EMPIRICAL BAYES TWO-STAGE PROCEDURES FOR SELECTING THE BEST BERNOULLI POPULATION COMPARED WITH A CONTROL

by

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

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

Re-Bin Rau
Department of Mathematics
Purdue University
West Lafayette, IN 47907-1395

Technical Report # 92-12C

Department of Statistics
Purdue University

March, 1992
Revised April, 1992
Revised May, 1992
Revised April, 1993

△

EMPIRICAL BAYES TWO-STAGE PROCEDURES FOR SELECTING THE BEST BERNOULLI POPULATION COMPARED WITH A CONTROL*

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

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

Re-Bin Rau
Department of Mathematics
Purdue University
West Lafayette, IN 47907-1395

April, 1993

Abstract

The problem of selecting the population with the largest probability of success from among $k(\geq 2)$ independent Bernoulli populations is investigated. The population to be selected must be as good as or better than a control. It is assumed that past observations are available when the current selection is made. Therefore, the empirical Bayes approach is employed. Combining useful information from the past data, an empirical Bayes two-stage selection procedure is developed. It is proved that the proposed empirical Bayes two-stage selection procedure is asymptotically optimal, having a rate of convergence of order $O(\exp(-cn))$, for some positive constant c, where n is the number of past observations at hand. A simulation study is also carried out to investigate the performance of the proposed empirical Bayes selection procedure for small to moderate values of n.

AMS 1980 Subject Classification: 62F07; 62C12
Keywords and Phrases: Asymptotic optimality; best population; empirical Bayes; rate of convergence; two-stage selection procedure.

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

1 Introduction

Consider k independent populations $\pi_1, \ldots, \pi_k$, where for each i, population π_i is characterized by the value of a parameter of interest, say θ_i. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of the parameters $\theta_1, \ldots, \theta_k$. It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. A population π_i with $\theta_i = \theta_{[k]}$ is called the best population. The problem of selecting the best population was studied in papers pioneered by Bechhofer (1954) using the indifference zone approach and by Gupta (1956) employing the subset selection approach. A discussion of these approaches and various modifications that have taken place since then, can be found in Gupta and Panchapakesan (1979,1985).

In many practical applications, one may not only be interested in the selection of the best population, but also desire that the quality of the selected population be good enough. For example, consider k different competing drugs developed for a certain ailment. Let θ_i be the success probability of curing the disease by using the drug π_i. We are interested in the selection of the drug associated with the highest success probability and desire the corresponding success probability to be at least equal to some required standard or control value. If there is no drug which achieves the required standard, one may not wish to select any. In the literature, Bechhofer and Turnbull (1978), Dunnett (1984) and Wilcox (1984) have considered such a selection goal and investigated selection procedures for selecting the best normal population compared with a control, respectively.

In this paper, we are concerned with the problem of selecting the best Bernoulli population provided it is as good as a specified standard. The Bernoulli model occurs in many fields, such as medicine, engineering, and sociology. A number of statistical procedures based on fixed sampling or sequential sampling rules have been studied in the literature for finding the best Bernoulli population. Sobel and Huyett (1957) have studied a fixed sample procedure through the indifference zone approach. Gupta and Sobel (1960) have studied this selection problem using the subset selection approach. Tamhane (1980) studied the problem of selecting the better Bernoulli treatment using a matched samples design. Sanchez (1987) investigated a modified least-failures sampling procedure for Bernoulli subset selection. Gupta and Huang (1976) and Jeyaratnam and Panchapakesan (1990) investigated certain selection procedures based on entropy functions. Bechhofer and Kulkarni (1982) and Kulkarni and Jennison (1986) studied sequential selection procedures. Yang (1989) treated this selection problem through a Bayesian approach. Gupta and Liang (1988, 1989) have developed empirical Bayes procedures for selecting the population associated with the highest success probability.

This paper deals with two-stage selection procedures for selecting the best Bernoulli population compared with a specified standard using the parametric empirical Bayes approach. The formulation of the selection problem is described in Section 2. A Bayes two-stage selection procedure is derived in Section 3. We then construct an empirical Bayes two-stage selection procedure in Section 4. The asymptotic optimality of the proposed empirical Bayes two-stage selection procedure is investigated in Section 5. Bayes and empirical Bayes two-stage selection procedures for a special case are studied in Section 6. It is proved that the proposed empirical Bayes two-stage selection procedures have a rate of convergence of order

$O(\exp(-cn))$ for some positive constant c, where n is the number of past observations at hand. A simulation study is also carried out to investigate the performance of the proposed empirical Bayes two-stage selection procedures for small to moderate values of n. The results of this study are described at the end of the paper.

2 Formulation of the Selection Problem

Consider k independent Bernoulli populations, say $\pi_1, \ldots, \pi_k$, with unknown success probabilities $\theta_1, \ldots, \theta_k$, respectively. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of the parameters $\theta_1, \ldots, \theta_k$. It is assumed that there is no prior information about the true pairing between the ordered and the unordered parameters. Any population associated with the largest $\theta_{[k]}$ is defined as the best population. For a given standard $\theta_0 (0 < \theta_0 < 1)$, population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. Our selection goal is to select a population which should be the best among the k competitors and good compared with the standard θ_0. If there is no such population, we select none.

A two-stage selection procedure is described as follows. First, we have M independent trials taken from each of the k Bernoulli populations. For each $i = 1, \ldots, k$, let X_i denote the number of successes among the M trials taken from the population π_i. Based on the observations $\underset{\sim}{X} = (X_1, \ldots, X_k)$, one decides whether the selection should be made immediately or not. If one decides to make the selection immediately, then based on the data one may select a population from among the k populations or one may select none in which case the k populations are excluded as bad populations. If one decides not to make the selection immediately, then one (potential) population is chosen, say population π_i, and further m trials are taken from this population. We let Y_i denote the number of successes among the m independent trials from population π_i. Then, based on the data $\underset{\sim}{X}$ and Y_i, one may decide to either select population π_i as the best population and consider π_i to be good, or select none and exclude all the k populations as bad populations.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | 0 \leq \theta_i \leq 1, i = 1, \ldots, k\}$ be the parameter space. Let $\underset{\sim}{a} = (a_0, a_1, \ldots, a_k)$ be an action, where $a_i = 0, 1; i = 0, 1, \ldots, k$ and $\sum_{i=0}^{k} a_i = 1$. When $a_i = 1$ for some $i = 1, \ldots, k$, it means that population π_i is selected as the best population and considered to be good. When $a_0 = 1$, it means that all k populations are excluded as bad populations. Also, let t denote a function associated with the termination action. When $t = 1$, it means that the selection is made immediately after $\underset{\sim}{X}$ is observed. When $t = 0$, it means that further m trials from some of the k populations are needed in order to make the selection. When $t = 0$, let $\underset{\sim}{\Delta} = (\Delta_1, \ldots, \Delta_k)$ be the identity action, where $\Delta_i = 0, 1, i = 1, \ldots, k$, and $\sum_{i=1}^{k} \Delta_i = 1$. When $\Delta_i = 1$, it means that the further m trials are taken from the population π_i. For the parameter $\underset{\sim}{\theta}$ and action $(\underset{\sim}{a}, t, \underset{\sim}{\Delta})$, the loss function $L(\underset{\sim}{\theta}, (\underset{\sim}{a}, t, \underset{\sim}{\Delta}))$ is defined to be:

$$L(\underset{\sim}{\theta}, (\underset{\sim}{a}, t, \underset{\sim}{\Delta})) = \max(\theta_{[k]}, \theta_0) - t \sum_{i=0}^{k} a_i \theta_i + Mkc_1$$

$$+(1-t)\left\{-\sum_{i=1}^{k}\Delta_i[a_i\theta_i+(1-a_i)\theta_0]+mc_2\right\}, \tag{2.1}$$

where $c_1 > 0$ is the cost for each trial taken at the first stage, and $c_2 > 0$ is the cost for each trial taken at the second stage.

Note that conditional on the parameter $\theta_i, X_i \sim B(M, \theta_i), Y_i \sim B(m, \theta_i)$ and X_i and Y_i are conditionally independent. We let $f_i(x|\theta_i)$ and $g_i(y|\theta_i)$ denote the conditional probability functions of X_i and Y_i, respectively. That is,

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

and

$$g_i(y|\theta_i) = \binom{m}{y}\theta_i^y(1-\theta_i)^{m-y}, y = 0, 1, \ldots, m.$$

It is assumed that for each $i = 1, \ldots, k, \theta_i$ is a realization of a random variable Θ_i which has a beta prior distribution with probability density function $h_i(\theta)$, where

$$h_i(\theta) = \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i\mu_i)\Gamma(\alpha_i(1-\mu_i))}\theta^{\alpha_i\mu_i-1}(1-\theta)^{\alpha_i(1-\mu_i)-1}, 0 < \theta < 1$$

where $0 < \mu_i < 1, \alpha_i > 0$, both μ_i and α_i are unknown. The random variables $\Theta_1, \ldots, \Theta_k$ are assumed to be mutually independent.

Let $\mathcal{X}$ be the sample space generated by $\underset{\sim}{X}$ and let $\mathcal{Y}$ be the sample space generated by $\underset{\sim}{Y} = (Y_1, \ldots, Y_k)$. A two-stage selection procedure, in general, consists of the following rules:

(a) Stopping rule τ: For each $\underset{\sim}{x} \in \mathcal{X}, \tau(\underset{\sim}{x})$ is the probability of terminating the sampling after observing $\underset{\sim}{x}$ and making a selection immediately based on $\underset{\sim}{x}$.

(b) Identity rule $\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k)$: For each $\underset{\sim}{x} \in \mathcal{X}, \delta_i(\underset{\sim}{x})$ is the probability of taking the additional m trials from the population π_i when the decision of going to the second-stage is made. Note that $\underset{\sim}{\delta}$ should satisfy that $\sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) = 1$ for all $\underset{\sim}{x} \in \mathcal{X}$.

(c) First-stage selection rule $\underset{\sim}{d}_1 = (d_{10}, d_{11}, \ldots, d_{1k})$: For each $\underset{\sim}{x} \in \mathcal{X}$, $d_{1i}(\underset{\sim}{x}), i = 1, \ldots, k$, is the probability of selecting the population π_i as the best and good, and $d_{10}(\underset{\sim}{x})$ is the probability of excluding all the k populations as bad and selecting none. Also, $\sum_{i=0}^{k} d_{1i}(\underset{\sim}{x}) = 1$ for all $\underset{\sim}{x} \in \mathcal{X}$.

(d) Second-stage selection rule $\underset{\sim}{d}_2 = (d_{20}, d_{21}, \ldots, d_{2k})$: For each $\underset{\sim}{x} \in \mathcal{X}, \underset{\sim}{y} \in \mathcal{Y}$, when the decision of going to the second-stage sampling from the population π_i is made, $d_{2i}(\underset{\sim}{x}, \underset{\sim}{y})$ is the probability of selecting the population π_i as the best and good, $i = 1, \ldots, k$. It should be noted that $d_{2i}(\underset{\sim}{x}, \underset{\sim}{y})$ depends on $\underset{\sim}{y}$ only through y_i since there are no observations from other populations $\pi_j, j \neq i$. Also, $d_{20}(\underset{\sim}{x}, \underset{\sim}{y}) = \sum_{i=1}^{k} \delta_i(\underset{\sim}{x})[1 - d_{2i}(\underset{\sim}{x}, \underset{\sim}{y})]$

is the probability of selecting none based on $\underset{\sim}{x}$ and the second-stage observation y_i for some $i = 1, \ldots, k$. For notational convenience, in the sequel, we may use either $d_{2i}(\underset{\sim}{x}, y)$ or $d_{2i}(\underset{\sim}{x}, y_i)$.

