y, $y = 0, 1, 2, \ldots$. We use $\varphi_{in}^*(x_i)$ to estimate $\varphi_i(x_i)$ and propose an empirical Bayes rule $\underset{\sim}{d}_n^* = (d_{1n}^*, \ldots, d_{nn}^*)$ as follows: For each $i = 1, \ldots, n$,

$$d_{in}^*(x_i|\underset{\sim}{x}(i)) \equiv d_{in}^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_{in}^*(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The performance of the preceding nonparametric empirical Bayes procedure will be discussed in Section 3.

2.2. A Parametric Empirical Bayes Rule

Here we assume that the prior distribution G is a member of gamma distribution family with unknown shape and scale parameters k and β, respectively. That is, G has a density function $g(\theta|k,\beta)$, where

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

Then, $X_1, \ldots, X_n$ are iid with marginal probability function $f(x) = \Gamma(x+k)\beta^k/[\Gamma(k)(1+\beta)^{x+k}x!]$, $x = 0, 1, 2, \ldots$. Also, $\varphi_i(x) = (x+k)/(1+\beta)$. A straight computation yields $\mu_1 \equiv E[X_i] = k/\beta$, $\mu_2 \equiv E[X_i^2] = (k+1)k/\beta^2 + k/\beta$. Thus, $\beta = \mu_1/(\mu_2 - \mu_1 - \mu_1^2)$ and $k = \mu_1^2/(\mu_2 - \mu_1 - \mu_1^2)$. Therefore, $\varphi_i(x) = [x(\mu_2 - \mu_1 - \mu_1^2) + \mu_1^2]/(\mu_2 - \mu_1^2)$.

For each $i = 1, \ldots, n$, let $\mu_{1n}(i) = \frac{1}{n-1}\sum_{\substack{j=1 \\ j \neq i}}^{n} X_j$ and $\mu_{2n}(i) = \frac{1}{n-1}\sum_{\substack{j=1 \\ j \neq i}}^{n} X_j^2$. That is, $\mu_{1n}(i)$ and $\mu_{2n}(i)$ are moment estimators of μ_1 and μ_2, respectively, based on $\underset{\sim}{X}(i)$. Note that it is possible that $\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \leq 0$ though $\mu_2 - \mu_1 - \mu_1^2 > 0$. Now, for each $i = 1, \ldots, n$ and $x_i = 0, 1, 2, \ldots$, define

$$\hat{\varphi}_{in}(x_i) = \begin{cases} \frac{x_i[\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i)] + \mu_{1n}^2(i)}{\mu_{2n}(i) - \mu_{1n}^2(i)} & \text{if } \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) > 0 \\ x_i & \text{otherwise.} \end{cases} \tag{2.5}$$

We then propose an empirical Bayes rule $\underset{\sim}{\hat{d}}_n = (\hat{d}_{1n}, \ldots, \hat{d}_{nn})$ as follows:

$$\hat{d}_{in}(x_i|\underset{\sim}{x}(i)) = \hat{d}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \hat{\varphi}_{in}(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

2.3. A Hierarchical Empirical Bayes Rule

Now, it is assumed that the prior distribution G is a gamma distribution with a known shape parameter k and an unknown scale parameter β. In this situation, the preceding parametric empirical Bayes approach can be applied here. However, since our purpose is to introduce the methods to incorporate data from different sources, a new method, called as hierarchical empirical Bayes, is used in the following.

Since β is a scale parameter, we assume that β has an improper prior $p(\beta) \propto \frac{1}{\beta}$, $\beta > 0$. Then the Bayes law yields the following posterior density function of β

$$p(\beta|x_1, \ldots, x_n) = \beta^{nk-1}(1+\beta)^{-b}\Gamma(b)/[\Gamma(nk)\Gamma(b-nk)],$$

where $b = nk + \sum_{j=1}^{n} x_j$. The posterior mean of β is

$$\beta_n = \begin{cases} \frac{nk}{\sum_{j=1}^{n} x_j - 1} & \text{if } \sum_{j=1}^{n} x_j \geq 2, \\ \infty & \text{if } \sum_{j=1}^{n} x_j \leq 1. \end{cases}$$

Now, for each $i = 1, \ldots, n$, define

$$\overline{\varphi}_{in}(x_i) = \begin{cases} (x_i + k)/(1 + \beta_n) & \text{if } \sum_{j=1}^{n} x_j \geq 2, \\ 0 & \text{if } \sum_{j=1}^{n} x_j \leq 1. \end{cases} \tag{2.7}$$

We then give an empirical Bayes rule $\overline{\underset{\sim}{d}}_n = (\overline{d}_{1n}, \ldots, \overline{d}_{nn})$ as follows:

$$\overline{d}_{in}(x_i | \underset{\sim}{x}(i)) = \overline{d}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{\varphi}_{in}(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

3. Asymptotic Optimality of the Proposed Empirical Bayes Rules

In this section, we investigate the asymptotic optimality of the proposed empirical Bayes rules.

