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

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

PURDUE UNIVERSITY

CENTER FOR STATISTICAL DECISION SCIENCES AND DEPARTMENT OF STATISTICS

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

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

* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620 at Purdue University.

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

by

Shanti S. Gupta and TaChen Liang

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

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 $0(\exp(-cn + \ln n))$ for some positive constant c, where the value of c varies depending on the empirical Bayes rule used. (KR)

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 the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964, DMS-8702620 at Purdue University.

1. Introduction

We consider a decision problem involving n components as follows. Let $\pi_1, \ldots, \pi_n$ denote n independent populations of the n components, respectively, where population π_i is characterized by a parameter θ_i, $i = 1, \ldots, n$. For the given decision problem, let a_i denote an action for the i-th component and let $L(\theta_i, a_i)$ be the corresponding loss function. Thus, $L^*(\underset{\sim}{\theta}, \underset{\sim}{a}) = \sum_{i=1}^{n} L(\theta_i, a_i)$ is the total loss where $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n)$ and $\underset{\sim}{a} = (a_1, \ldots, a_n)$. Suppose that for each $i = 1, \ldots, n$, the parameter θ_i is a realization of a random variable Θ_i, which has a prior distribution G_i over the parameter space Ω_i. Let X_i denote a random observation arising from population π_i with probability density function $f_i(x|\theta_i)$. Let d_i be a decision rule defined on the sample space $\mathcal{X}_i$ of X_i for the i-th component problem. Then, under some regularity conditions, the total Bayes risk of the decision rule $\underset{\sim}{d} = (d_1, \ldots, d_n)$ is:

$$r(\underset{\sim}{G}, \underset{\sim}{d}) = \sum_{i=1}^{n} r_i(G_i, d_i) \tag{1.1}$$

where $\underset{\sim}{G} = G_1 \times \ldots \times G_n$, and

$$\begin{aligned} r_i(G_i, d_i) &= \int_{\Omega_i} \int_{\mathcal{X}_i} L(\theta,\ d_i(x)) f_i(x|\theta) dx dG_i(\theta) \\ &= \int_{\mathcal{X}_i} \left[\int_{\Omega_i} L(\theta,\ d_i(x)) dG_i(\theta|x) \right] f_i(x) dx \end{aligned} \tag{1.2}$$

where $G_i(\theta|x)$ is the posterior distribution of Θ_i given $X_i = x$ and $f_i(x)$ is the marginal probability density function of X_i. Thus, for the i-th component problem, the Bayes rule is the one which minimizes $\int_{\Omega_i} L(\theta,\ d_i(x)) dG_i(\theta|x)$ among the class of decision rules for the i-th component decision problem. The overall minimum Bayes risk is

$$r(\underset{\sim}{G}, \underset{\sim}{d}_B) = \sum_{i=1}^{n} r_i(G_i, d_{iB})$$

where $\underset{\sim}{d}_B = (d_{1B}, \ldots, d_{nB})$ and d_{iB} is a Bayes rule for the i-th component decision problem, $i = 1, \ldots, n$.

When the prior distributions G_i, $i = 1, \ldots, n$, are unknown, the Bayes rule cannot be applied. However, in many situations, the n-component decision problems may share the same or similar properties. When this occurs, one may incorporate all the information obtained from different sources and make an appropriate decision for each of the n components. This idea is analogous to the empirical Bayes approach of Robbins (1956,1964). Thus, in the following, we let d_i denote a decision rule for the i–th component problem, where d_i is now defined on the sample space $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ of $\underset{\sim}{X} = (X_1, \ldots, X_n)$; also, denote $d_i(x_1, \ldots, x_n)$ by $d_i(x_i|\underset{\sim}{x}(i))$ where $\underset{\sim}{x}(i) = (x_1, \ldots, x_{i-1},\ x_{i+1}, \ldots, x_n)$. Then,

$$r_i(G_i, d_i) = E_i\left[\int_{\Omega_i}\int_{\mathcal{X}_i} L(\theta,\ d_i(x|\underset{\sim}{X}(i)) f_i(x|\theta) dx dG_i(\theta)\right],$$

where the expectation E_i is taken with respect to the marginal distribution of $\underset{\sim}{X}(i) = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. Since $r_i(G_i, d_{iB})$ is the minimum Bayes risk for the i–th component problem, $r_i(G_i, d_i) - r_i(G_i, d_{iB}) \geq 0$ for each $i = 1, \ldots, n$, and therefore, $r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B) = \sum_{i=1}^{n}[r_i(G_i, d_i) - r_i(G_i, d_{iB})] \geq 0$.