Under the preceding statistical model, the Bayes risk of the two-stage selection procedure $(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2)$ is denoted by $R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2)$. Then a straightforward computation yields the following:

$$
\begin{aligned}
& R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2) \\
= \; & C - \sum_{\underset{\sim}{x} \in \mathcal{X}} \tau(\underset{\sim}{x}) \left[\sum_{i=0}^{k} d_{1i}(\underset{\sim}{x}) \varphi_i(x_i | \alpha_i, \mu_i) \right] f(\underset{\sim}{x}) \qquad (2.2) \\
+ \; & \sum_{\underset{\sim}{x} \in \mathcal{X}} [1 - \tau(\underset{\sim}{x})] \left\{ mc_2 - \theta_0 + \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) \left[\sum_{y_i=0}^{m} d_{2i}(\underset{\sim}{x}, y_i)[\theta_0 - \psi_i(x_i, y_i | \alpha_i, \mu_i)] f_i(y_i | x_i, \alpha_i, \mu_i) \right] \right\} f(\underset{\sim}{x}) \\
= \; & \sum_{\underset{\sim}{x} \in \mathcal{X}} \tau(\underset{\sim}{x}) \left\{ \begin{array}{l} \sum_{i=0}^{k} d_{1i}(\underset{\sim}{x})[\theta_0 - \varphi_i(x_i | \alpha_i, \mu_i)] - mc_2 \\ - \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) \left[\sum_{y_i=0}^{m} d_{2i}(\underset{\sim}{x}, y_i)[\theta_0 - \psi_i(x_i, y_i | \alpha_i, \mu_i)] f_i(y_i | x_i, \alpha_i, \mu_i) \right] \end{array} \right\} f(\underset{\sim}{x}) \\
+ \; & \sum_{\underset{\sim}{x} \in \mathcal{X}} \left\{ mc_2 - \theta_0 + \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) \left[\sum_{y_i=0}^{m} d_{2i}(\underset{\sim}{x}, y_i)[\theta_0 - \psi_i(x_i, y_i | \alpha_i, \mu_i)] f_i(y_i | x_i, \alpha_i, \mu_i) \right] \right\} f(\underset{\sim}{x}) \\
+ \; & C, \qquad (2.3)
\end{aligned}
$$

where $\varphi_i(x_i | \alpha_i, \mu_i) = E[\Theta_i | X_i = x_i] = \frac{x_i + \alpha_i \mu_i}{M + \alpha_i}$ is the posterior mean of Θ_i given $X_i = x_i$ for each $i = 1, \ldots, k$, and $\varphi_0(x_0 | \alpha_0, \mu_0) \equiv \theta_0$; $\psi_i(x_i, y_i | \alpha_i, \mu_i) = E[\Theta_i | X_i = x_i, Y_i = y_i] = \frac{x_i + y_i + \alpha_i \mu_i}{M + m + \alpha_i}$ is the posterior mean of Θ_i given $(X_i, Y_i) = (x_i, y_i)$ for each $i = 1, \ldots, k; C = \int_\Omega \max(\theta_{[k]}, \theta_0) dH(\underset{\sim}{\theta}) + Mkc_1$ and $H(\underset{\sim}{\theta})$ is the joint distribution of $\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k)$; $f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i)$ and

$$f_i(x_i) = \int_0^1 f_i(x_i | \theta) h_i(\theta) d\theta = \binom{M}{x_i} \frac{\Gamma(\alpha_i)\Gamma(x_i + \alpha_i \mu_i)\Gamma(M - x_i + \alpha_i(1 - \mu_i))}{\Gamma(\alpha_i \mu_i)\Gamma(\alpha_i(1 - \mu_i))\Gamma(M + \alpha_i)}$$

is the marginal probability function of X_i; and $f_i(y_i | x_i, \alpha_i, \mu_i)$ is the marginal conditional probability function of Y_i given $X_i = x_i$. Again a direct computation yields

$$
\begin{aligned}
f_i(y_i | x_i, \alpha_i, \mu_i) &= \binom{m}{y_i} \frac{\Gamma(x_i + y_i + \alpha_i \mu_i)\Gamma(M + m - x_i - y_i + \alpha_i(1 - \mu_i))\Gamma(M + \alpha_i)}{\Gamma(x_i + \alpha_i \mu_i)\Gamma(M - x_i + \alpha_i(1 - \mu_i))\Gamma(M + m + \alpha_i)} \\
&= \left\{ \binom{m}{y_i} \prod_{j=0}^{y_i - 1} (x_i + j + \alpha_i \mu_i) \prod_{j=0}^{m - y_i - 1} (M - x_i + \alpha_i(1 - \mu_i) + j) \right\} \\
&\quad \times \left\{ \prod_{j=0}^{m-1} (M + j + \alpha_i)^{-1} \right\}
\end{aligned}
$$

where $\prod_{j=0}^{\ell} \equiv 1$ if $\ell = -1$.

3 Derivation of a Bayes Two-Stage Selection Procedure

In order to develop an empirical Bayes two-stage selection procedure, as a first step, we derive a Bayes two-stage selection procedure for the selection problem under consideration.

A First-Stage Selection Rule

For each $\underset{\sim}{x} \in \mathcal{X}$, let $I(\underset{\sim}{x}) = \{i | \varphi_i(x_i|\alpha_i, \mu_i) = \max_{1\le j\le k} \varphi_j(x_j|\alpha_j, \mu_j), i = 1, \dots, k\}$. Define $i^* \equiv i^*(\underset{\sim}{x}) = \min\{i | i \in I(\underset{\sim}{x})\}$. We then define a first-stage selection rule $\underset{\sim}{d}_1^B = (d_{10}^B, \dots, d_{1k}^B)$ as follows:

$$\begin{cases} \text{If } \varphi_{i^*}(x_{i^*}|\alpha_{i^*}, \mu_{i^*}) \ge \theta_0, \text{define } d_{1i^*}^B(\underset{\sim}{x}) = 1 \text{ and } d_{1j}^B(\underset{\sim}{x}) = 0 \text{ for } j \ne i^*. \\ \text{If } \varphi_{i^*}(x_{i^*}|\alpha_{i^*}, \mu_{i^*}) < \theta_0, \text{define } d_{10}^B(\underset{\sim}{x}) = 1 \text{ and } d_{1j}^B(\underset{\sim}{x}) = 0 \text{ for } j = 1, \dots, k. \end{cases} \tag{3.1}$$

A Second-Stage Selection Rule

We define a second-stage selection rule $\underset{\sim}{d}_2^B = (d_{20}^B, \dots, d_{2k}^B)$ as follows: For each $\underset{\sim}{x} \in \mathcal{X}, y \in \mathcal{Y}$, and $i = 1, \dots, k$, define

$$d_{2i}^B(\underset{\sim}{x}, \underset{\sim}{y}) \equiv d_{2i}^B(\underset{\sim}{x}, y_i) = \begin{cases} 1 & \text{if } \psi_i(x_i, y_i|\alpha_i, \mu_i) \ge \theta_0; \\ 0 & \text{otherwise;} \end{cases} \tag{3.2}$$

and

$$d_{20}^B(\underset{\sim}{x}, \underset{\sim}{y}) = \sum_{i=1}^{k} \delta_i^B(\underset{\sim}{x})[1 - d_{2i}^B(\underset{\sim}{x}, \underset{\sim}{y})],$$

where $\underset{\sim}{\delta}^B = (\delta_1^B, \dots, \delta_k^B)$ is the identity rule defined below.

An Identity Rule

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

$$T_i(\underset{\sim}{x}|\alpha_i, \mu_i) = \sum_{y_i=0}^{m} d_{2i}^B(\underset{\sim}{x}, y_i)[\theta_0 - \psi_i(x_i, y_i|\alpha_i, \mu_i)] f_i(y_i|x_i, \alpha_i, \mu_i). \tag{3.3}$$

Let $J(\underset{\sim}{x}) = \{j = 1, \dots, k | T_j(\underset{\sim}{x}|\alpha_j, \mu_j) = \min_{1\le i\le k} T_i(\underset{\sim}{x}|\alpha_i, \mu_i)\}$ and let $j^*(\underset{\sim}{x}) \equiv j^* = \min\{j | j \in J(\underset{\sim}{x})\}$. We then define an identity rule $\underset{\sim}{\delta}^B = (\delta_1^B, \dots, \delta_k^B)$ as follows:

$$\delta_j^B(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } j = j^*; \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

A Stopping Rule

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

$$Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) = \sum_{i=0}^{k} d_{1i}^{B}(\underset{\sim}{x})[\theta_0 - \varphi_i(x_i|\alpha_i, \mu_i)] - mc_2 - \sum_{i=1}^{k} \delta_i^B(\underset{\sim}{x}) T_i(\underset{\sim}{x}|\alpha_i, \mu_i), \tag{3.5}$$

where $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_k)$ and $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_k)$. We then define a stopping rule τ^B as follows:

$$\tau^B(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) \leq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

Then, we have the following result:

Theorem 3.1. The two-stage selection procedure $(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$ defined through (3.1)–(3.6) is a Bayes two-stage selection procedure.

Proof: Let $(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2)$ be any two-stage selection procedure. We only need to prove that $R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \geq 0$. Now,

$$R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) = I + II + III, \tag{3.7}$$

where

$$\begin{cases} I = R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1, \underset{\sim}{d}_2) - R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B), \\ II = R(\tau, \underset{\sim}{\delta}, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) - R(\tau, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B), \\ III = R(\tau, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B). \end{cases} \tag{3.8}$$

¿From (2.2),

$$I = \sum_{\underset{\sim}{x} \in \mathcal{X}} \tau(\underset{\sim}{x}) \left\{ \sum_{i=0}^{k} [d_{1i}^B(\underset{\sim}{x}) - d_{1i}(\underset{\sim}{x})] \varphi_i(x_i|\alpha_i, \mu_i) \right\} f(\underset{\sim}{x}) \tag{3.9}$$

$$+ \sum_{\underset{\sim}{x} \in \mathcal{X}} [1 - \tau(\underset{\sim}{x})] \left\{ \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) \left[\sum_{y_i=0}^{m} [d_{2i}(\underset{\sim}{x}, y_i) - d_{2i}^B(\underset{\sim}{x}, y_i)][\theta_0 - \psi_i(x_i, y_i|\alpha_i, \mu_i)] f_i(y_i|x_i, \alpha_i, \mu_i) \right] \right\} f(\underset{\sim}{x}).$$

By the definition of $\underset{\sim}{d}_1^B$ and $\underset{\sim}{d}_1$

$$\begin{aligned} \sum_{i=0}^{k} [d_{1i}^B(\underset{\sim}{x}) - d_{1i}(\underset{\sim}{x})] \varphi_i(x_i|\alpha_i, \mu_i) &= \max_{0 \leq i \leq k} (\varphi_i(x_i|\alpha_i, \mu_i)) - \textstyle\sum_{i=0}^{k} d_{1i}(\underset{\sim}{x}) \varphi_i(x_i|\alpha_i, \mu_i) \\ &\geq 0. \end{aligned}$$

Also, by the definition of $\underset{\sim}{d}_2^B$, for each $i = 1, \ldots, k$, we have,

$$[d_{2i}(\underset{\sim}{x}, y_i) - d_{2i}^B(\underset{\sim}{x}, y_i)][\theta_0 - \psi_i(x_i, y_i|\alpha_i, \mu_i)] \geq 0.$$

Hence, $I \geq 0$ since all the other terms in (3.9) are nonnegative.

¿From (2.2) again,

$$II = \sum_{\underset{\sim}{x} \in \mathcal{X}} [1 - \tau(\underset{\sim}{x})] \left\{ \sum_{i=1}^{k} [\delta_i(\underset{\sim}{x}) - \delta_i^B(\underset{\sim}{x})] T_i(\underset{\sim}{x}|\alpha_i, \mu_i) \right\} f(\underset{\sim}{x}) \geq 0, \tag{3.10}$$

since,

$$\begin{aligned} & \sum_{i=1}^{k} [\delta_i(\underset{\sim}{x}) - \delta_i^B(\underset{\sim}{x})] T_i(\underset{\sim}{x}|\alpha_i, \mu_i) \\ = & \sum_{i=1}^{k} \delta_i(\underset{\sim}{x}) T_i(\underset{\sim}{x}|\alpha_i, \mu_i) - \min_{1 \leq i \leq k} T_i(\underset{\sim}{x}|\alpha_i, \mu_i) \geq 0. \end{aligned} \tag{3.11}$$

Now, from (2.3)

$$III = \sum_{\underset{\sim}{x} \in \mathcal{X}} [\tau(\underset{\sim}{x}) - \tau^B(\underset{\sim}{x})] Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) f(\underset{\sim}{x}) \geq 0 \tag{3.12}$$

which holds by the definition of τ^B (see (3.6)).