Let $A(\theta_0) = \{x | \varphi(x) > \theta_0\}$ and $B(\theta_0) = \{x | \varphi(x) < \theta_0\}$. Define

$$M = \begin{cases} \min A(\theta_0) & \text{if } A(\theta_0) \neq \phi, \\ \infty & \text{otherwise,} \end{cases} \tag{3.1}$$

$$m = \begin{cases} \max B(\theta_0) & \text{if } B(\theta_0) \neq \phi, \\ -1 & \text{otherwise,} \end{cases} \tag{3.2}$$

where ϕ denotes the empty set.

By the increasing property of $\varphi(x)$ in the variable x, $m \leq M$; also $m < M$ if $A(\theta_0) \neq \phi$. Furthermore, $x \leq m$ iff $\varphi(x) < \theta_0$ and $y \geq M$ iff $\varphi(y) > \theta_0$. In the following, we consider only those priors G such that $\int_0^\infty \theta dG(\theta) < \infty$ and $m < \infty$. Note that the preceding requirements are always met if the prior distribution G is a member of gamma distribution family. Let $\underset{\sim}{d}_n = (d_{1n}, \ldots, d_{nn})$ be any of the three proposed empirical Bayes rules and let $(\varphi_{1n}(x_1),\ldots,\varphi_{nn}(x_n))$ be the corresponding empirical Bayes estimators. By the definitions of $\varphi^*_{in}(x_i)$, $\hat{\varphi}_{in}(x_i)$ and $\overline{\varphi}_{in}(x_i)$, $\varphi_{in}(x_i)$ is increasing in x_i when all the other variables x_j, $j \neq i$, are kept fixed. Thus, for each $i = 1,\ldots,n$,

$$\begin{aligned}
0 &\le r_i(G,d_{in}) - r_i(G,d_{iB}) \\
&= \sum_{x_i=0}^{m} [\theta_0 - \varphi(x_i)]P\{\varphi_{in}(x_i) \ge \theta_0\}f(x_i) + \sum_{x_i=M}^{\infty} [\varphi(x_i) - \theta_0]P\{\varphi_{in}(x_i) < \theta_0\}f(x_i) \\
&\le \sum_{x_i=0}^{m} [\theta_0 - \varphi(x_i)]P\{\varphi_{in}(m) \ge \theta_0\}f(x_i) + \sum_{x_i=M}^{\infty} [\varphi(x_i) - \theta_0]P\{\varphi_{in}(M) < \theta_0\}f(x_i) \\
&= b_1P\{\varphi_{in}(m) \ge \theta_0\} + b_2P\{\varphi_{in}(M) < \theta_0\},
\end{aligned} \tag{3.3}$$

where $0 \le b_1 = \sum_{x=0}^{m} [\theta_0 - \varphi(x)]f(x) < \infty$, $0 \le b_2 = \sum_{x=M}^{\infty} [\varphi(x) - \theta_0]f(x) < \infty$. The finiteness of both b_1 and b_2 is guaranteed by the assumption that $\int_0^\infty \theta dG(\theta) < \infty$.

From (3.3), we obtain:

$$\begin{aligned}
0 &\le r(G,\underset{\sim}{d}_n) - r(G) \\
&= \sum_{i=1}^{n}[r_i(G,d_{in}) - r_i(G,d_{iB})] \\
&\le \sum_{i=1}^{n}[b_1P\{\varphi_{in}(m) \ge \theta_0\} + b_2P\{\varphi_{in}(M) < \theta_0\}].
\end{aligned} \tag{3.4}$$

Therefore, it suffices to consider the asymptotic behavior of $P\{\varphi_{in}(m) \ge \theta_0\}$ and $P\{\varphi_{in}(M) < \theta_0\}$.

3.1 Asymptotic Optimality of $\underset{\sim}{d}^*_n$

We first present some useful results.

For each $i = 1,\ldots,n$ and $y = 0,1,\ldots,N_{in}$, let $\Psi_{in}(y) = \sum_{x=0}^{y} \varphi_{in}(x)W_{in}(x)$, $\Psi^*_{in}(y) = \sum_{x=0}^{y} \varphi^*_{in}(x)W_{in}(x)$ and $H_{in}(y) = \sum_{x=0}^{y} W_{in}(x)$ where $W_{in}(x)$, $x = 0,1,\ldots,N_{in}$, are the random weights defined in Section 3. From Barlow, et al. [1],

$$\Psi^*_{in}(y) \le \Psi_{in}(y) \text{ for all } y = 0,1,\ldots,N_{in}. \tag{3.5}$$