In certain compound decision problems, the average $\frac{1}{n}[r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B)]$ has been used as a measure of the performance of the decision rule $\underset{\sim}{d}$. The asymptotic behavior of $\frac{1}{n}[r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B)]$ has been investigated extensively; for example, see Vardeman (1978,1980), Gilliland and Hannan (1986) and Gilliland, Hannan and Huang (1976), among others. Many of the results indicate that $\frac{1}{n}[r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B)]$ tends to 0 as n tends to infinity. However, so far as we know, the asymptotic behavior of the regret value $r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B)$ has not been investigated since it seems that $r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B)$ might tend to infinity when n tends to infinity. Very surprisingly, we find that in certain compound empirical Bayes decision problems. $r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B) \to 0$ as $n \to \infty$. This result indicates the advantage of incorporating all the information from different sources for making a decision for each of the n component problems.

In this paper, we investigate the asymptotic optimality properties of certain empirical Bayes procedures for simultaneous testing problems. The regret value $r(\underset{\sim}{G}, \underset{\sim}{d}) - r(\underset{\sim}{G}, \underset{\sim}{d}_B)$ is used as a measure of the performance of the decision rule $\underset{\sim}{d}$. The general framework of the empirical Bayes decision problems under study is formulated in Section 2. Then, examples are given and used to illustrate how to incorporate information from different sources. For each of them, the corresponding convergence rate is investigated.

2. Formulation of the Empirical Bayes Decision Problem

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 (1955) and Gupta and Sobel (1958) have considered problems of selecting a subset containing all populations better than a control using some natural procedures. Lehmann (1961) and Spjøtvoll (1972) have treated the problem using methods from the theory of testing hypotheses. Randles and Hollander (1971), Gupta and Kim (1980), Miescke (1981) and Gupta and Miescke (1985) have derived optimal procedures via minimax or gamma–minimax approaches. The reader is referred to Gupta and Panchapakesan (1979,1985) 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 $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n) | f(x|\theta_i)$ 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:

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \sum_{i=1}^{n} a_i(\theta_0 - \theta_i) I(\theta_0 - \theta_i) + \sum_{i=1}^{n} (1 - a_i)(\theta_i - \theta_0) I(\theta_i - \theta_0) \tag{2.1}$$

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 $[0,1]^k$ 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

$$r(G, \underset{\sim}{d}) = \sum_{i=1}^{n} r_i(G, d_i) \tag{2.2}$$

where

$$r_i(G, d_i) = \int_{\mathcal{X}} [\theta_0 - \varphi_i(x_i)] d_i(\underset{\sim}{x}) \prod_{j=1}^{k} f(x_j) d\underset{\sim}{x} + C \tag{2.3}$$

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 $C = \int_{\mathcal{X}_i} \int_{\theta_0}^{\infty} (\theta - \theta_0) f(x|\theta) dG(\theta) dx$.

Since the value C is independent of the selection rule $\underset{\sim}{d}$, from (2.3), a Bayes rule, say $\underset{\sim}{d}_B = (d_{1B}, \ldots, d_{nB})$ is clearly given by

$$d_{iB}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_i(x_i) \geq \theta_0, \\ 0 & \text{otherwise}, \end{cases} \tag{2.4}$$

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 (1956,1964) 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:

$$d_{in}(x_i|\underset{\sim}{x}(i)) \equiv d_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_{in}(x_i|\underset{\sim}{x}(i)) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

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

$$r(G, \underset{\sim}{d}_n) = \sum_{i=1}^{n} r_i(G, d_{in}) \tag{2.6}$$

where

$$r_i(G, d_{in}) = E_i\left[\int_{\mathcal{X}_i} [\theta_0 - \varphi_i(x_i)] d_{in}(x_i|\underset{\sim}{X}(i)) f(x_i) dx_i\right] + C \tag{2.7}$$

where the expectation E_i is taken with respect to $\underset{\sim}{X}(i)$. 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 2.1

(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. The weak asymptotic optimality of compound decision rules has been studied in the literature by many authors, notably Vardeman (1978,1980), Gilliland and Hannan (1986), and Gilliland, Hannan and Huang (1976), 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 the following, we consider the problem of selecting good Poisson populations, and use this as an example to illustrate how to incorporate information from different sources for making decisions. Selection rules are constructed according to how much we know about the prior distribution G. The strong asymptotic optimality of the selection rules is investigated. The associated convergence rates of selection rules are also established.

