EMPIRICAL BAYES RULES FOR SELECTING THE BEST NORMAL POPULATION COMPARED WITH A CONTROL

by

Shanti S. Gupta
Purdue University

TaChen Liang
Wayne State University

and

Re-Bin Rau
Purdue University

Technical Report #93-1

Department of Statistics
Purdue University

January 1993
Revised April 1993

EMPIRICAL BAYES RULES FOR SELECTING THE BEST NORMAL 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

Abstract

The problem of selecting the population with the largest mean from among $k(\geq 2)$ independent normal 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. Accordingly, the empirical Bayes approach is employed. Combining useful information from the past data, empirical Bayes selection procedures are developed. It is proved that the proposed empirical Bayes selection procedures are asymptotically optimal, having a rate of convergence of order $O(\frac{(\ln n)^2}{n})$, 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 procedures for small to moderate values of n.

AMS 1991 Subject Classification: Primary 62F07; secondary 62C12, 62C10
Keywords and Phrases: Asymptotic optimality; best population; Bayes rule; empirical Bayes; rate of convergence.

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

1 Introduction

Consider k independent normal populations $\pi_1, \ldots, \pi_k$ with unknown means $\theta_1, \cdots, \theta_k$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered θ_i's. A population π_i with $\theta_i = \theta_{[k]}$ is called the best population. The problem of selecting the best population was studied in the pioneering works of Bechhofer (1954) and Gupta (1956), by using the indifference zone approach and the subset selection approach, respectively. Gupta and Panchapakesan (1979,1985) provide a comprehensive survey of the development in this research area.

In a practical situation, one may not only be interested in the selection of the best population, but also require the selected population to be good enough. For example, in medical studies, the performance of any proposed new treament must be better than a standard treatment before it can be accepted by medical practitioners. In the literature, Bechhofer and Turnbull (1978), Dunnett (1984) and Wilcox (1984) investigated procedures for selecting the best normal population compared with a control, respectively. Using the subset selection approach, Gupta and Sobel (1958) and Lehmann (1961) have made some contributions to this problem.

In this paper, we employ the empirical Bayes approach to select the best normal population provided it is as good as a specified standard. The empirical Bayes methodology was introduced by Robbins (1956, 1964). This empirical Bayes approach has been used in selection problems by several authors. Deely (1965) studied the empirical Bayes rule for selecting the best normal population. Recently, Gupta and Hsiao (1983), Gupta and Liang (1988,1989), and Gupta and Leu (1991) have investigated empirical Bayes procedures for several selection problems. Many such empirical Bayes selection procedures have been shown to be asymptotically optimal in the sense that the empirical Bayes risk converges to the minimum Bayes risk.

This paper deals with a single-stage selection procedure for selecting the best normal population compared with a specified standard using the parametric empirical Bayes approach. In Section 2, we describe the formulation of the selection problem, and derive a Bayes selection rule. In Section 3, we construct the empirical Bayes selection rules. In Section 4, the asymptotic optimality of the proposed empirical Bayes selection rules is investigated. It is shown that the empirical Bayes selection rules have a rate of convergence of order $O(\frac{(\ln n)^2}{n})$, where n is the number of past observations at hand. In Section 5, we present the results of the simulation study of the proposed empirical Bayes selection procedures for small to moderate values of n.

2 Formulation of the Selection Problem and a Bayes Selection Rule

Let $\pi_1, \ldots, \pi_k$ be k independent normal populations with unknown means $\theta_1, \ldots, \theta_k$, 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 the exact pairing between the ordered and the unordered parameters is unknown. A population π_i with $\theta_i = \theta_{[k]}$ is considered as the best population. Let θ_0 be a known control. A population π_i with $\theta_i \geq \theta_0$ is considered as a good population. Our goal is to derive empirical Bayes rules to select the best normal population which should also be good compared with the control θ_0. If there is no such population, we select none.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | \theta_i \in R, i = 1, \ldots, k\}$ 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 compared with the control θ_0. When $a_0 = 1$, it means that all k populations are excluded as bad populations. We consider the following loss function:

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \max(\theta_{[k]}, \theta_0) - \sum_{i=0}^{k} a_i \theta_i. \tag{2.1}$$

Thus, if $\theta_{[k]} > \theta_0$ and all populations are rejected then the loss is $\theta_{[k]} - \theta_0$. On the other hand, if $\theta_0 > \theta_{[k]}$ and population π_i is selected as the best and good then the loss is $\theta_0 - \theta_i$.

For each $i = 1, 2, \ldots, k$, let $X_{i1}, \cdots, X_{iM}$ be a sample of size M from a normal population π_i which has mean θ_i and variance σ_i^2. It is assumed that θ_i is a realization of a random variable Θ_i which has a $N(\mu_i, \tau_i^2)$ prior distribution with unknown parameters $(\mu_i, \tau_i^2), i = 1, \ldots, k$. The random variables $\Theta_1, \ldots, \Theta_k$ are assumed to be independent. We let $f_i(x_i \mid \theta_i)$ and $h_i(\theta_i | \mu_i, \tau_i^2)$ denote the conditional probability density of $X_i = \bar{X}_i = \frac{1}{M} \sum_{j=1}^{M} X_{ij}$ and the density of Θ_i, respectively. Let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ and let $\mathcal{X}$ be the sample space generated by $\underset{\sim}{X}$. A selection rule $\underset{\sim}{d} = (d_0, \ldots, d_k)$ is a mapping defined on the sample space $\mathcal{X}$. For every $\underset{\sim}{x} \in \mathcal{X}, d_i(\underset{\sim}{x}), i = 1, \ldots, k$, is the probability of selecting population π_i as the best and good, and $d_0(\underset{\sim}{x})$ is the probability of excluding all k populations as bad and selecting none. Also, $\sum_{i=0}^{k} d_i(\underset{\sim}{x}) = 1$, for all $\underset{\sim}{x} \in \mathcal{X}$.

Under the preceding statistical model, the Bayes risk of the selection rule $\underset{\sim}{d}$ is denoted by $R(\underset{\sim}{d})$. Then, a straightforward computation yields the following :

$$R(\underset{\sim}{d}) = -\int_{\mathcal{X}} \left[\sum_{i=0}^{k} d_i(\underset{\sim}{x}) \varphi_i(x_i) \right] f(\underset{\sim}{x}) d\underset{\sim}{x} + C, \tag{2.2}$$

where

$$\begin{cases} C = \int_{\Omega} \max(\theta_{[k]}, \theta_0) dH(\underset{\sim}{\theta}), \\ \varphi_0(x_0) \equiv \theta_0, \\ \varphi_i(x_i) = E(\Theta_i | x_i) = \frac{x_i \tau_i^2 + \frac{\sigma_i^2}{M} \mu_i}{\tau_i^2 + \frac{\sigma_i^2}{M}} : \text{ the posterior mean of } \Theta_i \text{ given } X_i = x_i,\ i \neq 0, \\ f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i), \quad f_i(x_i) = \int_R f_i(x_i | \theta_i) h_i(\theta_i | \mu_i, \tau_i^2) d\theta_i, \\ H(\underset{\sim}{\theta}) : \text{ the joint distribution of } \underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k). \end{cases} \tag{2.3}$$

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

$$\begin{cases} I(\underset{\sim}{x}) = \{i | \varphi_i(x_i) = \max_{0 \leq j \leq k} \varphi_j(x_j), i = 0, \ldots, k\}, \\ i^* \equiv i^*(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } I(\underset{\sim}{x}) = \{0\}; \\ \min\{i | i \in I(\underset{\sim}{x}),\ i \neq 0\} & \text{otherwise.} \end{cases} \end{cases} \tag{2.4}$$

Then a Bayes selection rule $\underset{\sim}{d}^B = (d_0^B, \ldots, d_k^B)$ is given as follows:

$$\begin{cases} d_{i^*}^B(\underset{\sim}{x}) &= 1, \\ d_j^B(\underset{\sim}{x}) &= 0 \quad \text{for } j \neq i^*. \end{cases} \tag{2.5}$$

3 The Empirical Bayes Selection Rules

Since the parameters $(\mu_i, \tau_i^2), i = 1, \ldots, k$, are unknown, it is not possible to apply the Bayes rule $\underset{\sim}{d}^B$ for the selection problem at hand. In the empirical Bayes framework, it is assumed that certain past data are available when the present selection is made. Let $X_{ijl}, j = 1, \ldots, M$, denote a sample of size M from π_i at time $l, l = 1, \ldots, n$. It is assumed that conditional on $(\theta_{il}, \sigma_i^2)$, $X_{ijl}, j = 1, \ldots, M$, follow a normal distribution $N(\theta_{il}, \sigma_i^2)$ and θ_{il} is a realization of a random variable Θ_{il} which has a normal distribution $N(\mu_i, \tau_i^2)$. It is also assumed that Θ_{il}, $i = 1, \ldots, k$, $l = 1, 2, \ldots$, are mutually independent. For ease of notation, we denote the current random observations $X_{ij\,n+1}$ by X_{ij}, $j = 1, \ldots, M$, $i = 1, \ldots, k$.

For population $\pi_i, i = 1, \ldots, k$, let $X_{i.l} = \bar{X}_{i.l}$ be the sample mean of the M observations obtained at time l, $X_i(n)$ be the overall sample mean of past data and let $S_i^2(n)$ be the overall sample variance of the past data. That is

$$\begin{cases} X_{i.l} &= \frac{1}{M} \sum_{j=1}^{M} X_{ijl}, \\ X_i(n) &= \frac{1}{n} \sum_{l=1}^{n} X_{i.l}, \\ S_i^2(n) &= \frac{1}{n-1} \sum_{l=1}^{n} (X_{i.l} - X_i(n))^2. \end{cases} \tag{3.1}$$

Also, let $v_i^2 = \tau_i^2 + \frac{\sigma_i^2}{M}$. Then, from the statistical model described before, $X_{i.1}, X_{i.2}, \ldots, X_{i.n}$ are marginally independent with a $N(\mu_i, v_i^2)$ distribution. Hence, $X_i(n)$ has a $N(\mu_i, \frac{v_i^2}{n})$ distribution and $\frac{n-1}{v_i^2} S_i^2(n)$ has a $\chi^2(n-1)$ distribution. By the strong law of large numbers, we have

$$\begin{cases} X_i(n) \longrightarrow \mu_i \text{ a.s. }, \\ S_i^2(n) \longrightarrow v_i^2 \text{ a.s. }. \end{cases} \tag{3.2}$$

3.1 Case 1: (μ_i, τ_i^2) unknown and σ_i^2 known, $i = 1, \ldots, k$

Consider the case where both (μ_i, τ_i^2) are unknown and σ_i^2 is known, $i = 1, \cdots, k$. Since $E(X_i(n)) = \mu_i$, $E(S_i^2(n) - \frac{\sigma_i^2}{M}) = \tau_i^2$ and it is possible that $S_i^2(n) - \frac{\sigma_i^2}{M} \leq 0$, we define μ_{in} and τ_{in}^2 as estimators of μ_i and τ_i^2, respectively, by the following:

$$\begin{cases} \mu_{in} &= X_i(n), \\ \tau_{in}^2 &= \max(S_i^2(n) - \frac{\sigma_i^2}{M}, 0). \end{cases} \tag{3.3}$$

Now, we define, for $i = 1, 2, \ldots, k$,

$$\begin{cases} v_{in}^2 &= \tau_{in}^2 + \frac{\sigma_i^2}{M}, \\ \varphi_{in}(x_i) &= \frac{x_i \tau_{in}^2 + \frac{\sigma_i^2}{M}\mu_{in}}{v_{in}^2}, \\ \varphi_{0n}(x_0) &\equiv \theta_0. \end{cases} \tag{3.4}$$

We use v_{in}^2 and $\varphi_{in}(x_i)$ to estimate v_i^2 and $\varphi_i(x_i)$, respectively.

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

$$\begin{cases} I_n(\underset{\sim}{x}) = \{i | \varphi_{in}(x_i) = \max_{0 \le j \le k} \varphi_{jn}(x_j), i = 0, \ldots, k\}, \\ i_n^* \equiv i_n^*(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } I_n(\underset{\sim}{x}) = \{0\}, \\ \min\{i | i \in I_n(\underset{\sim}{x}),\ i \neq 0\} & \text{otherwise.} \end{cases} \end{cases} \tag{3.5}$$

We then obtain an empirical Bayes selection rule $\underset{\sim}{d}^{*n} = (d_0^{*n}, \ldots, d_k^{*n})$ as follows:

$$\begin{cases} d_{i_n^*}^{*n}(\underset{\sim}{x}) &= 1, \\ d_j^{*n}(\underset{\sim}{x}) &= 0 \quad \text{for } j \neq i_n^*. \end{cases} \tag{3.6}$$

3.2 Case 2: (μ_i, τ_i^2) and σ_i^2 unknown, $i = 1, \cdots, k$.

When σ_i^2, $i = 1, \ldots, k$, are unknown, it is assumed that $M \geq 2$. For each $i = 1, \cdots, k$, at time l, let $W_{i.l}^2$ and $W_i^2(n)$ be the sample variance at time l and the overall (pooled) sample variance, respectively. That is

$$\begin{cases} W_{i.l}^2 &= \frac{1}{M-1} \sum_{j=1}^M (X_{ijl} - X_{i.l})^2, \\ W_i^2(n) &= \frac{1}{n} \sum_{l=1}^n W_{i.l}^2. \end{cases} \tag{3.7}$$

Then, $\frac{M-1}{\sigma_i^2} W_{i.1}^2, \cdots, \frac{M-1}{\sigma_i^2} W_{i.n}^2$ are i.i.d. having a $\chi^2(M-1)$ distribution and hence $\frac{n(M-1)}{\sigma_i^2} W_i^2(n)$ has a $\chi^2(n(M-1))$ distribution. From the above discussion and by the strong law of large numbers, we have

$$\begin{cases} X_i(n) \longrightarrow \mu_i \quad \text{a.s.}, \\ W_i^2(n) \longrightarrow \sigma_i^2 \quad \text{a.s.}, \\ S_i^2(n) \longrightarrow v_i^2 \quad \text{a.s.}, \\ S_i^2(n) - \frac{W_i^2(n)}{M} \longrightarrow \tau_i^2 \quad \text{a.s.}, \\ E(X_i(n)) = \mu_i, \quad E(S_i^2(n)) = v_i^2, \quad E(W_i^2(n)) = \sigma_i^2, \\ E(S_i^2(n) - \frac{W_i^2(n)}{M}) = v_i^2 - \frac{\sigma_i^2}{M} = \tau_i^2. \end{cases} \tag{3.8}$$

Since, it is possible that $S_i^2(n) - \frac{W_i^2(n)}{M} < 0$, we define $\hat{\mu}_{in}, \hat{\sigma}_{in}^2, \hat{v}_{in}^2$ and $\hat{\tau}_{in}^2$ as estimators of

μ_i, σ_i^2, v_i^2 and τ_i^2, respectively, by the following :

$$\begin{cases} \hat{\mu}_{in} &= X_i(n), \\ \hat{\sigma}_{in}^2 &= W_i^2(n), \\ \hat{v}_{in}^2 &= S_i^2(n), \\ \hat{\tau}_{in}^2 &= \max(\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M}, \ 0). \end{cases} \tag{3.9}$$

For $i = 1, 2, \cdots, k$, we define

$$\begin{cases} \hat{\varphi}_{in}(x_i) &= \frac{x_i \hat{\tau}_{in}^2 + \frac{\hat{\sigma}_{in}^2}{M}\hat{\mu}_{in}}{\hat{v}_{in}^2}, \\ \hat{\varphi}_{0n}(x_0) &\equiv \theta_0, \end{cases} \tag{3.10}$$

and use $\hat{\varphi}_{in}(x_i)$ as an estimator of $\varphi_i(x_i)$.

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

$$\begin{cases} \hat{I}_n(\underset{\sim}{x}) = \{i | \hat{\varphi}_{in}(x_i) = \max_{0 \le j \le k} \hat{\varphi}_{jn}(x_j), i = 0, \dots, k\}, \\ \hat{i}_n \equiv \hat{i}_n(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } \hat{I}_n(\underset{\sim}{x}) = \{0\}, \\ \min\{i | i \in \hat{I}_n(\underset{\sim}{x}), \ i \neq 0\} & \text{otherwise.} \end{cases} \end{cases} \tag{3.11}$$

We then have an empirical Bayes selection rule $\hat{\underset{\sim}{d}}^n = (\hat{d}_0^n, \dots, \hat{d}_k^n)$ as follows:

$$\begin{cases} \hat{d}_{\hat{i}_n}^n(\underset{\sim}{x}) &= 1, \\ \hat{d}_j^n(\underset{\sim}{x}) &= 0 \quad \text{for } j \neq \hat{i}_n. \end{cases} \tag{3.12}$$

4 Asymptotic Optimality of the Empirical Bayes Selection Rules

In this section, we prove two theorems (Theorem 4.1 and Theorem 4.2) concerning the asymptotic optimality of the preceding empirical Bayes rules.

Consider an empirical Bayes selection rule $\underset{\sim}{d}^n = (d_0^n, \dots, d_k^n)$. We denote the associated Bayes risk of this empirical Bayes rule by $R(\underset{\sim}{d}^n)$. Then, from (2.2),

$$R(\underset{\sim}{d}^n) = -\int_{\mathcal{X}} \left[\sum_{i=0}^{k} d_i^n(\underset{\sim}{x})\varphi_i(x_i)\right] f(\underset{\sim}{x})d\underset{\sim}{x} + C. \tag{4.1}$$

Also, $R(\underset{\sim}{d}^n) - R(\underset{\sim}{d}^B) \geq 0$, since $R(\underset{\sim}{d}^B)$ is the minimum Bayes risk. Thus, $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B) \geq 0$, where the expectation E_n is taken with respect to X_{ijl}, $i = 1, \dots, k$, $j = 1, \dots, M$ and $l = 1, \dots, n$. The nonnegative difference $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$ is generally used as a measure of the performance of the selection rule $\underset{\sim}{d}^n$.

Definition 4.1 A sequence of empirical Bayes rules $\{\underset{\sim}{d}^n\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order β_n if $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B) = O(\beta_n)$, where β_n is a sequence of positive numbers such that $\lim_{n\to\infty} \beta_n = 0$.

In order to investigate the asymptotic optimality of the proposed empirical Bayes selection rules, we introduce some useful lemmas.

Lemma 4.1 is part of Theorem 1 of Chernoff (1952).

Lemma 4.1 Suppose S_n is the sum of n independent observations $X_1, X_2, \ldots, X_n$ of a random variable X with moment generating function $M(t) = E(e^{tX})$. Let $m(a) = \inf_t E(e^{t(X-a)}) = \inf_t e^{-at}M(t)$. Then,

$$(a) \quad \text{If } E(X) > -\infty \text{ and } a \leq E(X) \text{ then } P(S_n \leq na) \leq [m(a)]^n,$$
$$(b) \quad \text{If } E(X) < +\infty \text{ and } a \geq E(X) \text{ then } P(S_n \geq na) \leq [m(a)]^n.$$

Corollary 4.1 Let X have a $\chi^2(1)$ distribution. Then, S_n has a $\chi^2(n)$ distribution and

$$(a) \quad P\{S_n \leq n(1-\eta)\} \leq \exp(-\frac{n}{2}g_1(\eta)) \text{ for any } \eta,\ 0 < \eta < 1,$$
$$(b) \quad P\{S_n \geq n(1+\eta)\} \leq \exp(-\frac{n}{2}g_2(\eta)) \text{ for any } \eta,\ \eta > 0;$$

where

$$\begin{aligned} g_1(\eta) &= -\eta - \ln(1-\eta) \text{ for any } \eta,\ 0 < \eta < 1, \\ g_2(\eta) &= \eta - \ln(1+\eta) \text{ for any } \eta,\ \eta > 0. \end{aligned}$$

Proof : The moment generating function of X is given by $M(t) = (1-2t)^{-\frac{1}{2}}$ for $t < \frac{1}{2}$ and hence $m(a) = \inf_t E(e^{t(X-a)}) = E(e^{\frac{a-1}{2a}(X-a)}) = [e^{(1-a)}a]^{\frac{1}{2}}$. Therefore, $m(1-\eta) = [e^{\eta}(1-\eta)]^{\frac{1}{2}} = e^{\frac{1}{2}(\eta+\ln(1-\eta))} = e^{-\frac{1}{2}(-\eta-\ln(1-\eta))} = e^{-\frac{1}{2}g_1(\eta)}$ and $m(1+\eta) = [e^{-\eta}(1+\eta)]^{\frac{1}{2}} = e^{-\frac{1}{2}(\eta-\ln(1+\eta))} = e^{-\frac{1}{2}g_2(\eta)}$. The results follow from Lemma 4.1. □

Remark 1. Observe that $g_1(0) = g_2(0) = 0$, $\frac{d}{d\eta}g_1(\eta) > 0$, for $0 < \eta < 1$, and $\frac{d}{d\eta}g_2(\eta) > 0$, for $\eta > 0$. Thus, $g_1(\eta)$ and $g_2(\eta)$ are positive and strictly increasing functions for $0 < \eta < 1$ and $\eta > 0$, respectively.

Remark 2. $\lim_{\eta\to 0} \frac{g_1(\eta)}{\eta^2} = \lim_{\eta\to 0} \frac{g_1'(\eta)}{2\eta} = \lim_{\eta\to 0} \frac{\frac{\eta}{1-\eta}}{2\eta} = \frac{1}{2}$. Similarly, $\lim_{\eta\to 0} \frac{g_2(\eta)}{\eta^2} = \frac{1}{2}$.

4.1 Case 1: (μ_i, τ_i^2) unknown and σ_i^2 known, $i = 1, \ldots, k$

Let P_n be the probability measure generated by the past random observations X_{ijl}, $i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$.