From Puri and Singh [15], the isotonic regression estimators $\varphi^*_{in}(x),\ x = 0, 1, \ldots, N_{in}$, can be rewritten as:

$$\varphi^*_{in}(x) = \min_{x \le y \le N_{in}} \left[\frac{\Psi_{in}(y) - \Psi^*_{in}(x-1)}{H_{in}(y) - H_{in}(x-1)} \right],\quad x = 0, 1, \ldots, N_{in}, \tag{3.6}$$

where $\Psi^*_{in}(-1) = H_{in}(-1) \equiv 0$. Thus, from (3.5) and (3.6),

$$\varphi^*_{in}(x) \ge \min_{x \le y \le N_{in}} \left[\frac{\Psi_{in}(y) - \Psi_{in}(x-1)}{H_{in}(y) - H_{in}(x-1)} \right],\quad x = 0, 1, \ldots, N_{in}, \tag{3.7}$$

where $\Psi_{in}(-1) \equiv 0$.

The following Lemma is taken from Liang [13].

Lemma 3.1. Let $\{a_m\}$ be a sequence of real numbers and let $\{b_m\}$ be a sequence of positive numbers such that $b_m \le 1$ and b_m is nonincreasing in m. Then, for each positive constant c,

$$\sup_{n \ge 1} \left| \sum_{m=1}^{n} a_m b_m \right| \ge (>) c \Rightarrow \sup_{n \ge 1} \left| \sum_{m=1}^{n} a_m \right| \ge (>) c.$$

Lemma 3.2. Define a function $Q(y) = \theta_0 \sum_{x=M}^{y} f(x) \frac{a(x+1)}{a(x)} - \sum_{x=M}^{y} f(x+1)$ on the set $\{y | y = M, M+1, \ldots\}$. Then, $Q(y)$ is a decreasing function of y. Hence $\max_{y \ge M} Q(y) = Q(M) = f(M) \frac{a(M+1)}{a(M)} [\theta_0 - \varphi(M)] < 0$.

Proof: $Q(y+1) - Q(y) = f(y+1) \frac{a(y+2)}{a(y+1)} [\theta_0 - \varphi(y+1)] < 0$ since $y+1 > M$ and thus $\varphi(y+1) \ge \varphi(M) > \theta_0$. Thus, $Q(y)$ is a decreasing function of y which leads to the result of this lemma. □

Theorem 3.3. $P\{\varphi^*_{in}(M) < \theta_0\} \le O(\exp(-\tau_1 n))$
where $\tau_1 = \min(2(Q(M) \min(1, \theta_0^{-1})/8)^2,\ \ln[F(M)]^{-1}) > 0$, and $F(\cdot)$ is the marginal distribution of X_i's.

Proof: $P\{\varphi^*_{in}(M) < \theta_0\}$

$$= P\{\varphi^*_{in}(M) < \theta_0,\ N_{in} < M\} + P\{\varphi^*_{in}(M) < \theta_0,\ N_{in} \ge M\}. \tag{3.8}$$

Now,

$$P\{\varphi^*_{in}(M) < \theta_0,\ N_{in} < M\} \le [F(M)]^{n-1} = O(\exp(-n \ln[F(M)]^{-1})). \tag{3.9}$$

Also, from (2.1)–(2.3), (3.7), Lemma 3.2, and by the definitions of $\Psi_{in}(y)$ and $H_{in}(y)$, straightforward computation yields the following:

$$\begin{aligned}
E &\equiv \{\varphi^*_{in}(M) < \theta_0,\ N_{in} \ge M\} \\
&\subset \{\Psi_{in}(y) - \Psi_{in}(M-1) < \theta_0[H_{in}(y) - H_{in}(M-1)] \text{ for some } y,\ M \le y \le N_{in}\} \\
&\subset \bigcup_{y \ge M} \left\{ \sum_{x=M}^{y} \Delta_{in}(x+1) - \theta_0 \sum_{x=M}^{y} \Delta_{in}(x) \frac{a(x+1)}{a(x)} < (\theta_0 - 1)\delta_n \sum_{x=M}^{y} a(x+1) \right. \\
&\qquad \left. + Q(M) \right\} \\
&\equiv E_1. \qquad (3.10)
\end{aligned}$$

where $\Delta_{in}(x) = f_{in}(x) - f(x)$. Since $a(x) \ge 0$ for all $x = 0, 1, \ldots$, and $\sum_{x=0}^{\infty} a(x) < \infty$ and $\delta_n = o(1)$, then, for sufficiently large n, $(\theta_0 - 1)\delta_n \sum_{x=M}^{y} a(x+1) + Q(M) < Q(M)/2 < 0$ for all $y \ge M$. Note that $a(x+1)/a(x) = (x+1)^{-1}$, which is positive, bounded above by 1, and decreasing in x for $x = 0, 1, 2, \ldots$. By the preceding facts and Lemma 3.1, we obtain:

$$\begin{aligned}
E_1 &\subset \bigcup_{y \ge M} \left\{ \left| \sum_{x=M}^{y} \Delta_{in}(x+1) \right| > -\frac{Q(M)}{4} \text{ or } \left| \sum_{x=M}^{y} \Delta_{in}(x) \frac{a(x+1)}{a(x)} \right| > -\frac{Q(M)}{4\theta_0} \right\} \\
&\subset \bigcup_{y \ge M} \left\{ \left| \sum_{x=M}^{y} \Delta_{in}(x+1) \right| > -\frac{Q(M)}{4} \text{ or } \left| \sum_{x=M}^{y} \Delta_{in}(x) \right| > -\frac{Q(M)}{4\theta_0} \right\} \\
&\subset \left\{ \sup_{y \ge 0} |F_{in}(y) - F(y)| > -Q(M)\min(1, \theta_0^{-1})/8 \right\} \qquad (3.11)
\end{aligned}$$

where $F_{in}(y)$ is the empirical distribution based on $\underset{\sim}{X}(i)$.

From (3.10) and (3.11), we obtain

$$\begin{aligned}
&P\{\varphi^*_{in}(M) < \theta_0,\ N_{in} \ge M\} \\
&\le P\{\sup_{y \ge 0} |F_{in}(y) - F(y)| > -Q(M)\min(1, \theta_0^{-1})/8\} \qquad (3.12) \\
&\le d\exp\{-2n(Q(M)\min(1, \theta_0^{-1})/8)^2\}
\end{aligned}$$

where the last inequality follows from Lemma 2.1 of Schuster [19].

Now, let $\tau_1 = \min(2(Q(M)\min(1, \theta_0^{-1})/8)^2, \ln[F(M)]^{-1})$. Clearly $\tau_1 > 0$. Combining (3.8), (3.9) and (3.12) gives the result of this theorem. □

Theorem 3.4. $P\{\varphi_{in}^*(m) \geq \theta_0\} \leq O(\exp(-\tau_2 n))$
where $\tau_2 = [R^*(m) \min(1, \theta_0^{-1})]^2/8 > 0$ and $R^*(m)$ is defined below.

Proof: From (2.1)–(2.3) and by the definition of $\varphi_{in}^*(m)$,

$$
\begin{aligned}
&\{\varphi_{in}^*(m) \geq \theta_0\} \\
\subset&\{\varphi_{in}(x) \geq \theta_0 \text{ for some } 0 \leq x \leq m\} \qquad (3.13)\\
\subset&\{a(x)\Delta_{in}(x+1) - \theta_0 a(x+1)\Delta_{in}(x) > R(x) - a(x)a(x+1)\delta_n[1-\theta_0] \text{ for some } 0 \leq x \leq m\},
\end{aligned}
$$

where $R(x) = -a(x)f(x+1) + \theta_0 a(x+1)f(x) = a(x+1)f(x)[-\varphi(x) + \theta_0] > 0$ since $\theta_0 - \varphi(x) \geq \theta_0 - \varphi(m) > 0$, by the definition of m and the fact that $0 \leq x \leq m$. Thus, $R^*(m) = \min_{0 \leq x \leq m} R(x) > 0$ and therefore, for sufficiently large n, $R(x) - a(x)a(x+1)\delta_n[1 - \theta_0] \geq R^*(m)/2$ since $\delta_n = o(1)$. Therefore, from (3.13) and by Theorem 1 of Hoeffding [11],

$$
\begin{aligned}
&P\{\varphi_{in}^*(m) \geq \theta_0\} \\
\leq&\sum_{x=0}^{m}[P\{\Delta_{in}(x+1) > R^*(m)/(4a(x))\} + P\{\Delta_{in}(x) < -R^*(m)/(4\theta_0 a(x+1))\}] \\
\leq&\sum_{x=0}^{m}[c\exp\{-2n[R^*(m)/(4a(x))]^2\} + c\exp\{-2n[R^*(m)/(4\theta_0 a(x+1))]^2\}] \\
=&O(\exp(-\tau_2 n)).
\end{aligned}
$$
□

Based on the preceding discussions, we have the following result.

Theorem 3.5. Assume that the prior distribution G is such that $\int_0^\infty \theta dG(\theta) < \infty$ and $m < \infty$. Then, for the empirical Bayes rule $\underset{\sim}{d}_n^*$, $0 \leq r(G, \underset{\sim}{d}_n^*) - r(G) \leq O(\exp(-\tau n + \ln n))$ where $\tau = \min(\tau_1, \tau_2) > 0$.