3. 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, $f(x_i) = \int_0^\infty e^{-\theta}\theta^{x_i}/(x!)dG(\theta) = a(x_i)h(x_i)$, 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.

3.1. 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 $N_{in} = \max\limits_{j \neq i} X_j - 1$. For each $x_i = 0, 1, \ldots, N_{in} + 1$, let

$$f_{in}(x_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} I_{\{x_i\}}(X_j), \tag{3.1}$$

$$h_{in}(x_i) = f_{in}(x_i)/a(x_i). \tag{3.2}$$

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

$$\varphi_{in}(x_i) = [h_{in}(x_i+1)+\delta_n]/[h_{in}(x_i)+\delta_n], \tag{3.3}$$

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

It is intuitive 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 $\varphi_{in}(x_i)$ may not possess the increasing property. Thus, we consider a smoothed version of $\varphi_{in}(x_i)$. Let $\{\varphi^*_{in}(x_i)\}_{x_i=0}^{N_{in}}$ be the isotonic regression of $\{\varphi_{in}(x_i)\}_{x_i=0}^{N_{in}}$ with random weights $\{W_{in}(x_i)\}_{x_i=0}^{N_{in}}$, where $W_{in}(x_i) = [h_{in}(x_i)+\delta_n]a(x_i+1)$. For $y > N_{in}$, define $\varphi^*_{in}(y) = \varphi^*_{in}(N_{in})$. Therefore, $\varphi^*_{in}(x_i)$ is nondecreasing in x_i, $x_i = 0,1,2,\ldots$. 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 $i = 1,\ldots,n$,

$$d^*_{in}(x_i|\underset{\sim}{x}(i)) \equiv d^*_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi^*_{in}(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The performance of the preceding nonparametric empirical Bayes procedure will be discussed in Section 4.

3.2. 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

$$\hat{\varphi}_{in}(x_i) = \begin{cases} \frac{x_i[\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i)] + \mu_{1n}^2(i)}{\mu_{2n}(i) - \mu_{1n}^2(i)} & \text{if } \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) > 0 \\ x_i & \text{otherwise.} \end{cases} \tag{3.5}$$

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

$$\hat{d}_{in}(x_i | \underset{\sim}{x}(i)) = \hat{d}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \hat{\varphi}_{in}(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

3.3. 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 $h(\beta) = \frac{1}{\beta}$, $\beta > 0$. Thus, conditional on β, $X_1, \ldots, X_n$ are iid with the probability function $f(x|\beta) = \int_0^\infty f(x|\theta) g(\theta|k, \beta) d\theta = \frac{\Gamma(x+k)\beta^k}{x!\Gamma(k)(1+\beta)^{x+k}}$, $x = 0, 1, 2, \ldots$. Therefore, $(X_1, \ldots, X_n)$ has a joint marginal probability function $\underset{\sim}{f}(x_1, \ldots, x_n)$ where

$$\underset{\sim}{f}(x_1, \ldots, x_n) = \int_0^\infty \prod_{i=1}^{n} f(x_i|\beta) h(\beta) d\beta = \prod_{j=1}^{n} \left[\frac{\Gamma(x_j + k)}{x_j! \Gamma(k)} \right] \int_0^\infty \frac{\beta^{nk-1}}{(1+\beta)^b} d\beta, \text{ where}$$

$b = nk + \sum_{j=1}^{n} x_j$. Thus, the posterior density function of β given $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ is

$$\begin{aligned} h(\beta | x_1, \ldots, x_n) &= \frac{f(x_1|\beta) \ldots f(x_n|\beta) h(\beta)}{\underset{\sim}{f}(x_1, \ldots, x_n)} \\ &= \frac{\beta^{nk-1}}{(1+\beta)^b} \left[\int_0^\infty \frac{\beta^{nk-1}}{(1+\beta)^b} d\beta \right]^{-1}, \end{aligned}$$

and the posterior mean of β given $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ is

$$\beta_n = E[\beta | x_1, \ldots, x_n] = \begin{cases} \frac{nk}{\sum_{j=1}^{n} x_j - 1} & \text{if } \sum_{j=1}^{n} x_j \geq 2, \\ \infty & \text{if } \sum_{j=1}^{n} x_j \leq 1. \end{cases}$$