The proof of the above theorem is completed by combining (3.7), (3.9), (3.10) and (3.12).

Remark: Note that by the definitions of $d_{1i}^B(\underset{\sim}{x}), d_{2i}^B(\underset{\sim}{x}, y_i), \varphi_i(x_i|\alpha_i, \mu_i)$ and $\psi_i(x_i, y_i|\alpha_i, \mu_i)$, $-1 \leq T_i(\underset{\sim}{x}|\alpha_i, \mu_i) \leq 0$ for all $i = 1, \ldots, k$ and for all $\underset{\sim}{x} \in \mathcal{X}$. Hence, $\theta_0 - 1 - mc_2 \leq Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) \leq 1 - mc_2$. Therefore, $Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) \leq 0$ if $mc_2 \geq 1$. In the following, we assume that c_2 is small enough so that $mc_2 < 1$.

4 The Proposed Empirical Bayes Two-Stage Selection Procedure

In the empirical Bayes framework, it is generally assumed that there are certain past observations available when the present selection should be made. At time $j = 1, 2, \ldots$, let X_{ij} denote the number of successes among the M trials taken at the first-stage sampling from population π_i. Let $\underset{\sim}{\Theta}_j = (\Theta_{1j}, \ldots, \Theta_{kj})$ be a random vector where Θ_{ij} stands for the (random) probability of success for each trial taken from population π_i at time j. We assume that $\underset{\sim}{\Theta}_j$, $j = 1, 2, \ldots$ are iid with a prior density $h(\underset{\sim}{\theta}_j) = \prod_{i=1}^{k} h_i(\theta_{ij})$, $\underset{\sim}{\theta}_j = (\theta_{1j}, \ldots, \theta_{kj})$. Therefore, conditional on $\Theta_{ij} = \theta_{ij}$, $X_{ij}|\theta_{ij} \sim B(M, \theta_{ij})$, and X_{ij} has a marginal probability function $f_i(x)$. Let $\underset{\sim}{X}_j = (X_{1j}, \ldots, X_{kj})$ denote the random observations of the first-stage sampling taken at time $j = 1, 2, \ldots$. We also let $\underset{\sim}{X}_{n+1} \equiv \underset{\sim}{X} = (X_1, \ldots, X_k)$ denote the random observations of the first-stage sampling taken at the present moment. Then, we have for each $i = 1, \ldots, k$,

$$\begin{aligned} E\left[\frac{X_{ij}}{M}\right] &= \mu_i \\ E\left[\left(\frac{X_{ij}}{M}\right)^2\right] &= \frac{\mu_i}{M} + \frac{(\alpha_i\mu_i + 1)\mu_i(M-1)}{M(\alpha_i + 1)} \equiv \nu_i \end{aligned} \tag{4.1}$$

A direct computation yields $\alpha_i = \frac{D_i}{C_i}$, where

$$\begin{cases} D_i = \mu_i - \nu_i, \\ C_i = \nu_i - \frac{\mu_i}{M} + \frac{\mu_i^2}{M} - \mu_i^2. \end{cases} \tag{4.2}$$

Note that $D_i = E[\frac{X_{ij}}{M}(1 - \frac{X_{ij}}{M})] > 0$ since $0 \leq \frac{X_{ij}}{M} \leq 1$ and X_{ij} is a non-degenerate random variable. Also $C_i > 0$ since $\alpha_i > 0$. The parameters μ_i and ν_i satisfy the following inequalities: $r_i < \nu_i < \mu_i$ where $r_i = \frac{\mu_i}{M} - \frac{\mu_i^2}{M} + \mu_i^2$. From (4.2), α_i can be viewed as a function of μ_i and ν_i for $\mu_i \in (0,1)$ and $\nu_i \in (r_i, \mu_i)$. If μ_i is kept fixed, α_i is decreasing in ν_i and $\lim_{\nu_i \to \mu_i} \alpha_i = 0$ and $\lim_{\nu_i \to r_i} \alpha_i = \infty$.

Let

$$\begin{cases} \mu_{in} = \frac{1}{n} \sum_{j=1}^{n} (X_{ij}/M), \\ \nu_{in} = \frac{1}{n} \sum_{j=1}^{n} (X_{ij}/M)^2. \end{cases} \tag{4.3}$$

Also, let

$$\begin{cases} C_{in} = \nu_{in} - \frac{\mu_{in}}{M} + \frac{\mu_{in}^2}{M} - \mu_{in}^2, \\ D_{in} = \mu_{in} - \nu_{in}. \end{cases} \tag{4.4}$$

Note that the moment estimators μ_{in} and ν_{in} are unbiased estimators of μ_i and ν_i, respectively. Though $C_i > 0$, it is possible that the estimator C_{in} may be nonpositive. Hence, we define

$$\alpha_{in} = \begin{cases} \frac{D_{in}}{C_{in}} & \text{if } C_{in} > 0, \\ \infty & \text{otherwise.} \end{cases} \tag{4.5}$$

Since $\lim_{\alpha_i \to \infty} \varphi_i(x_i|\alpha_i, \mu_i) = \mu_i$ and $\lim_{\alpha_i \to \infty} \psi_i(x_i, y_i|\alpha_i, \mu_i) = \mu_i$, we define empirical Bayes estimators $\varphi_{in}(x_i)$ and $\psi_{in}(x_i, y_i)$ for the posterior means $\varphi_i(x_i|\alpha_i, \mu_i)$ and $\psi_i(x_i, y_i|\alpha_i, \mu_i)$, respectively, as follows:

$$\varphi_{in}(x_i) = \begin{cases} \frac{x_i + \alpha_{in}\mu_{in}}{M + \alpha_{in}} & \text{if } C_{in} > 0, \\ \mu_{in} & \text{otherwise;} \end{cases} \tag{4.6}$$

and

$$\psi_{in}(x_i, y_i) = \begin{cases} \frac{x_i + y_i + \alpha_{in}\mu_{in}}{M + m + \alpha_{in}} & \text{if } C_{in} > 0, \\ \mu_{in} & \text{otherwise.} \end{cases} \tag{4.7}$$

Also, $\lim_{\alpha_i \to \infty} f_i(y_i|x_i, \alpha_i, \mu_i) = \binom{m}{y_i} \mu_i^y (1 - \mu_i)^{m - y_i}$. Therefore, we define an empirical Bayes estimator for the marginal conditional probability function $f_i(y_i|x_i, \alpha_i, \mu_i)$ as follows:

$$f_{in}(y_i|x_i) = \begin{cases} f_i(y_i|x_i, \alpha_{in}, \mu_{in}) & \text{if } C_{in} > 0, \\ \binom{m}{y_i} \mu_{in}^{y_i} (1 - \mu_{in})^{m - y_i}, & \text{otherwise.} \end{cases} \tag{4.8}$$

Now, we propose an empirical Bayes two-stage selection procedure $(\tau^{*n}, \delta^{*n}, d_1^{*n}, d_2^{*n})$ as follows:

Empirical Bayes First-Stage Selection Rule $\underset{\sim}{d}_1^{*n} = (d_{10}^{*n}, \dots, d_{1k}^{*n})$

For each $\underset{\sim}{x} \in \mathcal{X}$, let $I_n(\underset{\sim}{x}) = \{i = 1, \dots, k | \varphi_{in}(x_i) = \max_{1 \le j \le k} \varphi_{jn}(x_j)\}$. Define $i_n^* \equiv i_n^*(\underset{\sim}{x}) = \min\{i | i \in I_n(\underset{\sim}{x})\}$. Then, we define $\underset{\sim}{d}_1^{*n} = (d_{10}^{*n}, \dots, d_{1k}^{*n})$ as follows:

$$\begin{cases} \text{If } \varphi_{i_n^* n}(x_{i_n^*}) \ge \theta_0, \text{ define } d_{1i_n^*}^{*n}(\underset{\sim}{x}) = 1 \text{ and } d_{1i}^{*n}(\underset{\sim}{x}) = 0 \text{ for } i \ne i_n^*. \\ \text{If } \varphi_{i_n^* n}(x_{i_n^*}) < \theta_0, \text{ define } d_{10}^{*n}(\underset{\sim}{x}) = 1 \text{ and } d_{1i}^{*n}(\underset{\sim}{x}) = 0 \text{ for } i = 1, \dots k. \end{cases} \tag{4.9}$$

Empirical Bayes Second-Stage Selection Rule $\underset{\sim}{d}_2^{*n} = (d_{20}^{*n}, \dots, d_{2k}^{*n})$

We define $\underset{\sim}{d}_2^{*n}$ as follows: For each $\underset{\sim}{x} \in \mathcal{X}$ and $\underset{\sim}{y} \in \mathcal{Y}$, $i = 1, \dots, k$,

$$d_{2i}^{*n}(\underset{\sim}{x}, \underset{\sim}{y}) = d_{2i}^{*n}(\underset{\sim}{x}, y_i) = \begin{cases} 1 & \text{if } \psi_{in}(x_i, y_i) \ge \theta_0, \\ 0 & \text{otherwise;} \end{cases} \tag{4.10}$$

and

$$d_{20}^{*n}(\underset{\sim}{x}, \underset{\sim}{y}) = \sum_{i=1}^{k} \delta_i^{*n}(\underset{\sim}{x})[1 - d_{2i}^{*n}(\underset{\sim}{x}, \underset{\sim}{y})]$$

where $\underset{\sim}{\delta}^{*n} = (\delta_1^{*n}, \dots, \delta_k^{*n})$ is the empirical Bayes identity rule defined below.

Empirical Bayes Identity Rule $\underset{\sim}{\delta}^{*n} = (\delta_1^{*n}, \dots, \delta_k^{*n})$

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

$$T_{in}(\underset{\sim}{x}) = \sum_{y_i=0}^{m} d_{2i}^{*n}(\underset{\sim}{x}, y_i)[\theta_0 - \psi_{in}(x_i, y_i)] f_{in}(y_i | x_i). \tag{4.11}$$

Let $J_n(\underset{\sim}{x}) = \{j = 1, \dots, k | T_{jn}(\underset{\sim}{x}) = \min_{1 \le i \le k} T_{in}(\underset{\sim}{x})\}$ and let $j_n^*(\underset{\sim}{x}) \equiv j_n^* = \min\{j | j \in J_n(\underset{\sim}{x})\}$. Then, the empirical Bayes identity rule $\underset{\sim}{\delta}^{*n} = (\delta_1^{*n}, \dots, \delta_k^{*n})$ is defined as:

$$\delta_j^{*n}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } j = j_n^*, \\ 0 & \text{otherwise.} \end{cases} \tag{4.12}$$

Empirical Bayes Stopping Rule τ^{*n}

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

$$Q_n(\underset{\sim}{x}) = \sum_{i=0}^{k} d_{1i}^{*n}(\underset{\sim}{x})[\theta_0 - \varphi_{in}(x_i)] - mc_2 - \sum_{i=1}^{k} \delta_i^{*n}(\underset{\sim}{x}) T_{in}(\underset{\sim}{x}). \tag{4.13}$$

We may use $Q_n(\underset{\sim}{x})$ to estimate $Q(\underset{\sim}{x} | \underset{\sim}{\alpha}, \underset{\sim}{\mu})$ and define an empirical Bayes stopping rule τ^{*n} accordingly. That is, for each $\underset{\sim}{x} \in \mathcal{X}$, define

$$\tau^{*n}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } Q_n(\underset{\sim}{x}) \le 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.14}$$

5 Asymptotic Optimality of $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$

Consider an empirical Bayes two-stage selection procedure $(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$. Let $R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$ be the associated conditional Bayes risk (conditioning on the past observation $\underset{\sim}{X}_j, j = 1, \ldots, n$) and let $E_n R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$ be the overall Bayes risk of the empirical Bayes two-stage selection procedure $(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$, where the expectation E_n is taken with respect to the probability measure generated by $(\underset{\sim}{X}_j, j = 1, \ldots, n)$. Since $(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$ is the Bayes two-stage selection procedure, $R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \geq 0$ for all $\underset{\sim}{X}_j, j = 1, \ldots, n$, and for all n. Hence, $E_n R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \geq 0$ for all n. The nonnegative regret risk $E_n R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$ is always used as a measure of performance of the empirical Bayes two-stage selection procedure $(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$.