Proof: By (3.4), Theorem 3.3 and Theorem 3.4, we have

$$
\begin{aligned}
0 \leq r(G, \underset{\sim}{d}_n^*) - r(G) &\leq O(n \exp(-\tau n)) \\
&= O(\exp(-\tau n + \ln n)).
\end{aligned}
$$
□

3.2. Asymptotic Optimality of $\hat{\underset{\sim}{d}}_n$

We let $M_1(t)$ and $M_2(t)$ denote the moment generating functions of X_1 and X_1^2, respectively. For each real value a, define

$$m_1(a) = \inf_t e^{-at} M_1(t)$$
$$m_2(a) = \inf_t e^{-at} M_2(t)$$

where the infimum is taken with respect to real values of t.

Lemma 3.6. For any positive constant c,

$$0 \le m_i(\mu_i + c) < 1, \; 0 \le m_i(\mu_i - c) < 1 \quad \text{for } i = 1, 2,$$

where $\mu_1 = E[X_1]$ and $\mu_2 = E[X_1^2]$.

Proof: For the fixed real value a, consider the function

$$S_1(t) = e^{-at} M(t) = E[e^{t(X_1 - a)}].$$

We have

$$S_1^{(1)}(t) = E[(X_1 - a)e^{t(X_1 - a)}],$$
$$S_1^{(2)}(t) = E[(X_1 - a)^2 e^{t(X_1 - a)}],$$

where $S_1^{(j)}(t)$ denotes the j–th derivative of $S_1(t)$ with respect to t.

Since $S_1^{(2)}(t) > 0$ for all t, $S_1(t)$ is a convex function. Also, $S_1^{(1)}(0) = E[X_1 - a] <$ $(=,>)0$ iff $\mu_1 < (=,>)a$. Thus, as $\mu_1 < a$, $S_1^{(1)}(0) < 0$, which implies that $S_1(t)$ is strictly decreasing in a neighborhood of point zero. Also, $S_1(0) = 1$. Therefore, $m_1(a) < 1$ if $\mu_1 < a$. Similarly, we can also obtain the following result: $m_1(a) < 1$ if $\mu_1 > a$. Now, by the definition, $m_1(a) \ge 0$. These results yields that $0 \le m_1(\mu_1 + c) < 1$ and $0 \le m_1(\mu_1 - c) < 1$ for any positive constant c.

The results that $0 \le m_2(\mu_2 + c) < 1$ and $0 \le m_2(\mu_2 - c) < 1$ for any positive constant c follow from similar arguments. □

Lemma 3.7. For each $i = 1, \ldots, n$, let $\mu_{1n}(i)$ and $\mu_{2n}(i)$ be the moment estimators of μ_1 and μ_2, respectively, which are defined in Section 2. Then, for any positive constant c,

(a) $P\{\mu_{1n}(i) - \mu_1 \le -c\} \le [m_1(\mu_1 - c)]^{n-1}$,

(b) $P\{\mu_{1n}(i) - \mu_1 \ge c\} \le [m_1(\mu_1 + c)]^{n-1}$,

(c) $P\{\mu_{2n}(i) - \mu_2 \le -c\} \le [m_2(\mu_2 - c)]^{n-1}$ and

(d) $P\{\mu_{2n}(i) - \mu_2 \ge c\} \le [m_2(\mu_2 + c)]^{n-1}$.

Proof: This lemma is a direct application of Chernoff [2]. The proof can be completed by noting the fact that $0 < E[X_1] < \infty$ and $0 < E[X_1^2] < \infty$. □

Let $\mu = \mu_2 - \mu_1 - \mu_1^2$. Thus, $\mu > 0$, see Section 2. Define $A = \max(m_2(\mu_2 - \frac{\mu}{3}), m_1(\mu_1 + \frac{\mu}{3}), m_1(\mu_1 + \frac{\mu}{9\mu_1}), m_1(2\mu_1))$. By Lemma 3.6, $0 \le A < 1$.

Lemma 3.8. $P\{\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \le 0\} \le O(\exp(-\alpha_1 n))$ where $\alpha_1 = \begin{cases} -\ln A & \text{if } A > 0, \\ \infty & \text{if } A = 0. \end{cases}$

Proof: $P\{\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \le 0\}$

$$\begin{aligned} &= P\{[\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i)] - [\mu_2 - \mu_1 - \mu_1^2] \le -\mu\} \\ &\le P\left\{\mu_{2n}(i) - \mu_2 \le -\frac{\mu}{3}\right\} + P\left\{\mu_{1n}(i) - \mu_1 \ge \frac{\mu}{3}\right\} \\ &\quad + P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3}\right\}. \end{aligned}$$

By Lemma 3.7,

$$\begin{aligned} &P\left\{\mu_{2n}(i) - \mu_2 \le -\frac{\mu}{3}\right\} \le \left[m_2\left(\mu_2 - \frac{\mu}{3}\right)\right]^{n-1}, \\ &P\left\{\mu_{1n}(i) - \mu_1 \ge \frac{\mu}{3}\right\} \le \left[m_1\left(\mu_1 + \frac{\mu}{3}\right)\right]^{n-1}, \text{ and} \\ &P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3}\right\} \\ &= P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3},\ \mu_{1n}(i) < 2\mu_1\right\} + P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3},\ \mu_{1n}(i) \ge 2\mu_1\right\} \\ &\le P\left\{\mu_{1n}(i) - \mu_1 \ge \frac{\mu}{9\mu_1}\right\} + P\{\mu_{1n}(i) - \mu_1 \ge \mu_1\} \qquad (3.14) \\ &\le \left[m_1\left(\mu_1 + \frac{\mu}{9\mu_1}\right)\right]^{n-1} + [m_1(2\mu_1)]^{n-1}. \end{aligned}$$

Combining the preceding results, the lemma follows. □

Theorem 3.9. $P\{\hat{\varphi}_{in}(M) < \theta_0\} \le O(\exp(-\alpha_2 n))$ for some positive constant α_2.

Proof: $P\{\hat{\varphi}_{in}(M) < \theta_0\} = P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \le 0\}$

$$+P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) > 0\}, \tag{3.15}$$

where

$$\begin{aligned} &P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \le 0\} \\ &\le O(\exp(-\alpha_1 n)) \text{ by Lemma 3.8.} \end{aligned} \tag{3.16}$$

Now, let $q(M) = M(\mu_2 - \mu_1 - \mu_1^2) + \mu_1^2 - \theta_0(\mu_2 - \mu_1^2)$. By definition of M, $q(M) > 0$. Thus,

$$\begin{aligned} &P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) > 0\} \\ \le& P\{(M-\theta_0)\mu_{2n}(i) - M\mu_{1n}(i) - (M-1-\theta_0)\mu_{1n}^2(i) < 0\} \\ =& P\{(M-\theta_0)(\mu_{2n}(i) - \mu_2) - M(\mu_{1n}(i) - \mu_1) - (M-1-\theta_0)(\mu_{1n}^2(i) - \mu_1^2) < -q(M)\} \\ \le& P\left\{(M-\theta_0)(\mu_{2n}(i) - \mu_2) < -\frac{q(M)}{3}\right\} + P\left\{M(\mu_{1n}(i) - \mu_1) > \frac{q(M)}{3}\right\} \\ &+ P\left\{(M-1-\theta_0)(\mu_{1n}^2(i) - \mu_1^2) > \frac{q(M)}{3}\right\}. \end{aligned} \tag{3.17}$$

By Lemma 3.7,

$$P\left\{M(\mu_{1n}(i) - \mu_1) > \frac{q(M)}{3}\right\} \le \left[m_1\left(\mu_1 + \frac{q(M)}{3M}\right)\right]^{n-1}. \tag{3.18}$$

$$P\left\{(M-\theta_0)(\mu_{2n}(i) - \mu_2) < -\frac{q(M)}{3}\right\} \le \begin{cases} \left[m_2\left(\mu_2 - \frac{q(M)}{3(M-\theta_0)}\right)\right]^{n-1} & \text{if } M-\theta_0 > 0, \\ 0 & \text{if } M-\theta_0 = 0, \\ \left[m_2\left(\mu_2 + \frac{q(M)}{3(\theta_0-M)}\right)\right]^{n-1} & \text{if } M-\theta_0 < 0, \end{cases} \tag{3.19}$$

and analogous to (3.14),

$$\begin{aligned} &P\left\{(M-1-\theta_0)(\mu_{1n}^2(i) - \mu_1^2) > \frac{q(M)}{3}\right\} \\ \le& \begin{cases} \left[m_1\left(\mu_1 + \frac{q(M)}{6(M-1-\theta_0)\mu_1}\right)\right]^{n-1} + [m_1(2\mu_1)]^{n-1} & \text{if } M-1-\theta_0 > 0, \\ 0 & \text{if } M-1-\theta_0 = 0, \\ \left[m_1\left(\mu_1 + \frac{q(M)}{6(M-1-\theta_0)\mu_1}\right)\right]^{n-1} & \text{if } M-1-\theta_0 < 0. \end{cases} \end{aligned} \tag{3.20}$$

Combining (3.15)–(3.20), and by Lemma 3.6, it follows that there exists a positive constant, say α_2, such that $P\{\hat{\varphi}_{in}(M) < \theta_0\} \le O(\exp(-\alpha_2 n))$. □

Theorem 3.10. $P\{\hat{\varphi}_{in}(m) \geq \theta_0\} \leq O(\exp(-\alpha_3 n))$ for some positive constant α_3.

Proof: The proof is analogous to that of Theorem 3.9. We omit the details here. □

The following theorem is a direct result of (3.4) and Theorems 3.9 and 3.10.

Theorem 3.11. Let $\hat{d}_n$ be the empirical Bayes rule defined in Section 2. Assume that the prior distribution G is a member of the gamma distribution family. Then,

$$0 \leq r(G, \hat{d}_n) - r(G) \leq O(\exp(-\alpha n + \ln n)),$$

where $\alpha = \min(\alpha_2, \alpha_3) > 0$.

3.3. Asymptotic Optimality of $\bar{d}_n$.

Theorem 3.12. Let $\bar{d}_n$ be the empirical Bayes rule defined in Section 2. Assume that the prior distribution G is a member of gamma distribution family $\Gamma(k, \beta)$, where k is a known positive constant. Then,

$$0 \leq r(G, \bar{d}_n) - r(G) \leq O(\exp(-\gamma n + \ln n))$$

for some positive constant γ.

Note that the statistical model considered here is simpler than that of Section 3.2. Thus, the proof for Theorem 3.12 is analogous to and simpler than that for Theorem 3.11. We omit the details of the proof.

4. Small Sample Performance: Simulation Study

A Monte Carlo study was designed to investigate the performance of the three empirical Bayes procedures. We let the prior distribution G to be a gamma distribution with $k = 1$ and $\beta = 1$. With this specified prior distribution, $f(x) = 2^{-x-1}$ and $\varphi(x) = \frac{x+1}{2}$, $x = 0, 1, 2, \ldots$. Also, the minimum Bayes risk for each of the n component decision problems is $r_i(G, d_{iB}) = e^{-\theta_0} - 4^{-\theta_0}$, where $\theta_0 > 0$ is the known standard. Therefore, the total minimum Bayes risk $r(G) = \sum_{i=1}^{n} r_i(G, d_{iB}) = n[e^{-\theta_0} - 4^{-\theta_0}]$.

Let $\underset{\sim}{d}_n = (d_{1n}, \ldots, d_{nn})$ be any of the three proposed empirical Bayes procedures. Since $r(G, \underset{\sim}{d}_n) - r(G) = \sum_{i=1}^{n} [r_i(G, d_{in}) - r_i(G, d_{iB})] = n[r_1(G, d_{1n}) - r_1(G, d_{1B})]$, in the following, we have simulated the difference $r_1(G, d_{1n}) - r_1(G, d_{1B})$ by $R_1(G, d_{1n}) - r_1(G, d_{1B})$, which is the difference between the conditional Bayes risk of d_{1n} conditional on $\underset{\sim}{X}(1)$ and the minimum Bayes risk. We have then used $n[R_1(G, d_{1n}) - r_1(G, d_{1B})]$ as an estimator of the total difference $r(G, \underset{\sim}{d}_n) - r(G)$.

The simulation scheme used in this paper is described as follows:

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

(2) Based on the values $X_1, \ldots, X_n$, construct the empirical Bayes procedure d_{1n} and compute the conditional difference $D(d_{1n}) \equiv R_1(G, d_{1n}) - r_1(G, d_{1B})$. It should be noted that for the nonparametric empirical Bayes procedure $\underset{\sim}{d}_n^*$, the sequence $\{\delta_n\}_{n=1}^{\infty}$ is chosen such that $\delta_n = n^{-\frac{1}{2}}$.

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

Tables 1–3 list some simulation results on the performance of the three empirical Bayes procedure $\underset{\sim}{d}_n^*, \hat{\underset{\sim}{d}}_n$ and $\bar{\underset{\sim}{d}}_n$, respectively, for the case where $\theta_0 = 1.5$. The notation $\text{SE}(\overline{D}(d_{1n}))$ is used to denote the estimated standard errors of the corresponding estimate $\overline{D}(d_{1n})$.

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

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

Table 1. Performance of $\underset{\sim}{d}_n^*$ as described in Section 4.

n	$D(d_{1n}^*)$	SE($D(d_{1n}^*)$)	$nD(d_{1n}^*)$	$r(G)$
10	0.02419	0.00109	0.24190	0.98130
20	0.01877	0.00088	0.37541	1.96260
30	0.02493	0.00111	0.74800	2.94390
40	0.02787	0.00119	1.11474	3.92521
50	0.02680	0.00116	1.33987	4.90651
60	0.02553	0.00110	1.53189	5.88781
70	0.02603	0.00106	1.82214	6.86911
80	0.02508	0.00102	2.00718	7.85041
90	0.02407	0.00096	2.16587	8.83171
100	0.02404	0.00101	2.40478	9.81301
120	0.02306	0.00093	2.76767	11.77562
140	0.02149	0.00092	3.00801	13.73822
160	0.01961	0.00082	3.13834	15.70083
180	0.01897	0.00080	3.41499	17.66343
200	0.01810	0.00076	3.61943	19.62603
220	0.01738	0.00072	3.82250	21.58864
240	0.01726	0.00072	4.14188	23.55124
260	0.01634	0.00070	4.24938	25.51384
280	0.01589	0.00067	4.44938	27.47644
300	0.01530	0.00066	4.59140	29.43905

Table 2. Performance of $\hat{\underset{\sim}{d}}_n$ as described in Section 4.

n	$\overline{D}(\hat{d}_{1n})$	$\mathrm{SE}(\overline{D}(\hat{d}_{1n}))$	$n\overline{D}(\hat{d}_{1n})$	$r(G)$
10	0.01803	0.00118	0.18030	0.98130
20	0.01609	0.00108	0.32187	1.96260
30	0.01305	0.00096	0.39154	2.94390
40	0.01167	0.00091	0.46678	3.92521
50	0.00925	0.00099	0.46245	4.90651
60	0.00581	0.00062	0.34874	5.88781
70	0.00440	0.00050	0.30784	6.86911
80	0.00318	0.00041	0.25453	7.85041
90	0.00236	0.00036	0.21270	8.83171
100	0.00174	0.00030	0.17378	9.81301
120	0.00109	0.00024	0.13124	11.77562
140	0.00078	0.00020	0.10876	13.73822
160	0.00061	0.00017	0.09750	15.70083
180	0.00041	0.00014	0.07453	17.66343
200	0.00032	0.00013	0.06406	19.62603
220	0.00013	0.00006	0.02750	21.58864
240	0.00006	0.00004	0.01500	23.55124
260	0.00003	0.00003	0.00813	25.51384
280	0.	0.	0.	27.47644
300	0.	0.	0.	29.43905

Table 3. Performance of $\underset{\sim}{\overline{d}}_n$ as described in Section 4.

n	$\overline{D}(d_{1n})$	$\mathrm{SE}(\overline{D}(d_{1n}))$	$n\overline{D}(d_{1n})$	$r(G)$
10	0.00666	0.00044	0.06656	0.98130
20	0.00266	0.00029	0.05313	1.96260
30	0.00134	0.00020	0.04031	2.94390
40	0.00072	0.00015	0.02875	3.92521
50	0.00059	0.00013	0.02969	4.90651
60	0.00019	0.00007	0.01125	5.88781
70	0.	0.	0.	6.86911
80	0.	0.	0.	7.85041
90	0.00003	0.00003	0.00281	8.83171
100	0.	0.	0.	9.81301

REFERENCES

[1] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

[2] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23**, 493–507.

[3] Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096–1121.

[4] Gilliland, D.C. and Hannan, J. (1986). The finite state compound decision problem, equivalence and restricted risk components. *Adaptive Statistical Procedures and Related Topics* (Ed. J. Van Ryzin), IMS Lecture Notes–Monograph Series, Vol. 8, 129–145.

[5] Gilliland, D.C., Hannan, J. and Huang, J.S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. *Ann. Statist.* **4**, 1101–1112.

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

[7] Gupta, S.S. and Miescke, K.J. (1985). Minimax multiple t–test for comparing k normal populations with a control. *J. Statist. Plann. Inference* **12**, 161–169.

[8] Gupta, S.S. and Panchapakesan, S. (1979). *Multiple Decision Procedures.* Wiley, New York.

[9] Gupta, S.S. and Panchapakesan, S. (1985). Subset selection procedures: review and assessment. *Amer. J. Math. Management Sci.* **5**, 235–311.

[10] Gupta, S.S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.* **29**, 235–244.

[11] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables.

J. Amer. Statist. Assoc. **58**, 13–30.

[12] Lehmann, E.L. (1961). Some Model I problems of selection. *Ann. Math. Statist.* **32**, 990–1012.

[13] Liang, T. (1989). Empirical Bayes selection for the highest probability of success in negative binomial distributions. Technical Report #89–18C, Department of Statistics, Purdue University, West Lafayette, Indiana.

[14] Miescke, K.J. (1981). Γ–minimax selection procedures in simultaneous testing problems. *Ann. Statist.* **9**, 215–220.

[15] Puri, P.S. and Singh, H. (1990). On recursive formulas for isotonic regression useful for statistical inference under order restrictions. *J. Statist. Plann. Inference* **24**, 1–11.

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

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

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

[19] Schuster, E.F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40**, 1187–1195.

[20] Spjøtvoll, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Statist.* **43**, 398–411.

[21] Vardeman, S.B. (1978). Bounds on the empirical Bayes and compound risks of truncated version of Robbins's estimator of a binomial parameter. *J. Statist. Plann. Inference* **2**, 245–252.

[22] Vardeman, S.B. (1980). Admissible solutions of k–extended finite state set and sequence compound decision problems. *J. Multivariate Anal.* **10**, 426–441.