Now, for each $i = 1, \ldots, n$, and $x_i = 0, 1, 2, \ldots$, define

$$\overline{\varphi}_{in}(x_i) = \begin{cases} (x_i + k)/(1 + \beta_n) & \text{if } \sum_{j=1}^{n} x_j \geq 2, \\ 0 & \text{if } \sum_{j=1}^{n} x_j \leq 1. \end{cases} \tag{3.7}$$

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

$$\overline{d}_{in}(x_i | \underset{\sim}{x}(i)) = \overline{d}_{in}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{\varphi}_{in}(x_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

4. 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

$$M = \begin{cases} \min A(\theta_0) & \text{if } A(\theta_0) \neq \phi, \\ \infty & \text{otherwise,} \end{cases} \tag{4.1}$$

$$m = \begin{cases} \max B(\theta_0) & \text{if } B(\theta_0) \neq \phi, \\ -1 & \text{otherwise,} \end{cases} \tag{4.2}$$

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 $i = 1, \ldots, n$,

$$\begin{aligned}
0 &\leq r_i(G, d_{in}) - r_i(G, d_{iB}) \\
&= \sum_{x_i=0}^{m} [\theta_0 - \varphi(x_i)] P\{\varphi_{in}(x_i) \geq \theta_0\} f(x_i) + \sum_{x_i=M}^{\infty} [\varphi(x_i) - \theta_0] P\{\varphi_{in}(x_i) < \theta_0\} f(x_i) \\
&\leq \sum_{x_i=0}^{m} [\theta_0 - \varphi(x_i)] P\{\varphi_{in}(m) \geq \theta_0\} f(x_i) + \sum_{x_i=M}^{\infty} [\varphi(x_i) - \theta_0] P\{\varphi_{in}(M) < \theta_0\} f(x_i) \\
&= b_1 P\{\varphi_{in}(m) \geq \theta_0\} + b_2 P\{\varphi_{in}(M) < \theta_0\}. \qquad (4.3)
\end{aligned}$$

In (4.3), the probability measure P is computed with respect to $\underset{\sim}{X}(i)$. Also, $0 \leq b_1 = \sum_{x=0}^{m} [\theta_0 - \varphi(x)] f(x) < \infty$, $0 \leq b_2 = \sum_{x=M}^{\infty} [\varphi(x) - \theta_0] f(x) < \infty$. The finiteness of both b_1 and b_2 is guaranteed by the assumption that $\int_0^\infty \theta dQ(\theta) < \infty$.

From (4.3), we obtain:

$$\begin{aligned}
0 &\leq r(G, \underset{\sim}{d}_n) - r(G) \\
&= \sum_{i=1}^{n} [r_i(G, d_{in}) - r_i(G, d_{iB})] \\
&\leq \sum_{i=1}^{n} [b_1 P\{\varphi_{in}(m) \geq \theta_0\} + b_2 P\{\varphi_{in}(M) < \theta_0\}].
\end{aligned} \qquad (4.4)$$

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

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

We first present some useful results.

For each i = 1, ..., 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 $\Pi_{in}(y) = \sum_{x=0}^{y} W_{in}(x)$ where $W_{in}(x)$, $x = 0, 1, \ldots, N_{in}$, are the random weights defined in Section 3. From Barlow, et al. (1972),

$$\Psi^*_{in}(y) \leq \Psi_{in}(y) \text{ for all } y = 0, 1, \ldots, N_{in}. \qquad (4.5)$$

From Puri and Singh (1988), the isotonic regression estimators $\varphi^*_{in}(x)$, $x = 0, 1, \ldots, N_{in}$, can be rewritten as:

$$\varphi^*_{in}(x) = \min_{x \le y \le N_{in}} \left[\frac{\Psi_{in}(y) - \Psi^*_{in}(x-1)}{H_{in}(y) - H_{in}(x-1)} \right], \quad x = 0, 1, \ldots, N_{in}, \tag{4.6}$$

where $\Psi^*_{in}(-1) = H_{in}(-1) \equiv 0$. Thus, from (4.5) and (4.6),

$$\varphi^*_{in}(x) \ge \min_{x \le y \le N_{in}} \left[\frac{\Psi_{in}(y) - \Psi_{in}(x-1)}{H_{in}(y) - H_{in}(x-1)} \right], \quad x = 0, 1, \ldots, N_{in}, \tag{4.7}$$

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

The following Lemma is taken from Liang (1989).