Definition 5.1. A sequence of empirical Bayes two-stage selection procedures $\{(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order $\{\varepsilon_n\}$ if $E_n R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) = O(\varepsilon_n)$, where $\{\varepsilon_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \varepsilon_n = 0$.

In the following, we evaluate the asymptotic optimality of the proposed empirical Bayes two-stage selection procedure $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$.

By definitions of the two selection procedures $(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$ and $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$, we have:

$$R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n}) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) = I_n + II_n + III_n, \tag{5.1}$$

where

$$\begin{aligned} 0 \leq I_n &= R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n}) - R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \\ &= \sum_{\underset{\sim}{x} \in \mathcal{X}} \tau^{*n}(\underset{\sim}{x}) \{ [d_{1i^*}^B(\underset{\sim}{x}) - d_{1i^*}^{*n}(\underset{\sim}{x})][\varphi_{i^*}(x_{i^*}|\alpha_{i^*}, \mu_{i^*}) - \theta_0] \\ &\quad + [d_{1i_n^*}^B(\underset{\sim}{x}) - d_{1i_n^*}^{*n}(\underset{\sim}{x})][\varphi_{i_n^*}(x_{i_n^*}|\alpha_{i_n^*}, \mu_{i_n^*}) - \theta_0] \} f(\underset{\sim}{x}) \\ &\quad + \sum_{\underset{\sim}{x} \in \mathcal{X}} [1 - \tau^{*n}(\underset{\sim}{x})] \{ \sum_{i=1}^{k} \delta_i^{*n}(\underset{\sim}{x}) [\sum_{y_i=0}^{m} [d_{2i}^{*n}(\underset{\sim}{x}, y_i) - d_{2i}^B(\underset{\sim}{x}, y_i)][\theta_0 - \psi_i(x_i, y_i|\alpha_i, \mu_i)] \\ &\quad \times f_i(y_i|x_i, \alpha_i, \mu_i)]\} f(\underset{\sim}{x}); \end{aligned} \tag{5.2}$$

$$\begin{aligned} 0 \leq II_n &= R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) - R(\tau^{*n}, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \\ &= \sum_{\underset{\sim}{x} \in \mathcal{X}} [1 - \tau^{*n}(\underset{\sim}{x})] \{ [\delta_{j^*}^{*n}(\underset{\sim}{x}) - \delta_{j^*}^B(\underset{\sim}{x})] T_{j^*}(\underset{\sim}{x}|\alpha_{j^*}, \mu_{j^*}) \\ &\quad + [\delta_{j_n^*}^{*n}(\underset{\sim}{x}) - \delta_{j_n^*}^B(\underset{\sim}{x})] T_{j_n^*}(\underset{\sim}{x}|\alpha_{j_n^*}, \mu_{j_n^*}) \} f(\underset{\sim}{x}); \end{aligned} \tag{5.3}$$

and

$$\begin{aligned} 0 \leq III_n &= R(\tau^{*n}, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \\ &= \sum_{\underset{\sim}{x} \in \mathcal{X}} [\tau^{*n}(\underset{\sim}{x}) - \tau^B(\underset{\sim}{x})] Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) f(\underset{\sim}{x}). \end{aligned} \tag{5.4}$$

Let $I^c(\underset{\sim}{x}) = \{i = 1, \ldots, k | i \notin I(\underset{\sim}{x})\}$, the complement of $I(\underset{\sim}{x})$ relative to the set $\{1, \ldots, k\}$. Note that $i_n^* \in I(\underset{\sim}{x})$ iff $\varphi_{i_n^*}(x_{i_n^*}|\alpha_{i_n^*}, \mu_{i_n^*}) = \varphi_{i^*}(x_{i^*}|\alpha_{i^*}, \mu_{i^*})$ iff $[d_{1i^*}^B(\underset{\sim}{x}) - d_{1i^*}^{*n}(\underset{\sim}{x})][\varphi_{i^*}(x_{i^*}|\alpha_{i^*}, \mu_{i^*}) - \theta_0] = -[d_{1i_n^*}^B(\underset{\sim}{x}) - d_{1i_n^*}^{*n}(\underset{\sim}{x})][\varphi_{i_n^*}(x_{i_n^*}|\alpha_{i_n^*}, \mu_{i_n^*}) - \theta_0]$. Also, $i_n^* \notin I(\underset{\sim}{x})$ implies that there exists a $j \notin I(\underset{\sim}{x})$ such that $\varphi_{jn}(x_j) \geq \varphi_{in}(x_i)$ for all $i \in I(\underset{\sim}{x})$. Let

$$
\begin{aligned}
2b_1 &= \min_{\underset{\sim}{x} \in \mathcal{X}} \{\varphi_i(x_i|\alpha_i, \mu_i) - \varphi_j(x_j|\alpha_j, \mu_j) | i \in I(\underset{\sim}{x}), j \in I^c(\underset{\sim}{x})\} \\
\mathcal{Y}_i(\underset{\sim}{x}) &= \{0 \leq y_i \leq m | \psi_i(x_i, y_i|\alpha_i, \mu_i) \neq \theta_0\} \\
\text{and} \qquad b_2 &= \min_{\underset{\sim}{x} \in \mathcal{X}} \min_{1 \leq i \leq k} \min_{\mathcal{Y}_i(\underset{\sim}{x})} \{|\theta_0 - \psi_i(x_i, y_i|\alpha_i, \mu_i)|\}.
\end{aligned}
$$

Note that $b_1 > 0$ and $b_2 > 0$ since the sample space under consideration is finite.

Since $0 \leq \tau^{*n}(\underset{\sim}{x}) \leq 1$, $|\varphi_i(x_i|\alpha_i, \mu_i) - \theta_0| \leq 1$, $0 \leq \delta_i^{*n}(\underset{\sim}{x}) \leq 1$ and $|\psi_i(x_i, y_i|\alpha_i, \mu_i) - \theta_0| \leq 1$ for all $\underset{\sim}{x}, y_i$ and i, and $0 \leq f_i(y_i|x_i, \alpha_i, \mu_i) \leq 1$ we have

$$
\begin{aligned}
E_n I_n &\leq \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i \in I(\underset{\sim}{x})} \sum_{j \in I^c(\underset{\sim}{x})} P_n\{\varphi_{jn}(x_j) \geq \varphi_{in}(x_i)\} f(\underset{\sim}{x}) \\
&\quad + \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} \left[\sum_{y_i \in \mathcal{Y}_i(\underset{\sim}{x})} P_n\{d_{2i}^{*n}(\underset{\sim}{x}, y_i) \neq d_{2i}^B(\underset{\sim}{x}, y_i)\} f_i(y_i|x_i, \alpha_i, \mu_i) \right] f(\underset{\sim}{x}) \qquad (5.5) \\
&\leq k \sum_{\underset{\sim}{x} \in \mathcal{X}} \left[\sum_{i=1}^{k} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1\} \right] f(\underset{\sim}{x}) \\
&\quad + \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} \left[\sum_{y_i \in \mathcal{Y}_i(\underset{\sim}{x})} P_n\{|\psi_{in}(x_i, y_i) - \psi_i(x_i, y_i|\alpha_i, \mu_i)| \geq b_2\} f_i(y_i|x_i, \alpha_i, \mu_i) \right] f(\underset{\sim}{x}),
\end{aligned}
$$

where P_n is the probability measure generated by $(\underset{\sim}{X}_j, j = 1, \ldots, n)$. Let

$$2b_3 = \min_{\underset{\sim}{x} \in \mathcal{X}} \{T_j(\underset{\sim}{x}|\alpha_j, \mu_j) - T_i(\underset{\sim}{x}|\alpha_i, \mu_i) | i \in J(\underset{\sim}{x}), j \in J^c(\underset{\sim}{x})\}$$

where $J^c(\underset{\sim}{x})$ is the complement of the set $J(\underset{\sim}{x})$ relative to $\{1, \ldots, k\}$. Then, $b_3 > 0$. By noting that $-1 \leq T_i(\underset{\sim}{x}|\alpha_i, \mu_i) \leq 0$ for all i, we have

$$
\begin{aligned}
E_n II_n &\leq \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i \in J(\underset{\sim}{x})} \sum_{j \in J^c(\underset{\sim}{x})} P_n\{T_{jn}(\underset{\sim}{x}) \leq T_{in}(\underset{\sim}{x})\} f(\underset{\sim}{x}) \\
&\leq k \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} P_n\{|T_{in}(\underset{\sim}{x}) - T_i(\underset{\sim}{x}|\alpha_i, \mu_i)| \geq b_3\} f(\underset{\sim}{x}). \qquad (5.6)
\end{aligned}
$$

Finally, let $\mathcal{X}_1 = \{\underset{\sim}{x} \in \mathcal{X} | Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu}) \neq 0\}$ and let $b_4 = \min_{\underset{\sim}{x} \in \mathcal{X}_1} \{|Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu})|\}$. Then $b_4 > 0$. By noting that $|Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu})| \leq 2$, we have

$$
\begin{aligned}
E_n III_n &\leq 2 \sum_{\underset{\sim}{x} \in \mathcal{X}_1} P_n\{\tau^{*n}(\underset{\sim}{x}) \neq \tau^B(\underset{\sim}{x})\} f(\underset{\sim}{x}) \\
&\leq 2 \sum_{\underset{\sim}{x} \in \mathcal{X}_1} P_n\{|Q_n(\underset{\sim}{x}) - Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu})| \geq b_4\} f(\underset{\sim}{x}). \qquad (5.7)
\end{aligned}
$$

Therefore, from (5.1)–(5.7),

$$
\begin{aligned}
& E_n R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n}) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) \\
\leq \;& k \sum_{\underset{\sim}{x} \in \mathcal{X}} \left[\sum_{i=1}^{k} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1\} \right] f(\underset{\sim}{x}) \\
+ \;& \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} \left[\sum_{y_i=0}^{m} P_n\{|\psi_{in}(x_i, y_i) - \psi_i(x_i, y_i|\alpha_i, \mu_i)| \geq b_2\} f_i(x_i|y_i, \alpha_i, \mu_i) \right] f(\underset{\sim}{x}) \\
+ \;& k \sum_{\underset{\sim}{x} \in \mathcal{X}} \sum_{i=1}^{k} P_n\{|T_{in}(\underset{\sim}{x}) - T_i(\underset{\sim}{x}|\alpha_i, \mu_i)| \geq b_3\} f(\underset{\sim}{x}) \qquad (5.8) \\
+ \;& 2 \sum_{\underset{\sim}{x} \in \mathcal{X}_1} P_n\{|Q_n(\underset{\sim}{x}) - Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu})| \geq b_4\} f(\underset{\sim}{x}).
\end{aligned}
$$

Hence, in order to investigate the asymptotic optimality of the empirical Bayes two-stage selection procedure $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$, it suffices to study the asymptotic behavior of

$$
\begin{aligned}
& P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1\}, P_n\{|\psi_{in}(x_i, y_i) - \psi_i(x_i, y_i|\alpha_i, \mu_i)| \geq b_2\}, \\
& P_n\{|T_{in}(\underset{\sim}{x}) - T_i(\underset{\sim}{x}|\alpha_i, \mu_i)| \geq b_3\} \text{ and } P_n\{|Q_n(\underset{\sim}{x}) - Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu})| \geq b_4\} \text{ for } \underset{\sim}{x} \in \mathcal{X}_1,
\end{aligned}
$$

respectively. Now,

$$
\begin{aligned}
& \{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1\} \\
& \subset \{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1, C_{in} > 0\} \cup \{C_{in} \leq 0\}. \qquad (5.9)
\end{aligned}
$$

Since $\varphi_i(x_i|\alpha_i, \mu_i) = \frac{x_i + \alpha_i \mu_i}{M + \alpha_i}$ is a continuous function of (α_i, μ_i), when $C_{in} > 0$, there exists a positive constant $q_{i1}(x_i, \alpha_i, \mu_i)$, which depends on x_i, α_i and μ_i, such that

$$
\begin{aligned}
& \{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1, C_{in} > 0\} \qquad (5.10) \\
& \subset \{|\alpha_{in} - \alpha_i| \geq q_{i1}(x_i, \alpha_i, \mu_i), C_{in} > 0\} \cup \{|\mu_{in} - \mu_i| > q_{i1}(x_i, \alpha_i, \mu_i), C_{in} > 0\}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \geq b_1\} \leq \;& P_n\{|\alpha_{in} - \alpha_i| \geq q_{i1}(x_i, \alpha_i, \mu_i), C_{in} > 0\} \\
& + P_n\{|\mu_{in} - \mu_i| \geq q_{i1}(x_i, \alpha_i, \mu_i), C_{in} > 0\} \\
& + P_n\{C_{in} \leq 0\}.
\end{aligned}
$$

Similarly, for each x_i and y_i, there exists a positive constant $q_{i2}(x_i, y_i, \alpha_i, \mu_i)$ such that

$$
\begin{aligned}
& P_n\{|\psi_{in}(x_i, y_i) - \psi_i(x_i, y_i|\alpha_i, \mu_i)| \geq b_2\} \\
& \leq P_n\{|\alpha_{in} - \alpha_i| \geq q_{i2}(x_i, y_i, \alpha_i, \mu_i), C_{in} > 0\} \qquad (5.11) \\
& \quad + P_n\{|\mu_{in} - \mu_i| \geq q_{i2}(x_i, y_i, \alpha_i, \mu_i), C_{in} > 0\} + P_n\{C_{in} \leq 0\}.
\end{aligned}
$$

Next, we consider the term $P_n\{|T_{in}(\underset{\sim}{x}) - T_i(\underset{\sim}{x}|\alpha_i,\mu_i)| \geq b_3\}$. We first show that $T_i(\underset{\sim}{x}|\alpha_i,\mu_i)$ is a continuous function of (α_i,μ_i). From (3.3), $T_i(\underset{\sim}{x}|\alpha_i,\mu_i)$ is a summation of finite terms, namely, $d_{2i}^B(\underset{\sim}{x},y_i)[\theta_0 - \psi_i(x_i,y_i|\alpha_i,\mu_i)]f_i(y_i|x_i,\alpha_i,\mu_i)$, $y_i = 0,\ldots,m$. Therefore, it suffices to investigate the continuity property of each term. For this purpose, we introduce the following lemma.

Lemma 5.1. Let $p(t)$ be a continuous function of the variable t. Define

$$W(t) = \begin{cases} 1 & \text{if } p(t) \leq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $P(t) = W(t)p(t)$. Then $P(t)$ is a continuous function of t.

The proof is straightforward by using the definition of continuity. The detail is omitted here.

It is clear that $[\theta_0 - \psi_i(x_i,y_i|\alpha_i,\mu_i)]f_i(y_i|x_i,\alpha_i,\mu_i)$ is a continuous function of (α_i,μ_i). By Lemma 5.1 and the definition of $d_{2i}^B(\underset{\sim}{x},y_i)$, it can be seen that $T_i(\underset{\sim}{x}|\alpha_i,\mu_i)$ is a continuous function of (α_i,μ_i).

Then, applying an argument similar to what we used in the preceding, we can assert: There exists a positive constant $q_{i3}(x_i,\alpha_i,\mu_i)$ such that

$$\begin{aligned} &P_n\{|T_{in}(\underset{\sim}{x}) - T_i(\underset{\sim}{x}|\alpha_i,\mu_i)| \geq b_3\} \\ &\leq P_n\{|\alpha_{in} - \alpha_i| \geq q_{i3}(x_i,\alpha_i,\mu_i), C_{in} > 0\} \\ &\quad + P_n\{|\mu_{in} - \mu_i| \geq q_{i3}(x_i,\alpha_i,\mu_i), C_{in} > 0\} + P_n\{C_{in} \leq 0\}. \end{aligned} \tag{5.12}$$

Applying an argument analogous to the preceding one, we can obtain:

$$\begin{aligned} &P_n\{|Q_n(\underset{\sim}{x}) - Q(\underset{\sim}{x}|\underset{\sim}{\alpha},\underset{\sim}{\mu})| \geq b_4\} \\ &\leq P_n\{|Q_n(\underset{\sim}{x}) - Q(\underset{\sim}{x}|\underset{\sim}{\alpha},\underset{\sim}{\mu})| \geq b_4, C_{in} > 0 \text{ for all } i = 1,\ldots,k\} + \sum_{i=1}^{k} P\{C_{in} \leq 0\} \quad (5.13) \\ &\leq \sum_{i=1}^{k}[P_n\{|\alpha_{in} - \alpha_i| > q_4(\underset{\sim}{x},\underset{\sim}{\alpha},\underset{\sim}{\mu}), C_{in} > 0\} + P_n\{|\mu_{in} - \mu_i| > q_4(\underset{\sim}{x},\underset{\sim}{\alpha},\underset{\sim}{\mu}), C_{in} > 0\}] \\ &\quad + \sum_{i=1}^{k} P_n\{C_{in} \leq 0\}. \end{aligned}$$

where $q_4(\underset{\sim}{x},\underset{\sim}{\alpha},\underset{\sim}{\mu})$ is a positive constant depending on $\underset{\sim}{x}$, $\underset{\sim}{\alpha}$ and $\underset{\sim}{\mu}$.

¿From (5.10)-(5.13), it suffices to investigate the asymptotic behaviors of $P_n\{|\alpha_{in} - \alpha_i| > \varepsilon, C_{in} > 0\}$, $P_n\{|\mu_{in} - \mu_i| > \varepsilon, C_{in} > 0\}$ and $P_n\{C_{in} \leq 0\}$ for each $i = 1,\ldots,k$, where $\varepsilon > 0$.

The following lemma is from Gupta and Liang (1989).

Lemma 5.2. For $\varepsilon > 0$,

(a) $P_n\{|\mu_{in} - \mu_i| > \varepsilon\} = O(\exp(-2n\varepsilon^2))$.

(b) $P_n\{|C_{in} - C_i| > \varepsilon\} = O(\exp(-n\varepsilon^2/8))$.

(c) $P_n\{|D_{in} - D_i| > \varepsilon\} = O(\exp(-n\varepsilon^2/2))$.

(d) $P_n\{C_{in} \leq 0\} = O(\exp(-nC_i^2/8))$.

Lemma 5.3. For $\varepsilon > 0$,

$$P_n\{|\alpha_{in} - \alpha_i| > \varepsilon, C_{in} > 0\} = O(\exp(-\frac{n\varepsilon^2 C_i^2}{8}\min(1, \frac{1}{4(\alpha_i+\varepsilon)^2}))).$$

Proof: $P_n\{|\alpha_{in} - \alpha_i| > \varepsilon, C_{in} > 0\}$

$$= P_n\{\alpha_{in} - \alpha_i < -\varepsilon, C_{in} > 0\} + P_n\{\alpha_{in} - \alpha_i > \varepsilon, C_{in} > 0\}.$$

where $P_n\{\alpha_{in} - \alpha_i < -\varepsilon, C_{in} > 0\} = 0$ if $\alpha_i - \varepsilon \leq 0$. As $\alpha_i - \varepsilon > 0$, by the definition of α_{in}, Lemma 5.2, and an application of Bonferroni inequality, we obtain:

$$\begin{aligned}
&P_n\{\alpha_{in} - \alpha_i < -\varepsilon, C_{in} > 0\} \\
&\leq P_n\{(D_{in} - D_i) - (C_{in} - C_i)(\alpha_i - \varepsilon) < -\varepsilon C_i\} \\
&\leq P_n\{D_{in} - D_i < -\varepsilon C_i/2\} + P_n\{C_{in} - C_i > \varepsilon C_i/2(\alpha_i - \varepsilon)\} \\
&= O(\exp(-n\varepsilon^2 C_i^2/8)) + O(\exp(-n\varepsilon^2 C_i^2/32(\alpha_i - \varepsilon)^2)) \\
&= O(\exp(-\frac{n\varepsilon^2 C_i^2}{8}\min(1, \frac{1}{4(\alpha_i - \varepsilon)^2}))).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&P_n\{\alpha_{in} - \alpha_i > \varepsilon, C_{in} > 0\} \\
&\leq P_n\{(D_{in} - D_i) - (C_{in} - C_i)(\alpha_i + \varepsilon) > \varepsilon C_i\} \\
&\leq P_n\{D_{in} - D_i > \varepsilon C_i/2\} + P\{C_{in} - C_i < -\varepsilon C_i/2(\alpha_i + \varepsilon)\} \\
&= O(\exp(-\frac{n\varepsilon^2 C_i^2}{8}\min(1, \frac{1}{4(\alpha_i + \varepsilon)^2}))).
\end{aligned}$$

This completes the proof of Lemma 5.3.

Now, let

$$\begin{aligned}
&\varepsilon_1 = \min_{1\leq i\leq k}\min_{0\leq x_i\leq M}\{q_{i1}(x_i, \alpha_i, \mu_i)\}, \varepsilon_2 = \min_{1\leq i\leq k}\min_{(x_i,y_i)}\{q_{i2}(x_i, y_i, \alpha_i, \mu_i)\}, \\
&\varepsilon_3 = \min_{1\leq i\leq k}\min_{x_i}\{q_{i3}(x_i, \alpha_i, \mu_i)\}, \varepsilon_4 = \min_{\underset{\sim}{x}}\{q_4(\underset{\sim}{x}, \underset{\sim}{\alpha}, \underset{\sim}{\mu})\}, \varepsilon_5 = \frac{1}{2}\min_{1\leq i\leq k}\{\alpha_i\}
\end{aligned}$$

and let $\varepsilon^* = \min_{1\leq j\leq 5}\{\varepsilon_j\}$. Since the sample space $\mathcal{X} \times \mathcal{Y}$ is finite, we see that $\varepsilon_j > 0$ for $j = 1, \ldots, 5$, and therefore $\varepsilon^* > 0$.

In the following theorem we prove the asymptotic optimality of the empirical Bayes two-stage selection procedure $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$.

Theorem 5.1. Let $\{(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})\}_{n=1}^{\infty}$ be the sequence of empirical Bayes two-stage selection procedures constructed in Section 4. Then,

$$E_n R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n}) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B) = O(\exp(-c^* n))$$

where $c^* = \min(\frac{\varepsilon^{*2}}{8}, \min_{1\le i\le k} \frac{C_i^2}{8}, \min_{1\le i\le k}(\frac{\varepsilon^{*2}C_i^2}{8}\min(1, \frac{1}{9\alpha_i^2}))) > 0$.

Proof: By definitions of ε^* and c^*, from Lemmas 5.2 and 5.3, we have:

$$\begin{aligned}
&P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i|\alpha_i, \mu_i)| \ge b_1\} = O(\exp(-c^* n)),\\
&P_n\{|\psi_{in}(x_i, y_i) - \psi_i(x_i, y_i|\alpha_i, \mu_i)| \ge b_2\} = O(\exp(-c^* n)),\\
&P_n\{|T_{in}(\underset{\sim}{x}) - T_i(\underset{\sim}{x}|\alpha_i, \mu_i)| \ge b_3\} = O(\exp(-c^* n)),\\
\text{and}\quad &P_n\{|Q_n(\underset{\sim}{x}) - Q(\underset{\sim}{x}|\underset{\sim}{\alpha}, \underset{\sim}{\mu})| \ge b_4\} = O(\exp(-c^* n))
\end{aligned}$$

for all $\underset{\sim}{x}$ and y_i. Note that the bound is independent of $\underset{\sim}{x}$ and y_i. Replacing these results into (5.8), we obtain:

$$\begin{aligned}
&E_n R(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n}) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)\\
&\le O(\exp(-c^* n))k^2 \sum_{\underset{\sim}{x}\in\mathcal{X}} f(\underset{\sim}{x})\\
&\quad + O(\exp(-c^* n)) \sum_{\underset{\sim}{x}\in\mathcal{X}} \sum_{i=1}^{k} \left[\sum_{y_i=0}^{m} f_i(y_i|x_i, \alpha_i, \mu_i)\right] f(\underset{\sim}{x})\\
&\quad + O(\exp(-c^* n))k^2 \sum_{\underset{\sim}{x}\in\mathcal{X}} f(\underset{\sim}{x}) + 2O(\exp(-c^* n)) \sum_{\underset{\sim}{x}\in\mathcal{X}} f(\underset{\sim}{x})\\
&= O(\exp(-c^* n)).
\end{aligned}$$

Hence the proof of the theorem is completed.

6 A Special Case Where $(\alpha_1, \mu_1) =, \cdots, = (\alpha_k, \mu_k)$

In this section, it is assumed that $(\alpha_1, \mu_1) =, \cdots, = (\alpha_k, \mu_k) = (\alpha, \mu)$, and the values of the common parameters (α, μ) are unknown. Under this assumption, the Bayes two-stage selection procedure can be simplified. Also, $X_{ij}, j = 1, 2, \ldots, i = 1, \ldots, k$, are *iid*. Hence, we can construct a more efficient empirical Bayes two-stage selection procedure. First, we derive a Bayes two-stage selection procedure $(\hat{\tau}^B, \hat{\underset{\sim}{\delta}}^B, \hat{\underset{\sim}{d}}_1^B, \hat{\underset{\sim}{d}}_2^B)$.

A First-Stage Selection Rule $\hat{\underset{\sim}{d}}_1^B = (\hat{d}_{10}^B, \ldots, \hat{d}_{1k}^B)$

For each $\underset{\sim}{x} \in \mathcal{X}$, let $\hat{I}(\underset{\sim}{x}) = \{i = 1, \ldots, k | x_i = \max_{1\le j\le k} x_j\}$. Define $\hat{\imath} \equiv \hat{\imath}(\underset{\sim}{x}) = \min\{i | i \in \hat{I}(\underset{\sim}{x})\}$. We then define a first-stage selection rule $\hat{\underset{\sim}{d}}_1^B = (\hat{d}_{10}^B, \ldots, \hat{d}_{1k}^B)$ as follows:

$$\begin{cases} \text{If } \varphi_{\hat{\imath}}(x_{\hat{\imath}}|\alpha, \mu) \ge \theta_0, \text{ define } \hat{d}_{1\hat{\imath}}^B(\underset{\sim}{x}) = 1 \text{ and } \hat{d}_{1j}^B(\underset{\sim}{x}) = 0 \text{ for } j \ne \hat{\imath}. \\ \text{If } \varphi_{\hat{\imath}}(x_{\hat{\imath}}|\alpha, \mu) < \theta_0, \text{ define } \hat{d}_{10}^B(\underset{\sim}{x}) = 1 \text{ and } \hat{d}_{1j}^B(\underset{\sim}{x}) = 0 \text{ for } j = 1, \ldots, k. \end{cases} \tag{6.1}$$

A Second-Stage Selection Rule $\hat{\underset{\sim}{d}}_2^B = (\hat{d}_{20}^B, \ldots, \hat{d}_{2k}^B)$

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

$$\hat{d}_{2i}^B(\underset{\sim}{x}, y) \equiv \hat{d}_{2i}^B(\underset{\sim}{x}, y_i) = \begin{cases} 1 & \text{if } \psi_i(x_i, y_i|\alpha, \mu) \geq \theta_0, \\ 0 & \text{otherwise;} \end{cases} \tag{6.2}$$

and

$$\hat{d}_{20}^B(\underset{\sim}{x}, y) = \sum_{i=1}^{k} \hat{\delta}_i^B(\underset{\sim}{x})[1 - \hat{d}_{2i}^B(\underset{\sim}{x}, y)]$$

where $\hat{\underset{\sim}{\delta}}^B = (\hat{\delta}_1^B, \ldots, \hat{\delta}_k^B)$ is the identity rule defined as:

An Identity Rule $\hat{\underset{\sim}{\delta}}^B = (\hat{\delta}_1^B, \ldots, \hat{\delta}_k^B)$

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

$$\hat{\delta}_i^B(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } i = \hat{\imath}, \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

where $\hat{\imath}$ is the index defined before.

A Stopping Rule $\hat{\tau}^B$

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

$$\hat{Q}(\underset{\sim}{x}|\alpha, \mu) = -mc_2 + \hat{d}_{1\hat{\imath}}^B(\underset{\sim}{x})[\theta_0 - \varphi_{\hat{\imath}}(x_{\hat{\imath}}|\alpha, \mu)] - \sum_{y=0}^{m} \hat{d}_{2\hat{\imath}}^B(\underset{\sim}{x}, y)[\theta_0 - \psi_{\hat{\imath}}(x_{\hat{\imath}}, y|\alpha, \mu)] f_{\hat{\imath}}(y|x_{\hat{\imath}}, \alpha, \mu). \tag{6.4}$$

We then define a stopping rule $\hat{\tau}^B$ as follows:

$$\hat{\tau}^B(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \hat{Q}(\underset{\sim}{x}|\alpha, \mu) \leq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.5}$$

We claim the following result.

Theorem 6.1 Suppose that $(\alpha_1, \mu_1) = \ldots = (\alpha_k, \mu_k) = (\alpha, \mu)$. Then, the two-stage selection procedure $(\hat{\tau}^B, \hat{\underset{\sim}{\delta}}^B, \hat{\underset{\sim}{d}}_1^B, \hat{\underset{\sim}{d}}_2^B)$ defined through (6.1) – (6.5) is a Bayes two-stage selection procedure.

Theorem 6.1 is a consequence of the following lemma. Since the proofs for the lemma and the theorem involves only straightforward computation, the detail is omitted.

Lemma 6.1

(a) The marginally conditional probability function $f_i(y_i|x_i, \alpha, \mu)$ has monotone likelihood ratio in x_i and y_i. That is, for $y_i < y_i'$, $f_i(y_i'|x_i, \alpha, \mu)/f_i(y_i|x_i, \alpha, \mu)$ is increasing in x_i and for $x_i < x_i'$, $f_i(y_i|x_i', \alpha, \mu)/f_i(y_i/x_i, \alpha, \mu)$ is increasing in y_i.

(b) Let $\hat{T}_i(\underset{\sim}{x}|\alpha, \mu) = \sum_{y_i=0}^{m} \hat{d}_{2i}^B(\underset{\sim}{x}, y_i)[\theta_0 - \psi_i(x_i, y_i|\alpha, \mu)] f_i(y_i|x_i, \alpha, \mu)$ for $\underset{\sim}{x} \in \mathcal{X}, i = 1, \ldots, k$. Then $\hat{T}_i(\underset{\sim}{x}|\alpha, \mu)$ is nonincreasing in x_i.

The Proposed Empirical Bayes Two-Stage Selection Procedure $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}^n, \hat{\underset{\sim}{d}}_1^n, \hat{\underset{\sim}{d}}_2^n)$

Under the assumption that $(\alpha_1, \mu_1) = \ldots = (\alpha_k, \mu_k) = (\alpha, \mu)$, X_{ij}, $j = 1, 2, \ldots, i = 1, \ldots, k$ are *iid* and

$$\begin{aligned} E[X_{ij}/M] &= \mu, \\ E[(X_{ij}/M)^2] &= \frac{\mu}{M} + \frac{(\alpha\mu + 1)\mu(M-1)}{M(\alpha+1)} \equiv \nu, \end{aligned}$$

$\alpha = \frac{D}{C}$, where $D = \mu - \nu$, and $C = \nu - \frac{\mu}{M} + \frac{\mu^2}{M} - \mu^2$.

Let

$$\begin{aligned} \mu_n &= \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{X_{ij}}{M}, \\ \nu_n &= \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n}(\frac{X_{ij}}{M})^2, \\ C_n &= \nu_n - \frac{\mu_n}{M} + \frac{\mu_n^2}{M} - \mu_n^2, \\ D_n &= \mu_n - \nu_n. \end{aligned}$$

Define

$$\hat{\alpha}_n = \begin{cases} \frac{D_n}{C_n} & \text{if } C_n > 0, \\ \infty & \text{otherwise;} \end{cases} \tag{6.6}$$

$$\hat{\varphi}_{in}(x_i) = \begin{cases} \frac{x_i + \hat{\alpha}_n\mu_n}{M + \hat{\alpha}_n} & \text{if } C_n > 0, \\ \mu_n & \text{otherwise;} \end{cases} \tag{6.7}$$

$$\hat{\psi}_{in}(x_i, y_i) = \begin{cases} \frac{x_i + y_i + \hat{\alpha}_n\mu_n}{M + m + \hat{\alpha}_n} & \text{if } C_n > 0, \\ \mu_n & \text{otherwise;} \end{cases} \tag{6.8}$$

and

$$\hat{f}_{in}(y_i|x_i) = \begin{cases} f_i(y_i|x_i, \hat{\alpha}_n, \mu_n) & \text{if } C_n > 0, \\ \binom{m}{y}\mu_n^y(1-\mu_n)^{m-y} & \text{otherwise.} \end{cases} \tag{6.9}$$

We propose an empirical Bayes two-stage selection procedure $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}^n, \hat{\underset{\sim}{d}}_1^n, \hat{\underset{\sim}{d}}_2^n)$ as follows:

Empirical Bayes First-Stage Selection Rule $\hat{\underset{\sim}{d}}_1^n = (\hat{d}_{10}^n, \ldots, \hat{d}_{1k}^n)$

For each $\underset{\sim}{x} \in \mathcal{X}$, let $\hat{i} = \hat{i}(\underset{\sim}{x})$ be the index defined precedingly. Then, we define $\hat{\underset{\sim}{d}}_1^n$ as follows:

$$\begin{cases} \text{If } \hat{\varphi}_{\hat{i}n}(x_{\hat{i}}) \geq \theta_0, \text{ define } \hat{d}_{1\hat{i}}^n(\underset{\sim}{x}) = 1 \text{ and } \hat{d}_{1j}^n(\underset{\sim}{x}) = 0 \text{ for } j \neq \hat{i}. \\ \text{If } \hat{\varphi}_{\hat{i}n}(x_{\hat{i}}) < \theta_0, \text{ define } \hat{d}_{10}^n(\underset{\sim}{x}) = 1 \text{ and } \hat{d}_{1i}^n(\underset{\sim}{x}) = 0 \text{ for } i = 1, \ldots, k. \end{cases} \tag{6.10}$$

Empirical Bayes Second-Stage Selection Rule $\hat{\underset{\sim}{d}}_2^n = (\hat{d}_{20}^n, \ldots, \hat{d}_{2k}^n)$

We define $\hat{\underset{\sim}{d}}{}_2^n$ as follows: For each $\underset{\sim}{x}$ and $\underset{\sim}{y}$, $i = 1, \ldots, k$,

$$\hat{d}^n_{2i}(\underset{\sim}{x}, \underset{\sim}{y}) = \begin{cases} 1 & \text{if } \hat{\psi}_{in}(x_i, y_i) \geq \theta_0, \\ 0 & \text{otherwise;} \end{cases} \tag{6.11}$$

and

$$\hat{d}^n_{20}(\underset{\sim}{x}, \underset{\sim}{y}) = \sum_{i=1}^{k} \hat{\delta}^B_i(\underset{\sim}{x})[1 - \hat{d}^n_{2i}(\underset{\sim}{x}, \underset{\sim}{y})]$$

where $\hat{\underset{\sim}{\delta}}{}^B = (\hat{\delta}^B_1, \ldots, \hat{\delta}^B_k)$ is the Bayes identity rule given in (6.3).

Empirical Bayes Identity Rule

We use $\hat{\underset{\sim}{\delta}}{}^B$ as the identity rule. That is $\hat{\underset{\sim}{\delta}}{}^n = \hat{\underset{\sim}{\delta}}{}^B$. Therefore, $\hat{\underset{\sim}{\delta}}{}^n$ is independent of the past observations $X_{ij}, j = 1, \ldots, n, i = 1, \ldots, k$.

Empirical Bayes Stopping Rule $\hat{\tau}^n$

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

$$\hat{Q}_n(\underset{\sim}{x}) = -mc_2 + \hat{d}^n_{1i}(\underset{\sim}{x})[\theta_0 - \hat{\varphi}_{in}(x_i)] - \sum_{y=0}^{m} \hat{d}^n_{2i}(\underset{\sim}{x}, y)[\theta_0 - \hat{\psi}_{in}(x_i, y)]\hat{f}_{in}(y|x_i).$$

Then, define

$$\hat{\tau}^n(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \hat{Q}_n(\underset{\sim}{x}) \leq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.12}$$

Asymptotic Optimality of $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}{}^n, \hat{\underset{\sim}{d}}{}^n_1, \hat{\underset{\sim}{d}}{}^n_2)$

For the empirical Bayes two-stage selection procedure $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}{}^n, \hat{\underset{\sim}{d}}{}^n_1, \hat{\underset{\sim}{d}}{}^n_2)$, we can establish the following asymptotic optimality.

Theorem 6.2 Let $\{(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}{}^n, \hat{\underset{\sim}{d}}{}^n_1, \hat{\underset{\sim}{d}}{}^n_2)\}_{n=1}^{\infty}$ be the sequence of the empirical Bayes two-stage selection procedures constructed through (6.10) – (6.12). Then, under the assumption that $(\alpha_1, \mu_1) =, \cdots, = (\alpha_k, \mu_k) = (\alpha, \mu)$, we have

$$E_n R(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}{}^n, \hat{\underset{\sim}{d}}{}^n_1, \hat{\underset{\sim}{d}}{}^n_2) - R(\hat{\tau}^B, \hat{\underset{\sim}{\delta}}{}^B, \hat{\underset{\sim}{d}}{}^B_1, \hat{\underset{\sim}{d}}{}^B_2) = O(\exp(-c^{**}n))$$

for some positive constant c^{**}.

The proof for Theorem 6.2 is analogous to that of Theorem 5.1. We omit the detail here.

7 Small Sample Performance: Simulation Study

A simulation study was carried out to investigate the performance of the proposed empirical Bayes two-stage selection procedures for small to moderate values of n. We considered $k = 3$ Bernoulli populations π_1, π_2 and π_3. Recall that the random success probability Θ_i of population π_i has a beta prior distribution with the probability density function

$$h_i(\theta) = \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i\mu_i)\Gamma(\alpha_i(1-\mu_i))}\theta^{\alpha_i\mu_i - 1}(1-\theta)^{\alpha_i(1-\mu_i)-1},$$

$0 < \theta < 1$, where $0 < \mu_i < 1, \alpha_i > 0$.

For given past observations $(\underset{\sim}{X}_j = (X_{1j}, \ldots, X_{kj}), j = 1, 2, \ldots, n)$, $R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$ is the associated conditional Bayes risk of the proposed empirical Bayes two-stage selection procedure $(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$. Then, we use $R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$ as an estimator of the difference between $E_n R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$ and $R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$ where the expectation E_n is taken with respect to the probability measure generated by $(\underset{\sim}{X}_j, j = 1, \ldots, n)$.

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

(1) For each n and for each $i = 1, 2, 3$ generate independent random values $X_{i1}, X_{i2}, \ldots, X_{in}$ as follows:

$$\begin{cases} \text{for } j = 1, 2, \ldots, n \\ \text{(a) generate } \Theta_i \text{ from density } h_i(\theta). \\ \text{(b) generate } X_{ij} \text{ from Binomial } (M, \theta_i). \end{cases}$$

(2) Based on the past observation $(\underset{\sim}{X}_j, j = 1, \ldots, n)$ and the present observations $\underset{\sim}{X} = (X_1 \ldots, X_k)$ and $\underset{\sim}{Y} = (Y_1, \ldots, Y_k)$, construct the empirical Bayes two-stage selection procedure $(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n)$ and compute the conditional difference

$$D(n) = R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B).$$

(3) Steps (1) and (2) were repeated 400 times. The average of the conditional difference based on the 400 repetitions, which is denoted by $\bar{D}(n)$, is used as an estimator of the difference $E_n R(\tau^n, \underset{\sim}{\delta}^n, \underset{\sim}{d}_1^n, \underset{\sim}{d}_2^n) - R(\tau^B, \underset{\sim}{\delta}^B, \underset{\sim}{d}_1^B, \underset{\sim}{d}_2^B)$. Also, $SE(\bar{D}(n))$, the estimated standard error, is computed.

Note that $\bar{D}(n)$ corresponding to $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$ and $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}^n, \hat{\underset{\sim}{d}}_1^n, \hat{\underset{\sim}{d}}_2^n)$ are denoted by $\bar{D}^*(n)$ and $\hat{\bar{D}}(n)$, respectively. Tables 1 and 2 list some simulation results on the performance of the proposed empirical Bayes two-stage procedure $(\tau^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$, for the case where $\theta_0 = 0.7, c_2 = 0.05, M = m = 5, \alpha_1 = \alpha_2 = \alpha_3 = 3$. Also, Tables 3 lists the simulation results of $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}^n, \hat{\underset{\sim}{d}}_1^n, \hat{\underset{\sim}{d}}_2^n)$, for $\theta_0 = 0.7, c_2 = 0.05, M = m = 5, \alpha_1 = \alpha_2 = \alpha_3 = 3, \mu_1 = \mu_2 = \mu_3 = 0.6$.

The simulation results indicate that $\bar{D}(n)$ tends to zero quite fast as the value of n increases. ¿From Table 1 and Table 2, we learn that $\bar{D}^*(n)$ has roughly the same rate of convergence in both tables, for $n \leq 140$ but $\bar{D}^*(n)$ converges faster in Table 2 than in Table 1, for $n \geq 160$. By direct computation, the marginal means of X_i and Y_i are $M\mu_i$ and $m\mu_i$, respectively. The distances of the μ_i's in Table 1 $(\mu_1 = 0.67, \mu_2 = 0.69, \mu_3 = 0.71)$ are 0.02. and those of the μ_i's in Table 2 $(\mu_1 = 0.4, \mu_2 = 0.6, \mu_3 = 0.8)$ are 0.2. Therefore, the result is reasonable, because it is easier to identify the best population when the distances between the means of the populations are larger.

¿From Table 3, we learn that $\hat{\bar{D}}(n)$ decreases to zero very rapidly, in fact, $\hat{\bar{D}}(n) \equiv 0$ when $n \geq 140$. Since, in the special case, all the past observations from populations π_1, π_2 and π_3 are used together to estimate the parameters (α, μ), we get a more efficient empirical Bayes two-stage selection procedure $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}^n, \hat{\underset{\sim}{d}}_1^n, \hat{\underset{\sim}{d}}_2^n)$.

Table 1. Performance of $(\underset{\sim}{\tau}^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$ for $\mu_1 = 0.67, \mu_2 = 0.69$ and $\mu_3 = 0.71$

n	$\bar{D}^*(n)$	$SE(\bar{D}^*(n))$
20	3526.0389×10^{-5}	375.0017×10^{-5}
40	1000.4051×10^{-5}	191.7060×10^{-5}
60	297.4756×10^{-5}	86.3280×10^{-5}
80	118.5197×10^{-5}	5.5431×10^{-5}
100	108.1376×10^{-5}	5.1142×10^{-5}
120	103.5187×10^{-5}	5.1124×10^{-5}
140	98.0947×10^{-5}	4.8563×10^{-5}
160	91.3719×10^{-5}	4.7352×10^{-5}
180	88.5946×10^{-5}	4.5664×10^{-5}
200	87.1701×10^{-5}	4.5202×10^{-5}
250	80.1192×10^{-5}	4.2767×10^{-5}
300	73.2212×10^{-5}	4.1745×10^{-5}
350	65.3211×10^{-5}	3.8014×10^{-5}
400	61.3566×10^{-5}	3.6322×10^{-5}
450	58.3146×10^{-5}	3.4673×10^{-5}
500	56.1668×10^{-5}	3.6309×10^{-5}

Table 2. Performance of $(\underset{\sim}{\tau}^{*n}, \underset{\sim}{\delta}^{*n}, \underset{\sim}{d}_1^{*n}, \underset{\sim}{d}_2^{*n})$ for $\mu_1 = 0.4, \mu_2 = 0.6$ and $\mu_3 = 0.8$

n	$\bar{D}^*(n)$	$SE(\bar{D}^*(n))$
20	3572.4389×10^{-5}	397.5006×10^{-5}
40	1464.2346×10^{-5}	244.4725×10^{-5}
60	881.7971×10^{-5}	188.2158×10^{-5}
80	536.3852×10^{-5}	136.3745×10^{-5}
100	271.9296×10^{-5}	89.2091×10^{-5}
120	181.6412×10^{-5}	58.5219×10^{-5}
140	96.8247×10^{-5}	20.3857×10^{-5}
160	69.8568×10^{-5}	7.7019×10^{-5}
180	56.6766×10^{-5}	7.0904×10^{-5}
200	51.2190×10^{-5}	6.3550×10^{-5}
250	32.7545×10^{-5}	4.8583×10^{-5}
300	18.5442×10^{-5}	3.3724×10^{-5}
350	17.5653×10^{-5}	3.3141×10^{-5}
400	11.8336×10^{-5}	2.4134×10^{-5}
450	9.2903×10^{-5}	2.1138×10^{-5}
500	7.4519×10^{-5}	1.6726×10^{-5}

Table 3. Performance of $(\hat{\tau}^n, \hat{\underset{\sim}{\delta}}^n, \hat{\underset{\sim}{d}}_1^n, \hat{\underset{\sim}{d}}_2^n)$ for $\mu_1 = \mu_2 = \mu_3 = 0.6$

n	$\bar{D}^*(n)$	$SE(\bar{D}^*(n))$
20	299.2199×10^{-5}	64.3043×10^{-5}
40	118.6488×10^{-5}	20.1839×10^{-5}
60	62.1228×10^{-5}	14.9756×10^{-5}
80	43.0392×10^{-5}	12.5011×10^{-5}
100	15.5307×10^{-5}	7.6079×10^{-5}
120	7.6334×10^{-5}	5.3908×10^{-5}
140	0.0000×10^{-5}	0.0000×10^{-5}
160	0.0000×10^{-5}	0.0000×10^{-5}
180	0.0000×10^{-5}	0.0000×10^{-5}
200	0.0000×10^{-5}	0.0000×10^{-5}
250	0.0000×10^{-5}	0.0000×10^{-5}
300	0.0000×10^{-5}	0.0000×10^{-5}
350	0.0000×10^{-5}	0.0000×10^{-5}
400	0.0000×10^{-5}	0.0000×10^{-5}
450	0.0000×10^{-5}	0.0000×10^{-5}
500	0.0000×10^{-5}	0.0000×10^{-5}

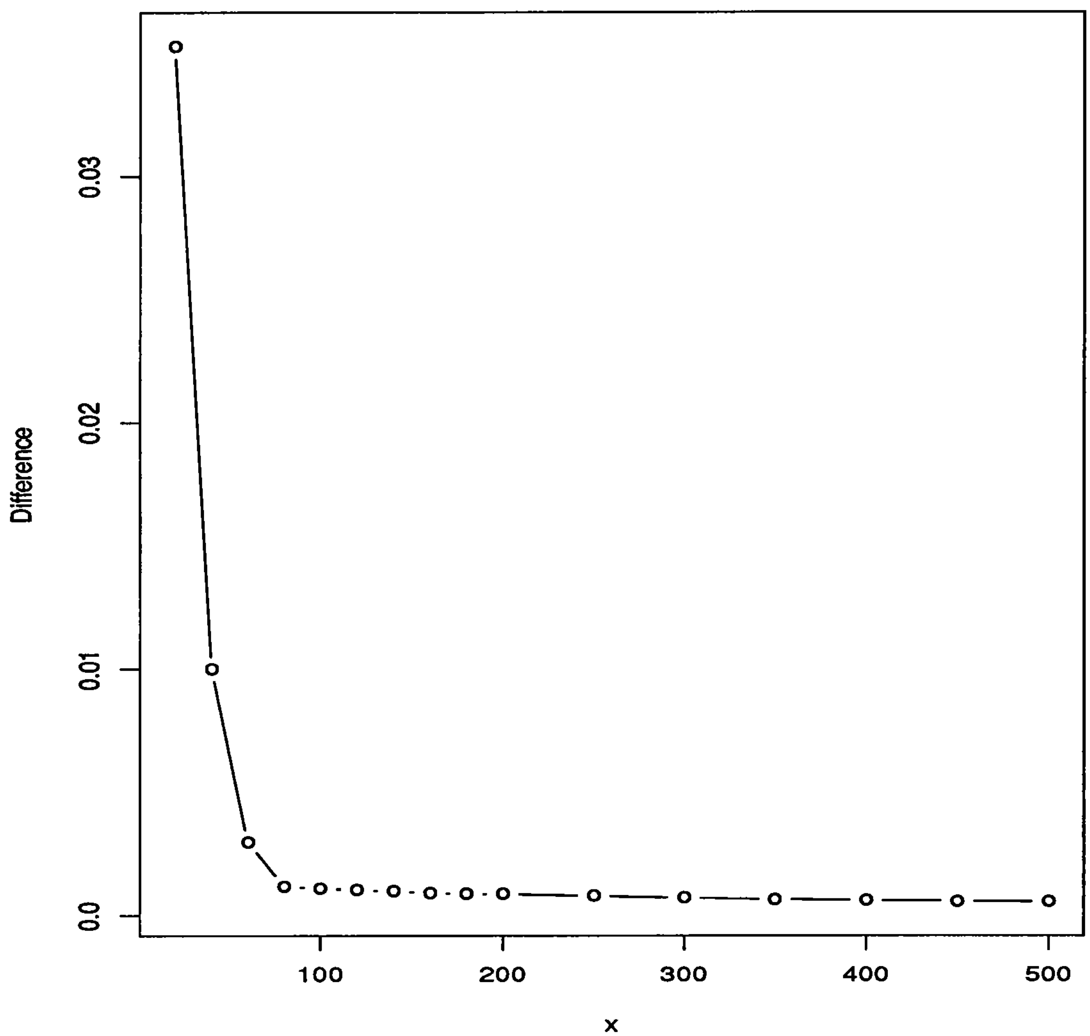

Figure 1: $\bar{D}^{*n}$ **vs** n **for Table 1**

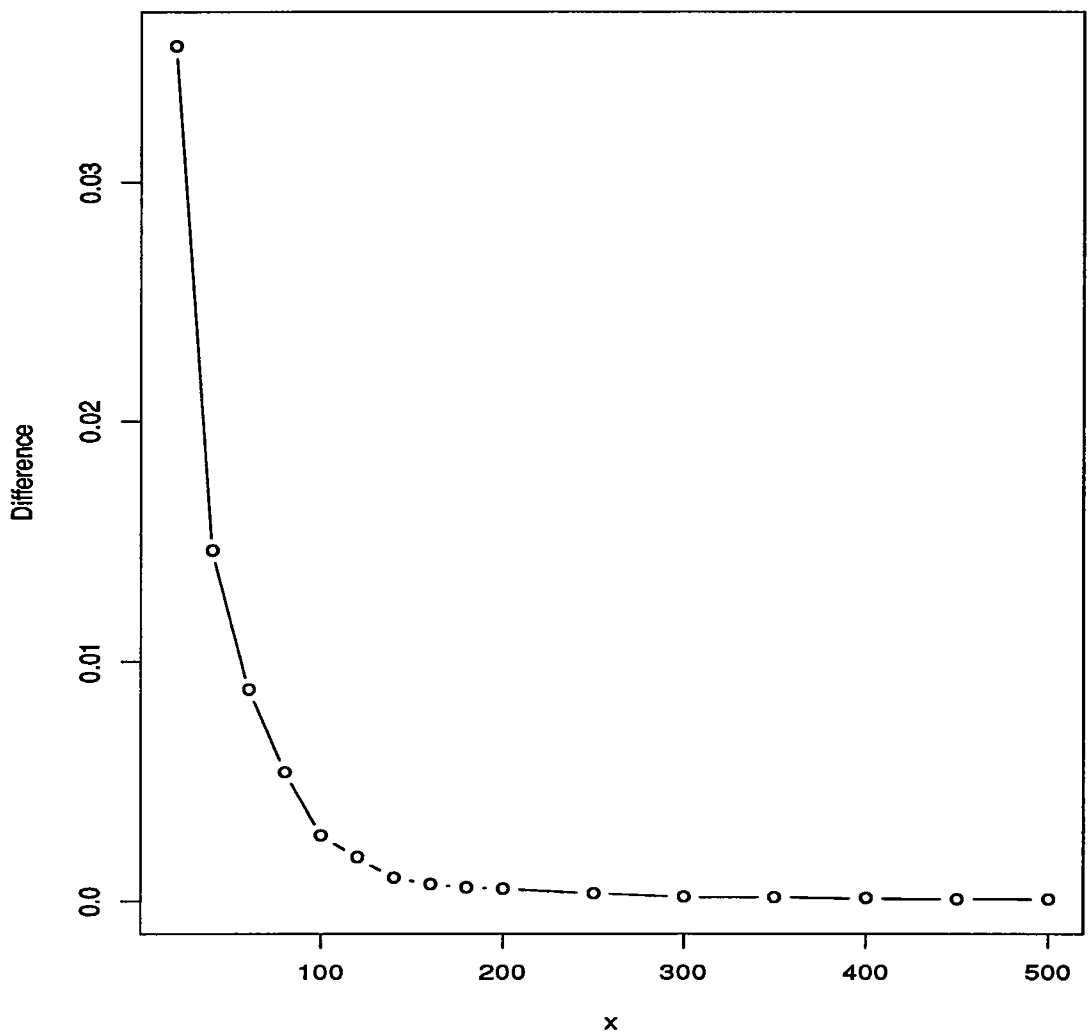

Figure 2: $\bar{D}^{*n}$ vs n **for Table 2**

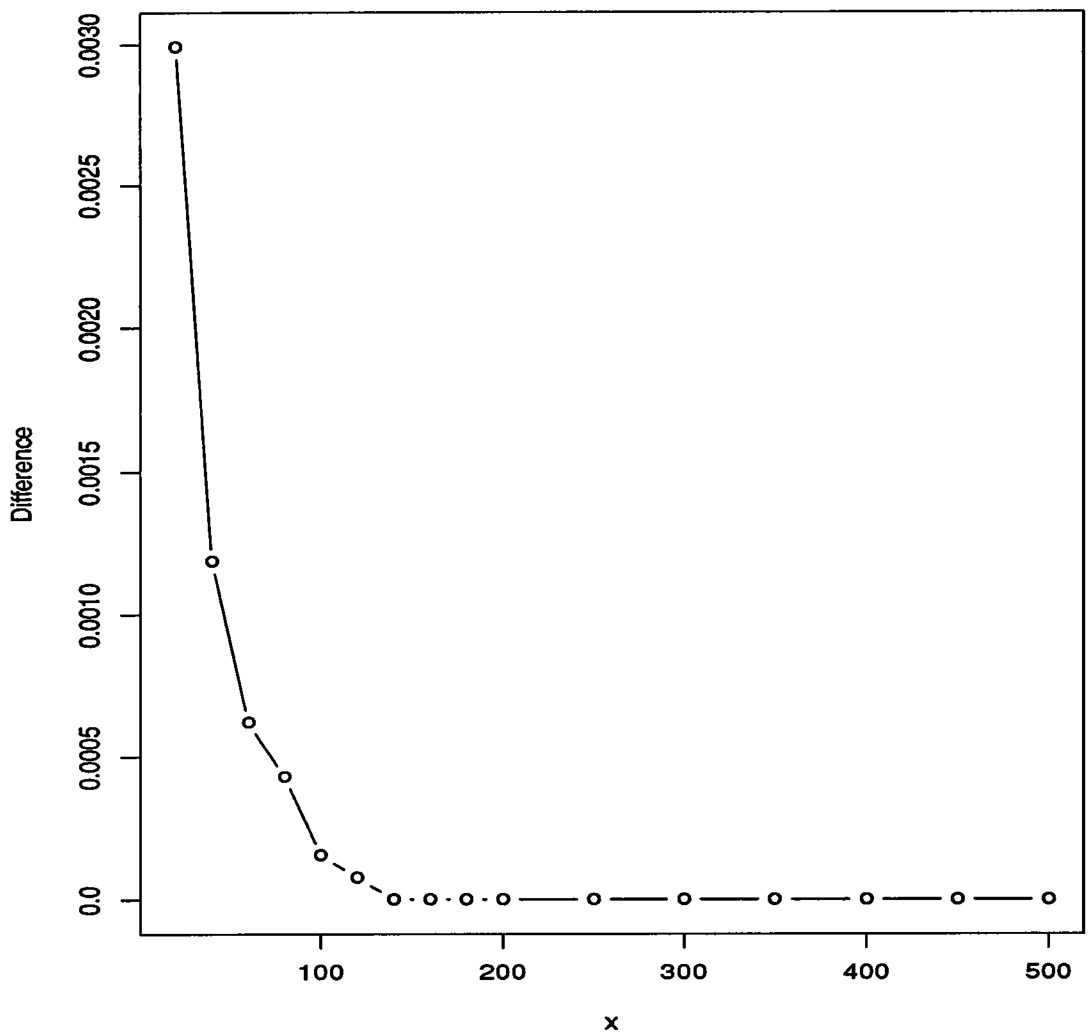

Figure 3: $\hat{\bar{D}}^n$ **vs** n **for Table 3**

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.

Bechhofer, R. E. and Kulkarni, R. V. (1982). Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. *Statistical Decision Theory and Related Topics - III*, (Eds. S. S. Gupta and J. O. Berger), Academic Press, New York, Vol. I, 61–108.

Bechhofer, R. E. and Turnbull, B. W. (1978). Two $(k+1)$-decision selection procedures for comparing k normal means with a specified standard. *J. Amer. Statist. Assoc.*, **73**, 385–392.

Dunnett, Charles W. (1984). Selection of the best treatment in comparison to a control with an application to a medical trial. *Design of Experiments: Ranking and Selection*, (Eds. A. C. Tamhane and T. Santner), Marcel Dekker, 47–66.

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

Gupta, S. S. and Huang, D. Y. (1976). Subset selection procedures for the entropy function associated with the binomial populations. *Sankhyā Ser. A*, **38**, 153–173.

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, Berlin-New York, **Vol. 1**, 213–224.

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

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

Gupta, S. S. and Panchapakesan, S. (1985). Subset selection procedures: review and assessment. *American Journal of Mathematical and Management Sciences*, American Sciences Press, **Vol. 5**, Nos. 3 & 4, 235–311.

Gupta, S. S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. *Contributions to Probability and Statistics*, (Eds. I. Olkin et.al.), Stanford University Press, Stanford, CA, 224–248.

Jeyaratnam, S. and Panchapakesan, S. (1989). Entropy based subset selection from Bernoulli populations. *Proceedings of the International Conference on Computing and Information,ICCI'89*, (Eds. R. Janicki and W. W. Koczkodaj), Canadian Scholars' Press, Toronto, **Vol. 2**, 202–204.

Kulkarni, R. V. and Jennison, C. (1986). Optimal properties of the Bechhofer-Kulkarni Bernoulli selection procedures. *Ann. Statist.*, **14**, 298–314.

Sanchez, S. M. (1987). A modified least-failures sampling procedure for Bernoulli subset selection. *Commun. Statist.-Theory and Methods*, **16**, 3609–3629.

Sobel, M. and Huyett, M. J. (1957). Selecting the best one of several binomial populations. *Bell System Tech. J.*, **36**, 537–576.

Tamhane, A. C. (1980). Selecting the better Bernoulli treatment using a matched samples design. *J. Royal Statist. Soc. B* **42**, 26–30.

Wilcox, R. R. (1984). Selecting the best population, provided it is better than a standard: the unequal variance case. *J. Amer. Statist. Assoc.*, **79**, 887–891.

Yang, H. M. (1989). On selecting the treatment with the largest probability of survival. *Commun. Statist.-Theory and Methods*, **18**, 639–659.