Lemma 4.2 Let μ_{in} and τ_{in}^2 be the estimators of μ_i and τ_i^2, respectively, as defined in (3.3). Also, let $g_1(\eta)$ and $g_2(\eta)$ be the functions defined in Corollary 4.1. Then, for any $c > 0$ and

$0 < c_{v_i} < v_i^2,\ i = 1, \ldots k$, we have

$$(a)\quad P_n\{|\mu_{in} - \mu_i| \geq c\} \leq \frac{2v_i}{\sqrt{2\pi}c}\frac{1}{\sqrt{n}}\exp(\frac{-c^2}{2v_i^2}n),$$

$$(b)\quad P_n\{|\tau_{in}^2 - \tau_i^2| \geq c_{v_i}\} \leq \exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})) + \exp(-\frac{n-1}{2}g_1(\frac{c_{v_i}}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{c_{v_i}}{v_i^2})).$$

<u>Proof</u> : (a) Note that $\mu_{in} = X_i(n)$ has a $N(\mu_i, \frac{v_i^2}{n})$ distribution and by the fact that $P\{Z \geq \eta\} < \frac{1}{\eta}\frac{exp(\frac{-\eta^2}{2})}{\sqrt{2\pi}}$, for any $\eta > 0$ and for a $N(0,1)$ distributed random variable Z, (see Pollard (1984) Appendix B) the result follows.

(b)

$$\begin{aligned}
& P_n\{|\tau_{in}^2 - \tau_i^2| \geq c_{v_i}\} \\
\leq\ & P_n\{S_i^2(n) - \frac{\sigma_i^2}{M} \leq 0\} + P_n\{|S_i^2(n) - \frac{\sigma_i^2}{M} - \tau_i^2| \geq c_{v_i}, S_i^2(n) - \frac{\sigma_i^2}{M} > 0\} \\
\leq\ & P_n\{|S_i^2(n) - v_i^2| \geq \tau_i^2\} + P_n\{|S_i^2(n) - v_i^2| \geq c_{v_i}\} \\
=\ & P_n\{|\frac{n-1}{v_i^2}S_i^2(n) - (n-1)| \geq (n-1)\frac{\tau_i^2}{v_i^2}\} + P_n\{|\frac{n-1}{v_i^2}S_i^2(n) - (n-1)| \geq (n-1)\frac{c_{v_i}}{v_i^2}\} \\
\leq\ & \exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})) \\
& + \exp(-\frac{n-1}{2}g_1(\frac{c_{v_i}}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{c_{v_i}}{v_i^2})).
\end{aligned}$$

The last inequality follows from Corollary 4.1 and the fact that $\frac{n-1}{v_i^2}S_i^2(n)$ has a $\chi^2(n-1)$ distribution. □

Lemma 4.3 Let $\varphi_i(x_i)$ and $\varphi_{in}(x_i)$ be defined as in (2.3) and (3.4), respectively. Then, for any $\varepsilon > 0$ and any $x_i \in R$, we have

$$(a)\quad P_n\{\varphi_{in}(x_i) - \varphi_i(x_i) > \varepsilon\} \leq P_n\{|\mu_{in} - \mu_i| > \frac{Mv_i^2\varepsilon}{2\sigma_i^2}\} + P_n\{|\tau_{in}^2 - \tau_i^2| > \frac{v_i^4\varepsilon}{2(\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2)}\},$$

$$(b)\quad P_n\{\varphi_{in}(x_i) - \varphi_i(x_i) < -\varepsilon\} \leq P_n\{|\mu_{in} - \mu_i| > \frac{Mv_i^2\varepsilon}{2\sigma_i^2}\} + P_n\{|\tau_{in}^2 - \tau_i^2| > \frac{v_i^4\varepsilon}{2(\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2)}\}.$$

<u>Proof</u> : We prove (a) only. The proof of (b) is similar to that of (a). Let $a = x_i, b = \frac{\sigma_i^2}{M}, y = \tau_i^2, z = \mu_i, y_n = \tau_{in}^2$ and $z_n = \mu_{in}$. Then, $y + b = v_i^2$ and we have

$$P_n\{\varphi_{in}(x_i) - \varphi_i(x_i) > \varepsilon\}$$

$$
\begin{aligned}
&= P_n\{\frac{ay_n+bz_n}{y_n+b}-\frac{ay+bz}{y+b}>\varepsilon\} \\
&= P_n\{[b(a-z)-\varepsilon(y+b)](y_n-y)+b(b+y)(z_n-z)>\varepsilon(y+b)^2\} \\
&= P_n\{[\frac{\sigma_i^2}{M}(x_i-\mu_i)-\varepsilon v_i^2](\tau_{in}^2-\tau_i^2)+\frac{\sigma_i^2}{M}v_i^2(\mu_{in}-\mu_i)>\varepsilon v_i^4\} \\
&\le P_n\{\frac{\sigma_i^2}{M}v_i^2(\mu_{in}-\mu_i)>\frac{1}{2}\varepsilon v_i^4\}+P_n\{[\frac{\sigma_i^2}{M}(x_i-\mu_i)-\varepsilon v_i^2](\tau_{in}^2-\tau_i^2)>\frac{1}{2}\varepsilon v_i^4\} \\
&\le P_n\{|\mu_{in}-\mu_i|>\frac{Mv_i^2\varepsilon}{2\sigma_i^2}\}+P_n\{|\tau_{in}^2-\tau_i^2|>\frac{v_i^4\varepsilon}{2(\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2)}\}.
\end{aligned}
$$

□

Since $\varphi_1(X_1),\dots,\varphi_k(X_k)$ are mutually independent, WLOG, we assume $\varphi_i(X_i)\neq\varphi_j(X_j)$, $\forall\, i\neq j$. This assumption does not change the Bayes risk $R(\underset{\sim}{d}^B)$ and the empirical Bayes risk $R(\underset{\sim}{d}^{*n})$ and hence the difference $E_n[R(\underset{\sim}{d}^{*n})]-R(\underset{\sim}{d}^B)$.

To investigate the convergence rate of $E_n[R(\underset{\sim}{d}^{*n})]-R(\underset{\sim}{d}^B)$, we state some facts :

1 If $i^*=0,\ \varphi_l(x_l)<\theta_0$ for all $l=1,\dots,k$. Then, if $i_n^*=j\neq 0$,

$$
\begin{aligned}
&P_n\{i^*=0,\ i_n^*=j\}=P_n\{\varphi_l(x_l)<\theta_0\ \forall\, l\neq 0,\ \varphi_{jn}(x_j)\ge\varphi_{ln}(x_l)\ \forall\, l\neq j\} \\
&\le P_n\{\varphi_j(x_j)<\theta_0,\ \varphi_{jn}(x_j)\ge\theta_0\} \\
&\le P_n\{\varphi_{jn}(x_j)-\varphi_j(x_j)>\theta_0-\varphi_j(x_j)\}.
\end{aligned}
$$

2 If $i_n^*=0,\ \varphi_{ln}(x_l)<\theta_0$ for all $l=1,\dots,k$. Then, if $i^*=i\neq 0$,

$$
\begin{aligned}
&P_n\{i^*=i,\ i_n^*=0\}=P_n\{\varphi_i(x_i)\ge\varphi_l(x_l)\ \forall\, l\neq i,\ \varphi_{ln}(x_l)<\theta_0\ \forall\, l\neq 0\} \\
&\le P_n\{\varphi_i(x_i)\ge\theta_0,\ \varphi_{in}(x_i)<\theta_0\} \\
&\le P_n\{\varphi_{in}(x_i)-\varphi_i(x_i)<-(\varphi_i(x_i)-\theta_0)\}.
\end{aligned}
$$

3 If $i^*=i\neq 0,\ i_n^*=j\neq 0$ and $i\neq j$, then

$$
\begin{aligned}
&P_n\{i^*=i,\ i_n^*=j\}=P_n\{\varphi_i(x_i)\ge\varphi_l(x_l)\ \forall\, l\neq i,\ \varphi_{jn}(x_j)\ge\varphi_{ln}(x_l)\ \forall\, l\neq j\} \\
&\le P_n\{\varphi_i(x_i)\ge\varphi_j(x_j),\ \varphi_{jn}(x_j)\ge\varphi_{in}(x_i)\} \\
&= P_n\{\varphi_{jn}(x_j)-\varphi_j(x_j)-[\varphi_{in}(x_i)-\varphi_i(x_i)]\ge\varphi_i(x_i)-\varphi_j(x_j),\ \varphi_i(x_i)\ge\varphi_j(x_j)\} \\
&\le P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{\varphi_i(x_i)-\varphi_j(x_j)}{2}\}+P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varphi_i(x_i)-\varphi_j(x_j)}{2}\}.
\end{aligned}
$$

¿From (2.2), (4.1) and by Facts 1, 2 and 3, we get

$$
\begin{aligned}
&E_n[R(\underset{\sim}{d}^{*n})]-R(\underset{\sim}{d}^B) \\
&= E_n\int_{\mathcal{X}}[d_{i^*}^B(\underset{\sim}{x})\varphi_{i^*}(x_{i^*})-d_{i_n^*}^{*n}(\underset{\sim}{x})\varphi_{i_n^*}(x_{i_n^*})]f(\underset{\sim}{x})d\underset{\sim}{x}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=0}^{k}\sum_{j=0}^{k} E_n \int_{\mathcal{X}} \mathrm{I}_{\{i^*=i, i_n^*=j\}}[\varphi_i(x_i) - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \\
&= \sum_{i=0}^{k}\sum_{j=0}^{k} \int_{\mathcal{X}} P_n\{i^* = i, i_n^* = j\}[\varphi_i(x_i) - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \\
&= \sum_{i=1}^{k} \int_{\mathcal{X}} P_n\{i^* = i, i_n^* = 0\}[\varphi_i(x_i) - \theta_0] f(\underset{\sim}{x}) d\underset{\sim}{x} \\
&\quad + \sum_{j=1}^{k} \int_{\mathcal{X}} P_n\{i^* = 0, i_n^* = j\}[\theta_0 - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \qquad (4.2) \\
&\quad + \sum_{i=1}^{k}\sum_{j=1}^{k} \int_{\mathcal{X}} P_n\{i^* = i, i_n^* = j\}[\varphi_i(x_i) - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \\
&\leq \sum_{i=1}^{k} \int_{R} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\} |\varphi_i(x_i) - \theta_0| f_i(x_i) dx_i \\
&\quad + \sum_{i=1}^{k}\sum_{j=1}^{k} \int_{R^2} \Bigg[P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > \frac{|\varphi_i(x_i) - \varphi_j(x_j)|}{2}\} \\
&\qquad\qquad + P_n\{|\varphi_{jn}(x_j) - \varphi_j(x_j)| > \frac{|\varphi_i(x_i) - \varphi_j(x_j)|}{2}\} \Bigg] \\
&\qquad\qquad \times |\varphi_i(x_i) - \varphi_j(x_j)| f_i(x_i) f_j(x_j) dx_i dx_j \\
&= I_n + II_n.
\end{aligned}
$$

Recall that $\varphi_i(x_i) = \frac{x_i\tau_i^2 + \frac{\sigma_i^2}{M}\mu_i}{\tau_i^2 + \frac{\sigma_i^2}{M}}$ and X_i is marginally $N(\mu_i, v_i^2)$ distributed. Therefore, $\varphi_i(X_i)$ is $N(\mu_i, \frac{\tau_i^4}{v_i^2})$ distributed. For $\varepsilon_n > 0$, and $i, j = 1, \ldots, k$, let

$$
\begin{cases} \mathcal{X}_i &= \{x_i | \; |\varphi_i(x_i) - \theta_0| \leq \varepsilon_n\}, \\ \mathcal{X}_{ij} &= \{(x_i, x_j) | \; |\varphi_i(x_i) - \varphi_j(x_j)| \leq \varepsilon_n\}. \end{cases} \tag{4.3}
$$

Then,

$$
\begin{aligned}
I_n &= \sum_{i=1}^{k} \int_{\mathcal{X}_i} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\} |\varphi_i(x_i) - \theta_0| f_i(x_i) dx_i \\
&\quad + \sum_{i=1}^{k} \int_{R-\mathcal{X}_i} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\} |\varphi_i(x_i) - \theta_0| f_i(x_i) dx_i \\
&\leq \sum_{i=1}^{k} \int_{\mathcal{X}_i} \varepsilon_n f_i(x_i) dx_i \qquad (4.4) \\
&\quad + \sum_{i=1}^{k} \int_{R} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > \varepsilon_n\} |\varphi_i(x_i) - \theta_0| f_i(x_i) dx_i \\
&\leq O(\varepsilon_n^2)
\end{aligned}
$$

$$+\sum_{i=1}^{k}\int_{R}P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}[|\varphi_i(x_i)-\mu_i|+|\mu_i-\theta_0|]f_i(x_i)dx_i,$$

where

$$\sum_{i=1}^{k}\int_{\mathcal{X}_i}\varepsilon_n f_i(x_i)dx_i=O(\varepsilon_n^2),$$

since

$$\int_{\{x_i|\ |\varphi_i(x_i)-\theta_0|\le\varepsilon_n\}}f_i(x_i)dx_i\le\frac{2v_i}{\sqrt{2\pi}\tau_i^2}\varepsilon_n,\quad i=1,\ldots,k.$$

Moreover, $\varphi_i(X_i)-\varphi_j(X_j)$ has a $N(\mu_i-\mu_j,\frac{\tau_i^4}{v_i^2}+\frac{\tau_j^4}{v_j^2})$ distribution. Therefore,

$$\begin{aligned}
II_n &= \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}_{ij}}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\\
&\quad+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\Big]|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j\\
&\quad+\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2-\mathcal{X}_{ij}}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\\
&\quad+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\Big]|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j\\
&\le \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}_{ij}}2\varepsilon_n f_i(x_i)f_j(x_j)dx_idx_j\\
&\quad+\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{\varepsilon_n}{2}\}\Big]\\
&\qquad\times|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j\\
&\le \sum_{i=1}^{k}\sum_{j=1}^{k}2\varepsilon_n\frac{1}{\sqrt{2\pi}}\frac{2\varepsilon_n}{\sqrt{\frac{\tau_i^4}{v_i^2}+\frac{\tau_j^4}{v_j^2}}} \qquad (4.5)\\
&\quad+\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{\varepsilon_n}{2}\}\Big]\\
&\qquad\times[|\varphi_i(x_i)-\mu_i|+|\varphi_j(x_j)-\mu_j|+|\mu_i-\mu_j|]\,f_i(x_i)f_j(x_j)dx_idx_j.
\end{aligned}$$

Since $X_1,X_2,\ldots,X_k$ are mutually independent and $E|\varphi_i(X_i)-\mu_i|<+\infty$, $i=1,\ldots,k$, also by (4.2), (4.4) and (4.5), it suffices to investigate the following two terms.

$$\begin{cases}\int_R P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}f_i(x_i)dx_i,\\ \int_R P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}|\varphi_i(x_i)-\mu_i|f_i(x_i)dx_i.\end{cases}\qquad(4.6)$$

Furthermore, by Lemma 4.2 and Lemma 4.3, we have

$$P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}$$

$$
\begin{aligned}
&\leq 2\left[P_n\{|\mu_{in}-\mu_i| > \frac{Mv_i^2\varepsilon_n}{4\sigma_i^2}\} \qquad \text{(by lemma 4.3)}\right.\\
&\quad \left. + P_n\{|\tau_{in}^2-\tau_i^2| > \frac{v_i^4\varepsilon_n}{4(\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2)}\}\right]\\
&\leq 2\left[\frac{8\sigma_i^2}{\sqrt{2\pi}Mv_i}\frac{1}{\varepsilon_n\sqrt{n}}\exp(\frac{-M^2v_i^2}{32\sigma_i^4}\varepsilon_n^2 n) \qquad \text{(by lemma 4.2)}\right.\\
&\quad +\exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})) \qquad (4.7)\\
&\quad \left. +\exp(-\frac{n-1}{2}g_1(\frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}))\right].
\end{aligned}
$$

Also, $|\varphi_i(x_i)-\mu_i| = \frac{\tau_i^2}{v_i^2}|x_i-\mu_i|$. Hence, if we let $\eta_n = \frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}$ then, from (4.6) and (4.7), it suffices to consider the rate of convergence of the following terms.

$$
\begin{cases}
II_a = \frac{8\sigma_i^2}{\sqrt{2\pi}Mv_i}\frac{1}{\varepsilon_n\sqrt{n}}\exp(\frac{-M^2v_i^2}{32\sigma_i^4}\varepsilon_n^2 n)+\exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})),\\
II_b = \int_R[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))]f_i(x_i)dx_i,\\
II_c = \int_R[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))]|x_i-\mu_i|f_i(x_i)dx_i.
\end{cases} \qquad (4.8)
$$

First, we consider the term II_a. For $i=1,\ldots,k$, note that $0<\frac{\tau_i^2}{v_i^2}<1$, hence, $g_1(\frac{\tau_i^2}{v_i^2})>0$ and $g_2(\frac{\tau_i^2}{v_i^2})>0$, by the Remark 1 of Corollary 4.1. Therefore,

$$
\exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})) \leq O(\exp(-c_1 n))
$$

where $c_1 = \frac{1}{2}\min_{1\leq i\leq k}\{g_1(\frac{\tau_i^2}{v_i^2}), g_2(\frac{\tau_i^2}{v_i^2})\}$. In the sequel, we let $\varepsilon_n = \frac{\ln n}{\sqrt{cn}}$, where $c = \min_{1\leq i\leq k}\{\frac{M^2v_i^2}{1024\sigma_i^4}\}$. Then,

$$
\frac{1}{\varepsilon_n\sqrt{n}}\exp(\frac{-M^2v_i^2}{32\sigma_i^4}\varepsilon_n^2 n) \leq O(\frac{1}{n\ln n}).
$$

Thus, from the above argument and (4.8), we have

$$
II_a \leq O(\frac{1}{n\ln n}). \qquad (4.9)
$$

Now, let us investigate the rate of convergence of II_b. For the same ε_n, we divide the integration of II_b into two parts by the set $\{|x_i-\mu_i| < \frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}$ and its complement. By Remark 1 and Remark 2 of Corollary 4.1 and for n sufficiently large, we have

$$
|x_i-\mu_i| < \frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}
$$

$$\Rightarrow \eta_n = \frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i| + \frac{\varepsilon_n}{2}v_i^2} > \frac{1}{2}\frac{1}{\sqrt{\frac{n}{128\ln n}}+1}$$

$$\Rightarrow \eta_n > \frac{1}{4}\frac{1}{\sqrt{\frac{n}{128\ln n}}} \Rightarrow g_1(\eta_n) > g_1(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})$$

$$\Rightarrow \exp(-\frac{n-1}{2}g_1(\eta_n)) < \exp(-\frac{n-1}{2}g_1(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})) \tag{4.10}$$

$$\leq \exp\left(-\frac{n-1}{2}(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})^2\frac{g_1(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})}{(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})^2}\right)$$

$$= O(\frac{1}{n}).$$

Similarly,

$$|x_i-\mu_i| < \frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}} \Rightarrow \exp(-\frac{n-1}{2}g_2(\eta_n)) \leq O(\frac{1}{n}). \tag{4.11}$$

Therefore,

$$\int_{\{|x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}}\left[\exp(-\frac{n-1}{2}g_1(\eta_n)) + \exp(-\frac{n-1}{2}g_2(\eta_n))\right] f_i(x_i)dx_i$$
$$\leq O(\frac{1}{n}). \tag{4.12}$$

Now, by using a similar argument as in the proof of Lemma 4.2(a), we have

$$EI_{\{|X_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} = P\{\frac{|X_i-\mu_i|}{v_i} \geq \frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}}\}$$
$$\leq 2\frac{1}{\frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}}}\frac{\exp(-\frac{1}{2}(\frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}})^2)}{\sqrt{2\pi}}$$
$$\leq O(\frac{1}{n\sqrt{\ln n}}).$$

Moreover, observe that $0 < \eta_n < \frac{1}{2}$, this implies that $g_1(\eta_n) > 0$ and $g_2(\eta_n) > 0$. Hence,

$$\int_{\{|x_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}}\left[\exp(-\frac{n-1}{2}g_1(\eta_n)) + \exp(-\frac{n-1}{2}g_2(\eta_n))\right] f_i(x_i)dx_i$$
$$\leq 2EI_{\{|X_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} \tag{4.13}$$
$$\leq O(\frac{1}{n\sqrt{\ln n}}).$$

¿From (4.8), (4.12) and (4.13), we get

$$II_b \le O(\frac{1}{n}). \tag{4.14}$$

Again, for the same ε_n, we divide II_c into two parts:

$$II_c = II_{c,1} + II_{c,2}, \tag{4.15}$$

where

$$\begin{aligned} II_{c,1} &= \int_{\{|x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} \left[\exp(-\frac{n-1}{2}g_1(\eta_n)) + \exp(-\frac{n-1}{2}g_2(\eta_n))\right] \\ &\qquad |x_i-\mu_i| f_i(x_i)dx_i, \\ II_{c,2} &= \int_{\{|x_i-\mu_i|\ge\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} \left[\exp(-\frac{n-1}{2}g_1(\eta_n)) + \exp(-\frac{n-1}{2}g_2(\eta_n))\right] \\ &\qquad |x_i-\mu_i| f_i(x_i)dx_i. \end{aligned}$$

By (4.10), (4.11) and $E|X_i-\mu_i| < +\infty$, we have

$$II_{c,1} \le O(\frac{1}{n}). \tag{4.16}$$

Also, recall that $g_1(\eta_n) > 0$ and $g_2(\eta_n) > 0$, then

$$\begin{aligned} II_{c,2} &\le 2\int_{\{|x_i-\mu_i|\ge\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} |x_i-\mu_i| f_i(x_i)dx_i \\ &\le 2v_i \int_{\{|z|\ge\frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}}\}} |z|d\Phi(z) \\ &\le \frac{4v_i}{\sqrt{2\pi}}\exp(-\frac{M^2v_i^2}{8\sigma_i^4}\frac{\ln n}{128c}) \qquad (4.17) \\ &\le O(\frac{1}{n}), \end{aligned}$$

where $\Phi(z)$ is the c.d.f. of the standard normal distribution.
Hence, from (4.15) – (4.17),

$$II_c \le O(\frac{1}{n}). \tag{4.18}$$

Therefore, from (4.4) – (4.6), (4.8), (4.9), (4.14), (4.18) and for the same ε_n, we have

$$I_n \le O(\varepsilon_n^2) = O(\frac{(\ln n)^2}{n}), \tag{4.19}$$

$$II_n \le O(\varepsilon_n^2) = O(\frac{(\ln n)^2}{n}). \tag{4.20}$$

By combining (4.2), (4.19) and (4.20), we have proved the following theorem.

Theorem 4.1 The empirical Bayes selection rule $\underset{\sim}{d}^{*n}(\underset{\sim}{x})$, defined in (3.6), is asymptotically optimal with convergence rate of order $O(\frac{(\ln n)^2}{n})$. That is, $E_n[R(\underset{\sim}{d}^{*n})] - R(\underset{\sim}{d}^B) \le O(\frac{(\ln n)^2}{n})$.

4.2 Case 2: (μ_i, τ_i^2) and σ_i^2 unknown, $i = 1, \cdots, k$.

Lemma 4.5 Let $\hat{\mu}_{in}, \hat{\sigma}_{in}^2$ and $\hat{v}_{in}^2$ be the estimators of μ_i, σ_i^2 and v_i^2, respectively, as defined in (3.9). Also, let $g_1(\eta)$ and $g_2(\eta)$ be the functions defined in Corollary 4.1. Then, for any $c > 0, 0 < c_{v_i} < v_i^2$, and $0 < c_{\sigma_i} < \sigma_i^2$, $i = 1, \ldots, k$, we have

$$(a)\quad P_n\{|\hat{\mu}_{in} - \mu_i| \geq c\} \leq \frac{2v_i}{\sqrt{2\pi}c}\frac{1}{\sqrt{n}}\exp(\frac{-c^2}{2v_i^2}n),$$

$$(b)\quad P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| \geq c_{\sigma_i}\} \leq \exp(-\frac{n(M-1)}{2}g_1(\frac{c_{\sigma_i}}{\sigma_i^2})) + \exp(-\frac{n(M-1)}{2}g_2(\frac{c_{\sigma_i}}{\sigma_i^2})),$$

$$(c)\quad P_n\{|\hat{v}_{in}^2 - v_i^2| \geq c_{v_i}\} \leq \exp(-\frac{n-1}{2}g_1(\frac{c_{v_i}}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{c_{v_i}}{v_i^2})).$$

Proof : (a) The proof is the same as in Lemma 4.2(a).

(b)

$$\begin{aligned} & P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| \geq c_{\sigma_i}\} \\ = \ & P_n\{|n(M-1)\frac{\hat{\sigma}_{in}^2}{\sigma_i^2} - n(M-1)| \geq n(M-1)\frac{c_{\sigma_i}}{\sigma_i^2}\} \\ \leq \ & \exp(-\frac{n(M-1)}{2}g_1(\frac{c_{\sigma_i}}{\sigma_i^2})) + \exp(-\frac{n(M-1)}{2}g_2(\frac{c_{\sigma_i}}{\sigma_i^2})). \end{aligned}$$

The last inequality follows by Corollary 4.1 and the fact that $n(M-1)\frac{\hat{\sigma}_{in}^2}{\sigma_i^2}$ has a $\chi^2(n(M-1))$ distribution.

(c) The proof is similar to that of (b), hence, we omit it. □

Lemma 4.6 Let $\varphi_i(x_i)$ and $\hat{\varphi}_{in}(x_i)$ be defined as in (2.3) and (3.10), respectively. Then, for any $\varepsilon > 0$, any $\kappa > 0$ and any $x_i \in R$, we have

$$\begin{aligned} (a) \quad & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon\} \\ \leq \ & P_n\{|\hat{\mu}_{in} - \mu_i| \geq \kappa\} + P_n\{|\hat{\mu}_{in} - \mu_i| > \frac{Mv_i^2\varepsilon}{5\sigma_i^2}\} + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\} \\ & + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{Mv_i^2\varepsilon}{5\kappa}\} + 2P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{\sigma_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\} \\ & + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\}, \end{aligned}$$

$$\begin{aligned} (b) \quad & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) < -\varepsilon\} \\ \leq \ & P_n\{|\hat{\mu}_{in} - \mu_i| \geq \kappa\} + P_n\{|\hat{\mu}_{in} - \mu_i| > \frac{Mv_i^2\varepsilon}{5\sigma_i^2}\} + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\} \\ & + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{Mv_i^2\varepsilon}{5\kappa}\} + 2P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{\sigma_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\} \\ & + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\}. \end{aligned}$$

Proof : We prove (a) only. The proof of (b) is similar to that of (a). Let $a = x_i, b = \frac{\sigma_i^2}{M}, y = \tau_i^2, z = \mu_i, b_n = \frac{\hat{\sigma}_{in}^2}{M}, y_n = \hat{\tau}_{in}^2$ and $z_n = \hat{\mu}_{in}$. Therefore, $y + b = v_i^2$ and $y_n + b_n = \hat{v}_{in}^2$ if $\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} \geq 0$. Therefore,

$$\begin{aligned} & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon\} \\ \leq\ & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon, \hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} \geq 0\} + P_n\{\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} < 0\}, \end{aligned}$$

where

$$\begin{aligned} & P_n\{\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} < 0\} \\ =\ & P_n\{(\hat{v}_{in}^2 - v_i^2) - (\frac{\hat{\sigma}_{in}^2}{M} - \frac{\sigma_i^2}{M}) < -\tau_i^2\} \\ \leq\ & P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\} \end{aligned}$$

and

$$\begin{aligned} & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon, \hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} \geq 0\} \\ \leq\ & P_n\{v_i^2(z_n - z)(b_n - b) - (a - z)v_i^2(b_n - b) + v_i^2 b(z_n - z) + [(a - z)b - \varepsilon v_i^2](\hat{v}_{in}^2 - v_i^2) > \varepsilon v_i^4\} \\ =\ & P_n\{v_i^2(z_n - z)(b_n - b) - [(a - z)b + \varepsilon v_i^2]\frac{v_i^2}{b}(b_n - b) + \frac{\varepsilon v_i^4}{b}(b_n - b) \\ & + v_i^2 b(z_n - z) + [(a - z)b - \varepsilon v_i^2](\hat{v}_{in}^2 - v_i^2) > \varepsilon v_i^4\} \\ \leq\ & P_n\{v_i^2|z_n - z||b_n - b| + (|a - z|b + \varepsilon v_i^2)\frac{v_i^2}{b}|b_n - b| \\ & + \frac{\varepsilon v_i^4}{b}|b_n - b| + v_i^2 b|z_n - z| + (|a - z|b + \varepsilon v_i^2)|\hat{v}_{in}^2 - v_i^2| > \varepsilon v_i^4\} \\ \leq\ & P_n\{|z_n - z||b_n - b| > \frac{\varepsilon v_i^2}{5}\} \\ & + P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a - z| + \varepsilon v_i^2}\} + P_n\{|b_n - b| > \frac{b}{5}\} \\ & + P_n\{|z_n - z| > \frac{\varepsilon v_i^2}{5b}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a - z| + \varepsilon v_i^2}\} \\ =\ & P_n\{|z_n - z| \geq \kappa,\ |z_n - z||b_n - b| > \frac{\varepsilon v_i^2}{5}\} + P_n\{|z_n - z| < \kappa,\ |z_n - z||b_n - b| > \frac{\varepsilon v_i^2}{5}\} \\ & + P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a - z| + \varepsilon v_i^2}\} + P_n\{|b_n - b| > \frac{b}{5}\} \\ & + P_n\{|z_n - z| > \frac{\varepsilon v_i^2}{5b}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a - z| + \varepsilon v_i^2}\} \\ \leq\ & P_n\{|z_n - z| \geq \kappa\} + P_n\{|b_n - b| > \frac{v_i^2\varepsilon}{5\kappa}\} \end{aligned}$$

$$
\begin{aligned}
&+P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a-z|+\varepsilon v_i^2}\} + P_n\{|b_n - b| > \frac{b}{5}\} \\
&+P_n\{|z_n - z| > \frac{v_i^2\varepsilon}{5b}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a-z|+\varepsilon v_i^2}\} \\
\le\ & P_n\{|z_n - z| \ge \kappa\} + P_n\{|z_n - z| > \frac{v_i^2\varepsilon}{5b}\} \\
&+P_n\{|b_n - b| > \frac{v_i^2\varepsilon}{5\kappa}\} + 2P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a-z|+\varepsilon v_i^2}\} \\
&+P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a-z|+\varepsilon v_i^2}\} \\
\le\ & P_n\{|\hat{\mu}_{in} - \mu_i| \ge \kappa\} + P_n\{|\hat{\mu}_{in} - \mu_i| > \frac{Mv_i^2\varepsilon}{5\sigma_i^2}\} \\
&+P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{Mv_i^2\varepsilon}{5\kappa}\} + 2P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{\sigma_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2}\} \\
&+P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2}\}.
\end{aligned}
$$

Hence, the result follows. □

Let $\{\underset{\sim}{\hat{d}}^n\}_{n=1}^{\infty}$ be the empirical Bayes rules defined in (3.12). Then,

$$E_n[R(\underset{\sim}{\hat{d}}^n)] - R(\underset{\sim}{d}^B) \le \hat{I}_n + \hat{II}_n. \tag{4.21}$$

where

$$
\begin{aligned}
\hat{I}_n &= \sum_{i=1}^{k}\int_R P_n\{|\hat{\varphi}_{in}(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\}|\varphi_i(x_i) - \theta_0|f_i(x_i)dx_i, \\
\hat{II}_n &= \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Big[P_n\{|\hat{\varphi}_{in}(x_i) - \varphi_i(x_i)| > \frac{|\varphi_i(x_i) - \varphi_j(x_j)|}{2}\} \\
&\qquad + P_n\{|\hat{\varphi}_{jn}(x_j) - \varphi_j(x_j)| > \frac{|\varphi_i(x_i) - \varphi_j(x_j)|}{2}\}\Big] \\
&\qquad \times|\varphi_i(x_i) - \varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j.
\end{aligned}
$$

By an argument similar to that of (4.4) and (4.5), it suffices to investigate the following two terms.

$$
\begin{cases}
\int_R P_n\{|\hat{\varphi}_{in}(x_i) - \varphi_i(x_i)| > \frac{\varepsilon_n}{2}\}f_i(x_i)dx_i, \\
\int_R P_n\{|\hat{\varphi}_{in}(x_i) - \varphi_i(x_i)| > \frac{\varepsilon_n}{2}\}|\varphi_i(x_i) - \mu_i|f_i(x_i)dx_i.
\end{cases} \tag{4.22}
$$

Moreover, by Lemma 4.6, we have

$$P_n\{|\hat{\varphi}_{in}(x_i) - \varphi_i(x_i)| > \frac{\varepsilon}{2}\}$$

$$\leq 2\Big[P_n\{|\hat{\mu}_{in}-\mu_i|\geq\kappa\}+P_n\{|\hat{\mu}_{in}-\mu_i|>\frac{Mv_i^2\frac{\varepsilon}{2}}{5\sigma_i^2}\}+P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{M\tau_i^2}{2}\}$$
$$+P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{Mv_i^2\frac{\varepsilon}{2}}{5\kappa}\}+2P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{\sigma_i^2}{5}\frac{\frac{\varepsilon}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon}{2}v_i^2}\} \qquad (4.23)$$
$$+P_n\{|\hat{v}_{in}^2-v_i^2|>\frac{\tau_i^2}{2}\}+P_n\{|\hat{v}_{in}^2-v_i^2|>\frac{v_i^2}{5}\frac{\frac{\varepsilon}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon}{2}v_i^2}\}\Big].$$

Let $\varepsilon=\varepsilon_n=\frac{\ln n}{\sqrt{c_*n}}$ and $\kappa\equiv\kappa_n=\sqrt{c_\kappa\ln n}$, where $c_*=\min_{1\leq i\leq k}\{\frac{M^2v_i^2}{6400\sigma_i^4}\}$ and $c_\kappa=\min_{1\leq i\leq k}\{4v_i\}$. Then, by using Lemma 4.5 and Remark 1 and Remark 2 of Corollary 4.1, the two terms concerning κ in (4.23) have the following convergence rate.

$$P_n\{|\hat{\mu}_{in}-\mu_i|\geq\kappa\} \leq O\left(\frac{1}{\sqrt{n\ln n}}\exp(-c_\kappa(\max_{1\leq j\leq k}\{2v_i^2\})^{-1}n\ln n)\right),$$
$$P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{Mv_i^2\frac{\varepsilon}{2}}{5\kappa}\} \leq O(\frac{1}{n}).$$

Again, by Lemma 4.5, we get

$$P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{M\tau_i^2}{2}\} \leq O\left(\exp(-\frac{M-1}{2}\min_{1\leq i\leq k}\{g_1(\frac{M\tau_i^2}{2\sigma_i^2}),g_2(\frac{M\tau_i^2}{2\sigma_i^2})\}n)\right),$$
$$P_n\{|\hat{v}_{in}^2-v_i^2|>\frac{\tau_i^2}{2}\} \leq O\left(\exp(-\frac{1}{2}\min_{1\leq i\leq k}\{g_1(\frac{\tau_i^2}{2v_i^2}),g_2(\frac{\tau_i^2}{2v_i^2})\}n)\right).$$

Now, by a proof of the rate of convergence analogous to that of (4.6), it can be shown that the two terms in (4.22) have a rate of convergence of order $O(\frac{(\ln n)^2}{n})$.

Hence, by the above argument, (4.21) and (4.22), we have the following theorem.

Theorem 4.2 The empirical Bayes selection rule $\hat{\underset{\sim}{d}}^n(\underset{\sim}{x})$, defined in (3.12), is asymptotically optimal with convergence rate of order $O(\frac{(\ln n)^2}{n})$. That is, $E_n[R(\hat{\underset{\sim}{d}}^n)]-R(\underset{\sim}{d}^B)\leq O(\frac{(\ln n)^2}{n})$.

5 Small Sample Performance: Simulation Study

We carried out a simulation study to investigate the performance of the empirical Bayes selection rules $\underset{\sim}{d}^{*n}(\underset{\sim}{x})$ and $\hat{\underset{\sim}{d}}^n(\underset{\sim}{x})$ defined in Sections 3.1 and 3.2, respectively. Recall that E and E_n are the expectations taken with respect to the probability measures generated by the current observation $\underset{\sim}{X}$ and the past observation X_{ijl} ($i=1,\ldots,k$, $j=1,\ldots,M$ and $l=1,\ldots,n$), respectively. In Definition 4.1 $E_n[R(\underset{\sim}{d}^n)]-R(\underset{\sim}{d}^B)$ is used as a measure of the performance of the empirical Bayes rule $\underset{\sim}{d}^n$. For any given current observation $\underset{\sim}{X}$ and any given past observation X_{ijl} ($i=1,\ldots,k$, $j=1,\ldots,M$ and $l=1,\ldots,n$), let

$$D^n(\underset{\sim}{X})=\sum_{i=0}^{k}[d_i^B(\underset{\sim}{X})-d_i^n(\underset{\sim}{X})]\varphi_i(X_i).$$

Then, from (4.2)

$$E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B) = EE_nD^n(\underset{\sim}{X}).$$

Therefore, by the law of large numbers, the sample mean of $D^n(\underset{\sim}{X})$, based on the observations of $\underset{\sim}{X}$ and X_{ijl} ($i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$), can be used as an estimator of $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$.

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

(1) For each $l = 1, \ldots, n$ and for each $i = 1, 2$ and 3, generate the independent past observations $X_{i1l}, \ldots, X_{iMl}$ by the following :

$$\begin{cases} (a) & \text{Generate } \Theta_{il} \text{ from a } N(\mu_i, \tau_i^2) \text{ prior distribution.} \\ (b) & \text{Generate random sample } X_{i1l}, X_{i2l}, \ldots, X_{iMl} \text{ from a } N(\theta_{il}, \sigma_i^2) \text{ distribution.} \end{cases}$$

(2) Generate the current observation $\underset{\sim}{X} = (X_1, \ldots, X_k)$, where X_i has a $N(\mu_i, \frac{\sigma_i^2}{M} + \tau_i^2)$ distribution and $X_1, \ldots, X_k$ are independent.

(3) Based on the past observation X_{ijl} ($i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$) and the current observation $\underset{\sim}{X}$, construct the Bayes rule $\underset{\sim}{d}^B$ and the empirical Bayes rule $\underset{\sim}{d}^n$ and compute $D^n(\underset{\sim}{X})$.

(4) Steps (1), (2) and (3) were repeated 4000 times. The average of $D^n(\underset{\sim}{X})$ based on the 4000 repetitions, which is denoted by $\bar{D}^n$, is used as an estimator of $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$. Also, $SE(\bar{D}^n)$, the estimated standard error, and $n\bar{D}^n$ are computed.

It should be mentioned that the same past observation X_{ijl} ($i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$) and the current observation $\underset{\sim}{X}$ were used for both rules $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$. Also, the term $\bar{D}^n$ corresponding to $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$ are denoted by $\bar{D}^{*n}$ and $\hat{\bar{D}}^n$, respectively.

Tables 1, 2, 3 and 4 list some simulation results on the performance of the proposed empirical Bayes rules $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$, for the case where $k = 3$ populations, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1.0, \tau_1^2 = 1.0, \tau_2^2 = 2.0, \tau_3^2 = 3.0, \theta_0 = 6.0$ and $M = 3$.

¿From the tables, we observe that the values of $\bar{D}^n$ decrease quite rapidly as n increases, for $n \leq 80$. Observe that the distances between the μ_i's are 0.2 in Tables 1 and 2 ($\mu_1 = 5.7, \mu_2 = 5.9, \mu_3 = 6.1$) and those in Tables 3 and 4 are 2 ($\mu_1 = 3.0, \mu_2 = 5.0, \mu_3 = 7.0$). 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. Also, the simulation results indicate that the values of $n\bar{D}^n$ are decreasing as well as oscillating as n increases. This supports Theorem 4.1 and Theorem 4.2 that the rate of convergence is at least of order $O(\frac{(\ln n)^2}{n})$.

Tables 5 and 6 also list some simulation results on the performance of the proposed empirical Bayes rules $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$, for the case where $k = 5$ populations $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1.0$, $\tau_1^2 = 1.0, \tau_2^2 = 2.0, \tau_3^2 = 3.0, \tau_4^2 = 4.0, \tau_5^2 = 5.0$, $\theta_0 = 6.0$ and $M = 3$. Observe that the pattern of convergence in Tables 5 and 6 is similar to that in Tables 1, 2, 3 and 4.

Table 1. Performance of $\underset{\sim}{d}^{*n}$ for $\mu_1 = 5.7, \mu_2 = 5.9$ and $\mu_3 = 6.1$

n	$\bar{D}^{*n}$	$n\bar{D}^{*n}$	$SE(\bar{D}^{*n})$
20	379.0539×10^{-5}	75.81×10^{-3}	50.1028×10^{-5}
40	142.7377×10^{-5}	57.09×10^{-3}	21.7963×10^{-5}
60	100.5828×10^{-5}	60.34×10^{-3}	17.6571×10^{-5}
80	67.4769×10^{-5}	53.98×10^{-3}	10.9439×10^{-5}
100	48.3055×10^{-5}	48.30×10^{-3}	8.5001×10^{-5}
120	43.2401×10^{-5}	51.88×10^{-3}	8.0966×10^{-5}
140	38.5272×10^{-5}	53.93×10^{-3}	7.8631×10^{-5}
160	30.0363×10^{-5}	48.05×10^{-3}	6.6650×10^{-5}
180	31.4014×10^{-5}	56.52×10^{-3}	6.8492×10^{-5}
200	28.5429×10^{-5}	57.08×10^{-3}	6.5033×10^{-5}
250	21.5604×10^{-5}	53.90×10^{-3}	5.2406×10^{-5}
300	18.6757×10^{-5}	56.02×10^{-3}	4.5120×10^{-5}
350	15.4411×10^{-5}	54.04×10^{-3}	4.1852×10^{-5}
400	13.0606×10^{-5}	52.24×10^{-3}	3.7982×10^{-5}
450	8.6897×10^{-5}	39.10×10^{-3}	2.8739×10^{-5}
500	6.6412×10^{-5}	33.20×10^{-3}	2.0115×10^{-5}
550	6.7118×10^{-5}	36.91×10^{-3}	1.9547×10^{-5}
600	6.7059×10^{-5}	40.23×10^{-3}	1.9551×10^{-5}
650	5.4907×10^{-5}	35.68×10^{-3}	1.6339×10^{-5}
700	5.5703×10^{-5}	38.99×10^{-3}	1.6803×10^{-5}
750	6.3512×10^{-5}	47.63×10^{-3}	1.9093×10^{-5}
800	4.7742×10^{-5}	38.19×10^{-3}	1.7197×10^{-5}
850	4.6138×10^{-5}	39.21×10^{-3}	1.7375×10^{-5}
900	4.1697×10^{-5}	37.52×10^{-3}	1.5612×10^{-5}
950	3.9447×10^{-5}	37.47×10^{-3}	1.6174×10^{-5}
1000	3.4231×10^{-5}	34.23×10^{-3}	1.3860×10^{-5}

Table 2. Performance of $\hat{\underset{\sim}{d}}^n$ for $\mu_1 = 5.7, \mu_2 = 5.9$ and $\mu_3 = 6.1$

n	$\hat{\hat{D}}^n$	$n\hat{\hat{D}}^n$	$SE(\hat{\hat{D}}^n)$
20	582.5521×10^{-5}	116.51×10^{-3}	83.4687×10^{-5}
40	173.0359×10^{-5}	69.21×10^{-3}	26.1816×10^{-5}
60	126.7646×10^{-5}	76.05×10^{-3}	21.5321×10^{-5}
80	55.6173×10^{-5}	44.49×10^{-3}	9.5628×10^{-5}
100	54.9795×10^{-5}	54.97×10^{-3}	9.5976×10^{-5}
120	46.5250×10^{-5}	55.83×10^{-3}	8.2954×10^{-5}
140	40.2595×10^{-5}	56.36×10^{-3}	7.7687×10^{-5}
160	42.4713×10^{-5}	67.95×10^{-3}	8.0760×10^{-5}
180	38.0874×10^{-5}	68.55×10^{-3}	6.9213×10^{-5}
200	30.8239×10^{-5}	61.64×10^{-3}	6.0143×10^{-5}
250	22.8861×10^{-5}	57.21×10^{-3}	4.9300×10^{-5}
300	22.0051×10^{-5}	66.01×10^{-3}	4.9433×10^{-5}
350	16.6170×10^{-5}	58.15×10^{-3}	4.2685×10^{-5}
400	18.9588×10^{-5}	75.83×10^{-3}	4.6921×10^{-5}
450	17.5423×10^{-5}	78.94×10^{-3}	4.5963×10^{-5}
500	13.3475×10^{-5}	66.73×10^{-3}	3.9649×10^{-5}
550	9.2189×10^{-5}	50.70×10^{-3}	3.2928×10^{-5}
600	10.2007×10^{-5}	61.20×10^{-3}	3.2321×10^{-5}
650	5.8174×10^{-5}	37.81×10^{-3}	1.8195×10^{-5}
700	6.1850×10^{-5}	43.29×10^{-3}	1.9211×10^{-5}
750	7.0418×10^{-5}	52.81×10^{-3}	2.0546×10^{-5}
800	4.8460×10^{-5}	38.76×10^{-3}	1.6057×10^{-5}
850	4.4542×10^{-5}	37.86×10^{-3}	1.5668×10^{-5}
900	4.8734×10^{-5}	43.86×10^{-3}	1.6305×10^{-5}
950	4.9814×10^{-5}	47.32×10^{-3}	1.6527×10^{-5}
1000	4.7434×10^{-5}	47.43×10^{-3}	1.6151×10^{-5}

Table 3. Performance of $\underset{\sim}{d}^{*n}$ for $\mu_1 = 3.0, \mu_2 = 5.0$ and $\mu_3 = 7.0$

n	$\bar{D}^{*n}$	$n\bar{D}^{*n}$	$SE(\bar{D}^{*n})$
20	234.4199×10^{-5}	46.88×10^{-3}	40.2088×10^{-5}
40	98.6160×10^{-5}	39.44×10^{-3}	19.5706×10^{-5}
60	54.9728×10^{-5}	32.98×10^{-3}	11.7389×10^{-5}
80	40.9964×10^{-5}	32.79×10^{-3}	10.0235×10^{-5}
100	29.8254×10^{-5}	29.82×10^{-3}	7.4748×10^{-5}
120	25.7806×10^{-5}	30.93×10^{-3}	6.7082×10^{-5}
140	17.4323×10^{-5}	24.40×10^{-3}	4.3719×10^{-5}
160	16.5850×10^{-5}	26.53×10^{-3}	4.2272×10^{-5}
180	18.7297×10^{-5}	33.71×10^{-3}	5.0930×10^{-5}
200	11.7901×10^{-5}	23.58×10^{-3}	3.5439×10^{-5}
250	9.3719×10^{-5}	23.42×10^{-3}	3.1085×10^{-5}
300	9.9277×10^{-5}	29.78×10^{-3}	3.2287×10^{-5}
350	5.5070×10^{-5}	19.27×10^{-3}	1.9322×10^{-5}
400	4.5086×10^{-5}	18.03×10^{-3}	1.7411×10^{-5}
450	5.0565×10^{-5}	22.75×10^{-3}	1.7304×10^{-5}
500	3.1582×10^{-5}	15.79×10^{-3}	1.3989×10^{-5}
550	2.1795×10^{-5}	11.98×10^{-3}	0.9076×10^{-5}
600	2.7548×10^{-5}	16.52×10^{-3}	0.9945×10^{-5}
650	2.4533×10^{-5}	15.94×10^{-3}	0.9479×10^{-5}
700	1.7479×10^{-5}	12.23×10^{-3}	0.7678×10^{-5}
750	2.5670×10^{-5}	19.25×10^{-3}	1.1224×10^{-5}
800	1.4740×10^{-5}	11.79×10^{-3}	0.7175×10^{-5}
850	1.9989×10^{-5}	16.99×10^{-3}	0.8077×10^{-5}
900	1.6339×10^{-5}	14.70×10^{-3}	0.7350×10^{-5}
950	1.6339×10^{-5}	15.52×10^{-3}	0.7350×10^{-5}
1000	1.5635×10^{-5}	15.63×10^{-3}	0.7317×10^{-5}

Table 4. Performance of $\hat{\underset{\sim}{d}}^n$ for $\mu_1 = 3.0, \mu_2 = 5.0$ and $\mu_3 = 7.0$

n	$\hat{\hat{D}}^n$	$n\hat{\hat{D}}^n$	$SE(\hat{\hat{D}}^n)$
20	322.3745×10^{-5}	64.47×10^{-3}	52.7135×10^{-5}
40	123.5071×10^{-5}	49.40×10^{-3}	24.3336×10^{-5}
60	75.0701×10^{-5}	45.04×10^{-3}	16.8086×10^{-5}
80	52.7460×10^{-5}	42.19×10^{-3}	11.0058×10^{-5}
100	31.7421×10^{-5}	31.74×10^{-3}	8.0309×10^{-5}
120	35.7179×10^{-5}	42.86×10^{-3}	8.4613×10^{-5}
140	22.1338×10^{-5}	30.98×10^{-3}	7.1240×10^{-5}
160	20.8428×10^{-5}	33.34×10^{-3}	4.9779×10^{-5}
180	17.2954×10^{-5}	31.13×10^{-3}	4.6061×10^{-5}
200	13.7309×10^{-5}	27.46×10^{-3}	3.8921×10^{-5}
250	13.0443×10^{-5}	32.61×10^{-3}	3.5459×10^{-5}
300	10.4571×10^{-5}	31.37×10^{-3}	3.0632×10^{-5}
350	8.3921×10^{-5}	29.37×10^{-3}	2.7565×10^{-5}
400	6.0642×10^{-5}	24.25×10^{-3}	2.0972×10^{-5}
450	3.8821×10^{-5}	17.46×10^{-3}	1.4264×10^{-5}
500	3.3148×10^{-5}	16.57×10^{-3}	1.2925×10^{-5}
550	4.1512×10^{-5}	22.83×10^{-3}	1.5390×10^{-5}
600	3.4319×10^{-5}	20.59×10^{-3}	1.3113×10^{-5}
650	2.8374×10^{-5}	18.44×10^{-3}	1.2210×10^{-5}
700	2.8374×10^{-5}	19.86×10^{-3}	1.2210×10^{-5}
750	1.4740×10^{-5}	11.05×10^{-3}	0.7175×10^{-5}
800	1.4740×10^{-5}	11.79×10^{-3}	0.7175×10^{-5}
850	1.4740×10^{-5}	12.52×10^{-3}	0.7175×10^{-5}
900	2.0011×10^{-5}	18.01×10^{-3}	0.8900×10^{-5}
950	1.4740×10^{-5}	14.00×10^{-3}	0.7175×10^{-5}
1000	1.4037×10^{-5}	14.03×10^{-3}	0.7141×10^{-5}

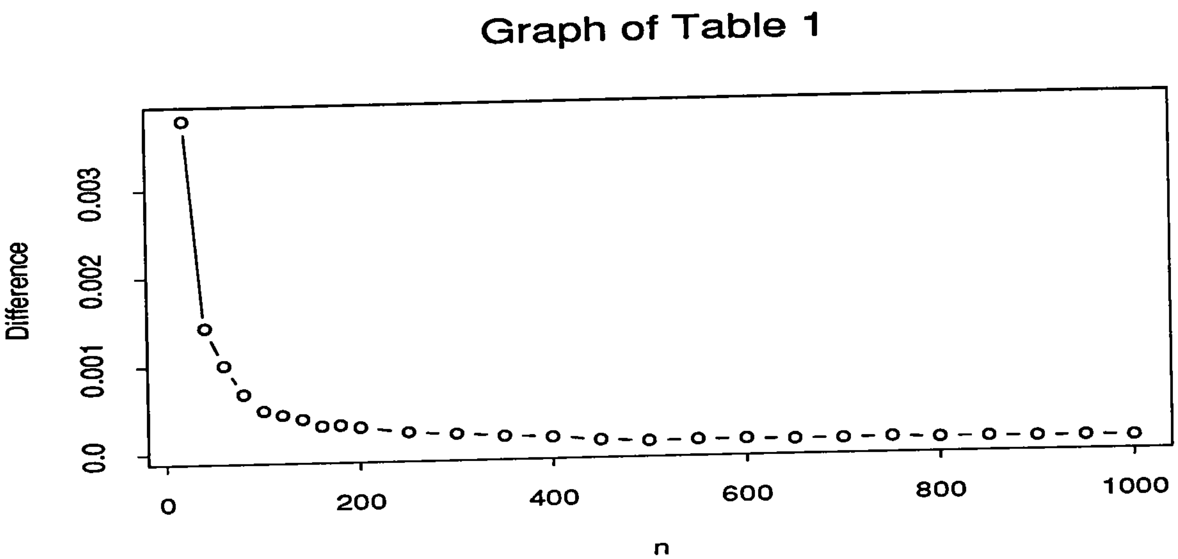
Table 5. Performance of $\underset{\sim}{d}^{*n}$ for $\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0, \mu_4 = 7.0$ and $\mu_5 = 9.0$

n	$\bar{D}^{*n}$	$n\bar{D}^{*n}$	$SE(\bar{D}^{*n})$
20	78.0680×10^{-5}	15.61×10^{-3}	18.5037×10^{-5}
40	34.8640×10^{-5}	13.94×10^{-3}	8.4872×10^{-5}
60	15.9108×10^{-5}	9.54×10^{-3}	5.4989×10^{-5}
80	14.6761×10^{-5}	11.74×10^{-3}	4.6919×10^{-5}
100	11.7843×10^{-5}	11.78×10^{-3}	4.2531×10^{-5}
120	9.9412×10^{-5}	11.92×10^{-3}	3.4225×10^{-5}
140	6.0193×10^{-5}	8.42×10^{-3}	2.5179×10^{-5}
160	4.0314×10^{-5}	6.45×10^{-3}	2.1292×10^{-5}
180	2.0986×10^{-5}	3.77×10^{-3}	0.9941×10^{-5}
200	3.6327×10^{-5}	7.26×10^{-3}	1.5376×10^{-5}
250	1.5373×10^{-5}	3.84×10^{-3}	0.7277×10^{-5}
300	1.5300×10^{-5}	4.59×10^{-3}	0.7715×10^{-5}
350	1.6649×10^{-5}	5.82×10^{-3}	0.8965×10^{-5}
400	1.5256×10^{-5}	6.10×10^{-3}	0.8869×10^{-5}
450	1.1824×10^{-5}	5.32×10^{-3}	0.4642×10^{-5}
500	2.2978×10^{-5}	11.48×10^{-3}	1.2156×10^{-5}
550	0.8211×10^{-5}	4.51×10^{-3}	0.3541×10^{-5}
600	0.9623×10^{-5}	5.77×10^{-3}	0.4396×10^{-5}
650	0.7880×10^{-5}	5.12×10^{-3}	0.3886×10^{-5}
700	0.6514×10^{-5}	4.56×10^{-3}	0.3109×10^{-5}
750	0.6514×10^{-5}	4.88×10^{-3}	0.3109×10^{-5}
800	0.5996×10^{-5}	4.79×10^{-3}	0.3066×10^{-5}
850	1.2694×10^{-5}	10.79×10^{-3}	0.6916×10^{-5}
900	1.2694×10^{-5}	11.42×10^{-3}	0.6916×10^{-5}
950	1.1384×10^{-5}	10.81×10^{-3}	0.6791×10^{-5}
1000	0.5204×10^{-5}	5.20×10^{-3}	0.2820×10^{-5}

Table 6. Performance of $\hat{\underset{\sim}{d}}^n$ for $\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0, \mu_4 = 7.0$ and $\mu_5 = 9.0$

n	$\hat{\hat{D}}^n$	$n\hat{\hat{D}}^n$	$SE(\hat{\hat{D}}^n)$
20	104.1815×10^{-5}	20.83×10^{-3}	21.5616×10^{-5}
40	43.0341×10^{-5}	17.21×10^{-3}	9.6348×10^{-5}
60	32.0583×10^{-5}	19.23×10^{-3}	9.1339×10^{-5}
80	21.8620×10^{-5}	17.48×10^{-3}	6.5085×10^{-5}
100	16.8074×10^{-5}	16.80×10^{-3}	4.9772×10^{-5}
120	9.9662×10^{-5}	11.95×10^{-3}	3.4675×10^{-5}
140	7.8557×10^{-5}	10.99×10^{-3}	3.1872×10^{-5}
160	5.2283×10^{-5}	8.36×10^{-3}	2.3531×10^{-5}
180	8.2048×10^{-5}	14.76×10^{-3}	3.2318×10^{-5}
200	8.1503×10^{-5}	16.30×10^{-3}	3.1371×10^{-5}
250	5.0492×10^{-5}	12.62×10^{-3}	2.3132×10^{-5}
300	2.8145×10^{-5}	8.44×10^{-3}	1.4592×10^{-5}
350	3.1544×10^{-5}	11.04×10^{-3}	1.5650×10^{-5}
400	5.0784×10^{-5}	20.31×10^{-3}	2.4860×10^{-5}
450	0.8631×10^{-5}	3.88×10^{-3}	0.3760×10^{-5}
500	2.0451×10^{-5}	10.22×10^{-3}	1.1867×10^{-5}
550	0.7693×10^{-5}	4.23×10^{-3}	0.3503×10^{-5}
600	0.8903×10^{-5}	5.34×10^{-3}	0.5521×10^{-5}
650	0.4701×10^{-5}	3.05×10^{-3}	0.2333×10^{-5}
700	1.5939×10^{-5}	11.15×10^{-3}	1.1476×10^{-5}
750	0.4701×10^{-5}	3.52×10^{-3}	0.2333×10^{-5}
800	0.4701×10^{-5}	3.76×10^{-3}	0.2333×10^{-5}
850	0.3821×10^{-5}	3.24×10^{-3}	0.2161×10^{-5}
900	0.3821×10^{-5}	3.43×10^{-3}	0.2161×10^{-5}
950	0.3821×10^{-5}	3.63×10^{-3}	0.2161×10^{-5}
1000	0.3821×10^{-5}	3.82×10^{-3}	0.2161×10^{-5}

5Date: Tue, 27 Apr 93 14:52:59 EST



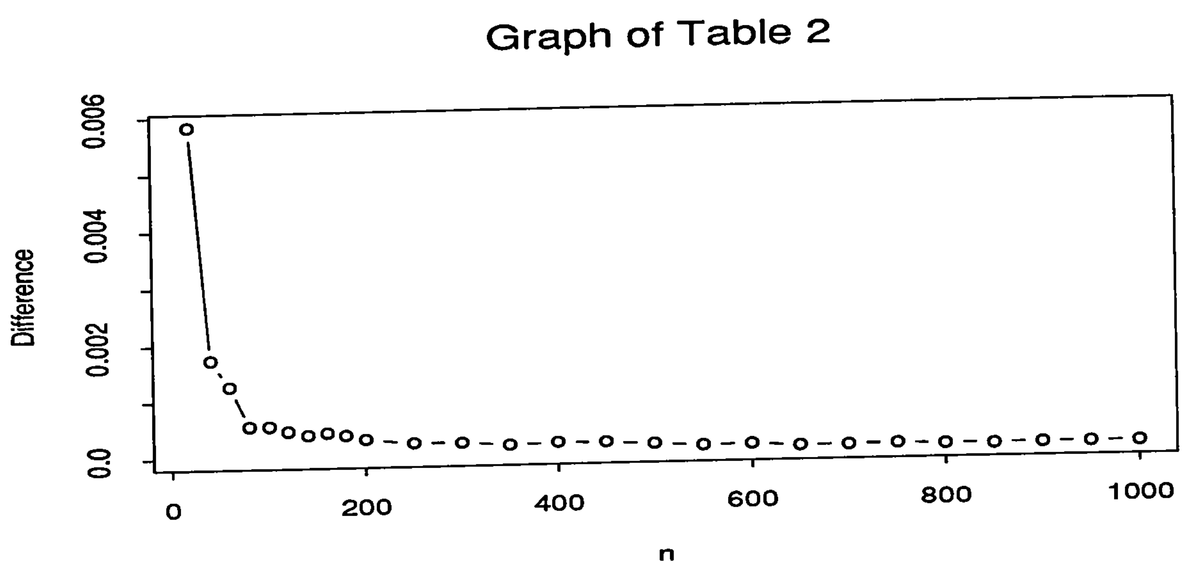


Figure 1: $\bar{D}^{*n}$ vs n and $\hat{\hat{D}}^{n}$ vs n for Table 1 and Table 2

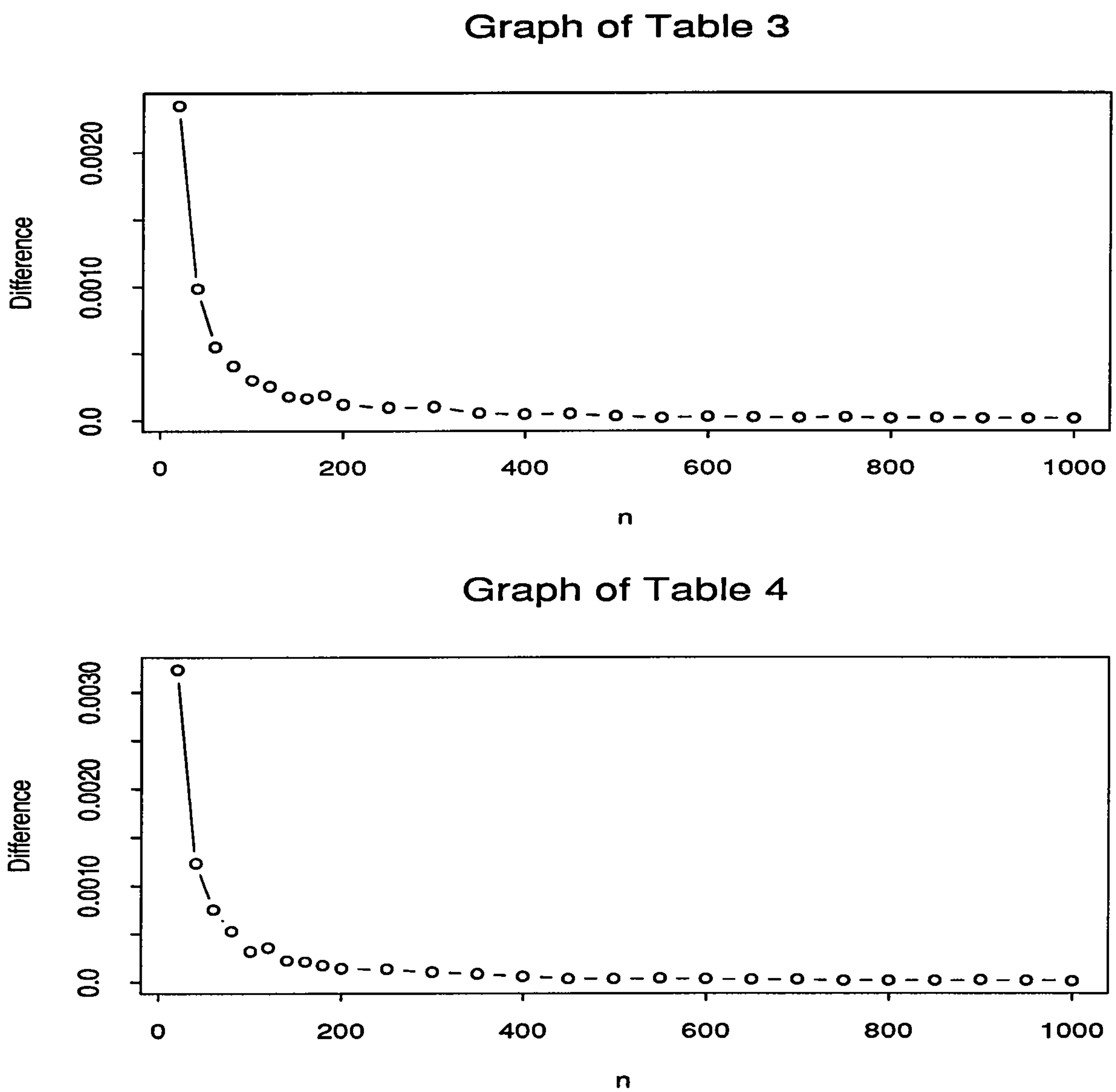


Figure 2: $\bar{D}^{*n}$ vs n and $\hat{\bar{D}}^{n}$ vs n for **Table 3 and Table 4**

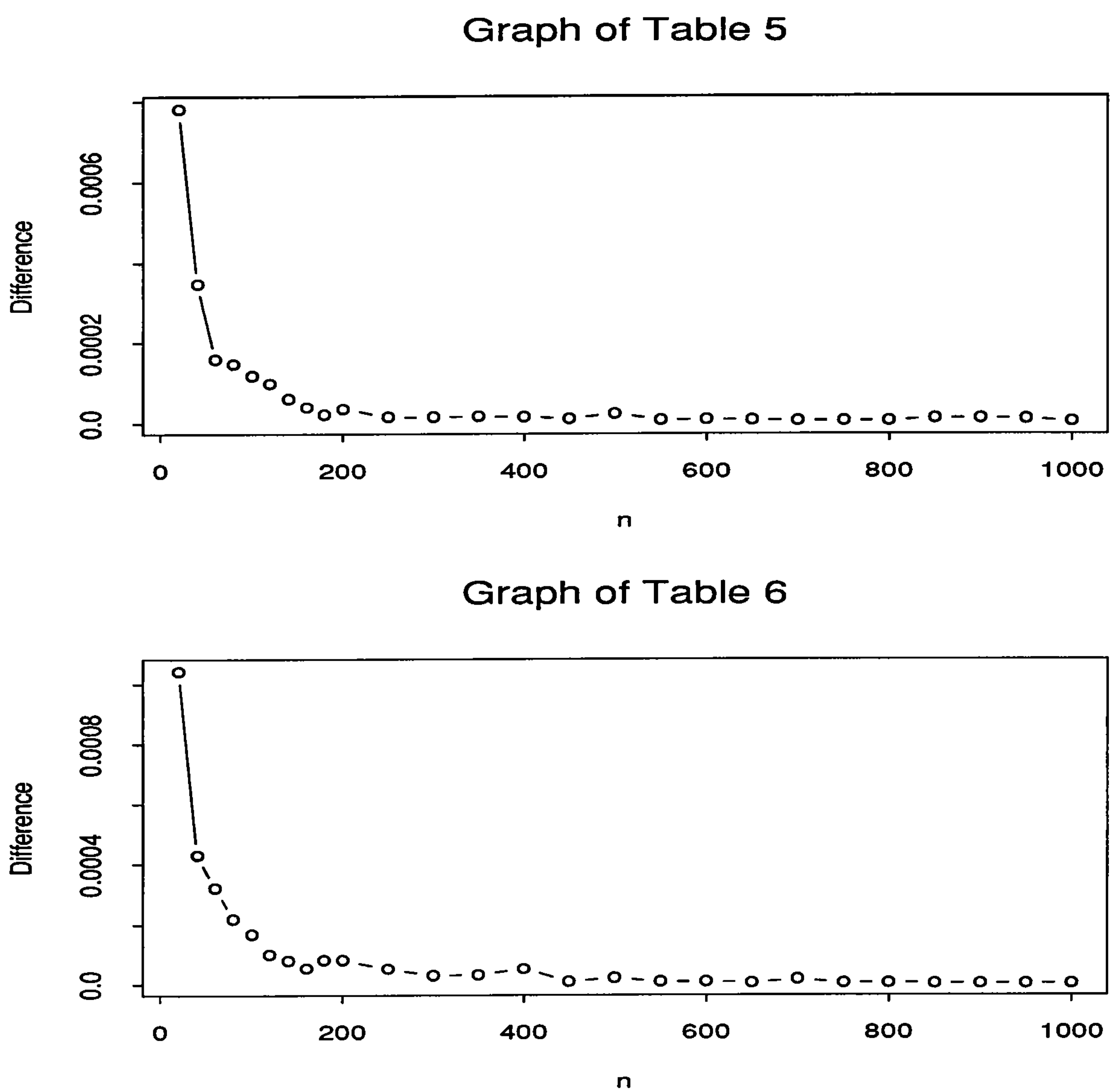


Figure 3: $\bar{D}^{*n}$ vs n and $\hat{\bar{D}}^n$ vs n for Table 5 and Table 6

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 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.

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.

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

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 Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Inference*, **8**, 87-101.

Gupta, S. S. and Leu, L. Y. (1991). On Bayes and empirical Bayes rules for selecting good populations. In *Proceedings of the 1990 Taipei Symposium in Statistics,* (Eds. M. T. Chao and P. E. Cheng), Institute of Statistical Science, Academia Sinica, Taipei, 167-182.

Gupta, S. S. and Liang, T. (1988). Empirical Bayes rules for selecting the best binomial population. *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. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.*, **29**, 235–244.

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

Pollard, David (1984). *Convergence of Stochastic Processes.* Springer-Verlag.

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

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

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.

From nlucas Thu Jan 14 08:38:42 1993
Received: by pop.stat.purdue.edu (4.1/Purdue_CC)
id AA28321; Thu, 14 Jan 93 08:38:39 EST
Date: Thu, 14 Jan 93 08:38:39 EST
From: nlucas (Norma Lucas)
Message-Id: <9301141338.AA28321@pop.stat.purdue.edu>
To: epperson@pop.stat.purdue.edu, hieberd@pop.stat.purdue.edu,
nlucas@pop.stat.purdue.edu, seele@pop.stat.purdue.edu
Subject: paper by rau, gupta, etc.
Status: R

There are 3 files in my main pop account.

rau.tex
rau1.tex – fig1.post
rau2.tex – fig2.post

Rau said to get it to latex it might be better to put it under a directory Rau. But you can try it as we normally do tech reports, and if it doesn't work, then you might try creating a directory all by itself. If it still doesn't latex, then see Mark. It is a paper with 2 figure files.

Thanks Norma

This is to be
93-1

```
From mds@snap.stat.purdue.edu Thu Jan 14 10:13:00 1993
Received: from snap.stat.purdue.edu by pop.stat.purdue.edu (4.1/Purdue_CC)
        id AA00116; Thu, 14 Jan 93 10:12:57 EST
Received: by snap.stat.purdue.edu (AIX 3.2/UCB 5.64/4.03)
          id AA41017; Thu, 14 Jan 1993 10:12:53 -0500
From: mds@snap.stat.purdue.edu (Mark Senn)
Message-Id: <9301141512.AA41017@snap.stat.purdue.edu>
To: seele@pop.stat.purdue.edu (Teena Seele)
Subject: Re: technical report
In-Reply-To: Your message of Thu, 14 Jan 93 10:04:23 EST.
             <9301141504.AA29865@pop.stat.purdue.edu>
Date: Thu, 14 Jan 93 10:12:53 -0500
Status: R
```

> I just sent a message to Rau. Maybe since it's his files he will know and save you some time.
>
> Thanks,
>
> Teena

I took a quick peek at it anyway.

He'll need to send you the "fig1.post" and "fig2.post" files.

Mark

From seele Thu Jan 14 09:34:19 1993
To: mds@stat.purdue.edu
Subject: technical report

Hi Mark,

I'm working on a technical report for Gupta/Liang/Rau and Rau typed the paper in latex. He has a couple of figures in it and I can't get it to latex. Could you take a look at the file for me? It's in TechnicalReports/1993/TR93-1. rau.tex is the main file and rau1.tex and rau2.tex are the figures.

Thanks,

Teena

Teena

\006

From seele Thu Jan 14 10:03:52 1993
To: raur
Subject: Technical Report 93-1

Hi,

I'm trying to latex your technical report file. I'm getting the following error message. Do you know what it might mean? I've also asked Mark Senn to help me, but he's so busy I'm not sure when he will get to it. Thanks,

Teena

```
**********************************88

psfig/tex 1.4 / TeXPS
psfig: searching fig1.post for bounding box
! FATAL ERROR: cannot open fig1.post.
\bb@missing ...ATAL ERROR: cannot open \@p@sfile }
                                                  \fi \loop \read \ps@strea...

\compute@bb ...no@bbtrue \fi \ifno@bb \bb@missing
                                                  \fi \ifno@bb \errmessage {...

\psfig ...@parms \parse@ps@parms {#1} \compute@bb
                                                  \compute@handw \compute@re...

\centerline #1->\@@line {\hss #1
                                \hss }
l.2532 ....post,height=6.2truein,width=6.5truein}}

?
```

From seele Thu Jan 14 10:04:23 1993
To: mds@stat.purdue.edu
Subject: technical report

I just sent a message to Rau. Maybe since it's his files he will know and save you some time.

Thanks,

From seele Thu Jan 14 09:34:19 1993
To: mds@stat.purdue.edu
Subject: technical report

Hi Mark,

I'm working on a technical report for Gupta/Liang/Rau and Rau typed the paper in latex. He has a couple of figures in it and I can't get it to latex. Could you take a look at the file for me? It's in TechnicalReports/1993/TR93-1. rau.tex is the main file and rau1.tex and rau2.tex are the figures.

Thanks,

Teena

EMPIRICAL BAYES RULES FOR SELECTING THE BEST NORMAL 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

Abstract

The problem of selecting the population with the largest mean from among $k(\geq 2)$ independent normal 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. Accordingly, the empirical Bayes approach is employed. Combining useful information from the past data, empirical Bayes selection procedures are developed. It is proved that the proposed empirical Bayes selection procedures are asymptotically optimal, having a rate of convergence of order $O(\frac{(\ln n)^2}{n})$, 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 procedures for small to moderate values of n.

AMS 1991 Subject Classification: Primary 62F07; secondary 62C12, 62C10
Keywords and Phrases: Asymptotic optimality; best population; Bayes rule; empirical Bayes; rate of convergence.

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

1 Introduction

Consider k independent normal populations $\pi_1, \ldots, \pi_k$ with unknown means $\theta_1, \cdots, \theta_k$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered θ_i. A population π_i with $\theta_i = \theta_{[k]}$ is called the best population. The problem of selecting the best population was studied in the pioneering papers, Bechhofer (1954) and Gupta (1956), by using the indifference zone approach and the subset selection approach, respectively. Gupta and Panchapakesan (1979) provide a comprehensive survey of the development in this research area.

In a practical situation, one may not only be interested in the selection of the best population, but also require the selected population to be good enough. For example, in medical studies, the performance of any proposed new treament must be better than a standard treatment before it can be accepted by medical practitioners. In the literature, Bechhofer and Turnbull (1978), Dunnett (1984) and Wilcox (1984) investigated procedures for selecting the best normal population compared with a control, respectively. Using the subset selection approach, Gupta and Sobel (1958) and Lehmann (1961) have made some contributions to this problem.

In this paper, we employ the empirical Bayes approach to select the best normal population provided it is as good as a specified standard. The empirical Bayes methodology was introduced by Robbins (1956, 1964). This empirical Bayes approach has been used in selection problems by several authors. Deely (1965) studied the empirical Bayes rule for selecting the best normal population. Recently, Gupta and Hsiao (1983), Gupta and Liang (1988,1989), and Gupta and Leu (1991) have investigated empirical Bayes procedures for several selection problems. Many such empirical Bayes selection procedures have been shown to be asymptotically optimal in the sense that the empirical Bayes risk converges to the minimum Bayes risk.

This paper deals with a single-stage selection procedure for selecting the best normal population compared with a specified standard using the parametric empirical Bayes approach. In Section 2, we describe the formulation of the selection problem, and derive a Bayes selection rule. In Section 3, we construct the empirical Bayes selection rules. In Section 4, the asymptotic optimality of the proposed empirical Bayes selection rules is investigated. It is shown that the empirical Bayes selection rules have a rate of convergence of order $O(\frac{(\ln n)^2}{n})$, where n is the number of past observations at hand. In Section 5, we present the results of the simulation study of the proposed empirical Bayes selection procedures for small to moderate values of n.

2 Formulation of the Selection Problem and the Bayes Selection Rule

Let $\pi_1, \ldots, \pi_k$ be k independent normal populations with unknown means $\theta_1, \ldots, \theta_k$, 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 the exact pairing between the ordered and the unordered parameters is unknown. A population π_i with $\theta_i = \theta_{[k]}$ is considered as the best population. Let θ_0 be a known control. A population π_i with $\theta_i \geq \theta_0$ is considered as a good population. Our goal is to derive empirical Bayes rules to select the best normal population which should also be good compared with the control θ_0. If there is no such population, we select none.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | \theta_i \in R, i = 1, \ldots, k\}$ 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 compared with the control θ_0. When $a_0 = 1$, it means that all k populations are excluded as bad populations. We consider the following loss function:

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \max(\theta_{[k]}, \theta_0) - \sum_{i=0}^{k} a_i \theta_i. \tag{2.1}$$

Thus, if $\theta_{[k]} > \theta_0$ and all populations are rejected then the loss is $\theta_{[k]} - \theta_0$. On the other hand, if $\theta_0 > \theta_{[k]}$ and population π_i is selected as the best and good then the loss is $\theta_0 - \theta_i$.

For each $i = 1, 2, \ldots, k$, let $X_{i1}, \cdots, X_{iM}$ be a sample of size M from a normal population π_i which has mean θ_i and variance σ_i^2. It is assumed that θ_i is a realization of a random variable Θ_i which has a $N(\mu_i, \tau_i^2)$ prior distribution with unknown parameters $(\mu_i, \tau_i^2), i = 1, \ldots, k$. The random variables $\Theta_1, \ldots, \Theta_k$ are assumed to be independent. We let $f_i(x_i \mid \theta_i)$ and $h_i(\theta_i | \mu_i, \tau_i^2)$ denote the conditional probability density of $X_i = \bar{X}_i = \frac{1}{M} \sum_{j=1}^{M} X_{ij}$ and the density of Θ_i, respectively. Let $\underset{\sim}{X} = (X_1, \ldots, X_k)$ and let $\mathcal{X}$ be the sample space generated by $\underset{\sim}{X}$. A selection rule $\underset{\sim}{d} = (d_0, \ldots, d_k)$ is a mapping defined on the sample space $\mathcal{X}$. For every $\underset{\sim}{x} \in \mathcal{X}, d_i(\underset{\sim}{x}), i = 1, \ldots, k$, is the probability of selecting population π_i as the best and good, and $d_0(\underset{\sim}{x})$ is the probability of excluding all k populations as bad and selecting none. Also, $\sum_{i=0}^{k} d_i(\underset{\sim}{x}) = 1$, for all $\underset{\sim}{x} \in \mathcal{X}$.

Under the preceding statistical model, the Bayes risk of the selection rule $\underset{\sim}{d}$ is denoted by $R(\underset{\sim}{d})$. Then, a straightforward computation yields the following :

$$R(\underset{\sim}{d}) = -\int_{\mathcal{X}} \left[\sum_{i=0}^{k} d_i(\underset{\sim}{x}) \varphi_i(x_i) \right] f(\underset{\sim}{x}) d\underset{\sim}{x} + C, \tag{2.2}$$

where

$$\begin{cases} C = \int_{\Omega} \max(\theta_{[k]}, \theta_0) dH(\underset{\sim}{\theta}), \\ \varphi_0(x_0) \equiv \theta_0, \\ \varphi_i(x_i) = E(\Theta_i | x_i) = \dfrac{x_i \tau_i^2 + \frac{\sigma_i^2}{M} \mu_i}{\tau_i^2 + \frac{\sigma_i^2}{M}} : \text{ the posterior mean of } \Theta_i \text{ given } X_i = x_i, \; i \neq 0, \\ f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(x_i), \quad f_i(x_i) = \int_R f_i(x_i | \theta_i) h_i(\theta_i | \mu_i, \tau_i^2) d\theta_i, \\ H(\underset{\sim}{\theta}) : \text{ the joint distribution of } \underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_k). \end{cases} \tag{2.3}$$

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

$$\begin{cases} I(\underset{\sim}{x}) = \{i | \varphi_i(x_i) = \max_{0 \leq j \leq k} \varphi_j(x_j), i = 0, \ldots, k\}, \\ i^* \equiv i^*(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } I(\underset{\sim}{x}) = \{0\}; \\ \min\{i | i \in I(\underset{\sim}{x}), \; i \neq 0\} & \text{otherwise.} \end{cases} \end{cases} \tag{2.4}$$

Then a Bayes selection rule $\underset{\sim}{d}^B = (d_0^B, \ldots, d_k^B)$ is given as follows:

$$\begin{cases} d_{i^*}^B(\underset{\sim}{x}) = 1, \\ d_j^B(\underset{\sim}{x}) = 0 \quad \text{for } j \neq i^*. \end{cases} \tag{2.5}$$

3 The Empirical Bayes Selection Rules

Since the parameters $(\mu_i, \tau_i^2), i = 1, \ldots, k$, are unknown, it is not possible to apply the Bayes rule $\underset{\sim}{d}^B$ for the selection problem at hand. In the empirical Bayes framework, it is assumed that certain past data are available when the present selection is made. Let $X_{ijl}, j = 1, \ldots, M$, denote a sample of size M from π_i at time $l, l = 1, \ldots, n$. It is assumed that conditional on $(\theta_{il}, \sigma_i^2)$, $X_{ijl}, j = 1, \ldots, M$, follow a normal distribution $N(\theta_{il}, \sigma_i^2)$ and θ_{il} is a realization of a random variable Θ_{il} which has a normal distribution $N(\mu_i, \tau_i^2)$. It is also assumed that Θ_{il}, $i = 1, \ldots, k$, $l = 1, 2, \ldots$, are mutually independent. For ease of notation, we denote the current random observations $X_{ij\,n+1}$ by X_{ij}, $j = 1, \ldots, M$, $i = 1, \ldots, k$.

For population $\pi_i, i = 1, \ldots, k$, let $X_{i.l} = \bar{X}_{i.l}$ be the sample mean of the M observations obtained at time l, $X_i(n)$be the overall sample mean of past data and let $S_i^2(n)$ be the overall sample variance of the past data. That is

$$\begin{cases} X_{i.l} &= \frac{1}{M}\sum_{j=1}^M X_{ijl}, \\ X_i(n) &= \frac{1}{n}\sum_{l=1}^n X_{i.l}, \\ S_i^2(n) &= \frac{1}{n-1}\sum_{l=1}^n (X_{i.l} - X_i(n))^2. \end{cases} \tag{3.1}$$

Also, let $v_i^2 = \tau_i^2 + \frac{\sigma_i^2}{M}$. Then, from the statistical model described before, $X_{i.1}, X_{i.2}, \ldots, X_{i.n}$ are marginally independent with a $N(\mu_i, v_i^2)$ distribution. Hence, $X_i(n)$ has a $N(\mu_i, \frac{v_i^2}{n})$ distribution and $\frac{n-1}{v_i^2}s_i^2(n)$ has a $\chi^2(n-1)$ distribution. By the strong law of large numbers, we have

$$\begin{cases} X_i(n) \longrightarrow \mu_i \text{ a.s. }, \\ S_i^2(n) \longrightarrow v_i^2 \text{ a.s. }. \end{cases} \tag{3.2}$$

3.1 Case 1: (μ_i, τ_i^2) unknown and σ_i^2 known, $i = 1, \ldots, k$

Consider the case where both (μ_i, τ_i^2) are unknown and σ_i^2 is known, $i = 1, \cdots, k$. Since $E(X_i(n)) = \mu_i$, $E(S_i^2(n) - \frac{\sigma_i^2}{M}) = \tau_i^2$ and it is possible that $S_i^2(n) - \frac{\sigma_i^2}{M} \leq 0$, we define μ_{in} and τ_{in}^2 as estimators of μ_i and τ_i^2, respectively, by the following:

$$\begin{cases} \mu_{in} = X_i(n), \\ \tau_{in}^2 = \max(S_i^2(n) - \frac{\sigma_i^2}{M},\ 0). \end{cases} \tag{3.3}$$

Now, we define, for $i = 1, 2, \ldots, k$,

$$\begin{cases} v_{in}^2 = \tau_{in}^2 + \frac{\sigma_i^2}{M}, \\ \varphi_{in}(x_i) = \frac{x_i\tau_{in}^2 + \frac{\sigma_i^2}{M}\mu_{in}}{v_{in}^2}, \\ \varphi_{0n}(x_0) \equiv \theta_0. \end{cases} \tag{3.4}$$

We use v_{in}^2 and $\varphi_{in}(x_i)$ to estimate v_i^2 and $\varphi_i(x_i)$, respectively.

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

$$\begin{cases} I_n(\underset{\sim}{x}) = \{i | \varphi_{in}(x_i) = \max_{0 \le j \le k} \varphi_{jn}(x_j), i = 0, \ldots, k\}, \\ i_n^* \equiv i_n^*(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } I_n(\underset{\sim}{x}) = \{0\}, \\ \min\{i | i \in I_n(\underset{\sim}{x}),\ i \neq 0\} & \text{otherwise.} \end{cases} \end{cases} \tag{3.5}$$

We then obtain an empirical Bayes selection rule $\underset{\sim}{d}^{*n} = (d_0^{*n}, \ldots, d_k^{*n})$ as follows:

$$\begin{cases} d_{i_n^*}^{*n}(\underset{\sim}{x}) = 1, \\ d_j^{*n}(\underset{\sim}{x}) = 0 \quad \text{for } j \neq i_n^*. \end{cases} \tag{3.6}$$

3.2 Case 2: (μ_i, τ_i^2) and σ_i^2 unknown, $i = 1, \cdots, k$.

When σ_i^2, $i = 1, \ldots, k$, are unknown, it is assumed that $M \geq 2$. For each $i = 1, \cdots, k$, at time l, let $W_{i.l}^2$ and $W_i^2(n)$ be the sample variance at time l and the overall (pooled) sample variance, respectively. That is

$$\begin{cases} W_{i.l}^2 = \frac{1}{M-1} \sum_{j=1}^M (X_{ijl} - X_{i.l})^2, \\ W_i^2(n) = \frac{1}{n} \sum_{l=1}^n W_{i.l}^2. \end{cases} \tag{3.7}$$

Then, $\frac{M-1}{\sigma_i^2} W_{i.1}^2, \cdots, \frac{M-1}{\sigma_i^2} W_{i.n}^2$ are i.i.d. having a $\chi^2(M-1)$ distribution and hence $\frac{n(M-1)}{\sigma_i^2} W_i^2(n)$ has a $\chi^2(n(M-1))$ distribution. From the above discussion and by the strong law of large numbers, we have

$$\begin{cases} X_i(n) \longrightarrow \mu_i \quad \text{a.s.}, \\ W_i^2(n) \longrightarrow \sigma_i^2 \quad \text{a.s.}, \\ S_i^2(n) \longrightarrow v_i^2 \quad \text{a.s.}, \\ S_i^2(n) - \frac{W_i^2(n)}{M} \longrightarrow \tau_i^2 \quad \text{a.s.}, \\ E(X_i(n)) = \mu_i, \quad E(S_i^2(n)) = v_i^2, \quad E(W_i^2(n)) = \sigma_i^2, \\ E(S_i^2(n) - \frac{W_i^2(n)}{M}) = v_i^2 - \frac{\sigma_i^2}{M} = \tau_i^2. \end{cases} \tag{3.8}$$

Since, it is possible that $S_i^2(n) - \frac{W_i^2(n)}{M} < 0$, we define $\hat{\mu}_{in}, \hat{\sigma}_{in}^2, \hat{v}_{in}^2$ and $\hat{\tau}_{in}^2$ as estimators of μ_i, σ_i^2, v_i^2 and τ_i^2, respectively, by the following :

$$\begin{cases} \hat{\mu}_{in} = X_i(n), \\ \hat{\sigma}_{in}^2 = W_i^2(n), \\ \hat{v}_{in}^2 = S_i^2(n), \\ \hat{\tau}_{in}^2 = \max(\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M},\ 0). \end{cases} \tag{3.9}$$

For $i = 1, 2, \cdots, k$, we define

$$\begin{cases} \hat{\varphi}_{in}(x_i) = \frac{x_i \hat{\tau}_{in}^2 + \frac{\hat{\sigma}_{in}^2}{M} \hat{\mu}_{in}}{\hat{v}_{in}^2}, \\ \hat{\varphi}_{0n}(x_0) \equiv \theta_0, \end{cases} \tag{3.10}$$

and use $\hat{\varphi}_{in}(x_i)$ as an estimator of $\varphi_i(x_i)$.

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

$$\begin{cases} \hat{I}_n(\underset{\sim}{x}) = \{i | \hat{\varphi}_{in}(x_i) = \max_{0 \leq j \leq k} \hat{\varphi}_{jn}(x_j), i = 0, \ldots, k\}, \\ \hat{i}_n \equiv \hat{i}_n(\underset{\sim}{x}) = \begin{cases} 0 & \text{if } \hat{I}_n(\underset{\sim}{x}) = \{0\}, \\ \min\{i | i \in \hat{I}_n(\underset{\sim}{x})\ i \neq 0\} & \text{otherwise.} \end{cases} \end{cases} \tag{3.11}$$

We then have an empirical Bayes selection rule $\hat{\underset{\sim}{d}}^n = (\hat{d}_0^n, \ldots, \hat{d}_k^n)$ as follows:

$$\begin{cases} \hat{d}_{\hat{i}_n}^n(\underset{\sim}{x}) = 1, \\ \hat{d}_j^n(\underset{\sim}{x}) = 0 \quad \text{for } j \neq \hat{i}_n. \end{cases} \tag{3.12}$$

4 Asymptotic Optimality of the Empirical Bayes Selection Rules

In this section, we prove two theorems (Theorem 4.1 and Theorem 4.2) concerning the asymptotic optimality of the preceding empirical Bayes rules.

Consider an empirical Bayes selection rule $\underset{\sim}{d}^n = (d_0^n, \ldots, d_k^n)$. We denote the associated Bayes risk of this empirical Bayes rule by $R(\underset{\sim}{d}^n)$. Then, from (2.2),

$$R(\underset{\sim}{d}^n) = -\int_{\mathcal{X}} \left[\sum_{i=0}^{k} d_i^n(\underset{\sim}{x}) \varphi_i(x_i) \right] f(\underset{\sim}{x}) d\underset{\sim}{x} + C. \tag{4.1}$$

Also, $R(\underset{\sim}{d}^n) - R(\underset{\sim}{d}^B) \geq 0$, since $R(\underset{\sim}{d}^B)$ is the minimum Bayes risk. Thus, $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B) \geq 0$, where the expection E_n is taken with respect to X_{ijl}, $i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$. The nonnegative difference $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$ is generally used as a measure of the performance of the selection rule $\underset{\sim}{d}^n$.

Definition 4.1 A sequence of empirical Bayes rules $\{\underset{\sim}{d}^n\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order β_n if $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B) = O(\beta_n)$, where β_n is a sequence of positive numbers such that $\lim_{n \to \infty} \beta_n = 0$.

In order to investigate the asymptotic optimality of the proposed empirical Bayes selection rules, we introduce some useful lemmas.

Lemma 4.1 is part of Theorem 1 of Chernoff (1952).

Lemma 4.1 Suppose S_n is the sum of n independent observations $X_1, X_2, \ldots, X_n$ of a random variable X with moment generating function $M(t) = E(e^{tX})$. Let $m(a) = \inf_t E(e^{t(X-a)}) = \inf_t e^{-at} M(t)$. Then,

(a) If $E(X) > -\infty$ and $a \leq E(X)$ then $P(S_n \leq na) \leq [m(a)]^n$,
(b) If $E(X) < +\infty$ and $a \geq E(X)$ then $P(S_n \geq na) \leq [m(a)]^n$.

Corollary 4.1 Let X have a $\chi^2(1)$ distribution. Then, S_n has a $\chi^2(n)$ distribution and

$$(a)\quad P\{S_n \le n(1-\eta)\} \le \exp(-\frac{n}{2}g_1(\eta)) \text{ for any } \eta,\ 0<\eta<1,$$
$$(b)\quad P\{S_n \ge n(1+\eta)\} \le \exp(-\frac{n}{2}g_2(\eta)) \text{ for any } \eta,\ \eta>0;$$

where

$$\begin{aligned} g_1(\eta) &= -\eta - \ln(1-\eta) \text{ for any } \eta,\ 0<\eta<1, \\ g_2(\eta) &= \eta - \ln(1+\eta) \text{ for any } \eta,\ \eta>0. \end{aligned}$$

Proof : The moment generating function of X is given by $M(t) = (1-2t)^{-\frac{1}{2}}$ for $t < \frac{1}{2}$ and hence $m(a) = \inf_t E(e^{t(X-a)}) = E(e^{\frac{a-1}{2a}(X-a)}) = [e^{(1-a)}a]^{\frac{1}{2}}$. Therefore, $m(1-\eta) = [e^{\eta}(1-\eta)]^{\frac{1}{2}} = e^{\frac{1}{2}(\eta+\ln(1-\eta))} = e^{-\frac{1}{2}(-\eta-\ln(1-\eta))} = e^{-\frac{1}{2}g_1(\eta)}$ and $m(1+\eta) = [e^{-\eta}(1+\eta)]^{\frac{1}{2}} = e^{-\frac{1}{2}(\eta-\ln(1+\eta))} = e^{-\frac{1}{2}g_2(\eta)}$. The results follow from Lemma 4.1. □

Remark 1. Observe that $g_1(0) = g_2(0) = 0$, $\frac{d}{d\eta}g_1(\eta) > 0$, for $0<\eta<1$, and $\frac{d}{d\eta}g_2(\eta) > 0$, for $\eta > 0$. Thus, $g_1(\eta)$ and $g_2(\eta)$ are positive and strictly increasing functions for $0<\eta<1$ and $\eta>0$, respectively.

Remark 2. $\lim_{\eta\to 0}\frac{g_1(\eta)}{\eta^2} = \lim_{\eta\to 0}\frac{g_1'(\eta)}{2\eta} = \lim_{\eta\to 0}\frac{\frac{\eta}{1-\eta}}{2\eta} = \frac{1}{2}$. Similarly, $\lim_{\eta\to 0}\frac{g_2(\eta)}{\eta^2} = \frac{1}{2}$.

4.1 Case 1: (μ_i, τ_i^2) unknown and σ_i^2 known, $i = 1, \ldots, k$

Let P_n be the probability measure generated by the past random observations X_{ijl}, $i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$.

Lemma 4.2 Let μ_{in} and τ_{in}^2 be the estimators of μ_i and τ_i^2, respectively, as defined in (3.3). Also, let $g_1(\eta)$ and $g_2(\eta)$ be the functions defined in Corollary 4.1. Then, for any $c>0$, we have

$$\begin{aligned} (a)\ P_n\{|\mu_{in}-\mu_i| \ge c\} &\le \frac{2v_i}{\sqrt{2\pi}c}\frac{1}{\sqrt{n}}\exp(\frac{-c^2}{2v_i^2}n), \\ (b)\ P_n\{|\tau_{in}^2-\tau_i^2| \ge c\} &\le \exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})) \\ &\quad + \exp(-\frac{n-1}{2}g_1(\frac{c}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{c}{v_i^2})). \end{aligned}$$

Proof : (a) Note that $\mu_{in} = X_i(n)$ has a $N(\mu_i, \frac{v_i^2}{n})$ distribution and by the fact that $P\{Z \ge \eta\} < \frac{1}{\eta}\frac{exp(\frac{-\eta^2}{2})}{\sqrt{2\pi}}$, for any $\eta > 0$ and for a $N(0,1)$ distributed random variable Z, (see Pollard (1984) Appendix B) the result follows.

(b)

$$P_n\{|\tau_{in}^2 - \tau_i^2| \ge c\}$$

$$
\begin{aligned}
&\le P_n\{S_i^2(n)-\frac{\sigma_i^2}{M}\le 0\}+P_n\{|S_i^2(n)-\frac{\sigma_i^2}{M}-\tau_i^2|\ge c, S_i^2(n)-\frac{\sigma_i^2}{M}>0\}\\
&\le P_n\{|S_i^2(n)-v_i^2|\ge \tau_i^2\}+P_n\{|S_i^2(n)-v_i^2|\ge c\}\\
&= P_n\{|\frac{n-1}{v_i^2}S_i^2(n)-(n-1)|\ge (n-1)\frac{\tau_i^2}{v_i^2}\}+P_n\{|\frac{n-1}{v_i^2}S_i^2(n)-(n-1)|\ge (n-1)\frac{c}{v_i^2}\}\\
&\le \exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2}))\\
&\quad +\exp(-\frac{n-1}{2}g_1(\frac{c}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{c}{v_i^2})).
\end{aligned}
$$

The last inequality follows from Corollary 4.1 and the fact that $\frac{n-1}{v_i^2}S_i^2(n)$ has a $\chi^2(n-1)$ distribution. □

Lemma 4.3 Let $\varphi_i(x_i)$ and $\varphi_{in}(x_i)$ be defined as in (2.3) and (3.4), respectively. Then, for any $\varepsilon>0$ and any $x_i\in R$, we have

$$
\begin{aligned}
(a)\quad & P_n\{\varphi_{in}(x_i)-\varphi_i(x_i)>\varepsilon\}\le P_n\{|\mu_{in}-\mu_i|>\frac{Mv_i^2\varepsilon}{2\sigma_i^2}\}\\
&\qquad +P_n\{|\tau_{in}^2-\tau_i^2|>\frac{v_i^4\varepsilon}{2(\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2)}\},\\
(b)\quad & P_n\{\varphi_{in}(x_i)-\varphi_i(x_i)<-\varepsilon\}\le P_n\{|\mu_{in}-\mu_i|>\frac{Mv_i^2\varepsilon}{2\sigma_i^2}\}\\
&\qquad +P_n\{|\tau_{in}^2-\tau_i^2|>\frac{v_i^4\varepsilon}{2(\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2)}\}.
\end{aligned}
$$

<u>Proof</u> : We prove (a) only. The proof of (b) is similar to that of (a). Let $a=x_i, b=\frac{\sigma_i^2}{M}, y=\tau_i^2, z=\mu_i, y_n=\tau_{in}^2$ and $z_n=\mu_{in}$. Then, $y+b=v_i^2$ and we have

$$
\begin{aligned}
&P_n\{\varphi_{in}(x_i)-\varphi_i(x_i)>\varepsilon\}\\
&= P_n\{\frac{ay_n+bz_n}{y_n+b}-\frac{ay+bz}{y+b}>\varepsilon\}\\
&= P_n\{[b(a-z)-\varepsilon(y+b)](y_n-y)+b(b+y)(z_n-z)>\varepsilon(y+b)^2\}\\
&= P_n\{[\frac{\sigma_i^2}{M}(x_i-\mu_i)-\varepsilon v_i^2](\tau_{in}^2-\tau_i^2)+\frac{\sigma_i^2}{M}v_i^2(\mu_{in}-\mu_i)>\varepsilon v_i^4\}\\
&\le P_n\{\frac{\sigma_i^2}{M}v_i^2(\mu_{in}-\mu_i)>\frac{1}{2}\varepsilon v_i^4\}+P_n\{[\frac{\sigma_i^2}{M}(x_i-\mu_i)-\varepsilon v_i^2](\tau_{in}^2-\tau_i^2)>\frac{1}{2}\varepsilon v_i^4\}\\
&\le P_n\{|\mu_{in}-\mu_i|>\frac{Mv_i^2\varepsilon}{2\sigma_i^2}\}+P_n\{|\tau_{in}^2-\tau_i^2|>\frac{v_i^4\varepsilon}{2(\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2)}\}.
\end{aligned}
$$

□

Since $\varphi_1(X_1),\ldots,\varphi_k(X_k)$ are mutually independent, WLOG, we assume $\varphi_i(X_i)\ne\varphi_j(X_j)$, $\forall\, i\ne j$. This assumption does not change the Bayes risk $R(\underset{\sim}{d}^B)$ and the empirical Bayes risk $R(\underset{\sim}{d}^{*n})$ and hence the difference $E_n[R(\underset{\sim}{d}^{*n})]-R(\underset{\sim}{d}^B)$.

To investigate the convergence rate of $E_n[R(\underset{\sim}{d}^{*n})] - R(\underset{\sim}{d}^B)$, we state some facts :

1. As $i^* = 0,\ i_n^* = j \neq 0,\ \varphi_l(x_l) < \theta_0$ for all $l = 1, \ldots, k$. Then

$$\begin{aligned} & P_n\{i^* = 0,\ i_n^* = j\} = P_n\{\varphi_l(x_l) < \theta_0\ \forall\, l \neq 0,\ \varphi_{jn}(x_j) \geq \varphi_{ln}(x_l)\ \forall\, l \neq j\} \\ \leq\ & P_n\{\varphi_j(x_j) < \theta_0,\ \varphi_{jn}(x_j) \geq \theta_0\} \\ \leq\ & P_n\{\varphi_{jn}(x_j) - \varphi_j(x_j) > \theta_0 - \varphi_j(x_j)\}. \end{aligned}$$

2. As $i^* = i \neq 0,\ i_n^* = 0,\ \varphi_{ln}(x_l) < \theta_0$ for all $l = 1, \ldots, k$. Then

$$\begin{aligned} & P_n\{i^* = i,\ i_n^* = 0\} = P_n\{\varphi_i(x_i) \geq \varphi_l(x_l)\ \forall\, l \neq i,\ \varphi_{ln}(x_l) < \theta_0\ \forall\, l \neq 0\} \\ \leq\ & P_n\{\varphi_i(x_i) \geq \theta_0,\ \varphi_{in}(x_i) < \theta_0\} \\ \leq\ & P_n\{\varphi_{in}(x_i) - \varphi_i(x_i) < -(\varphi_i(x_i) - \theta_0)\}. \end{aligned}$$

3. As $i^* = i \neq 0,\ i_n^* = j \neq 0$ and $i \neq j$. Then

$$\begin{aligned} & P_n\{i^* = i,\ i_n^* = j\} = P_n\{\varphi_i(x_i) \geq \varphi_l(x_l)\ \forall\, l \neq i,\ \varphi_{jn}(x_j) \geq \varphi_{ln}(x_l)\ \forall\, l \neq j\} \\ \leq\ & P_n\{\varphi_i(x_i) \geq \varphi_j(x_j),\ \varphi_{jn}(x_j) \geq \varphi_{in}(x_i)\} \\ =\ & P_n\{\varphi_{jn}(x_j) - \varphi_j(x_j) - [\varphi_{in}(x_i) - \varphi_i(x_i)] \geq \varphi_i(x_i) - \varphi_j(x_j),\ \varphi_i(x_i) \geq \varphi_j(x_j)\} \\ \leq\ & P_n\{|\varphi_{jn}(x_j) - \varphi_j(x_j)| > \frac{\varphi_i(x_i) - \varphi_j(x_j)}{2}\} + P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > \frac{\varphi_i(x_i) - \varphi_j(x_j)}{2}\}. \end{aligned}$$

From (2.2), (4.1) and by facts 1, 2 and 3, we get

$$\begin{aligned} & E_n[R(\underset{\sim}{d}^{*n})] - R(\underset{\sim}{d}^B) \\ =\ & E_n \int_{\mathcal{X}} [d_{i^*}^B(\underset{\sim}{x})\varphi_{i^*}(x_{i^*}) - d_{i_n^*}^{*n}(\underset{\sim}{x})\varphi_{i_n^*}(x_{i_n^*})] f(\underset{\sim}{x}) d\underset{\sim}{x} \\ =\ & \sum_{i=0}^{k}\sum_{j=0}^{k} E_n \int_{\mathcal{X}} \mathrm{I}_{\{i^*=i, i_n^*=j\}} [\varphi_i(x_i) - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \\ =\ & \sum_{i=0}^{k}\sum_{j=0}^{k} \int_{\mathcal{X}} P_n\{i^* = i, i_n^* = j\}[\varphi_i(x_i) - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \\ =\ & \sum_{i=1}^{k} \int_{\mathcal{X}} P_n\{i^* = i, i_n^* = 0\}[\varphi_i(x_i) - \theta_0] f(\underset{\sim}{x}) d\underset{\sim}{x} \\ & + \sum_{j=1}^{k} \int_{\mathcal{X}} P_n\{i^* = 0, i_n^* = j\}[\theta_0 - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \qquad (4.2) \\ & + \sum_{i=1}^{k}\sum_{j=1}^{k} \int_{\mathcal{X}} P_n\{i^* = i, i_n^* = j\}[\varphi_i(x_i) - \varphi_j(x_j)] f(\underset{\sim}{x}) d\underset{\sim}{x} \\ \leq\ & \sum_{i=1}^{k} \int_{R} P_n\{|\varphi_{in}(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\} |\varphi_i(x_i) - \theta_0| f_i(x_i) dx_i \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Bigg[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)| > \frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\} \\
&\qquad +P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)| > \frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\Bigg] \\
&\qquad |\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j \\
= \;& I_n + II_n.
\end{aligned}
$$

Recall that $\varphi_i(x_i) = \frac{x_i\tau_i^2+\frac{\sigma_i^2}{M}\mu_i}{\tau_i^2+\frac{\sigma_i^2}{M}}$ and X_i is marginally $N(\mu_i, v_i^2)$ distributed. Therefore, $\varphi_i(X_i)$ is $N(\mu_i, \frac{\tau_i^4}{v_i^2})$ distributed. For $\varepsilon_n > 0$, $i = 1, \ldots, k$ and $j = 1, \ldots, k$, let

$$
\begin{cases} \mathcal{X}_i = \{x_i |\; |\varphi_i(x_i)-\theta_0| \le \varepsilon_n\}, \\ \mathcal{X}_{ij} = \{(x_i, x_j) |\; |\varphi_i(x_i)-\varphi_j(x_j)| \le \varepsilon_n\}. \end{cases} \tag{4.3}
$$

Then,

$$
\begin{aligned}
I_n =\;& \sum_{i=1}^{k}\int_{\mathcal{X}_i} P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)| > |\varphi_i(x_i)-\theta_0|\}|\varphi_i(x_i)-\theta_0|f_i(x_i)dx_i \\
&+\sum_{i=1}^{k}\int_{R-\mathcal{X}_i} P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)| > |\varphi_i(x_i)-\theta_0|\}|\varphi_i(x_i)-\theta_0|f_i(x_i)dx_i \\
\le\;& \sum_{i=1}^{k}\int_{\mathcal{X}_i}\varepsilon_n f_i(x_i)dx_i \qquad\qquad (4.4)\\
&+\sum_{i=1}^{k}\int_{R} P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)| > \varepsilon_n\}|\varphi_i(x_i)-\theta_0|f_i(x_i)dx_i \\
\le\;& O(\varepsilon_n^2) \\
&+\sum_{i=1}^{k}\int_{R} P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)| > \frac{\varepsilon_n}{2}\}[|\varphi_i(x_i)-\mu_i| + |\mu_i-\theta_0|]f_i(x_i)dx_i,
\end{aligned}
$$

where

$$
\sum_{i=1}^{k}\int_{\mathcal{X}_i}\varepsilon_n f_i(x_i)dx_i = O(\varepsilon_n^2),
$$

since

$$
\int_{\{x_i|\;|\varphi_i(x_i)-\theta_0|\le\varepsilon_n\}} f_i(x_i)dx_i \le \frac{2v_i}{\sqrt{2\pi}\tau_i^2}\varepsilon_n, \quad i = 1, \ldots, k.
$$

Moreover, $\varphi_i(X_i) - \varphi_j(X_j)$ has a $N(\mu_i - \mu_j, \frac{\tau_i^4}{v_i^2} + \frac{\tau_j^4}{v_j^2})$ distribution. Therefore,

$$
\begin{aligned}
II_n =\;& \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}_{ij}}\Bigg[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)| > \frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\} \\
&+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)| > \frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\Bigg]\,|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2-\mathcal{X}_{ij}}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\\
&\quad+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\Big]\,|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j\\
\le\ & \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}_{ij}}2\varepsilon_n f_i(x_i)f_j(x_j)dx_idx_j\\
&+\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{\varepsilon_n}{2}\}\Big]\\
&\qquad|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j\\
\le\ & \sum_{i=1}^{k}\sum_{j=1}^{k}2\varepsilon_n\frac{1}{\sqrt{2\pi}}\frac{2\varepsilon_n}{\sqrt{\frac{\tau_i^4}{v_i^2}+\frac{\tau_j^4}{v_j^2}}} \qquad (4.5)\\
&+\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Big[P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}+P_n\{|\varphi_{jn}(x_j)-\varphi_j(x_j)|>\frac{\varepsilon_n}{2}\}\Big]\\
&\qquad[|\varphi_i(x_i)-\mu_i|+|\varphi_j(x_j)-\mu_j|+|\mu_i-\mu_j|]\,f_i(x_i)f_j(x_j)dx_idx_j.
\end{aligned}
$$

Since $\varphi_1(X_1),\varphi_2(X_2),\ldots,\varphi_k(X_k)$ are mutually independent and $E|\varphi_i(X_i)-\mu_i|<+\infty$, $i=1,\ldots,k$, also by (4.2), (4.4) and (4.5), it suffices to investigate the following two terms.

$$
\begin{cases}\int_R P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}f_i(x_i)dx_i,\\ \int_R P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}|\varphi_i(x_i)-\mu_i|f_i(x_i)dx_i.\end{cases} \qquad (4.6)
$$

Furthermore, by Lemma 4.2 and Lemma 4.3, we have

$$
\begin{aligned}
&P_n\{|\varphi_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}\\
\le\ & 2\Big[P_n\{|\mu_{in}-\mu_i|>\frac{Mv_i^2\varepsilon_n}{4\sigma_i^2}\} \qquad \text{(by lemma 4.3)}\\
&+P_n\{|\tau_{in}^2-\tau_i^2|>\frac{v_i^4\varepsilon_n}{4(\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2)}\}\Big]\\
\le\ & 2\Big[\frac{8\sigma_i^2}{\sqrt{2\pi}Mv_i}\frac{1}{\varepsilon_n\sqrt{n}}\exp(\frac{-M^2v_i^2}{32\sigma_i^4}\varepsilon_n^2 n) \qquad \text{(by lemma 4.2)}\\
&+\exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})) \qquad (4.7)\\
&+\exp(-\frac{n-1}{2}g_1(\frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}))\Big].
\end{aligned}
$$

Also, $|\varphi_i(x_i)-\mu_i|=\frac{\tau_i^2}{v_i^2}|x_i-\mu_i|$. Hence, if we let $\eta_n=\frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}$ then, from (4.6) and

(4.7), it suffices to consider the rate of convergence of the following terms.

$$\left\{\begin{array}{lcl} II_a & = & \frac{8\sigma_i^2}{\sqrt{2\pi}Mv_i}\frac{1}{\varepsilon_n\sqrt{n}}\exp(\frac{-M^2v_i^2}{32\sigma_i^4}\varepsilon_n^2 n)+\exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2})), \\ II_b & = & \int_R[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))]f_i(x_i)dx_i, \\ II_c & = & \int_R[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))]|x_i-\mu_i|f_i(x_i)dx_i. \end{array}\right. \tag{4.8}$$

First, we consider the term II_a. For $i=1,\ldots,k$, note that $0<\frac{\tau_i^2}{v_i^2}<1$, hence, $g_1(\frac{\tau_i^2}{v_i^2})>0$ and $g_2(\frac{\tau_i^2}{v_i^2})>0$, by the Remark 1 of Corollary 4.1. Therefore,

$$\exp(-\frac{n-1}{2}g_1(\frac{\tau_i^2}{v_i^2}))+\exp(-\frac{n-1}{2}g_2(\frac{\tau_i^2}{v_i^2}))\le O(\exp(-c_1n))$$

where $c_1=\frac{1}{2}\min_{1\le i\le k}\{g_1(\frac{\tau_i^2}{v_i^2}),g_2(\frac{\tau_i^2}{v_i^2})\}$. In the sequel, we let $\varepsilon_n=\frac{\ln n}{\sqrt{cn}}$, where $c=\min_{1\le i\le k}\{\frac{M^2v_i^2}{1024\sigma_i^4}\}$. Then,

$$\frac{1}{\varepsilon_n\sqrt{n}}\exp(\frac{-M^2v_i^2}{32\sigma_i^4}\varepsilon_n^2 n)\le O(\frac{1}{n\ln n}).$$

Thus, from the above argument and (4.8), we have

$$II_a\le O(\frac{1}{n\ln n}). \tag{4.9}$$

Now, let us investigate the rate of convergence of II_b. For the same ε_n, we divide the integration of II_b into two parts by the set $\{|x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}$ and its complement. By Remark 1 and Remark 2 of Corollary 4.1 and for n sufficiently large, we have

$$\begin{aligned} & |x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}} \\ & \Rightarrow \eta_n=\frac{1}{2}\frac{\frac{\varepsilon_n}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon_n}{2}v_i^2}>\frac{1}{2}\frac{1}{\sqrt{\frac{n}{128\ln n}}+1} \\ & \Rightarrow \eta_n>\frac{1}{4}\frac{1}{\sqrt{\frac{n}{128\ln n}}}\Rightarrow g_1(\eta_n)>g_1(\frac{1}{4}\sqrt{\frac{128\ln n}{n}}) \\ & \Rightarrow \exp(-\frac{n-1}{2}g_1(\eta_n))<\exp(-\frac{n-1}{2}g_1(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})) \\ & \qquad\qquad\qquad\qquad \le\exp\left(-\frac{n-1}{2}(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})^2\frac{g_1(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})}{(\frac{1}{4}\sqrt{\frac{128\ln n}{n}})^2}\right) \\ & \qquad\qquad\qquad\qquad =O(\frac{1}{n}). \end{aligned} \tag{4.10}$$

Similarly,

$$|x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\Rightarrow\exp(-\frac{n-1}{2}g_2(\eta_n))\le O(\frac{1}{n}). \tag{4.11}$$

Therefore,

$$\int_{\{|x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}}\left[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))\right]f_i(x_i)dx_i$$
$$\leq O(\frac{1}{n}). \tag{4.12}$$

Now, by using a similar argument as in the proof of Lemma 4.2(a), we have

$$EI_{\{|X_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} = P\{\frac{|X_i-\mu_i|}{v_i}\geq\frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}}\}$$
$$\leq 2\frac{1}{\frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}}}\frac{\exp(-\frac{1}{2}(\frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}})^2)}{\sqrt{2\pi}}$$
$$\leq O(\frac{1}{n\sqrt{\ln n}}).$$

Moreover, observe that $0<\eta_n<\frac{1}{2}$, this implies that $g_1(\eta_n)>0$ and $g_2(\eta_n)>0$. Hence,

$$\int_{\{|x_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}}\left[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))\right]f_i(x_i)dx_i$$
$$\leq 2EI_{\{|X_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} \tag{4.13}$$
$$\leq O(\frac{1}{n\sqrt{\ln n}}).$$

From (4.8), (4.12) and (4.13), we get

$$II_b\leq O(\frac{1}{n}). \tag{4.14}$$

Again, for the same ε_n, we divide II_c into two parts:

$$II_c = II_{c,1}+II_{c,2}, \tag{4.15}$$

where

$$II_{c,1} = \int_{\{|x_i-\mu_i|<\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}}\left[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))\right]|x_i-\mu_i|f_i(x_i)dx_i,$$
$$II_{c,2} = \int_{\{|x_i-\mu_i|\geq\frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}}\left[\exp(-\frac{n-1}{2}g_1(\eta_n))+\exp(-\frac{n-1}{2}g_2(\eta_n))\right]|x_i-\mu_i|f_i(x_i)dx_i.$$

By $(4.10), (4.11)$ and $E|X_i - \mu_i| < +\infty$, we have

$$II_{c,1} \leq O(\frac{1}{n}). \tag{4.16}$$

Also, recall that $g_1(\eta_n) > 0$ and $g_2(\eta_n) > 0$, then

$$\begin{aligned} II_{c,2} &\leq 2\int_{\{|x_i-\mu_i| \geq \frac{Mv_i^2}{2\sigma_i^2}\varepsilon_n\sqrt{\frac{n}{128\ln n}}\}} |x_i - \mu_i| f_i(x_i) dx_i \\ &\leq 2v_i \int_{\{|z| \geq \frac{Mv_i}{2\sigma_i^2}\sqrt{\frac{\ln n}{128c}}\}} |z| d\Phi(z) \\ &\leq \frac{4v_i}{\sqrt{2\pi}} \exp(-\frac{M^2 v_i^2}{8\sigma_i^4}\frac{\ln n}{128c}) \\ &\leq O(\frac{1}{n}), \end{aligned} \tag{4.17}$$

where $\Phi(z)$ is the c.d.f. of the standard normal distribution.
Hence, from $(4.15) - (4.17)$,

$$II_c \leq O(\frac{1}{n}). \tag{4.18}$$

Therefore, from $(4.4) - (4.6), (4.8), (4.9), (4.14), (4.18)$ and for the same ε_n, we have

$$I_n \leq O(\varepsilon_n^2) = O(\frac{(\ln n)^2}{n}), \tag{4.19}$$

$$II_n \leq O(\varepsilon_n^2) = O(\frac{(\ln n)^2}{n}). \tag{4.20}$$

By combining $(4.2), (4.19)$ and (4.20), we have proved the following theorem.

Theorem 4.1 The empirical Bayes selection rule $\underset{\sim}{d}^{*n}(\underset{\sim}{x})$, defined in (3.6), is asymptotically optimal with convergence rate of order $O(\frac{(\ln n)^2}{n})$. That is, $E_n[R_n(\underset{\sim}{d}^{*n})] - R(\underset{\sim}{d}^B) \leq O(\frac{(\ln n)^2}{n})$.

4.2 Case 2: (μ_i, τ_i^2) and σ_i^2 unknown, $i = 1, \cdots, k$.

Lemma 4.5 Let $\hat{\mu}_{in}, \hat{\sigma}_{in}^2$ and $\hat{v}_{in}^2$ be the estimators of μ_i, σ_i^2 and v_i^2, respectively, as defined in (3.9). Also, let $g_1(\eta)$ and $g_2(\eta)$ be the functions defined in Corollary 4.1. Then, for any $c > 0$, we have

$$\begin{aligned} (a)\ P_n\{|\hat{\mu}_{in} - \mu_i| \geq c\} &\leq \frac{2v_i}{\sqrt{2\pi}c}\frac{1}{\sqrt{n}}\exp(\frac{-c^2}{2v_i^2}n), \\ (b)\ P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| \geq c\} &\leq \exp(-\frac{n(M-1)}{2}g_1(\frac{c}{\sigma_i^2})) + \exp(-\frac{n(M-1)}{2}g_2(\frac{c}{\sigma_i^2})), \\ (c)\ P_n\{|\hat{v}_{in}^2 - v_i^2| \geq c\} &\leq \exp(-\frac{n-1}{2}g_1(\frac{c}{v_i^2})) + \exp(-\frac{n-1}{2}g_2(\frac{c}{v_i^2})). \end{aligned}$$

Proof : (a) The proof is the same as in Lemma 4.2(a).

(b)

$$\begin{aligned}
& P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| \geq c\} \\
= \; & P_n\{|n(M-1)\frac{\hat{\sigma}_{in}^2}{\sigma_i^2} - n(M-1)| \geq n(M-1)\frac{c}{\sigma_i^2}\} \\
\leq \; & \exp(-\frac{n(M-1)}{2} g_1(\frac{c}{\sigma_i^2})) + \exp(-\frac{n(M-1)}{2} g_2(\frac{c}{\sigma_i^2})).
\end{aligned}$$

The last inequality follows by Corollary 4.1 and the fact that $n(M-1)\frac{\hat{\sigma}_{in}^2}{\sigma_i^2}$ has a $\chi^2(n(M-1))$ distribution.

(c) The proof is similar to that of (b), hence, we omit it. □

Lemma 4.6 Let $\varphi_i(x_i)$ and $\hat{\varphi}_{in}(x_i)$ be defined as in (2.3) and (3.10), respectively. Then, for any $\varepsilon > 0$, any $\kappa > 0$ and any $x_i \in R$, we have

$$\begin{aligned}
(a) \quad & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon\} \\
\leq \; & P_n\{|\hat{\mu}_{in} - \mu_i| > \kappa\} + P_n\{|\hat{\mu}_{in} - \mu_i| > \frac{M v_i^2 \varepsilon}{5\sigma_i^2}\} + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\} \\
& + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M v_i^2 \varepsilon}{5\kappa}\} + 2P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{\sigma_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\} \\
& + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\}, \\
(b) \quad & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) < -\varepsilon\} \\
\leq \; & P_n\{|\hat{\mu}_{in} - \mu_i| > \kappa\} + P_n\{|\hat{\mu}_{in} - \mu_i| > \frac{M v_i^2 \varepsilon}{5\sigma_i^2}\} + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\} \\
& + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M v_i^2 \varepsilon}{5\kappa}\} + 2P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{\sigma_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\} \\
& + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i - \mu_i| + \varepsilon v_i^2}\}.
\end{aligned}$$

Proof : We prove (a) only. The proof of (b) is similar to that of (a). Let $a = x_i, b = \frac{\sigma_i^2}{M}, y = \tau_i^2, z = \mu_i, b_n = \frac{\hat{\sigma}_{in}^2}{M}, y_n = \hat{\tau}_{in}^2$ and $z_n = \hat{\mu}_{in}$. Therefore, $y + b = v_i^2$ and $y_n + b_n = \hat{v}_{in}^2$ if $\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} \geq 0$. Therefore,

$$\begin{aligned}
& P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon\} \\
\leq \; & P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon, \hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} \geq 0\} + P_n\{\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} < 0\},
\end{aligned}$$

where

$$P_n\{\hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} < 0\}$$

$$
\begin{aligned}
&= P_n\{(\hat{v}_{in}^2 - v_i^2) - (\frac{\hat{\sigma}_{in}^2}{M} - \frac{\sigma_i^2}{M}) < -\tau_i^2\} \\
&\leq P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} + P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\}
\end{aligned}
$$

and

$$
\begin{aligned}
& P_n\{\hat{\varphi}_{in}(x_i) - \varphi_i(x_i) > \varepsilon, \hat{v}_{in}^2 - \frac{\hat{\sigma}_{in}^2}{M} \geq 0\} \\
\leq\ & P_n\{v_i^2(z_n - z)(b_n - b) - (a-z)v_i^2(b_n - b) + v_i^2 b(z_n - z) + [(a-z)b - \varepsilon v_i^2](\hat{v}_{in}^2 - v_i^2) > \varepsilon v_i^4\} \\
=\ & P_n\{v_i^2(z_n - z)(b_n - b) - [(a-z)b + \varepsilon v_i^2]\frac{v_i^2}{b}(b_n - b) + \frac{\varepsilon v_i^4}{b}(b_n - b) \\
& \quad + v_i^2 b(z_n - z) + [(a-z)b - \varepsilon v_i^2](\hat{v}_{in}^2 - v_i^2) > \varepsilon v_i^4\} \\
\leq\ & P_n\{v_i^2|z_n - z||b_n - b| + (|a-z|b + \varepsilon v_i^2)\frac{v_i^2}{b}|b_n - b| \\
& + \frac{\varepsilon v_i^4}{b}|b_n - b| + v_i^2 b|z_n - z| + (|a-z|b + \varepsilon v_i^2)|\hat{v}_{in}^2 - v_i^2| > \varepsilon v_i^4\} \\
\leq\ & P_n\{|z_n - z||b_n - b| > \frac{\varepsilon v_i^2}{5}\} \\
& + P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} + P_n\{|b_n - b| > \frac{b}{5}\} \\
& + P_n\{|z_n - z| > \frac{\varepsilon v_i^2}{5b}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} \\
=\ & P_n\{|z_n - z| \geq \kappa,\ |z_n - z||b_n - b| > \frac{\varepsilon v_i^2}{5}\} + P_n\{|z_n - z| < \kappa,\ |z_n - z||b_n - b| > \frac{\varepsilon v_i^2}{5}\} \\
& + P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} + P_n\{|b_n - b| > \frac{b}{5}\} \\
& + P_n\{|z_n - z| > \frac{\varepsilon v_i^2}{5b}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} \\
\leq\ & P_n\{|z_n - z| \geq \kappa\} + P_n\{|b_n - b| > \frac{v_i^2\varepsilon}{5\kappa}\} \\
& + P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} + P_n\{|b_n - b| > \frac{b}{5}\} \\
& + P_n\{|z_n - z| > \frac{v_i^2\varepsilon}{5b}\} + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} \\
\leq\ & P_n\{|z_n - z| \geq \kappa\} + P_n\{|z_n - z| > \frac{v_i^2\varepsilon}{5b}\} \\
& + P_n\{|b_n - b| > \frac{v_i^2\varepsilon}{5\kappa}\} + 2P_n\{|b_n - b| > \frac{b}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\} \\
& + P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{v_i^2}{5}\frac{\varepsilon v_i^2}{b|a-z| + \varepsilon v_i^2}\}
\end{aligned}
$$

$$
\begin{aligned}
\leq\ & P_n\{|\hat{\mu}_{in}-\mu_i|\geq\kappa\}+P_n\{|\hat{\mu}_{in}-\mu_i|>\frac{Mv_i^2\varepsilon}{5\sigma_i^2}\} \\
&+P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{Mv_i^2\varepsilon}{5\kappa}\}+2P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{\sigma_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2}\} \\
&+P_n\{|\hat{v}_{in}^2-v_i^2|>\frac{v_i^2}{5}\frac{\varepsilon v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\varepsilon v_i^2}\}.
\end{aligned}
$$

Hence, the result follows. □

Let $\{\underset{\sim}{\hat{d}}^n\}_{n=1}^{\infty}$ be the empirical Bayes rules defined in (3.12). Then,

$$
E_n[R(\underset{\sim}{\hat{d}}^n)]-R(\underset{\sim}{d}^B)\leq\hat{I}_n+\hat{I}I_n. \tag{4.21}
$$

where

$$
\begin{aligned}
\hat{I}_n &= \sum_{i=1}^{k}\int_R P_n\{|\hat{\varphi}_{in}(x_i)-\varphi_i(x_i)|>|\varphi_i(x_i)-\theta_0|\}|\varphi_i(x_i)-\theta_0|f_i(x_i)dx_i, \\
\hat{I}I_n &= \sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2}\Big[P_n\{|\hat{\varphi}_{in}(x_i)-\varphi_i(x_i)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\} \\
&\qquad +P_n\{|\hat{\varphi}_{jn}(x_j)-\varphi_j(x_j)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}\Big] \\
&\qquad |\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j.
\end{aligned}
$$

By an argument similar to that of (4.4) and (4.5), it suffices to investigate the following two terms.

$$
\begin{cases}
\int_R P_n\{|\hat{\varphi}_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}f_i(x_i)dx_i, \\
\int_R P_n\{|\hat{\varphi}_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon_n}{2}\}|\varphi_i(x_i)-\mu_i|f_i(x_i)dx_i.
\end{cases} \tag{4.22}
$$

Moreover, by Lemma 4.6, we have

$$
\begin{aligned}
& P_n\{|\hat{\varphi}_{in}(x_i)-\varphi_i(x_i)|>\frac{\varepsilon}{2}\} \\
\leq\ & 2\Big[P_n\{|\hat{\mu}_{in}-\mu_i|>\kappa\}+P_n\{|\hat{\mu}_{in}-\mu_i|>\frac{Mv_i^2\frac{\varepsilon}{2}}{5\sigma_i^2}\}+P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{M\tau_i^2}{2}\} \\
&+P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{Mv_i^2\frac{\varepsilon}{2}}{5\kappa}\}+2P_n\{|\hat{\sigma}_{in}^2-\sigma_i^2|>\frac{\sigma_i^2}{5}\frac{\frac{\varepsilon}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon}{2}v_i^2}\} \\
&+P_n\{|\hat{v}_{in}^2-v_i^2|>\frac{\tau_i^2}{2}\}+P_n\{|\hat{v}_{in}^2-v_i^2|>\frac{v_i^2}{5}\frac{\frac{\varepsilon}{2}v_i^2}{\frac{\sigma_i^2}{M}|x_i-\mu_i|+\frac{\varepsilon}{2}v_i^2}\}\Big].
\end{aligned} \tag{4.23}
$$

Let $\varepsilon=\varepsilon_n=\frac{\ln n}{\sqrt{c_* n}}$ and $\kappa\equiv\kappa_n=\sqrt{c_\kappa\ln n}$, where $c_*=\min_{1\leq i\leq k}\{\frac{M^2v_i^2}{6400\sigma_i^4}\}$ and $c_\kappa=\min_{1\leq i\leq k}\{4v_i\}$. Then, by using Lemma 4.5 and Remark 1 and Remark 2 of Corollary 4.1, the two terms concerning κ in (4.23) have the following convergence rate.

$$P_n\{|\hat{\mu}_{in} - \mu_i| > \kappa\} \leq O(\frac{1}{\sqrt{n \ln n}} \exp(-c_\kappa(\max_{1\leq j\leq k}\{2v_i^2\})^{-1} n \ln n),$$

$$P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{Mv_i^2 \frac{\varepsilon}{2}}{5\kappa}\} \leq O(\frac{1}{n}).$$

Again, by Lemma 4.5, we get

$$P_n\{|\hat{\sigma}_{in}^2 - \sigma_i^2| > \frac{M\tau_i^2}{2}\} \leq O\left(\exp(-\frac{M-1}{2}\min_{1\leq i\leq k}\{g_1(\frac{M\tau_i^2}{2\sigma_i^2}), g_2(\frac{M\tau_i^2}{2\sigma_i^2})\}n)\right),$$

$$P_n\{|\hat{v}_{in}^2 - v_i^2| > \frac{\tau_i^2}{2}\} \leq O\left(\exp(-\frac{1}{2}\min_{1\leq i\leq k}\{g_1(\frac{\tau_i^2}{2v_i^2}), g_2(\frac{\tau_i^2}{2v_i^2})\}n)\right).$$

Now, by a proof of the rate of convergence analogous to that of (4.6), it can be shown that the two terms in (4.22) have a rate of convergence of order $O(\frac{(\ln n)^2}{n})$.

Hence, by the above argument, (4.21) and (4.22), we have the following theorem.

Theorem 4.2 The empirical Bayes selection rule $\hat{\underset{\sim}{d}}^n(\underset{\sim}{x})$, defined in (3.12), is asymptotically optimal with convergence rate of order $O(\frac{(\ln n)^2}{n})$. That is, $E_n[R_n(\hat{\underset{\sim}{d}}^n)] - R(\underset{\sim}{d}^B) \leq O(\frac{(\ln n)^2}{n})$.

5. Small Sample Performance: Simulation Study

We carried out a simulation study to investigate the performance of the empirical Bayes selection rules $\underset{\sim}{d}^{*n}(\underset{\sim}{x})$ and $\hat{\underset{\sim}{d}}^n(\underset{\sim}{x})$ defined in Sections 3.1 and 3.2, respectively. We considered $k = 3$ populations π_1, π_2 and π_3. Recall that E and E_n are the expectations taken with respect to the probability measures generated by the current observation $\underset{\sim}{X}$ and the past observation X_{ijl} $(i = 1, \ldots, k,\ j = 1, \ldots, M$ and $l = 1, \ldots, n)$, respectively. In definition 4.1 $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$ is used as a measure of the performance of the empirical Bayes rule $\underset{\sim}{d}^n$. For any given current observation $\underset{\sim}{X}$ and any given past observation X_{ijl} $(i = 1, \ldots, k,\ j = 1, \ldots, M$ and $l = 1, \ldots, n)$, let

$$D^n(\underset{\sim}{X}) = \sum_{i=0}^{k}[d_i^B(\underset{\sim}{X}) - d_i^n(\underset{\sim}{X})]\varphi_i(X_i).$$

Then, from (4.3)

$$E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B) = EE_nD^n(\underset{\sim}{X}).$$

Therefore, by the law of large numbers, the sample mean of $D^n(\underset{\sim}{X})$, based on the observations of $\underset{\sim}{X}$ and X_{ijl} $(i = 1, \ldots, k,\ j = 1, \ldots, M$ and $l = 1, \ldots, n)$, can be used as an estimator of $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$.

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

(1) For each $l = 1, \ldots, n$ and for each $i = 1, 2$ and 3, generate the independent past observations $X_{i1l}, \ldots, X_{iMl}$ by the following :

(a) Generate Θ_{il} from a $N(\mu_i, \tau_i^2)$ prior distribution.
(b) Generate random sample $X_{i1l}, X_{i2l}, \ldots, X_{iMl}$ from a $N(\theta_{il}, \sigma_i^2)$ distribution.

(2) Generate the current observation $\underset{\sim}{X} = (X_1, \ldots, X_k)$, where X_i has a $N(\mu_i, \frac{\sigma_i^2}{M} + \tau_i^2)$ distribution and $X_1, \ldots, X_k$ are independent.

(3) Based on the past observation X_{ijl} ($i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$) and the current observation $\underset{\sim}{X}$, construct the Bayes rule $\underset{\sim}{d}^B$ and the empirical Bayes rule $\underset{\sim}{d}^n$ and compute $D^n(\underset{\sim}{X})$.

(4) Steps (1), (2) and (3) were repeated 2000 times. The average of $D^n(\underset{\sim}{X})$ based on the 2000 repetitions, which is denoted by $\bar{D}^n$, is used as an estimator of $E_n[R(\underset{\sim}{d}^n)] - R(\underset{\sim}{d}^B)$. Also, $SE(\bar{D}^n)$, the estimated standard error, and $n\bar{D}^n$ are computed.

It should be mentioned that the same past observation X_{ijl} ($i = 1, \ldots, k$, $j = 1, \ldots, M$ and $l = 1, \ldots, n$) and the current observation $\underset{\sim}{X}$ were used for both rules $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$. Also, the term $\bar{D}^n$ corresponding to $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$ are denoted by $\bar{D}^{*n}$ and $\hat{\bar{D}}^n$, respectively.

Tables 1, 2, 3 and 4 list some simulation results on the performance of the proposed empirical Bayes rules $\underset{\sim}{d}^{*n}$ and $\hat{\underset{\sim}{d}}^n$, for the case where $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1.0, \tau_1^2 = 1.0, \tau_2^2 = 2.0, \tau_3^2 = 3.0, \theta_0 = 6.0$ and $M = 3$.

From the tables, we learn that the values of $\bar{D}^n$ decrease quite rapidly as n increases, for $n \leq 100$. In tables 3 and 4, the value of $\bar{D}^n$ are almost all 0 when $n \geq 650$ and $n \geq 800$, respectively. Observe that the distances between the μ_i's are 2 in Tables 1 and 2 ($\mu_1 = 3.0, \mu_2 = 5.0, \mu_3 = 7.0$) and those in Tables 3 and 4 are 4 ($\mu_1 = 0.0, \mu_2 = 4.0, \mu_3 = 8.0$). 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. Also, the simulation results indicate that the values of $n\bar{D}^n$ are decreasing as well as oscillating as n increases. This supports Theorem 4.1 and Theorem 4.2 that the rate of convergence is at least of order $O(\frac{(\ln n)^2}{n})$.

Table 1. Performance of $\underset{\sim}{d}^{*n}$ for $\mu_1 = 3.0, \mu_2 = 5.0$ and $\mu_3 = 7.0$

n	$\bar{D}^{*n}$	$n\bar{D}^{*n}$	$SE(\bar{D}^{*n})$
20	309.3678×10^{-5}	61.87×10^{-3}	75.8224×10^{-5}
40	85.3516×10^{-5}	34.14×10^{-3}	29.9093×10^{-5}
60	60.0714×10^{-5}	36.04×10^{-3}	14.8358×10^{-5}
80	51.3486×10^{-5}	41.07×10^{-3}	17.7981×10^{-5}
100	35.1271×10^{-5}	35.12×10^{-3}	15.6088×10^{-5}
120	19.5562×10^{-5}	23.46×10^{-3}	6.4159×10^{-5}
140	16.6746×10^{-5}	23.34×10^{-3}	5.8600×10^{-5}
160	20.1251×10^{-5}	32.20×10^{-3}	6.5213×10^{-5}
180	17.6404×10^{-5}	31.75×10^{-3}	5.9192×10^{-5}
200	11.3668×10^{-5}	22.73×10^{-3}	4.3587×10^{-5}
250	10.4540×10^{-5}	26.13×10^{-3}	4.2630×10^{-5}
300	8.9227×10^{-5}	26.76×10^{-3}	3.3842×10^{-5}
350	4.3252×10^{-5}	15.13×10^{-3}	2.2729×10^{-5}
400	4.8568×10^{-5}	19.42×10^{-3}	2.4191×10^{-5}
450	5.2380×10^{-5}	23.57×10^{-3}	2.4485×10^{-5}
500	3.3443×10^{-5}	16.72×10^{-3}	1.5542×10^{-5}
550	3.6253×10^{-5}	19.93×10^{-3}	1.7340×10^{-5}
600	3.7534×10^{-5}	22.52×10^{-3}	1.7955×10^{-5}
650	2.5596×10^{-5}	16.63×10^{-3}	1.3424×10^{-5}
700	3.3443×10^{-5}	23.41×10^{-3}	1.5542×10^{-5}
750	3.3662×10^{-5}	25.24×10^{-3}	1.5543×10^{-5}
800	3.3443×10^{-5}	26.75×10^{-3}	1.5542×10^{-5}
850	3.3443×10^{-5}	28.42×10^{-3}	1.5542×10^{-5}
900	2.1784×10^{-5}	19.60×10^{-3}	1.2874×10^{-5}
950	2.4315×10^{-5}	23.09×10^{-3}	1.2588×10^{-5}
1000	1.9673×10^{-5}	19.67×10^{-3}	1.1705×10^{-5}

Table 2. Performance of $\hat{\underset{\sim}{d}}^n$ for $\mu_1 = 3.0, \mu_2 = 5.0$ and $\mu_3 = 7.0$

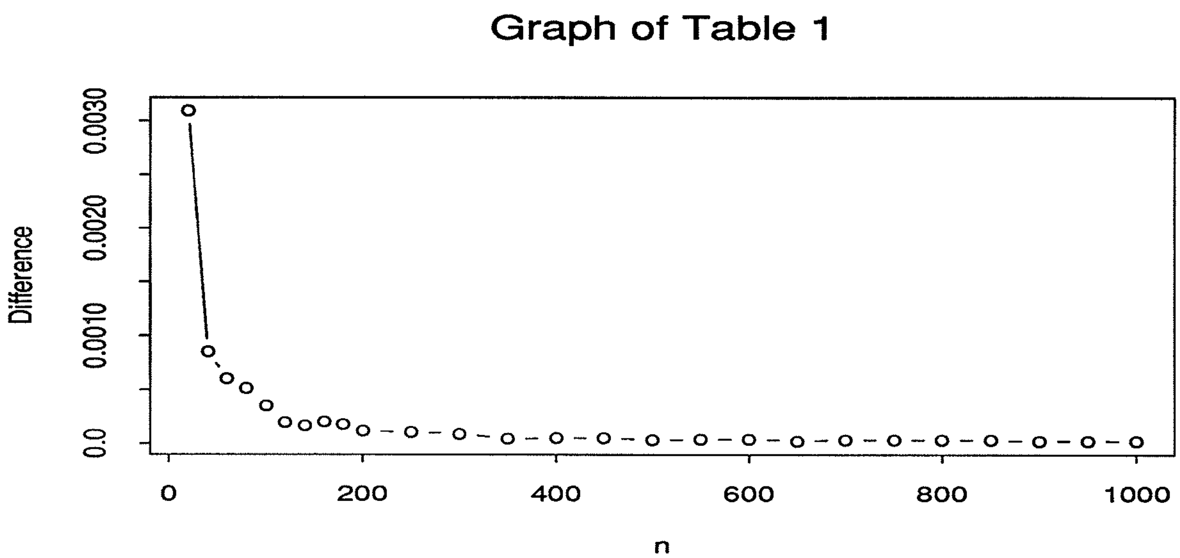
n	$\hat{\hat{D}}^n$	$n\hat{\hat{D}}^n$	$SE(\hat{\hat{D}}^n)$
20	345.6666×10^{-5}	69.13×10^{-3}	81.7556×10^{-5}
40	97.6678×10^{-5}	39.06×10^{-3}	32.7335×10^{-5}
60	91.4959×10^{-5}	54.89×10^{-3}	23.1381×10^{-5}
80	57.5668×10^{-5}	46.05×10^{-3}	18.6208×10^{-5}
100	50.5301×10^{-5}	50.53×10^{-3}	17.8186×10^{-5}
120	16.0868×10^{-5}	19.30×10^{-3}	5.7544×10^{-5}
140	17.3240×10^{-5}	24.25×10^{-3}	6.3963×10^{-5}
160	24.5363×10^{-5}	39.25×10^{-3}	7.9382×10^{-5}
180	17.0878×10^{-5}	30.75×10^{-3}	5.5962×10^{-5}
200	10.8352×10^{-5}	21.67×10^{-3}	4.2796×10^{-5}
250	14.9260×10^{-5}	37.31×10^{-3}	6.3393×10^{-5}
300	9.6272×10^{-5}	28.88×10^{-3}	3.4559×10^{-5}
350	9.9422×10^{-5}	34.79×10^{-3}	3.4691×10^{-5}
400	4.4898×10^{-5}	17.95×10^{-3}	2.1748×10^{-5}
450	6.3836×10^{-5}	28.72×10^{-3}	2.8822×10^{-5}
500	2.3513×10^{-5}	11.75×10^{-3}	1.2105×10^{-5}
550	3.5451×10^{-5}	19.49×10^{-3}	1.6993×10^{-5}
600	4.8291×10^{-5}	28.97×10^{-3}	2.0490×10^{-5}
650	4.4580×10^{-5}	28.97×10^{-3}	1.9281×10^{-5}
700	3.2860×10^{-5}	23.00×10^{-3}	1.5156×10^{-5}
750	2.5815×10^{-5}	19.36×10^{-3}	1.3425×10^{-5}
800	2.5815×10^{-5}	20.65×10^{-3}	1.3425×10^{-5}
850	2.9308×10^{-5}	24.91×10^{-3}	1.5111×10^{-5}
900	1.6687×10^{-5}	15.01×10^{-3}	0.9852×10^{-5}
950	1.8872×10^{-5}	17.92×10^{-3}	1.1184×10^{-5}
1000	1.8872×10^{-5}	18.87×10^{-3}	1.1184×10^{-5}

Table 3. Performance of $\underset{\sim}{d}^{*n}$ for $\mu_1 = 0.0, \mu_2 = 4.0$ and $\mu_3 = 8.0$

n	$\bar{D}^{*n}$	$n\bar{D}^{*n}$	$SE(\bar{D}^{*n})$
20	210.8993×10^{-5}	42.17×10^{-3}	53.4561×10^{-5}
40	88.0526×10^{-5}	35.22×10^{-3}	23.6048×10^{-5}
60	39.0124×10^{-5}	23.40×10^{-3}	13.7470×10^{-5}
80	39.1506×10^{-5}	31.32×10^{-3}	13.9124×10^{-5}
100	22.3508×10^{-5}	22.35×10^{-3}	9.3681×10^{-5}
120	27.0754×10^{-5}	32.49×10^{-3}	11.3922×10^{-5}
140	14.8855×10^{-5}	20.83×10^{-3}	6.4664×10^{-5}
160	17.2913×10^{-5}	27.66×10^{-3}	6.8967×10^{-5}
180	14.0296×10^{-5}	25.25×10^{-3}	6.2722×10^{-5}
200	14.0296×10^{-5}	28.05×10^{-3}	6.2722×10^{-5}
250	12.6434×10^{-5}	31.60×10^{-3}	5.9333×10^{-5}
300	20.0851×10^{-5}	60.25×10^{-3}	10.6554×10^{-5}
350	10.7009×10^{-5}	37.45×10^{-3}	5.7736×10^{-5}
400	7.4840×10^{-5}	29.96×10^{-3}	4.1339×10^{-5}
450	3.3130×10^{-5}	14.90×10^{-3}	1.9359×10^{-5}
500	2.3934×10^{-5}	11.96×10^{-3}	1.7041×10^{-5}
550	2.3934×10^{-5}	13.16×10^{-3}	1.7041×10^{-5}
600	1.3404×10^{-5}	8.04×10^{-3}	1.3404×10^{-5}
650	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
700	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
750	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
800	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
850	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
900	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
950	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
1000	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}

Table 4. Performance of $\hat{\underset{\sim}{d}}^n$ for $\mu_1 = 0.0, \mu_2 = 4.0$ and $\mu_3 = 8.0$

n	$\hat{\hat{D}}^n$	$n\hat{\hat{D}}^n$	$SE(\hat{\hat{D}}^n)$
20	232.8929×10^{-5}	46.57×10^{-3}	55.3256×10^{-5}
40	151.5610×10^{-5}	60.62×10^{-3}	42.0964×10^{-5}
60	46.8879×10^{-5}	28.13×10^{-3}	17.4218×10^{-5}
80	38.8236×10^{-5}	31.05×10^{-3}	14.8289×10^{-5}
100	32.0151×10^{-5}	32.01×10^{-3}	12.2051×10^{-5}
120	51.0811×10^{-5}	61.29×10^{-3}	17.5057×10^{-5}
140	26.3587×10^{-5}	36.90×10^{-3}	11.1613×10^{-5}
160	14.6873×10^{-5}	23.49×10^{-3}	6.6145×10^{-5}
180	14.6873×10^{-5}	26.43×10^{-3}	6.6145×10^{-5}
200	11.0007×10^{-5}	22.00×10^{-3}	5.4955×10^{-5}
250	14.5926×10^{-5}	36.48×10^{-3}	6.2433×10^{-5}
300	11.7237×10^{-5}	35.17×10^{-3}	5.8626×10^{-5}
350	10.7009×10^{-5}	37.45×10^{-3}	5.7736×10^{-5}
400	11.6206×10^{-5}	46.48×10^{-3}	5.8454×10^{-5}
450	8.0087×10^{-5}	36.03×10^{-3}	4.5993×10^{-5}
500	8.0087×10^{-5}	40.04×10^{-3}	4.5993×10^{-5}
550	9.9906×10^{-5}	54.94×10^{-3}	4.9867×10^{-5}
600	4.6979×10^{-5}	28.18×10^{-3}	2.5449×10^{-5}
650	3.2191×10^{-5}	20.92×10^{-3}	2.3073×10^{-5}
700	1.3404×10^{-5}	9.38×10^{-3}	1.3404×10^{-5}
750	1.9491×10^{-5}	14.61×10^{-3}	1.9491×10^{-5}
800	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
850	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
900	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
950	0.0000×10^{-5}	0.00×10^{-3}	0.0000×10^{-5}
1000	0.5590×10^{-5}	5.59×10^{-3}	0.5590×10^{-5}



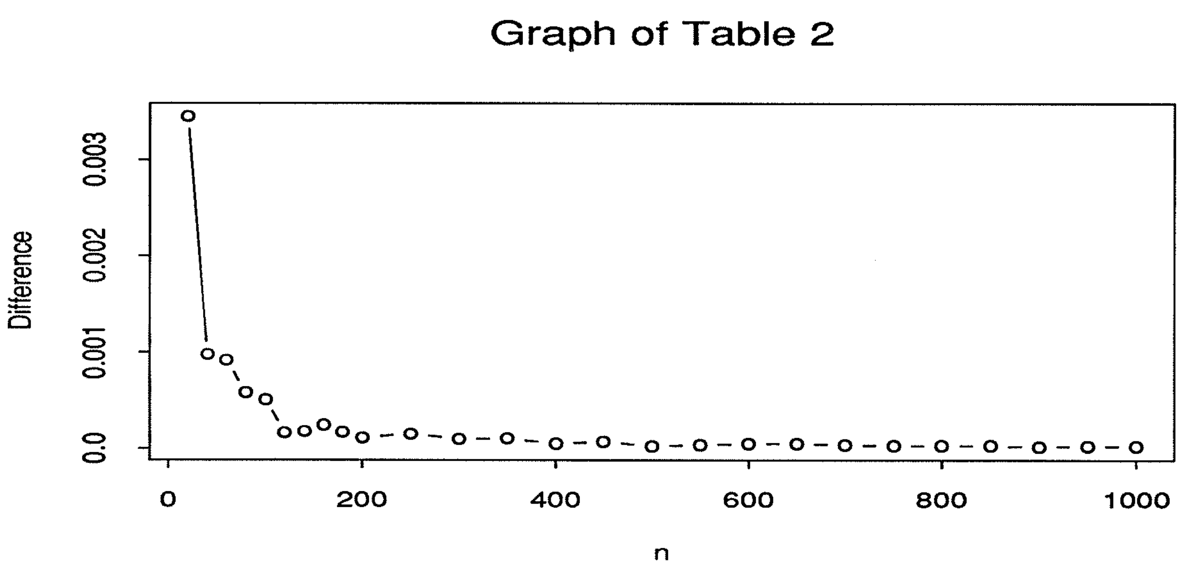


Figure 1: $\bar{D}^{*n}$ **vs** n **and** $\hat{\bar{D}}^{n}$ **vs** n **for Table 1 and Table 2**

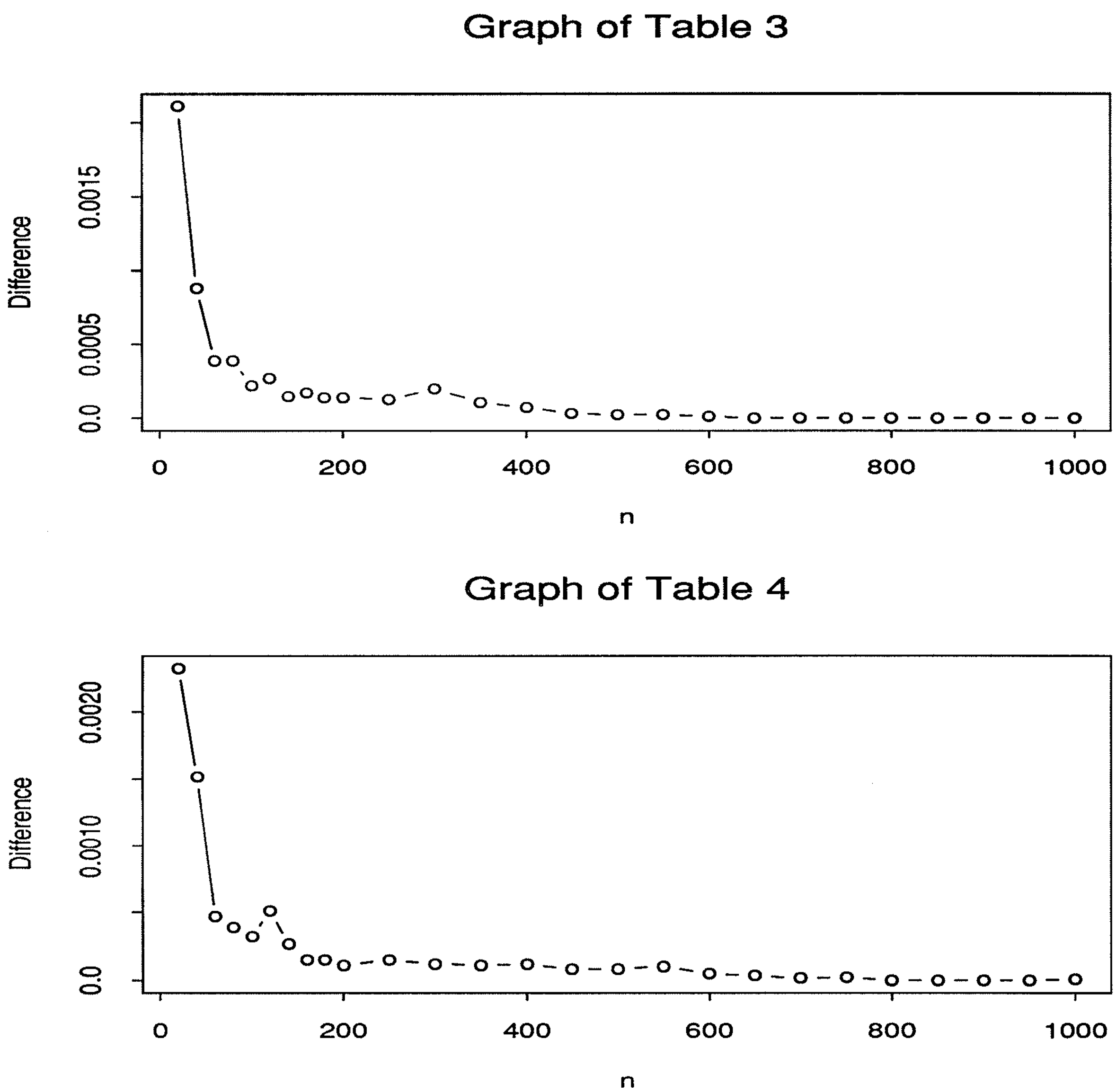


Figure 2: $\bar{D}^{*n}$ vs n and $\hat{\hat{D}}^{n}$ vs n **for Table 3 and Table 4**

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 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.

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.

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

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 Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Inference*, **8**, 87-101.

Gupta, S.S. and Leu, L.Y. (1991). On Bayes and empirical Bayes rules for selecting good populations. In *Proceedings of the 1990 Taipei Symposium in Statistics,* (Eds. M.T. Chao and P.E. Cheng), Institute of Statistical Science, Academia Sinica, Taipei, 167-182.

Gupta, S. S. and Liang, T. (1988). Empirical Bayes rules for selecting the best binomial population. *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 Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.*, **29**, 235–244.

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

Pollard, David (1984). *Convergence of Stochastic Processes.* Springer-Verlag.

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

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

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.