Lemma 4.1. 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 4.2. 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 4.3. $P\{\varphi^*_{in}(M) < \theta_0\} \le O(\exp(-\tau_1 n))$
where $\tau_1 = \min(2(Q(M) \max(1, \theta_0^{-1})/8)^2, \ln[F(M)]^{-1}) > 0$.

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

$$= P\{\varphi^*_{in}(M) < \theta_0, \ N_{in} < M\} + P\{\varphi^*_{in}(M) < \theta_0, \ N_{in} \ge M\}. \tag{4.8}$$

Now,

$$P\{\varphi^*_{in}(M) < \theta_0, \ N_{in} < M\} \le [F(M)]^{n-1} = O(\exp(-n \ln[F(M)]^{-1})), \tag{4.9}$$

where $F(\cdot)$ is the marginal distribution of X_i, and the inequality is obtained by the definition of N_{in}.

Also, from (3.1)-(3.3), (4.7), Lemma 4.2, and by the definitions of $\Psi_{in}(y)$ and $H_{in}(y)$, straightforward computation yields the following:

$$
\begin{aligned}
E &\equiv \{\varphi^*_{in}(M) < \theta_0,\ N_{in} \ge M\} \\
&\subset \{\Psi_{in}(y) - \Psi_{in}(M-1) < \theta_0[H_{in}(y) - H_{in}(M-1)] \text{ for some } y,\ M \le y \le N_{in}\} \\
&\subset \Bigg\{ \sum_{x=M}^{y} [f_{in}(x+1) - f(x+1)] - \theta_0 \sum_{x=M}^{y} [f_{in}(x) - f(x)] \frac{a(x+1)}{a(x)} < (\theta_0 - 1)\delta_n \sum_{x=M}^{y} a(x+1) \\
&\qquad + Q(M) \text{ for some } y \ge M \Bigg\} \\
&\equiv E_1. \qquad (4.10)
\end{aligned}
$$

Since $a(x) \ge 0$ for all $x = 0, 1, \ldots, \sum_{x=0}^{\infty} a(x) < \infty$ 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 4.1, we obtain:

$$
\begin{aligned}
E_1 &\subset \bigcup_{y \ge M} \left\{ \left| \sum_{x=M}^{y} [f_{in}(x+1) - f(x+1)] \right| > -\frac{Q(M)}{4} \text{ or } \left| \sum_{x=M}^{y} [f_{in}(x) - f(x)] \frac{a(x+1)}{a(x)} \right| > -\frac{Q(M)}{4\theta_0} \right\} \\
&\subset \bigcup_{y \ge M} \left\{ \left| \sum_{x=M}^{y} [f_{in}(x+1) - f(x+1)] \right| > -\frac{Q(M)}{4} \text{ or } \left| \sum_{x=M}^{y} [f_{in}(x) - f(x)] \right| > -\frac{Q(M)}{4\theta_0} \right\} \\
&\subset \left\{ \sup_{y \ge 0} |F_{in}(y) - F(y)| > -Q(M) \max(1, \theta_0^{-1})/8 \right\} \qquad (4.11)
\end{aligned}
$$

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

From (4.10) and (4.11), we obtain

$$
\begin{aligned}
& P\{\varphi^*_{in}(M) < \theta_0,\ N_{in} \ge M\} \\
&\le P\{\sup_{y \ge 0} |F_{in}(y) - F(y)| > -Q(M) \max(1, \theta_0^{-1})/8\} \qquad (4.12) \\
&\le d \exp\{-2n(Q(M) \max(1, \theta_0^{-1})/8)^2\}
\end{aligned}
$$

where the last inequality follows from Lemma 2.1 of Schuster (1969).

Now, let $\tau_1 = \min(2(Q(M)\max(1,\theta_0^{-1})/8)^2, \ln[F(M)]^{-1})$. Clearly $\tau_1 > 0$. Combining (4.8), (4.9) and (4.12) gives the result of this theorem.

