Ann. Inst. Statist. Math.
Vol. 51, No. 2, 281–299 (1999)

EMPIRICAL BAYES PROCEDURES FOR SELECTING THE BEST POPULATION WITH MULTIPLE CRITERIA*

WEN-TAO HUANG[1] AND YAO-TSUNG LAI[2]**

[1] *Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan, R.O.C.*
[2] *Graduate Institute of Mathematics, Tamkang University, Tamsui, Taiwan, R.O.C.*

(Received December 26, 1996; revised February 4, 1998)

Abstract. Consider k ($k \geq 2$) populations whose mean θ_i and variance σ_i^2 are all unknown. For given control values θ_0 and σ_0^2, we are interested in selecting some population whose mean is the largest in the qualified subset in which each mean is larger than or equal to θ_0 and whose variance is less than or equal to σ_0^2. In this paper we focus on the normal populations in details. However, the analogous method can be applied for the cases other than normal in some situations. A Bayes approach is set up and an empirical Bayes procedure is proposed which has been shown to be asymptotically optimal with convergence rate of order $O(\ln^2 n/n)$. A simulation study is carried out for the performance of the proposed procedure and it is found satisfactory.

Key words and phrases: Best population, multiple criteria, asymptotical optimality, empirical Bayes rule, convergence rate.

1. Introduction

In many practical occasions, an experimenter often faces with the situation of testing for homogeneity. And when the hypothesis of homogeneity is rejected, the experimenter often needs to rank priority of several categories or treatments under consideration according to his goal. This concerns the multiple comparison of ranking and selection which has been developed in last forty years. Readers are referred to Gupta and Panchapakesan (1979), for instance, among others.

In this area of ranking and selection, most literature are concerned with one criterion, for example, a population is considered as the best if it is associated with some largest (or the smallest) parameter in a finite set of populations. In many situations, it may not satisfy the experimenter's demand. For example, in industrial statistics, one needs not only to attain its largest target, but on the other

* This research was partially supported by NSC 86-2115-M-001-028 and NSC 86-2115-M-001-018.

** Now at the Department of Mathematics, Tamsui Oxford University College, Tamsui, Taiwan, R.O.C.

hand, one also needs to keep the variation of product under control. Under this circumstance, a single criterion selection of potential treatments does not meet our requirement. Recently, Gupta *et al.* (1994) consider selecting the best normal population compared with a control. It involves two criteria for selection, however, they belong to same character and only the location parameter is concerned. Since few literature concerns such situation, therefore, it necessitates to develop some methodology in this area for multiple criteria. In this paper, we formulate the problem in a more general setting, however, we focus on the case of normal populations. We consider two main different quantities, i.e. mean and variance for our main concern. Since the set of vectors of mean and variance is not linearly ordered in usual sense, it involves some problems in utility theory which needs further development for applications in this area. These statistical quantities, mean and variance, indicate two measures which are completely different in nature. Accordingly, instead of combining these two quantities to be a single value so that linear order is possible, we rather consider the order to be lexicographic which is more practical and significant and meets our requirement in many situations. It is noted that in the lexicographic order, it assumes implicitly some priority of its order. As can be seen in the next section, the first criterion in Definition 2.1 is more concerned than others. This can be explained in many occasions. For instance, according to the Taguchi's methodology in industrial statistics, one needs first to reduce variation of product to an acceptable level and then maximize its target (see, for example, Taguchi (1987)).

In Section 2, we formulate the problem and develop the Bayes framework. In Section 3, we propose an empirical Bayes procedure and in Section 4, we study the large sample behavior of the proposed empirical Bayes rule. It is shown that the proposed empirical Bayes selection rule has a rate of convergence of order $O(\ln^2 n/n)$, where n is the number of past observation at hand. In the last section some Monte Carlo simulation results are given to show the performance of the proposed procedure.

2. Formulation of problem and a Bayes selection rule

Suppose there are k populations (treatments or designs etc.) $\pi_1, \ldots, \pi_k$ such that π_i has distribution function $F_i(x)$ whose mean and variance are respectively, θ_i and σ_i^2, $i = 1, \ldots, k$. We are interested in identifying which population has the largest mean, however, its variance should not be large. More exactly, let θ_0 and σ_0^2 be two controls and it is expected to identify the population corresponding to the largest mean for which its mean is no less than θ_0 and its associated variance should be no larger than σ_0^2. We define it in details as follows.

DEFINITION 2.1. Let $\pi_1, \ldots, \pi_k$ be k populations such that π_i has mean θ_i and variance σ_i^2, $i = 1, \ldots, k$. Let θ_0 and σ_0^2 be two control values (prefixed). Define $S = \{\pi_i \mid \sigma_i^2 \leq \sigma_0^2\}$. A population π_i is called σ-qualified, if $\pi_i \in S$. A population π_i is considered as the best σ-qualified, if it simultaneously satisfies the following conditions:

(i) $\pi_i \in S$,

(ii) $\theta_i \geq \theta_0$ and

(iii) $\theta_i = \max_{\pi_j \in S} \theta_j$.

Consider a parameter vector $(\theta_i,\ \sigma_i^2)$ corresponding to π_i as a point in the upper half plane $R \times R^+$. Under restrictions of (i) and (ii) in Definition 2.1, we can order those k points. In other words, the qualified set, that is the set of pairs $\{(\theta_i, \sigma_i^2)\}$ in the domain defined by (i) and (ii), is linerly ordered such that (θ_i, σ_i^2) precedes (θ_j, σ_j^2), if $\theta_i \geq \theta_j$. In this sense, the best σ-qualified population is the one whose corresponding parameter vector in the qualified set has the largest order. Of course, when the qualified set is empty, none is the best σ-qualified.

When σ_0^2 is taken to be $+\infty$, the restriction of (i) is removed and under this situation, our selection problem becomes the classical selection problem with a control (see, for example, Gupta *et al.* (1994)). And if furthermore, θ_0 is taken to be $-\infty$, the condition (ii) is again removed and it becomes the usual classical selection problem.

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)$ and $\Omega = \{(\theta_i, \sigma_i^2) \mid -\infty < \theta_i < +\infty, \sigma_i > 0, i = 1, \ldots, k\}$ be the parameter space. Let $\boldsymbol{a} = (a_0, a_1, \ldots, a_k)$ denote an action, where $a_i = 0, 1$; $i = 0, 1, \ldots, k$ and $\sum_{i=0}^k a_i = 1$. Let $\mathcal{A}$ denote the action space, the set of all such $\boldsymbol{a}$. If $a_i = 1$, for some $i = 1, \ldots, k$, it means that population π_i is selected as the best σ-qualified. When $a_0 = 1$, it means no population is considered as the best σ-qualified, i.e. none in k populations satisfies both restricitions (i) and (ii) in Definition 2.1.

For the sake of convenience, corresponding to $\boldsymbol{\theta}$, we define a new parameter vector $\boldsymbol{\theta}' = (\theta_1', \ldots, \theta_k')$ and θ_0' as follows.

DEFINITION 2.2. For a given positive δ and for $i = 0, 1, \ldots, k$, define

$$\theta_i' = \begin{cases} \theta_0 - \delta & \text{if } \ \sigma_i^2 > \sigma_0^2, \\ \theta_i & \text{otherwise.} \end{cases}$$

Accordingly, those populations which do not meet the requirement (i) will also fail to meet the requirement (ii) in Definition 2.1 in terms of the transformed parameter θ_i'.

Choice of the threshold value δ is up to the decision maker. However, it can be clearly seen from the following loss (Definition 2.3) that when δ is large, the associated penalty becomes large, and thus the decision maker should take small δ. To see some behavior of risk for different choices of δ, it is referred to some simulation results in Section 5. It is usually recommended to take value of δ no larger than 10^{-3}.

In a decision-theoretic approach, we consider the following loss function.

DEFINITION 2.3. For parameter $\boldsymbol{\theta}, \boldsymbol{\sigma}$ (equivalently, $\boldsymbol{\theta}', \boldsymbol{\sigma}$), if action $\boldsymbol{a}$ is taken, a loss $L(\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{a})$ is incurred and which is defined by

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\sigma}; \boldsymbol{a}) &= L(\boldsymbol{\theta}', \boldsymbol{\sigma}; \boldsymbol{a}) \\ &= \alpha \left[\max(\theta_{[k]}', \theta_0) - \sum_{i=0}^{k} a_i \theta_i' \right] \end{aligned} \tag{2.1}$$

$$+ (1-\alpha)\sum_{i=0}^{k} \gamma a_i \left(\frac{\sigma_i}{\sigma_0} - 1\right) I_{\{\sigma_i > \sigma_0\}}$$

for prefixed α $(0 \le \alpha \le 1)$ and $\gamma(> 0)$, where $\theta'_{[k]} = \max_{1\le i\le k} \theta'_i$ and $\theta'_0 = \theta_0$.

Remark 1. (1) Obviously, the loss of an action consists of two parts: one is caused by either requirement (ii) or (iii) of Definition 2.1 in terms of means, and the other is caused by the requirement (i) in terms of variances. The value of α and $1-\alpha$ shows the ratio of penalty of the first part to the second part. This ratio is up to decision maker to determine which part of loss is more serious to him. It should be pointed out that few literature has considered the situation of combining different kinds of penalties together. There is no universal theory so far to define a loss function that is most convincing. The loss form in the second part of (2.1) is taken for granted to be the ratio rather than a difference, it is usually because they are scale parameters. As a matter of fact, the second term of (2.1) can be extended in a general form including the difference form, i.e.

$$(1-\alpha)\sum_{i=0}^{k} \gamma a_i (\sigma_i - \sigma_0) I_{\{\sigma_i > \sigma_0\}}.$$

This is explained in Remark 2.

(2) The positive value γ is determined by decision maker to adjust different scale of penalty for variance comparing to penalty for mean. This value can also be considered by decision maker to show its seriousness for the penalty for variance. If $\gamma = |\theta_0|$, then the term $(\frac{\sigma_i}{\sigma_0} - 1)\gamma = \frac{|\theta_0|}{\sigma_0}(\sigma_i - \sigma_0)$ which is the difference between σ_i and σ_0 multiplied by an adjustment factor $(|\theta_0|/\sigma_0)$. If a decision maker takes $\gamma = 1$, this means adjustness is not needed.

(3) If π_i is considered as the best σ-qualified population but, the requirement (i) is not met, then the penalty is given by $\alpha[\max(\theta'_{[k]}, \theta_0) - (\theta_0 - \delta)] + (1-\alpha)(\frac{\sigma_i}{\sigma_0} - 1)\gamma$. It is easy to see that the major punishment is caused by variance. A decision maker is quite flexible to adjust values of δ, α and γ for his penalty.

(4) If π_i is the best σ-qualified, but action $a_0 = 1$ is taken, that is none is taken as the best, then the penalty is $\alpha(\theta'_{[k]} - \theta_0)$.

(5) If π_j is in fact the best σ-qualified, and π_i is σ-qualified and selected as the best σ-qualified, then the penalty is $\alpha(\theta'_{[k]} - \theta_i)$.

In this paper, we focus on the problem of selecting the best σ-qualified normal population. Analogous method can be also applied for situations other than normal. In this paper, we consider a Bayes approach for the problem of selecting the best σ-qualified normal population.

Here we want to point out that we have two criteria and some preference is implicitly assumed. It is well known that some difficulty may exist in combining them. The priority of preference is assumed to be the order of conditions in Definition 2.1. As a matter of fact, we take some position to consider lexicographic preference on the quantity of variance first and then the mean preference. When this is combined in the loss of Definition 2.3, it is implicitly assumed that the second part of loss is more concerned especially, when γ is taken to be large. Accordingly, to make problem simple and more clear to be seen, we permit no perturbation on the quantity of variance. However, on the other hand, we permit the mean parameter to have some perturbation so that we consider some structure on the location parameter of mean, i.e. some prior is assumed on the mean.

For each $i = 1, \ldots, k$, let $X_{i1}, \ldots, X_{iM}$ be a sample of size M from a normal population π_i with mean θ_i and variance σ_i^2. The observed value of X_{ij} is denoted by x_{ij}. It is assumed that θ_i is a realization of random variable Θ_i with a normal prior distribution $N(\mu_i, \tau_i^2)$, where μ_i and τ_i^2 are respectively unknown mean and variance, $i = 1, \ldots, k$. The random variables $\Theta_1, \ldots, \Theta_k$ are assumed to be mutually independent.

For convenience, we denote $x_i = \frac{1}{M}\sum_{j=1}^{M} x_{ij}$. Let $f_i(x_i \mid \theta_i)$ and $h_i(\theta_i \mid \mu_i, \tau_i^2)$ denote the conditional probability density function of X_i and Θ_i, respectively, $i = 1, \ldots, k$. Let $\boldsymbol{x} = (x_1, \ldots, x_k)$ and χ be the sample space generated by $\boldsymbol{x}$. A selection rule $\boldsymbol{d} = (d_0, d_1, \ldots, d_k)$ is a mapping defined on the sample space χ into the $k+1$ product space $[0,1] \times [0,1] \times \cdots \times [0,1]$ such that $\sum_{i=0}^{k} d_i(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \chi$. For every $\boldsymbol{x} \in \chi$, $d_i(\boldsymbol{x})$ dentoes the probability of selecting population π_i as the best σ-qualified, $i = 1, \ldots, k$; and $d_0(\boldsymbol{x})$ denotes the probability that none is selected.

Under the preceding formulation, the Bayes risk of a selection rule $\boldsymbol{d}$, denoted by $r(\boldsymbol{d})$, is given by

$$\begin{aligned}
r(\boldsymbol{d}) &= E_{\boldsymbol{\theta}} E_{\boldsymbol{x}} L(\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{d}) \\
&= \alpha \int_{\Omega}\int_{\chi} \max(\theta'_{[k]}, \theta_0) f(\boldsymbol{x} \mid \boldsymbol{\theta}) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2) d\boldsymbol{x} d\boldsymbol{\theta} \\
&\quad - \alpha \int_{\Omega}\int_{\chi} \sum_{i=0}^{k} d_i(\boldsymbol{x}) \theta'_i f(\boldsymbol{x} \mid \boldsymbol{\theta}) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2) d\boldsymbol{x} d\boldsymbol{\theta} \\
&\quad + (1-\alpha) \int_{\Omega}\int_{\chi} \sum_{i=0}^{k} \gamma d_i(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1\right) I_{\{\sigma_i > \sigma_0\}} f(\boldsymbol{x} \mid \boldsymbol{\theta}) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2) d\boldsymbol{x} d\boldsymbol{\theta} \\
&= \alpha J_1 - \alpha J_2 + (1-\alpha) J_3, \qquad \text{say.}
\end{aligned}$$

Then, a straightforward computation yields the following

$$J_1 = \int_{\Omega} \max(\theta'_{[k]}, \theta_0) h(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2) d\boldsymbol{\theta} = C$$

for some constant C, and

$$\begin{aligned}
J_2 &= \int_{\chi}\int_{\Omega} \sum_{i=0}^{k} d_i(\boldsymbol{x}) \theta'_i h(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\tau}^2) f(\boldsymbol{x}) d\boldsymbol{\theta} d\boldsymbol{x} \\
&= \int_{\chi} \sum_{i=0}^{k} d_i(\boldsymbol{x}) \phi'_i(x_i) f(\boldsymbol{x}) d\boldsymbol{x}
\end{aligned}$$

where

$$\phi_i'(x_i) = \begin{cases} \theta_0 & \text{if} \quad i = 0, \\ \theta_0 - \delta & \text{if} \quad \sigma_i > \sigma_0, \\ \phi_i(x_i) & \text{otherwise}. \end{cases} \tag{2.2}$$

$f(\boldsymbol{x}) = \prod_{i=1}^k f_i(x_i)$ is the marginal probability density function of $\boldsymbol{X}$, $h(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\tau}^2)$ is the posterior probability density function of $\boldsymbol{\Theta}$ given $\boldsymbol{X} = \boldsymbol{x}$, and $\phi_i(x_i) = E(\Theta_i \mid x_i) = (x_i\tau_i^2 + \frac{\sigma_i^2}{M}\mu_i)/(\tau_i^2 + \frac{\sigma_i^2}{M})$, $i = 1, \ldots, k$, denoting that $\phi_0'(x_0) = \phi_0(x_0) = \theta_0$.

Remark 2. It is easy to see that if the second term defined in (2.1) is extended to become

$$(1-\alpha)\sum_{i=0}^{k} \gamma a_i u(\sigma_i/\sigma_0) I_{\{\sigma_i > \sigma_0\}}$$

when $u(\cdot)$ is a positive non-decreasing function, the quantity J_3 defined in $r(\boldsymbol{d})$ is unchanged and so is $r(\boldsymbol{d})$. Hence, the loss function defined in (2.1) can be extended to a bigger class.

Hence, for some constant C,

$$\begin{aligned} r(\boldsymbol{d}) = \alpha C - \alpha \int_{\chi} \sum_{i=0}^{k} d_i(\boldsymbol{x})\phi_i'(x_i) f(\boldsymbol{x}) d\boldsymbol{x} \\ + (1-\alpha)\int_{\chi} \sum_{i=0}^{k} \gamma d_i(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1\right) I_{\{\sigma_i > \sigma_0\}} f(\boldsymbol{x}) d\boldsymbol{x}. \end{aligned} \tag{2.3}$$

For each $\boldsymbol{x} \in \chi$, let

$$Q(\boldsymbol{x}) = \left\{ i \mid \phi_i'(x_i) = \max_{0 \le j \le k} \phi_j'(x_j), i = 0, 1, \ldots, k \right\}. \tag{2.4}$$

Then, define

$$i^* = i^*(\boldsymbol{x}) = \begin{cases} 0 & \text{if} \quad Q(\boldsymbol{x}) = \{0\}, \\ \min\{i \mid i \in Q(\boldsymbol{x}), i \neq 0\} & \text{otherwise}. \end{cases} \tag{2.5}$$

Then, according to (2.3), (2.4) and (2.5), it can be derived that a Bayes selection rule $\boldsymbol{d}^B = (d_0^B, d_1^B, \ldots, d_k^B)$ is given as follows

$$\begin{cases} d_{i^*}^B(\boldsymbol{x}) = 1, \\ d_j^B(\boldsymbol{x}) = 0, \quad \text{for} \quad j \neq i^*. \end{cases} \tag{2.6}$$

Hence,

$$r(\boldsymbol{d}^B) = \alpha C - \alpha \int_{\chi} \sum_{i=0}^{k} d_i^B(\boldsymbol{x})\phi_i'(x_i) f(\boldsymbol{x}) d\boldsymbol{x}. \tag{2.7}$$

3. The empirical Bayes selection rule

Since $\phi_i'(x_i)$ involves the unknown parameters σ_i^2 and (μ_i, τ_i^2), $i = 1, \ldots, k$, hence, the proposed Bayes rule $\boldsymbol{d}^B$ is not applicable. However, based on the past data, these unknown parameters can be estimated and hence a decision can be made if one more observation is taken. Let X_{ijt} denote a sample of size M from population π_i with a normal distribution $N(\theta_{it}, \sigma_i^2)$ at time t $(t = 1, \ldots, n)$, $j = 1, \ldots, M$, and θ_{it} is a realization of a random variable Θ_{it} which is an independent copy of Θ_i with a normal distribution $N(\mu_i, \tau_i^2)$, $t = 1, \ldots, n$. It is assumed that Θ_{it}, $i = 1, \ldots, k$, $t = 1, 2, \ldots, n$ are mutually independent. For our convenience, we denote the current random sample X_{ijn+1} by X_{ij}, for $j = 1, \ldots, M$, $i = 1, \ldots, k$.

For each π_i, $i = 1, 2, \ldots, k$, we will estimate the unknown parameters μ_i, τ_i^2 and σ_i^2 based on the past data X_{ijt}, $j = 1, \ldots, M$, $t = 1, \ldots, n$. We denote

$$\tag{3.1} \begin{cases} X_{i.t} = \dfrac{1}{M}\displaystyle\sum_{j=1}^{M} X_{ijt}, \qquad X_i(n) = \dfrac{1}{n}\displaystyle\sum_{t=1}^{n} X_{i.t}, \\ S_i^2(n) = \dfrac{1}{n-1}\displaystyle\sum_{t=1}^{n}(X_{i.t} - X_i(n))^2, \\ W_{i.t}^2 = \dfrac{1}{M-1}\displaystyle\sum_{j=1}^{M}(X_{ijt} - X_{i.t})^2, \qquad W_i^2(n) = \dfrac{1}{n}\displaystyle\sum_{t=1}^{n} W_{i.t}^2. \end{cases}$$

Also, let $\nu_i^2 = \tau_i^2 + \frac{\sigma_i^2}{M}$. Then, it is well-known that $X_{i.t}$, $t = 1, \ldots, n$, are i.i.d. with a normal distribution $N(\mu_i, \nu_i^2)$ and hence $X_i(n)$ has a normal distribution $N(\mu_i, \frac{\nu_i^2}{n})$ and $\frac{(n-1)}{\nu_i^2} S_i^2(n)$ has a chi-square distribution $\chi^2(n-1)$ and $(M-1)W_{i.t}^2/\sigma_i^2$, $t = 1, \ldots, n$ are i.i.d. with chi-square distribution $\chi^2(M-1)$ and $\frac{n(M-1)}{\sigma_i^2} W_i^2(n)$ has a chi-square distribution $\chi^2(n(M-1))$. From the above discussion and by the Strong Law of Large Number, we have

$$\tag{3.2} \begin{cases} X_i(n) \to \mu_i \ \text{a.s.}, \qquad W_i^2(n) \to \sigma_i^2 \ \text{a.s.}, \qquad S_i^2(n) \to \nu_i^2 \ \text{a.s.}, \\ S_i^2(n) - \dfrac{W_i^2(n)}{M} \to \tau_i^2 \ \text{a.s.}, \qquad E[X_i(n)] = \mu_i, \qquad E[S_i^2(n)] = \nu_i^2, \\ E[W_i^2(n)] = \sigma_i^2, \qquad E\left[S_i^2(n) - \dfrac{W_i^2(n)}{M}\right] = \nu_i^2 - \dfrac{\sigma_i^2}{M} = \tau_i^2. \end{cases}$$

For ease of notation and for the sake of nonnegativeness of variance, we define μ_{in}, σ_{in}^2, ν_{in}^2 and τ_{in}^2 as estimators of μ_i, σ_i^2, ν_i^2 and τ_i^2, respectively, by the following. Note that these estimators μ_{in}, σ_{in}^2, ν_{in}^2, and τ_{in}^2 have been used by many authors such as Ghosh and Meeden (1986), Ghosh and Lahiri (1987) and Gupta *et al.* (1994), among others.

$$\tag{3.3} \begin{aligned} &\mu_{in} = X_i(n), \qquad \sigma_{in}^2 = W_i^2(n), \qquad \nu_{in}^2 = S_i^2(n), \\ &\tau_{in}^2 = \max\left(\nu_{in}^2 - \frac{\sigma_{in}^2}{M}, 0\right). \end{aligned}$$

Also, for $i = 0, 1, \ldots, k$, we define

$$\phi_{in}(x_i) = \begin{cases} \theta_0 & \text{if} \quad i = 0, \\ \left(x_i \tau_{in}^2 + \dfrac{\sigma_{in}^2}{M} \mu_{in} \right) / \nu_{in}^2 & \text{otherwise} \end{cases} \tag{3.4}$$

and then define

$$\phi'_{in}(x_i) = \begin{cases} \theta_0 & \text{if} \quad i = 0, \\ \theta_0 - \delta & \text{if} \quad \sigma_{in}^2 > \sigma_0^2, \\ \phi_{in}(x_i) & \text{otherwise.} \end{cases} \tag{3.5}$$

We consider $\phi_{in}(x_i)$ to be an estimate of $\phi_i(x_i)$, and $\phi'_{in}(x_i)$ to be an estimate of $\phi'_i(x_i)$. For each $\boldsymbol{x} \in \chi$, let

$$Q_n(\boldsymbol{x}) = \left\{ i \mid \phi'_{in}(x_i) = \max_{0 \le j \le k} \phi'_{jn}(x_j), i = 0, 1, \ldots, k \right\}. \tag{3.6}$$

Again, define

$$i_n^* = i_n^*(\boldsymbol{x}) = \begin{cases} 0 & \text{if} \quad Q_n(\boldsymbol{x}) = \{0\}, \\ \min\{i \mid i \in Q_n(\boldsymbol{x}), i \neq 0\} & \text{otherwise.} \end{cases} \tag{3.7}$$

We then have an empirical Bayes selection rule $\boldsymbol{d}^{*n} = (d_0^{*n}, d_1^{*n}, \ldots, d_k^{*n})$ as follows:

$$\begin{cases} d_{i_n^*}^{*n}(\boldsymbol{x}) = 1, & \\ d_j^{*n}(\boldsymbol{x}) = 0, & \text{for} \quad j \neq i_n^*. \end{cases} \tag{3.8}$$

4. Some large sample properties

In this section, we study the asymptotic optimality of the proposed empirical Bayes rule. Before we start to investigate the asymptotic property, we need some preliminary tools.

4.1 *Some preliminary lemmas*

We need the following Lemma 4.1.1, due to Gupta *et al.* (1994), which is obtained by applying the well-known inequalities due to Chernoff (1952).

LEMMA 4.1.1. *Let S_n be a random variable having a $\chi^2(n)$ distribution. Then,*

(1) $P\{S_n \le n(1-\eta)\} \le \exp(-\frac{n}{2} g_1(\eta))$, *for any η, $0 < \eta < 1$,*
(2) $P\{S_n \ge n(1+\eta)\} \le \exp(-\frac{n}{2} g_2(\eta))$, *for any η, $\eta > 0$,*

where

$$g_1(\eta) = -\eta - \ln(1-\eta) \quad \textit{for any} \quad \eta, \ 0 < \eta < 1,$$
$$g_2(\eta) = \eta - \ln(1+\eta) \quad \textit{for any} \quad \eta, \ \eta > 0.$$

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

LEMMA 4.1.2. *Let σ_{in}^2 be the estimator of σ_i^2 defined in* (3.3). *Assume $\sigma_i^2 \neq \sigma_0^2$. Then,*

(1) $P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \leq \sigma_0^2\} \leq \exp(-\frac{n(M-1)}{2}(\frac{\sigma_0^2-\sigma_i^2}{\sigma_i^2} - \ln\frac{\sigma_0^2}{\sigma_i^2}))$,

(2) $P_n\{\sigma_{in}^2 \leq \sigma_0^2, \sigma_i^2 > \sigma_0^2\} \leq \exp(-\frac{n(M-1)}{2}(-\frac{\sigma_i^2-\sigma_0^2}{\sigma_i^2} - \ln\frac{\sigma_0^2}{\sigma_i^2}))$.

PROOF. (1) Since $\frac{n(M-1)\sigma_{in}^2}{\sigma_i^2}$ has a distribution $\chi^2(n(M-1))$, by Lemma 4.1.1, we get

$$\begin{aligned} P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \leq \sigma_0^2\} &= P_n\left\{\frac{n(M-1)\sigma_{in}^2}{\sigma_i^2} > n(M-1)\left(1 + \frac{\sigma_0^2 - \sigma_i^2}{\sigma_i^2}\right)\right\} \\ &\leq \exp\left(-\frac{n(M-1)}{2}\left(\frac{\sigma_0^2 - \sigma_i^2}{\sigma_i^2} - \ln\frac{\sigma_0^2}{\sigma_i^2}\right)\right). \end{aligned}$$

(2) The proof is analogous to (1). □

4.2 *Convergence rate of the empirical Bayes selection rule*

Consider an empirical Bayes selection rule $\boldsymbol{d}^n = (d_0^n, d_1^n, \ldots, d_k^n)$ and denote its associated Bayes risk by $r(\boldsymbol{d}^n)$. From (2.3) and (2.7), we have

$$\begin{aligned} (4.1)\quad r(\boldsymbol{d}^n) - r(d^B) &= \alpha \int_{\mathcal{X}} \sum_{i=0}^{k} [d_i^B(\boldsymbol{x}) - d_i^n(\boldsymbol{x})]\phi_i'(x_i) f(\boldsymbol{x}) d\boldsymbol{x} \\ &\quad + (1-\alpha) \int_{\mathcal{X}} \sum_{i=0}^{k} \gamma d_i^n(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1\right) I_{\{\sigma_i > \sigma_0\}} f(\boldsymbol{x}) d\boldsymbol{x} \\ &= \alpha \int_{\mathcal{X}} \sum_{i=0}^{k} \sum_{j=0}^{k} I_{\{i^*=i, i_n=j\}} [\phi_i'(x_i) - \phi_j'(x_j)] f(\boldsymbol{x}) d\boldsymbol{x} \\ &\quad + (1-\alpha) \int_{\mathcal{X}} \sum_{i=0}^{k} I_{\{i_n=i\}} \left(\frac{\sigma_i}{\sigma_0} - 1\right) \gamma I_{\{\sigma_i > \sigma_0\}} f(\boldsymbol{x}) d\boldsymbol{x}. \end{aligned}$$

Obviously, $r(\boldsymbol{d}^n) - r(\boldsymbol{d}^B) \geq 0$, since $r(\boldsymbol{d}^B)$ is the minimum Bayes risk. Thus, $E_n[r(\boldsymbol{d}^n)] - r(\boldsymbol{d}^B) \geq 0$, where the expectation E_n is taken with respect to the past observations X_{ijt}, $i = 1, \ldots, k$, $j = 1, \ldots, M$ and $t = 1, \ldots, n$. The non-negative difference $E_n[r(\boldsymbol{d}^n)] - r(\boldsymbol{d}^B)$ can be used to measure the performance of the selection rule $\boldsymbol{d}^n$.

DEFINITION 4.2.1. A sequence of empirical Bayes selection rule $\{\boldsymbol{d}^n\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order β_n, if $E_n[r(\boldsymbol{d}^n)] - r(\boldsymbol{d}^B) = O(\beta_n)$, where β_n is a sequence of positive numbers such that $\lim_{n\to\infty} \beta_n = 0$.

Then we have the following

THEOREM 4.2.1. *Assume $\sigma_i^2 \neq \sigma_0^2$, for all $i = 1, \ldots, k$. The empirical Bayes selection rule $\boldsymbol{d}^{*n}(\boldsymbol{x})$, defined in (3.7) and (3.8), is asymptotically optimal with convergence rate of order $O(\ln^2 n/n)$. That is*

$$E_n[r(\boldsymbol{d}^{*n})] - r(\boldsymbol{d}^B) = O(\ln^2 n/n).$$

A detailed proof is given in the Appendix. It should be pointed out that the normality property is not the necessary conditions for the quantities on the right hand side of (A.4) (i.e. $I_1 \sim I_8$) to hold.

5. Simulation study

In order to investigate the performance of proposed empirical Bayes selection rule $\boldsymbol{d}^{*n}(\boldsymbol{x})$ defined in Section 3, we carried out a simulation study which is summarized in this section. The quantity $E_n[r(\boldsymbol{d}^{*n})] - r(\boldsymbol{d}^B)$, mentioned in Definition 4.2.1, is used as a measure of performance of the empirical Bayes selection rule $\boldsymbol{d}^{*n}$.

For a given current observations $\boldsymbol{x}$ and given past observation x_{ijt}, let

$$\begin{aligned} D^n(\boldsymbol{x}) &= \alpha \sum_{i=0}^{k} [d_i^{*n}(\boldsymbol{x}) - d_i^B(\boldsymbol{x})]\phi_i'(\boldsymbol{x}) + (1-\alpha) \sum_{i=0}^{k} \gamma d_i^{*n}(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_o} - 1\right) I_{\{\sigma_i > \sigma_0\}} \\ &= \alpha[\phi_{i^*}'(\boldsymbol{x}) - \phi_{i_n^*}'(\boldsymbol{x})] + (1-\alpha)\left(\frac{\sigma_{i_n^*}}{\sigma_0} - 1\right) \gamma I_{\{\sigma_{i_n^*} > \sigma_0\}}. \end{aligned}$$

Then, from (4.1)

$$E_n[r(\boldsymbol{d}^{*n})] - r(\boldsymbol{d}^B) = EE_n D^n(\boldsymbol{X}).$$

Therefore, the sample mean of $D^n(\boldsymbol{x})$ based on the observations of $\boldsymbol{x}$ and x_{ijt}, $i = 1, \ldots, k$, $j = 1, \ldots, M$, $t = 1, \ldots, n$, can be used as an estimator of $E_n[r(\boldsymbol{d}^{*n})] - r(\boldsymbol{d}^B)$.

We briefly explain the simulation scheme as follows:

(1) For each time, $t = 1, \ldots, n$ and for each population π_i, $i = 1, \ldots, k$, generate observations $x_{i1t}, \ldots, x_{iMt}$ by the following way.

a. Take a value θ_{it} according to distribution $N(\mu_i, \tau_i^2)$.

b. For given θ_{it} and σ_i^2, generate random samples $x_{i1t}, \ldots, x_{iMt}$ according to distribution $N(\theta_{it}, \sigma_i^2)$.

(2) Based on the samples x_{ijt}, $i = 1, \ldots, k$, $j = 1, \ldots, M$, $t = 1, \ldots, n$, estimate the unknown parameters σ_i^2, μ_i, τ_i^2 according to (3.3) and they are denoted by σ_{in}^2, μ_{in}, τ_{in}^2 respectively.

(3) For population π_i, $i = 1, 2, \ldots, k$, repeat step (1) with $t = n + 1$ and take its sample mean as our current sample x_i. Thus the current sample vector is given by $\boldsymbol{x} = (x_1, \ldots, x_k)$.

Table 1. Behavior of empirical Bayes rules.

n	f_n	$\bar{D}_n$	$n\bar{D}_n$	$SE(D_n)$
20	0.8810	2.5792×10^{-2}	5.1585×10^{-1}	1.2990×10^{-2}
40	0.9448	4.6918×10^{-3}	1.8767×10^{-1}	1.6007×10^{-3}
60	0.9597	2.3430×10^{-3}	1.4058×10^{-1}	7.0444×10^{-4}
80	0.9663	1.1835×10^{-3}	9.4676×10^{-2}	1.2697×10^{-4}
100	0.9698	7.5871×10^{-4}	7.5871×10^{-2}	4.6151×10^{-5}
150	0.9737	5.3432×10^{-4}	8.0148×10^{-2}	2.4140×10^{-5}
200	0.9777	4.8255×10^{-4}	9.6510×10^{-2}	1.9511×10^{-5}
250	0.9795	3.2959×10^{-4}	8.2398×10^{-2}	1.0599×10^{-5}
300	0.9867	1.5689×10^{-4}	4.7066×10^{-2}	3.0435×10^{-6}
350	0.9843	2.0822×10^{-4}	7.2875×10^{-2}	6.1888×10^{-6}
400	0.9860	1.7324×10^{-4}	6.9297×10^{-2}	8.8127×10^{-6}
450	0.9877	1.4610×10^{-4}	6.5745×10^{-2}	3.8426×10^{-6}
500	0.9860	1.6198×10^{-4}	8.0991×10^{-2}	3.8001×10^{-6}
600	0.9898	1.0635×10^{-4}	6.3812×10^{-2}	2.2434×10^{-6}
700	0.9885	9.5706×10^{-5}	6.6995×10^{-2}	1.6320×10^{-6}
800	0.9908	8.0349×10^{-5}	6.4279×10^{-2}	1.2273×10^{-6}
900	0.9915	8.7131×10^{-5}	7.8418×10^{-2}	1.8236×10^{-6}
1000	0.9902	8.1926×10^{-5}	8.1926×10^{-2}	1.2413×10^{-6}

Table 2. Entries of $\bar{D}_n$ associated with various values of n and δ.

δ	$n = 20$	$n = 60$	$n = 100$
10^{-4}	2.3905×10^{-2}	2.3432×10^{-3}	9.5235×10^{-4}
10^{-3}	2.3924×10^{-2}	2.3442×10^{-3}	9.5243×10^{-4}
10^{-2}	2.4107×10^{-2}	2.3547×10^{-3}	9.5318×10^{-4}
10^{-1}	2.5937×10^{-2}	2.4597×10^{-3}	9.6068×10^{-4}
10^{0}	4.4237×10^{-2}	3.5097×10^{-3}	1.0357×10^{-3}
10^{1}	2.2724×10^{-1}	1.4010×10^{-2}	1.7857×10^{-3}

(4) For given values of δ, α and γ and control values σ_0^2 and θ_0, based on the current sample vector, determine the Bayes selection rule $\boldsymbol{d}^B$ and the empirical Bayes selection rule $\boldsymbol{d}^{*n}$ according to (2.6) and (3.8). Then, compute $D_n(\boldsymbol{x})$.

(5) Repeat step (1) through step (4) six thousand times, and then take its average denoted by $\bar{D}_n$ which is used as an estimate of $E_n[r(\boldsymbol{d}^{*n})] - r(\boldsymbol{d}^B)$. In addition, $SE(\bar{D}_n)$, the estimated standard error and $n\bar{D}_n$ are computed.

In this section Tables 1–2 and Figs. 1–3 are given and another Tables 3–4 which tabulate values associated with Figs. 1–2 are given in the Appendix.

In the following, we take $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 1$ for Tables 1–4. In Table 1, we take $k = 4$, ($\mu_1 = 7.5$, $\sigma_1 = 1.2$), ($\mu_2 = 8.5$, $\sigma_2 = 1.2$), ($\mu_3 = 7.5$, $\sigma_3 = 1.8$), ($\mu_4 = 8.5$, $\sigma_4 = 1.8$), $\delta = 0.001$, $\alpha = 0.5$, $\gamma = 1$, $\theta_0 = 8$, $\sigma_0 = 1.5$. The relative

Table 3. Entries of $\bar{D}_n$ associated with various values of σ_0 and γ.

σ_0	$\gamma = 1$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$
1.1	1.2961×10^{-2}	2.7901×10^{-2}	4.6575×10^{-2}	8.3923×10^{-2}
1.2	1.3681×10^{-1}	1.4353×10^{-1}	1.5194×10^{-1}	1.6874×10^{-1}
1.3	7.1387×10^{-2}	9.2490×10^{-2}	1.1887×10^{-1}	1.7162×10^{-1}
1.4	2.5649×10^{-1}	2.6768×10^{-1}	2.8167×10^{-1}	3.0965×10^{-1}
1.5	2.2343×10^{-1}	2.5029×10^{-1}	2.8388×10^{-1}	3.5104×10^{-1}
1.6	7.0133×10^{-1}	7.0133×10^{-1}	7.0133×10^{-1}	7.0133×10^{-1}
1.7	2.6216×10^{-1}	2.6216×10^{-1}	2.6216×10^{-1}	2.6216×10^{-1}
1.8	5.8612×10^{-2}	5.8612×10^{-2}	5.8612×10^{-2}	5.8612×10^{-2}

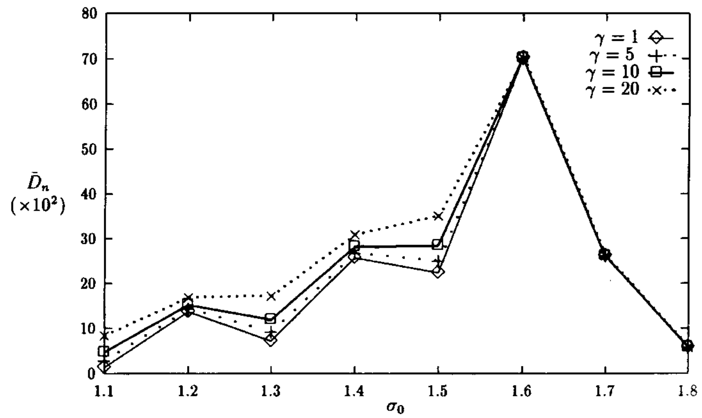

Fig. 1. Plot of Table 3.

frequency that the proposed empirical Bayes selection rule coincides with that of the Bayes selection rule is computed and denoted by f_n. It can be seen from Table 1 that values of $\bar{D}_n$ decrease quite rapidly as n increases. The performance of the proposed empirical Bayes rules behave satisfactorily when $n \geq 40$. Comparing to the situations of different unequal values of τ_i's, it is noted that the behaviors of f_n and $\bar{D}_n$ are robust. To save space, the results of different values of τ_i's are not tabulated.

In Table 2, we study some behavior of $\bar{D}_n$ with respect to δ and n. Take the same values of (μ_i, σ_i), π_i, α, γ and σ_0 as those in Table 1, associated with various values of n and δ, the values of $\bar{D}_n$ are tabulated. In this table, it can be seen that values of $\bar{D}_n$ are quite stable (for fixed n) when $\delta \leq 10^{-2}$. For given δ, $\bar{D}_n$ is quite robust in the sense when τ_i are unequal for $\delta \leq 10^{-1}$.

In Table 3 (see the Appendix), we take ($\mu_1 = 8$, $\sigma_1 = 1.0$), ($\mu_2 = 9$, $\sigma_2 = 1.2$), ($\mu_3 = 10$, $\sigma_3 = 1.4$) and ($\mu_4 = 13$, $\sigma_4 = 1.6$) and also $\theta_0 = 8.5$, $\alpha = 0.5$, $\delta = 0.001$, and $n = 50$. Associated with various values of σ_0 and γ, values of $\bar{D}_n$ are tabulated. Take values of σ_0 as x-axis and values of $\bar{D}_n$ as y-axis, associated with various

Table 4. Entries of $\bar{D}_n$ associated with various values of σ_0 and α.

σ_0	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.8$
1.1	1.2321×10^{-1}	8.3923×10^{-2}	4.4641×10^{-2}
1.2	1.0783×10^{-1}	1.6874×10^{-1}	2.2965×10^{-1}
1.3	1.9527×10^{-1}	1.7162×10^{-1}	1.4798×10^{-1}
1.4	1.9100×10^{-1}	3.0965×10^{-1}	4.2829×10^{-1}
1.5	3.0162×10^{-1}	3.5104×10^{-1}	4.0047×10^{-1}
1.6	2.8053×10^{-1}	7.0133×10^{-1}	1.1221×10^{0}
1.7	1.0486×10^{-1}	2.6216×10^{-1}	4.1945×10^{-1}
1.8	2.3445×10^{-2}	5.8612×10^{-2}	9.3780×10^{-2}

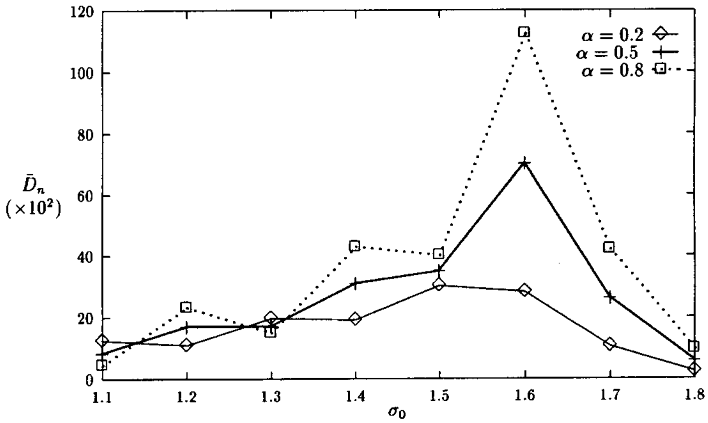

Fig. 2. Plot of Table 4.

values of γ, these are plotted in Fig. 1. Comparing to situations for unequal τ_i's, it is to be noted that the procedure is not so robust.

In Table 4, we take the same values of (μ_i, σ_i), θ_0, δ and n as those given in Table 3 and take $\gamma = 20$. Values of $\bar{D}_n$ associated with values of σ_0 and α are tabulated. Take values of σ_0 as x-axis and values of $\bar{D}_n$ as y-axis, these paris of vectors are plotted in Fig. 2 associated with different values of α at critical values of $\sigma_0 = 1.2$, 1.4 and 1.6 due to loss of over or under estimates of variances. This loss will be enlarged due to unequal of τ_i.

Finally, in Fig. 3, for $n = 50$, taking $\delta = 0.001$, $\theta_0 = 8$, $\sigma_0 = 1.5$, for given values of γ, we study behavior of $\bar{D}_n$ with respect to α. It can be seen from the loss function that it should be a linear function of α. As a matter of fact, when $\gamma = 1(2)9$, they are respectively given by

$$\bar{D}_n = 0.0049256\alpha + 0.001,$$
$$\bar{D}_n = 0.0029256\alpha + 0.003,$$
$$\bar{D}_n = 0.0009256\alpha + 0.005,$$

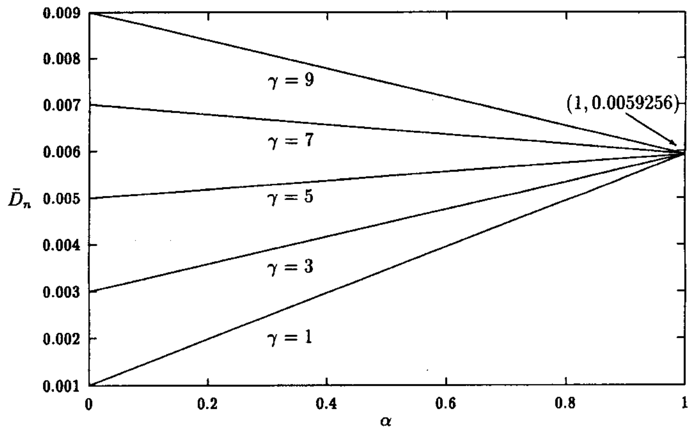

Fig. 3. Linear relation of $\bar{D_n}$ with respect to α.

$$\bar{D}_n = -0.0010744\alpha + 0.007$$

and

$$\bar{D}_n = -0.00307442\alpha + 0.009.$$

So, in general, we can have

$$\bar{D}_n = (0.0059256 - 0.001\gamma)\alpha + 0.001\gamma$$

under same set of data.

6. Conclusion

Most literature in decision theory considers unique criterion. In this paper we consider a selection problem with three criteria in which both location and scale parameters are involved. A class of loss functions is introduced and various behaviors of the risk function have been studied. In this loss function, there are two parameters α and γ which can be adjusted by the decision maker to fit one's requirement. The Bayes risk is sensitive to the variation of control σ_0 when σ_0 is closed to some σ_i, especially when τ_i's are unequal.

As is well-known, in many practical situations, it often involves more than one parameter and also it often needs more than one criterion to fit one's demand. Therefore, the framework in this paper is more applicable in many practical situations.

It is worthwhile and important to consider a general location-scale model. More study is needed for this model since it covers a big class of distributions which are quite useful in practical applications. This will be one of our next subjects for further study.

Acknowledgements

We are very grateful for the careful reading and helpful comments of two referees which have led to improve the presentation of this paper.

Appendix

PROOF OF THEOREM 4.2.1. In order to investigate the convergence rate of $E_n[r(\boldsymbol{d}^n)] - r(\boldsymbol{d}^B)$, we decompose the first term of (4.1) in following situations.

Case 1. When none is the best σ-qualified ($i^* = 0$), however, population π_j is selected as the best σ-qualified ($i_n^* = j \neq 0$).

Since $i^* = 0$, we have $\phi_j'(x_j) < \theta_0$. Hence,

$$\begin{cases} \text{either } \phi_j(x_j) < \theta_0 \text{ and } \sigma_j^2 \leq \sigma_0^2 \text{ (i.e. } \phi_j'(x_j) = \phi_j(x_i)\text{); or} \\ \sigma_j^2 > \sigma_0^2 \text{ (i.e. } \phi_j'(x_j) = \theta_0 - \delta\text{).} \end{cases}$$

Now, since $i_n^* = j \neq 0$, we have $\phi_{jn}(x_j) \geq \theta_0$ and $\sigma_{jn}^2 \leq \sigma_0^2$.

Combining above facts, we get

$$\begin{aligned} &E_n I_{\{i^*=0, i_n^*=j\neq 0\}}[\phi_0'(x_0) - \phi_j'(x_j)] \\ &\quad = P_n\{i^* = 0, i_n^* = j \neq 0\}|\phi_j'(x_j) - \theta_0| \\ &\quad \leq P_n\{\phi_{jn}(x_j) \geq \theta_0, \phi_j(x_j) < \theta_0\}|\phi_j(x_j) - \theta_0| + P_n\{\sigma_{jn}^2 \leq \sigma_0^2, \sigma_j^2 > \sigma_0^2\}\delta \\ &\quad \leq P_n\{\phi_{jn}(x_j) - \phi_j(x_j) > |\phi_j(x_j) - \theta_0|\}|\phi_j(x_j) - \theta_0| \\ &\qquad + P_n\{\sigma_{jn}^2 \leq \sigma_0^2, \sigma_j^2 > \sigma_0^2\}\delta. \end{aligned}$$

Case 2. When π_i is in fact the best σ-qualified ($i^* = i \neq 0$), but none is selected ($i_n^* = 0$).

Since $i^* = i \neq 0$, we have $\phi_i'(x_i) \geq \theta_0$. Hence,

$$\phi_i(x_i) \geq \theta_0 \quad \text{and} \quad \sigma_i^2 \leq \sigma_0^2 \quad \text{(i.e. } \phi_i'(x_i) = \phi_i(x_i)\text{).}$$

Now, since $i_n^* = 0$, we have $\phi_{in}'(x_i) < \theta_0$. Hence,

$$\begin{cases} \text{either } \phi_{in}(x_i) < \theta_0\text{; or} \\ \sigma_{in}^2 > \sigma_0^2. \end{cases}$$

Combining above facts, we get

$$\begin{aligned} &E_n I_{\{i^*=i\neq 0, i_n^*=0\}}[\phi_i'(x_i) - \phi_0'(x_0)] \\ &\quad = P_n\{i^* = i \neq 0, i_n^* = 0\}|\phi_i(x_i) - \theta_0| \\ &\quad \leq [P_n\{\phi_i(x_i) \geq \theta_0, \phi_{in}(x_i) < \theta_0\} + P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \leq \sigma_0^2\}]|\phi_i(x_i) - \theta_0| \\ &\quad \leq [P_n\{\phi_{in}(x_i) - \phi_i(x_i) < -|\phi_i(x_i) - \theta_0|\} \\ &\qquad\qquad + P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \leq \sigma_0^2\}]|\phi_i(x_i) - \theta_0|. \end{aligned}$$

Case 3. When π_i is in fact the best σ-qualified ($i^* = i \neq 0$), but π_j is selected as the best σ-qualified ($i_n^* = j \neq i$ and $j \neq 0$).

Since $i^* = i \neq 0$, we have $\phi_i'(x_i) \geq \phi_j'(x_j)$ and $\sigma_i^2 \leq \sigma_0^2$. Hence,

$$\text{(A.1)} \qquad \begin{cases} \text{either } \phi_i(x_i) \geq \phi_j(x_j),\ \sigma_i^2 \leq \sigma_0^2 \text{ and } \sigma_j^2 \leq \sigma_0^2 \\ \qquad \text{(i.e. } \phi_i'(x_i) = \phi_i(x_i) \text{ and } \phi_j'(x_j) = \phi_j(x_j)\text{); or} \\ \sigma_i^2 \leq \sigma_0^2 \text{ and } \sigma_j^2 > \sigma_0^2 \text{ (i.e. } \phi_i'(x_i) = \phi_i(x_i) \text{ and } \phi_j'(x_j) = \theta_0 - \delta). \end{cases}$$

On the other hand, since $i_n^* = j \neq i$ and $j \neq 0$, we have

$$\phi_{jn}'(x_j) \geq \phi_{in}'(x_i) \qquad \text{and} \qquad \sigma_{jn}^2 \leq \sigma_0^2.$$

Hence,

$$\text{(A.2)} \qquad \begin{cases} \text{either } \phi_{jn}(x_j) \geq \phi_{in}(x_i),\ \sigma_{jn}^2 \leq \sigma_0^2 \text{ and } \sigma_{in}^2 \leq \sigma_0^2\text{; or} \\ \sigma_{jn}^2 \leq \sigma_0^2 \text{ and } \sigma_{in}^2 > \sigma_0^2. \end{cases}$$

Combining (A.1) and (A.2), all situations can be classified into four categories. Hence, we get

$$\begin{aligned} &E_n I_{\{i^*=i\neq 0, i_n^*=j\neq i \text{ and } j\neq 0\}}[\phi_i'(x_i) - \phi_j'(x_j)] \\ &\quad \leq P_n\{\phi_i(x_i) \geq \phi_j(x_j), \phi_{jn}(x_j) \geq \phi_{in}(x_i)\}|\phi_i(x_i) - \phi_j(x_j)| \\ &\qquad + P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \leq \sigma_0^2\}|\phi_i(x_i) - \phi_j(x_j)| \\ &\qquad + P_n\{\sigma_{jn}^2 \leq \sigma_0^2, \sigma_j^2 > \sigma_0^2\}[|\phi_i(x_i) - \theta_0| + \delta] \\ &\qquad + P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_{jn}^2 \leq \sigma_0^2, \sigma_i^2 \leq \sigma_0^2, \sigma_j^2 > \sigma_0^2\}[|\phi_i(x_i) - \theta_0| + \delta] \\ &\quad \leq \left[P_n\left\{\phi_{in}(x_i) - \phi_i(x_i) > \frac{1}{2}|\phi_i(x_i) - \phi_j(x_j)|\right\}\right. \\ &\qquad\quad \left. + P_n\left\{\phi_{jn}(x_j) - \phi_j(x_j) > \frac{1}{2}|\phi_i(x_i) - \phi_j(x_j)|\right\}\right]|\phi_i(x_i) - \phi_j(x_j)| \\ &\qquad + P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \leq \sigma_0^2\}|\phi_i(x_i) - \phi_j(x_j)| \\ &\qquad + P_n\{\sigma_{jn}^2 \leq \sigma_0^2, \sigma_j^2 > \sigma_0^2\}[|\phi_i(x_i) - \theta_0| + \delta] \\ &\qquad + P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_{jn}^2 \leq \sigma_0^2, \sigma_i^2 \leq \sigma_0^2, \sigma_j^2 > \sigma_0^2\}[|\phi_i(x_i) - \theta_0| + \delta]. \end{aligned}$$

Next, we also note that some part of the second term of (4.1) can be dominated by some quantity. To be more exactly,

$$\text{(A.3)} \qquad E_n I_{\{i_n=i\neq 0\}}\left(\frac{\sigma_i}{\sigma_0} - 1\right)\gamma I_{\{\sigma_i>\sigma_0\}} \leq P_n\{\sigma_{in}^2 \leq \sigma_0^2, \sigma_i^2 > \sigma_0^2\}\left|\frac{\sigma_i}{\sigma_0} - 1\right|\gamma.$$

From (4.1), (A.3) and by Cases 1, 2 and 3, we get

$$\text{(A.4)} \qquad E_n[r(\boldsymbol{d}^{*n})] - r(\boldsymbol{d}^B)$$

$$
\begin{aligned}
&= \alpha \sum_{i=0}^{k} \sum_{j=0}^{k} E_n \int_x I_{\{i^*=i, i_n^*=j\}} [\phi_i'(x_i) - \phi_j'(x_j)] f(\boldsymbol{x}) d\boldsymbol{x} \\
&\quad + (1-\alpha) \sum_{i=0}^{k} E_n \int_x I_{\{i_n^*=i\}} \left(\frac{\sigma_i}{\sigma_0} - 1 \right) \gamma I_{\{\sigma_i > \sigma_0\}} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \alpha \sum_{j=1}^{k} \int_R E_n I_{\{i^*=0, i_n^*=j\}} [\theta_0 - \phi_j'(x_j)] f_j(x_j) dx_j \\
&\quad + \alpha \sum_{i=1}^{k} \int_R E_n I_{\{i^*=i, i_n^*=0\}} [\phi_i'(x_i) - \theta_0] f_i(x_i) dx_i \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \iint_{R^2} E_n I_{\{i^*=i, i_n^*=j\}} \\
&\qquad\qquad \times [\phi_i'(x_i) - \phi_j'(x_j)] f_i(x_i) f_j(x_j) dx_i dx_j \\
&\quad + (1-\alpha) \int_R \sum_{i=1}^{k} E_n I_{\{i_n^*=i\}} \left(\frac{\sigma_i}{\sigma_0} - 1 \right) \gamma I_{\{\sigma_i > \sigma_0\}} f_i(x_i) dx_i \\
&\le \alpha \sum_{i=1}^{k} \int_R P_n\{|\phi_{in}(x_i) - \phi_i(x_i)| > |\phi_i(x_i) - \theta_0|\} \\
&\qquad \times |\phi_i(x_i) - \theta_0| f_i(x_i) dx_i \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \\
&\quad \times \iint_{R^2} \left[P_n \left\{ |\phi_{in}(x_i) - \phi_i(x_i)| > \frac{1}{2} |\phi_i(x_i) - \theta_0| \right\} \right. \\
&\qquad\qquad \left. + P_n \left\{ |\phi_{jn}(x_j) - \phi_j(x_j)| > \frac{1}{2} |\phi_i(x_i) - \phi_j(x_j)| \right\} \right] \\
&\qquad \times |\phi_i(x_i) - \phi_j(x_j)| f_i(x_i) f_j(x_j) dx_i dx_j \\
&\quad + \alpha \sum_{j=1}^{k} P_n\{\sigma_{jn}^2 \le \sigma_0^2, \sigma_j^2 > \sigma_0^2\} \delta \\
&\quad + \alpha \sum_{i=1}^{k} \int_R P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \le \sigma_0^2\} |\phi_i(x_i) - \theta_0| f_i(x_i) dx_i \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \iint_{R^2} P_n\{\sigma_{in}^2 > \sigma_0^2, \sigma_i^2 \le \sigma_0^2\} \\
&\qquad\qquad \times |\phi_i(x_i) - \phi_j(x_j)| f_i(x_i) f_j(x_j) dx_i dx_j \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \iint_{R^2} P_n\{\sigma_{jn}^2 \le \sigma_0^2, \sigma_j^2 > \sigma_0^2\} \\
&\qquad\qquad \times [|\phi_i(x_i) - \theta_0| + \delta] f_i(x_i) f_j(x_j) dx_i dx_j
\end{aligned}
$$

$$+\alpha\sum_{i=1}^{k}\sum_{j=1}^{k}\iint_{R^2}P_n\{\sigma_{in}^2>\sigma_0^2,\sigma_{jn}^2\le\sigma_0^2,\sigma_i^2\le\sigma_0^2,\sigma_j^2>\sigma_0^2\}$$
$$\times[|\phi_i(x_i)-\theta_0|+\delta]f_i(x_i)f_j(x_j)dx_idx_j$$
$$+(1-\alpha)\sum_{i=1}^{k}P_n\{\sigma_{in}^2\le\sigma_0^2,\sigma_i^2>\sigma_0^2\}\left|\frac{\sigma_i}{\sigma_0}-1\right|\gamma$$
$$=\alpha(I_1+I_2+I_3+I_4+I_5+I_6+I_7)+(1-\alpha)\gamma I_8,\qquad \text{say.}$$

Gupta *et al.* (1994) have proved that

$$\text{(A.5)}\qquad \int_R P_n\{|\phi_{in}(x_i)-\phi_i(x_i)|>|\phi_i(x_i)-\theta_0|\}|\phi_i(x_i)-\theta_0|f_i(x_i)dx_i = O(\ln^2 n/n),$$

which belongs to type of I_1, and

$$\text{(A.6)}\qquad \int_{R^2} P_n\{|\phi_{in}(x_i)-\phi_i(x_i)|>|\phi_i(x_i)-\phi_j(x_j)|\} \times|\phi_i(x_i)-\phi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j = O(\ln^2 n/n),$$

which belongs to type of I_2.

By Lemma 4.1.2, we have immediately

$$\text{(A.7)}\qquad P_n\{\sigma_{in}^2\le\sigma_0^2,\sigma_i^2>\sigma_0^2\}=O(\exp(-c_0n)),$$

and

$$\text{(A.8)}\qquad P_n\{\sigma_{in}^2>\sigma_0^2,\sigma_i^2\le\sigma_0^2\}=O(\exp(-c_0n)),$$

where $c_0=\max_{1\le i\le k}\frac{M-1}{2}|\frac{\sigma_0^2-\sigma_i^2}{\sigma_i^2}-\ln\frac{\sigma_0^2}{\sigma_i^2}|>0$, for $i=1,\dots,k$.

Recall that $\phi_i(x_i)=\frac{x_i\tau_i^2+\frac{\sigma_i^2}{M}\mu_i}{\nu_i}$ and X_i is marginally $N(\mu_i,\nu_i^2)$ distributed. Therefore, $\phi_i(X_i)$ is $N(\mu_i,\frac{\tau_i^4}{\nu_i^2})$ distributed, $i=1,\dots,k$, and $\phi_i(x_i)-\mu_i=\frac{\tau_i^2}{\nu_i^2}(x_i-\mu_i)$. Hence,

$$\text{(A.9)}\qquad \int_R|\phi_i(x_i)-\theta_0|f_i(x_i)dx_i \le\int_R\frac{\tau_i^2}{\nu_i^2}|x_i-\mu_i|f_i(x_i)dx_i+\int_R|\mu_i-\theta_0|f_i(x_i)dx_i =\frac{2\tau_i^2}{\sqrt{2\pi}\nu_i}+|\mu_i-\theta_0|<+\infty.$$

Also, X_i and X_j are mutually independent, for all $i\ne j$, then $\phi_i(X_i)-\phi_j(X_j)$ is $N(\mu_i-\mu_j,\frac{\tau_i^4}{\nu_i^2}+\frac{\tau_j^4}{\nu_j^2})$ distributed. Similarly as in case of (A.9), we get

$$\text{(A.10)}\qquad \iint_{R^2}|\phi_i(x_i)-\phi_j(x_j)|f_i(x_i)f_j(x_j)dx_idx_j<+\infty.$$

Additionally, δ and $|\sigma_i - \sigma_0|$ are fitite. Combining (A.7)$\sim$(A.10), it is easy to see that I_3, I_4, I_5, I_6, I_7 and I_8 all converge with rate of order $1/n$. Finally, by combining (A.5) and (A.6), we complete the proof. $\square$

References

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.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples, *J. Amer. Statist. Assoc.*, **82**, 1153–1162.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling, *J. Amer. Statist. Assoc.*, **81**, 1058–1062.

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

Gupta, S. S., Liang, T. and Rau, R.-B. (1994). Empirical Bayes rules for selecting the best normal population compared with a control, *Statist. Decisions*, **12**, 125–147.

Taguchi, G. (1987). *System of Experimental Design*, White Plan: UNIPUB/Kraus International Publication.