Theorem 4.4. $P\{\varphi_{in}^*(m) \geq \theta_0\} \leq O(\exp(-\tau_2 n))$
where $\tau_2 = [R^*(m)\ \min(1,\theta_0^{-1})]^2/8 > 0$ and $R^*(m)$ is defined below.

Proof: From (3.1)-(3.3) and by the definition of $\varphi_{in}^*(m)$,

$$\begin{aligned}&\{\varphi_{in}^*(m) \geq \theta_0\}\\ \subset&\{\varphi_{in}(x) \geq \theta_0 \text{ for some } 0 \leq x \leq m\} \qquad (4.13)\\ \subset&\{a(x)\Delta_{in}(x+1) - \theta_0 a(x+1)\Delta_{in}(x) > R(x) - a(x)a(x+1)\delta_n[1-\theta_0] \text{ for some } 0 \leq x \leq m\},\end{aligned}$$

where $\Delta_{in}(x) = f_{in}(x) - f(x)$, $R(x) = -a(x)f(x+1) + \theta_0 a(x+1)f(x) = a(x+1)f(x)[-\varphi(x) + \theta_0] > 0$ 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 (4.13) and by Theorem 1 of Hoeffding (1963),

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

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

Theorem 4.5. 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 (4.4), Theorem 4.3 and Theorem 4.4, we have

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

4.2. Asymptotic Optimality of $\hat{d}_n$

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 4.6. 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 4.7. 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 3. 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 (1952). 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 3. 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 4.6, $0 \le A < 1$.

<u>Lemma 4.8</u>. $P\{\mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \le 0\} \le O(\exp(-\alpha_1 n))$
where $\alpha_1 = \begin{cases} -\ln A & \text{if } A > 0, \\ \infty & \text{if } A = 0. \end{cases}$

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

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

By Lemma 4.7,

$$\begin{aligned} &P\left\{\mu_{2n}(i) - \mu_2 \le -\frac{\mu}{3}\right\} \le \left[m_2\left(\mu_2 - \frac{\mu}{3}\right)\right]^{n-1}, \\ &P\left\{\mu_{1n}(i) - \mu_1 \ge \frac{\mu}{3}\right\} \le \left[m_1\left(\mu_1 + \frac{\mu}{3}\right)\right]^{n-1}, \text{ and} \\ &P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3}\right\} \\ &= P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3},\ \mu_{1n}(i) < 2\mu_1\right\} + P\left\{\mu_{1n}^2(i) - \mu_1^2 \ge \frac{\mu}{3},\ \mu_{1n}(i) \ge 2\mu_1\right\} \\ &\le P\left\{\mu_{1n}(i) - \mu_1 \ge \frac{\mu}{9\mu_1}\right\} + P\{\mu_{1n}(i) - \mu_1 \ge \mu_1\} \qquad (4.14) \\ &\le \left[m_1\left(\mu_1 + \frac{\mu}{9\mu_1}\right)\right]^{n-1} + [m_1(2\mu_1)]^{n-1}. \end{aligned}$$

Combining the preceding results, the lemma follows.

<u>Theorem 4.9</u>. $P\{\hat{\varphi}_{in}(M) < \theta_0\} \le O(\exp(-\alpha_2 n))$ for some positive constant α_2.

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

$$+P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) > 0\}, \tag{4.15}$$

where

$$\begin{aligned} &P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) \leq 0\} \\ \leq\ &O(\exp(-\alpha_1 n)) \text{ by Lemma 4.8.} \end{aligned} \tag{4.16}$$

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,

$$\begin{aligned} &P\{\hat{\varphi}_{in}(M) < \theta_0,\ \mu_{2n}(i) - \mu_{1n}(i) - \mu_{1n}^2(i) > 0\} \\ \leq &P\{(M-\theta_0)\mu_{2n}(i) - M\mu_{1n}(i) - (M-1-\theta_0)\mu_{1n}^2(i) < 0\} \\ = &P\{(M-\theta_0)(\mu_{2n}(i) - \mu_2) - M(\mu_{1n}(i) - \mu_1) - (M-1-\theta_0)(\mu_{1n}^2(i) - \mu_1^2) < -q(M)\} \\ \leq &P\left\{(M-\theta_0)(\mu_{2n}(i) - \mu_2) < -\frac{q(M)}{3}\right\} + P\left\{M(\mu_{1n}(i) - \mu_1) > \frac{q(M)}{3}\right\} \qquad (4.17) \\ &\quad + P\left\{(M-1-\theta_0)(\mu_{1n}^2(i) - \mu_1^2) > \frac{q(M)}{3}\right\}. \end{aligned}$$

By Lemma 4.7,

$$P\left\{M(\mu_{1n}(i) - \mu_1) > \frac{q(M)}{3}\right\} \leq \left[m_1\left(\mu_1 + \frac{q(M)}{3M}\right)\right]^{n-1}. \tag{4.18}$$

$$P\left\{(M-\theta_0)(\mu_{2n}(i) - \mu_2) < -\frac{q(M)}{3}\right\} \leq \begin{cases} \left[m_2\left(\mu_2 - \frac{q(M)}{3(M-\theta_0)}\right)\right]^{n-1} & \text{if } M-\theta_0 > 0, \\ 0 & \text{if } M-\theta_0 = 0, \\ \left[m_2\left(\mu_2 + \frac{q(M)}{3(\theta_0-M)}\right)\right]^{n-1} & \text{if } M-\theta_0 < 0, \end{cases} \tag{4.19}$$

and analogous to (4.14),

$$\begin{aligned} &P\left\{(M-1-\theta_0)(\mu_{1n}^2(i) - \mu_1^2) > \frac{q(M)}{3}\right\} \\ \leq &\begin{cases} \left[m_1\left(\mu_1 + \frac{q(M)}{q(M-1-\theta_0)\mu_1}\right)\right]^{n-1} + [m_1(2\mu_1)]^{n-1} & \text{if } M-1-\theta_0 > 0, \\ 0 & \text{if } M-1-\theta_0 = 0, \\ \left[m_1\left(\mu_1 + \frac{q(M)}{6(M-1-\theta_0)\mu_1}\right)\right]^{n-1} & \text{if } M-1-\theta_0 < 0. \end{cases} \end{aligned} \tag{4.20}$$

Combining (4.15)–(4.20), and by Lemma 4.6, it follows that there exists a positive constant, say α_2, such that $P\{\hat{\varphi}_{in}(M) < \theta_0\} \leq O(\exp(-\alpha_2 n))$.

Theorem 4.10. $P\{\hat{\varphi}_{in}(m) \geq \theta_0\} \leq O(\exp(-\alpha_3 n))$ for some positive constant α_3.

Proof: The proof is analogous to that of Theorem 4.9. We omit the detail here.

The following theorem is a direct result of (4.4) and Theorems 4.9 and 4.10.

Theorem 4.11. Let $\hat{\underset{\sim}{d}}_n$ be the empirical Bayes rule defined in Section 3. Assume that the prior distribution G is a member of the gamma distribution family. Then,

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

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

4.3. Asymptotic Optimality of $\bar{\underset{\sim}{d}}_n$.

Theorem 4.12. Let $\bar{\underset{\sim}{d}}_n$ be the empirical Bayes rule defined in Section 3. Assume that the prior distribution G is a member of gamma distribution family $\Gamma(k, \beta)$, where k is a known positive constant. Then,

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

for some positive constant γ.

Note that the statistical model considered here is simpler than that of Section 4.2. Thus, the proof for Theorem 4.12 is analogous to and simpler than that for Theorem 4.11. We omit the detail here.

REFERENCES

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

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.

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

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.

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.

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.

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.

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

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

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.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

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

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.

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

Puri, P.S. and Singh, H. (1988). On recursive formulas for isotonic regression useful for statistical inference under order restrictions. To appear in *J. Statist. Plann. Inference.*

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

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

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

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

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

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.

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

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #89-22C	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 47907	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170, NSF Grant DMS-8606964 DMS-8702620
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
ON THE ASYMPTOTIC OPTIMALITY OF CERTAIN EMPIRICAL BAYES SIMULTANEOUS TESTING PROCEDURES

12. PERSONAL AUTHOR(S)
Shanti S. Gupta and TaChen Liang

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	July 1989	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Asymptotic optimality; isotonic regression; nonparametric empirical Bayes; parametric empirical Bayes; hierarchical empirical Bayes; n-component decision problem

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

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.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Shanti S. Gupta	22b. TELEPHONE (Include Area Code) 317-494-6031 / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED