Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 2129–2143
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Some empirical Bayes rules for selecting the best population with multiple criteria

Wen-Tao Huang[a,*], Yi-Ping Chang[b]

[a]Department of Management Sciences and Decision Making, Tamkang University, Tamsui, Taiwan, ROC
[b]Department of Business Mathematics, Soochow University, Taipei, Taiwan, ROC

Available online 6 September 2005

Abstract

Consider k ($k \geqslant 2$) normal populations whose mean θ_i and variance σ_i^2 are all unknown. Let η_i be some function of θ_i and σ_i^2 and η_i is the parameter of main interest. For given control values η_0 and σ_0^2, we want to select some population whose associated value of η_i the largest and also it is larger than η_0 and whose associated variance is less than or equal to σ_0^2. An empirical Bayes selection rule is proposed which has been shown to be asymptotically optimal with convergence rate of order $O((\ln N)^2/N)$, where N is the minimum number of past observations at hand in each population. A simulation study is also carried out for the performance of the proposed empirical Bayes selection rule, and it is found satisfactory.

MSC: 62F07; 62C12

Keywords: Best population; Multiple criteria; Empirical Bayes rule; Asymptotic optimality; Convergence rate

1. Introduction

Let $\pi_1, \ldots, \pi_k$ be k ($k \geqslant 2$) normal populations where observations X_{ij} from π_i are independently distributed as $\mathcal{N}(\theta_i, \sigma_i^2)$ ($j = 1, \ldots, M_i, i = 1, \ldots, k$). All the means θ_i and variances σ_i^2 are unknown. When θ_i is the parameter of main interest, the problem of selecting the best population was studied in papers pioneered by Bechhofer (1954) using the indifference zone approach and by Gupta (1956, 1965) employing the subset selection approach. A discussion of these approaches and various related topics are referred to Gupta and Panchapakesan (1979) among others.

* Corresponding author. Tel.: +886 2 8631 3221; fax: +886 2 8631 3214.
E-mail address: wentao@mail.tku.edu.tw (W.-T. Huang).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2005.08.032

Let η_i be some function of θ_i and σ_i^2. Consider a selection criterion which is defined in terms of η_i. The population which is associated with the largest (or the smallest depending on a statistician's goal) η_i is called the best. For our convenience, here we call the parameter η_i the parameter of selection criterion. We are interested in selecting the best population. In most known results in the theory of ranking and selection, parameter of selection criteria are commonly focused on either θ_i or σ_i^2. However, in many practical situations, the pth quantile of a population, for example, is an important statistical quantity to be considered. Also, the quantity of signal-to-noise ratio is an important indication for some characteristic in practical application, particularly, in the area of industrial statistics for example.

On the other hand, most literature are concerned with one criterion. 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 also one needs to keep the variation of quality of product under control. Under this circumstance, a single criterion for selecting potential product does not meet our requirement. Gupta et al. (1994) considered selecting the population associated with the largest mean which is larger than a control. It involves two criteria for selection, however, they belong to same character and only the location parameter is concerned. For this reason, Huang et al. (1998), and Huang and Lai (1999) considered selecting the population associated with the largest mean under constraint of both mean and variance. In this paper, we are concerned with the problem of selecting the population associated with the largest parameter of η_i under some constraints.

In Section 2, we formulate the problem and develop the Bayes framework. In Section 3, we propose an empirical Bayes selection rule and in Section 4, we study the large sample behavior of the proposed rule. It is shown that the proposed empirical Bayes selection rule has a rate of convergence of order $O((\ln N)^2/N)$, where N is the minimum number of past observations at hand in each population.

2. Formulation of problem and a Bayes selection rule

Suppose there are k ($k \geqslant 2$) normal populations whose mean θ_i and variance σ_i^2 are all unknown. We are interested in identifying some population which is associated with the largest quantity η_i, some function of θ_i and σ_i^2, and whose variance should not be large. In this paper, we consider the quantity η_i to be a linear function of θ_i, i.e. $\eta_i = g_1(\sigma_i^2)\theta_i + g_2(\sigma_i^2)$ such that $(\theta_i, \sigma_i^2) \to (\eta_i, \sigma_i^2)$ is a one-to-one and onto mapping, where g_1 and g_2 are some functions of σ_i^2. So, domain of g_1 or g_2 is $(0, \infty)$. For example, if $g_1(\sigma_i^2) = 1$ and $g_2(\sigma_i^2) = \Phi^{-1}(p)\sigma_i$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function, then $\eta_i = \theta_i + \Phi^{-1}(p)\sigma_i$ is the pth quantile of the population π_i. If $g_1(\sigma_i^2) = 1/\sigma_i$ and $g_2(\sigma_i^2) = 0$, then $\eta_i = \theta_i/\sigma_i$ is the signal-to-noise ratio (or standardized mean) of the population π_i. Let η_0 and σ_0^2 be two control values (prefixed) and we are desired to identify the population corresponding to the largest quantity η_i such that η_i is no less than η_0 and its associated variance should be no larger than σ_0^2. For exact formulation, we introduce the following definition which is mainly due to Huang and Lai (1999).

Definition 2.1. Defined $S = \{\pi_i \mid \sigma_i^2 \leqslant \sigma_0^2\}$. 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:

(a) $\pi_i \in S$,

(b) $\eta_i > \eta_0$,
(c) $\eta_i = \max_{\pi_j \in S} \eta_j$.

Remark 2.1. Suppose $\eta_i = g_1(\sigma_i^2)\theta_i + g_2(\sigma_i^2)$ for some monotone functions of $g_1(\cdot)$ and $g_2(\cdot)$. If we take $g_1 = 1$ and $g_2 = 0$, then the criteria given by Definition 2.1 become exactly the same criteria considered in Huang and Lai (1999).

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$, and $\boldsymbol{\Omega}$ be the parameter space of $\boldsymbol{\eta}$. Let $\boldsymbol{a} = (a_0, a_1, \ldots, a_k)$ 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 σ-qualified. When $a_0 = 1$, it means that no population is considered as the best σ-qualified. Let $\mathscr{A} = \{\boldsymbol{a}\}$ denote the action space. For our convenience, corresponding to $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$, we define $\boldsymbol{\eta}' = (\eta'_0, \eta'_1, \ldots, \eta'_k)$ as follows.

Definition 2.2. For $i = 0, 1, \ldots, k$, define

$$\eta'_i = \begin{cases} \eta_0 & \text{if } i = 0 \text{ or } \sigma_i > \sigma_0, \\ \eta_i & \text{otherwise.} \end{cases}$$

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

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

$$\begin{aligned} L(\boldsymbol{\eta}, \boldsymbol{\sigma}; \boldsymbol{a}) &= L(\boldsymbol{\eta}', \boldsymbol{\sigma}; \boldsymbol{a}) \\ &= \alpha \left\{ \max(\eta'_{[k]}, \eta_0) - \sum_{i=0}^{k} a_i \eta'_i \right\} + (1 - \alpha) \sum_{i=0}^{k} a_i \left(\frac{\sigma_i}{\sigma_0} - 1 \right) \\ &\quad \times I(\sigma_i > \sigma_0) \end{aligned} \tag{2.1}$$

for prefixed $\alpha (0 \leqslant \alpha \leqslant 1)$, where $\eta'_{[k]} = \max_{1 \leqslant i \leqslant k} \eta'_i$ and $I(\cdot)$ is the indicator function.

The constant α in the loss is determined by a decision maker which is used as a weight ratio of the loss incurred due to failure of correct decision in terms of quantity $\boldsymbol{\eta}$ with respect to variance control. It also can be viewed as an adjustment of a loss due to incorrect decision concerning the quantity $\boldsymbol{\eta}$ against that of the quantity of variance. For further properties of loss $L(\boldsymbol{\eta}, \boldsymbol{\sigma}; \boldsymbol{a})$ defined in (2.1), it is referred to Huang and Lai (1999).

This paper mainly focuses on selecting the best σ-qualified normal population using empirical Bayes approach. To make problem more clear, we consider some prior distribution on the mean, but we permit no perturbation on the quantity of variance.

For each $i = 1, \ldots, k$, let $X_{i1}, \ldots, X_{iM_i}$ be a sample of size M_i from a normal population π_i with mean θ_i and variance σ_i^2. For convenience, we denote $X_i = \sum_{j=1}^{M_i} X_{ij}/M_i$. Let x_{ij} and x_i be the observed values of X_{ij} and X_i, respectively. It is assumed that for each $i = 1, \ldots, k$, θ_i is a realization of random variable Θ_i which has a normal prior distribution $\mathscr{N}(\mu_i, \tau_i^2)$, where $-\infty < \mu_i < \infty$ and $\tau_i^2 > 0$ both unknown. The random variables $\Theta_1, \ldots, \Theta_k$ are assumed to be mutually independent. It is obvious that the conditional posterior distribution of Θ_i given $X_i = x_i$ is a normal distribution with mean $E(\Theta_i \mid x_i) = (x_i\tau_i^2 + \frac{\sigma_i^2}{M_i}\mu_i)/(\tau_i^2 + \sigma_i^2/M_i)$ and variance $Var(\Theta_i \mid x_i) = (\tau_i^2\sigma_i^2)/(\sigma_i^2 + M_i\tau_i^2)$.

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)$, $\mathbf{X} = (X_1 \ldots X_k)$, $\boldsymbol{x} = (x_1, \ldots, x_k)$, and $\mathscr{X}$ be the sample space generated by $\boldsymbol{x}$. A selection rule $\boldsymbol{d} = (d_0, d_1, \ldots, d_k)$ is a mapping from the sample space $\mathscr{X}$ to the set $\{0, 1\}$ such that $\sum_{i=0}^k d_i(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \mathscr{X}$. That is, $\boldsymbol{d} \in \mathscr{A}$.

Define

$$\psi_i(x_i) = E(\eta_i \mid x_i) = E\{g_1(\sigma_i^2)\Theta_i + g_2(\sigma_i^2) \mid x_i\} = g_1(\sigma_i^2)\frac{x_i\tau_i^2 + \frac{\sigma_i^2}{M_i}\mu_i}{\tau_i^2 + \frac{\sigma_i^2}{M_i}} + g_2(\sigma_i^2)$$

and

$$\psi_i'(x_i) = \begin{cases} \eta_0 & \text{if } i = 0 \text{ or } \sigma_i^2 > \sigma_0^2, \\ \psi_i(x_i) & \text{otherwise.} \end{cases}$$

Analogous to arguments in Huang and Lai (1999), the Bayes risk of a selection rule $\boldsymbol{d}$, denoted by $r(\boldsymbol{d})$, is given by

$$r(\boldsymbol{d}) = \alpha \int_{\boldsymbol{\Omega}} \max(\eta_{[k]}', \eta_0) h(\boldsymbol{\eta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2)\, \mathrm{d}\boldsymbol{\eta} - \alpha \int_{\mathscr{X}} \sum_{i=0}^{k} d_i(\boldsymbol{x})\psi_i'(x_i) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$
$$+ (1-\alpha) \int_{\mathscr{X}} \sum_{i=0}^{k} d_i(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1\right) I(\sigma_i > \sigma_0) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}, \tag{2.2}$$

where $f(\boldsymbol{x})$ is the marginal probability density function of $\mathbf{X}$, and $h(\boldsymbol{\eta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2)$ is the conditional probability density function of $\boldsymbol{\eta}$. Note that, the first term in (2.2) is a constant.

For each $\boldsymbol{x} \in \mathscr{X}$, let

$$Q = \{i \mid \sigma_i^2 \leqslant \sigma_0^2\} \cup \{0\}, \tag{2.3}$$

$$Q'(\boldsymbol{x}) = \left\{ i \mid \psi_i'(x_i) = \max_{0 \leqslant j \leqslant k} \psi_j'(x_j), i \in Q \right\} \tag{2.4}$$

and

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

Analogous to arguments in Huang and Lai (1999), it can be derived that a Bayes selection rule $\boldsymbol{d}^{\mathrm{B}} = (d_0^{\mathrm{B}}, d_1^{\mathrm{B}}, \ldots, d_k^{\mathrm{B}})$ is given as follows:

$$d_j^{\mathrm{B}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } j = i^*, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

Hence,

$$r(\boldsymbol{d}^{\mathrm{B}})=\alpha\int_{\boldsymbol{\Omega}}\max(\eta'_{[k]},\eta_0)h(\boldsymbol{\eta}\mid\boldsymbol{\mu},\boldsymbol{\tau}^2)\,\mathrm{d}\boldsymbol{\eta}-\alpha\int_{\mathscr{X}}\sum_{i=0}^{k}d_i^{\mathrm{B}}(\boldsymbol{x})\psi'_i(x_i)f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Note that, combining (2.3)–(2.6), if $\sigma_i^2>\sigma_0^2$ then the population π_i is not selected.

3. The empirical Bayes selection rule

Since $\psi'_i(x_i)$ involves the unknown parameters σ_i^2 and (μ_i,τ_i^2), $i=1,\ldots,k$, hence, the proposed Bayes selection rule $\boldsymbol{d}^{\mathrm{B}}$ defined by (2.6) is not applicable. However, based on the past data, these unknown parameters can be estimated and a decision can be made if one more observation (current data) is taken. Let X_{ijt} denote a sample of size M_i from population π_i with a normal distribution $\mathscr{N}(\theta_{it},\sigma_i^2)$ at time t $(t=1,\ldots,n_i)$, $j=1,\ldots,M_i$, and θ_{it} is a realization of a random variable Θ_{it} which is an independent copy of Θ_i with a normal distribution $\mathscr{N}(\mu_i,\tau_i^2)$, $t=1,\ldots,n_i$, $i=1,\ldots,k$. For convenience, we denote the current random sample X_{ijn_i+1} by X_{ij} for $j=1,\ldots,M_i$, $i=1,\ldots,k$.

For each π_i, $i=1,\ldots,k$, we estimate the unknown parameters μ_i, τ_i^2, and σ_i^2 based on the past data X_{ijt}, $j=1,\ldots,M_i$, $t=1,\ldots,n_i$. Let $X_{i\cdot t}=1/M_i\sum_{j=1}^{M_i}X_{ijt}$, $X_i(n_i)=1/n_i\sum_{t=1}^{n_i}X_{i\cdot t}$, $S_i^2(n_i)=1/(n_i-1)\sum_{t=1}^{n_i}(X_{i\cdot t}-X_i(n_i))^2$, $W_{i\cdot t}^2=1/(M_i-1)\sum_{j=1}^{M_i}(X_{ijt}-X_{i\cdot t})^2$, $W_i^2(n_i)=1/n_i\sum_{t=1}^{n_i}W_{i\cdot t}^2$, and $v_i^2=\tau_i^2+\sigma_i^2/M_i$. We denote $\hat{\mu}_i$, $\hat{\sigma}_i^2$, $\hat{v}_i^2$, and $\hat{\tau}_i^2$ the estimators of μ_i, σ_i^2, v_i^2, and τ_i^2, respectively, and which are defined by

$$\hat{\mu}_i=X_i(n_i),\quad \hat{\sigma}_i^2=W_i^2(n_i),\quad \hat{v}_i^2=S_i^2(n_i),\quad\text{and}\quad \hat{\tau}_i^2=\max\left(\hat{v}_i^2-\frac{\hat{\sigma}_i^2}{M_i},0\right). \tag{3.1}$$

These consistent estimators $\hat{\mu}_i$, $\hat{\sigma}_i^2$, $\hat{v}_i^2$, and $\hat{\tau}_i^2$ have been applied by several authors such as Ghosh and Meeden (1986), Ghosh and Lahiri (1987), Gupta et al. (1994) and Huang and Lai (1999), among others.

Also, let $\boldsymbol{n}=(n_1,\ldots,n_k)$, we define

$$Q_{\boldsymbol{n}}=\{i\mid\hat{\sigma}_i^2\leqslant\sigma_0^2\}\cup\{0\} \tag{3.2}$$

and

$$\hat{\psi}'_i(x_i)=\begin{cases}\eta_0 & \text{if } i=0 \text{ or } \hat{\sigma}_i^2>\sigma_0^2,\\ \hat{\psi}_i(x_i) & \text{otherwise,}\end{cases} \tag{3.3}$$

where

$$\hat{\psi}_i(x_i)=g_1(\hat{\sigma}_i^2)\frac{x_i\hat{\tau}_i^2+\frac{\hat{\sigma}_i^2}{M_i}\hat{\mu}_i}{\hat{v}_i^2}+g_2(\hat{\sigma}_i^2), \tag{3.4}$$

$i=1,\ldots,k$. Here we consider $\hat{\psi}_i(x_i)$ and $\hat{\psi}'_i(x_i)$ as estimates of $\psi_i(x_i)$ and $\psi'_i(x_i)$ respectively. For each $\boldsymbol{x}\in\mathscr{X}$, let

$$Q'_{\boldsymbol{n}}(\boldsymbol{x})=\left\{i\mid\hat{\psi}'_i(x_i)=\max_{0\leqslant j\leqslant k}\hat{\psi}'_j(x_j),\ i\in Q_{\boldsymbol{n}}\right\}. \tag{3.5}$$

Again, define

$$i_{\boldsymbol{n}}^* \equiv i_{\boldsymbol{n}}^*(\boldsymbol{x}) = \begin{cases} 0 & \text{if } Q'_{\boldsymbol{n}}(\boldsymbol{x}) = \{0\}, \\ \min\{i \mid i \in Q'_{\boldsymbol{n}}(\boldsymbol{x}), i \neq 0\} & \text{otherwise.} \end{cases} \tag{3.6}$$

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

$$d_j^{*\boldsymbol{n}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } j = i_{\boldsymbol{n}}^*, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Hence,

$$r(\boldsymbol{d}^{*\boldsymbol{n}}) = \alpha \int_{\boldsymbol{\Omega}} \max(\eta'_{[k]}, \eta_0) h(\boldsymbol{\eta} \mid \boldsymbol{\mu}, \boldsymbol{\tau}^2)\, \mathrm{d}\boldsymbol{\eta} - \alpha \int_{\mathscr{X}} \sum_{i=0}^{k} d_i^{*\boldsymbol{n}}(\boldsymbol{x}) \psi'_i(x_i) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$
$$+ (1-\alpha) \int_{\mathscr{X}} \sum_{i=0}^{k} d_i^{*\boldsymbol{n}}(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1 \right) I(\sigma_i > \sigma_0) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

Combining (3.2)–(3.7), we note that if $\hat{\sigma}_i^2 > \sigma_0^2$, then the population π_i is not selected.

4. Asymptotic optimality of empirical Bayes selection rule

In this section, we study the asymptotic optimality of the proposed empirical Bayes selection rule. Consider an empirical Bayes selection rule $\boldsymbol{d}^{\boldsymbol{n}} = (d_0^{\boldsymbol{n}}, d_1^{\boldsymbol{n}}, \ldots, d_k^{\boldsymbol{n}})$ and denote its associated Bayes risk by $r(\boldsymbol{d}^{\boldsymbol{n}})$. Obviously, $r(\boldsymbol{d}^{\boldsymbol{n}}) - r(\boldsymbol{d}^{\mathrm{B}}) \geqslant 0$, since $r(\boldsymbol{d}^{\mathrm{B}})$ is the minimum Bayes risk. Thus, $E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}}) \geqslant 0$ for all $\boldsymbol{n}$, where $E_{\boldsymbol{n}}$ is taken with respect to the probability measure generated by X_{ijt}, $i = 1, \ldots, k$, $j = 1, \ldots, M_i$, and $t = 1, \ldots, n_i$. The nonnegative regret risk $E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}})$ is generally used as a measure of the performance of the selection rule $\boldsymbol{d}^{\boldsymbol{n}}$.

Definition 4.1. An empirical Bayes selection rule $\boldsymbol{d}^{\boldsymbol{n}}$ is said to be asymptotically optimal of order β_N if $E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}}) = O(\beta_N)$, where $N = \min\{n_i \mid 1 \leqslant i \leqslant k\}$ and $\{\beta_N\}$ is a sequence of positive numbers such that $\lim_{N\to\infty} \beta_N = 0$.

Theorem 4.1. *Assume $\sigma_i^2 \neq \sigma_0^2$, for all $i = 1, \ldots, k$. Suppose both g_1 and g_2 are Lipschitz continuous. Then, the empirical Bayes selection rule $\boldsymbol{d}^{*\boldsymbol{n}}$, defined by* (3.7), *is asymptotically optimal of order $O((\ln N)^2/N)$. That is, $E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{*\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}}) = O((\ln N)^2/N)$.*

Proof. Let $f_i(x_i)$ be the marginal probability density function of X_i and let $P_{\boldsymbol{n}}(\hat{\sigma}_i^2 > \sigma_0^2; \sigma_i^2 < \sigma_0^2)$ denote the quantity of $P_{\boldsymbol{n}}(\hat{\sigma}_i^2 > \sigma_0^2)$ under the condition that $\sigma_i^2 < \sigma_0^2$. Then, we have

$$E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{*\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}})$$
$$= \alpha \int_{\mathscr{X}} \sum_{i=0}^{k} \{d_i^{\mathrm{B}}(\boldsymbol{x}) - d_i^{*\boldsymbol{n}}(\boldsymbol{x})\} \psi'_i(x_i) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$
$$+ (1-\alpha) \int_{\mathscr{X}} \sum_{i=0}^{k} d_i^{*\boldsymbol{n}}(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1 \right) I(\sigma_i > \sigma_0) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

$$
\begin{aligned}
&\leqslant \alpha \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}(|\hat{\psi}_i(x_i) - \psi_i(x_i)| > |\psi_i(x_i) - \eta_0|)|\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \int\int_{R^2} [P_{\boldsymbol{n}}\{|\hat{\psi}_i(x_i) - \psi_i(x_i)| > \tfrac{1}{2}|\psi_i(x_i) - \psi_j(x_j)|\} \\
&\quad + P_{\boldsymbol{n}}\{|\hat{\psi}_j(x_j) - \psi_j(x_j)| > \tfrac{1}{2}|\psi_i(x_i) - \psi_j(x_j)|\}]|\psi_i(x_i) - \psi_j(x_j)| \\
&\quad \times f_i(x_i) f_j(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\
&\quad + \alpha \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}(\hat{\sigma}_i^2 > \sigma_0^2;\, \sigma_i^2 < \sigma_0^2)|\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \int\int_{R^2} P_{\boldsymbol{n}}(\hat{\sigma}_i^2 > \sigma_0^2;\, \sigma_i^2 < \sigma_0^2)|\psi_i(x_i) - \psi_j(x_j)| \\
&\quad \times f_i(x_i) f_j(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \int\int_{R^2} P_{\boldsymbol{n}}(\hat{\sigma}_j^2 \leqslant \sigma_0^2;\, \sigma_j^2 > \sigma_0^2)|\psi_i(x_i) - \eta_0| f_i(x_i) f_j(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\
&\quad + \alpha \sum_{i=1}^{k} \sum_{j=1}^{k} \int\int_{R^2} P_{\boldsymbol{n}}(\hat{\sigma}_i^2 > \sigma_0^2,\, \hat{\sigma}_j^2 \leqslant \sigma_0^2;\, \sigma_i^2 < \sigma_0^2,\, \sigma_j^2 > \sigma_0^2) \\
&\quad \times |\psi_i(x_i) - \eta_0| f_i(x_i) f_j(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\
&\quad + (1-\alpha) \sum_{i=1}^{k} P_{\boldsymbol{n}}(\hat{\sigma}_i^2 \leqslant \sigma_0^2;\, \sigma_i^2 > \sigma_0^2) \left| \frac{\sigma_i}{\sigma_0} - 1 \right| \\
&= \alpha(I_1 + I_2 + I_3 + I_4 + I_5 + I_6) + (1-\alpha) I_7, \quad \text{say.}
\end{aligned} \tag{4.1}
$$

Let

$$
\phi_i(x_i) = E(\Theta_i \mid x_i) = \frac{x_i \tau_i^2 + \frac{\sigma_i^2}{M_i}\mu_i}{\tau_i^2 + \frac{\sigma_i^2}{M_i}} = \frac{x_i \tau_i^2 + \frac{\sigma_i^2}{M_i}\mu_i}{v_i^2}
$$

and

$$
\hat{\phi}_i(x_i) = \frac{x_i \hat{\tau}_i^2 + \frac{\hat{\sigma}_i^2}{M_i}\hat{\mu}_i}{\hat{v}_i^2},
$$

then $\psi_i(x_i) = g_1(\sigma_i^2)\phi_i(x_i) + g_2(\sigma_i^2)$ and $\hat{\psi}_i(x_i) = g_1(\hat{\sigma}_i^2)\hat{\phi}_i(x_i) + g_2(\hat{\sigma}_i^2)$. Using the inequality that

$$
\begin{aligned}
&|g_1(\hat{\sigma}_i^2)\hat{\phi}_i(x_i) - g_1(\sigma_i^2)\phi_i(x_i)| \\
&\quad = |g_1(\hat{\sigma}_i^2)(\hat{\phi}_i(x_i) - \phi_i(x_i)) + (g_1(\hat{\sigma}_i^2) - g_1(\sigma_i^2))\phi_i(x_i)| \\
&\quad \leqslant |g_1(\hat{\sigma}_i^2)||(\hat{\phi}_i(x_i) - \phi_i(x_i)| + |(g_1(\hat{\sigma}_i^2) - g_1(\sigma_i^2)||\phi_i(x_i)|.
\end{aligned}
$$

Rewrite I_1 as

$$\begin{aligned}
I_1 &= \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|\hat{\psi}_i(x_i) - \psi_i(x_i)| > |\psi_i(x_i) - \eta_0|\}|\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\leqslant \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|g_1(\hat{\sigma}_i^2)||\hat{\phi}_i(x_i) - \phi_i(x_i)| > |\psi_i(x_i) - \eta_0|/3\}|\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\quad + \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|g_1(\hat{\sigma}_i^2) - g_1(\sigma_i^2)||\phi_i(x_i)| > |\psi_i(x_i) - \eta_0|/3\}|\psi_i(x_i) - \eta_0| \\
&\quad \times f_i(x_i)\,\mathrm{d}x_i \\
&\quad + \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|g_2(\hat{\sigma}_i^2) - g_2(\sigma_i^2)| > |\psi_i(x_i) - \eta_0|/3\}|\psi_i(x_i) - \eta_0| \\
&\quad \times f_i(x_i)\,\mathrm{d}x_i \\
&= I_{11} + I_{12} + I_{13}, \quad \text{say.}
\end{aligned}$$

Now,

$$\begin{aligned}
I_{11} &\leqslant \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|g_1(\hat{\sigma}_i^2)||\hat{\phi}_i(x_i) - \phi_i(x_i)| > |\psi_i(x_i) - \eta_0|/3,\ \sigma_i^2/2 \leqslant \hat{\sigma}_i^2 \leqslant 2\sigma_i^2\} \\
&\quad \times |\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\quad + \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|g_1(\hat{\sigma}_i^2)||\hat{\phi}_i(x_i) - \phi_i(x_i)| > |\psi_i(x_i) - \eta_0|/3, \\
&\quad \hat{\sigma}_i^2 < \sigma_i/2 \text{ or } \hat{\sigma}_i^2 > 2\sigma_i^2\}|\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&= I_{111} + I_{112}, \quad \text{say.}
\end{aligned}$$

Using Corollary 4.1 of Gupta et al. (1994), we have

$$I_{112} = O(\exp(-cN))$$

for some $c > 0$. Since g_1 is Lipschitz continuous, we have

$$\begin{aligned}
I_{111} &\leqslant \sum_{i=1}^{k} \int_R P_{\boldsymbol{n}}\{|\hat{\phi}_i(x_i) - \phi_i(x_i)| > |\psi_i(x_i) - \eta_0|/(3c_{1i}),\ \sigma_i^2/2 \leqslant \hat{\sigma}_i^2 \leqslant 2\sigma_i^2\} \\
&\quad \times |\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i,
\end{aligned}$$

where $c_{1i} = \sup_{\sigma_i^2/2 \leqslant \sigma^2 \leqslant 2\sigma_i^2} |g_1(\sigma^2)|$ and $0 < c_{1i} < \infty$. Then, following a discussion analogous to Gupta et al. (1994), one can obtain that $I_{111} \leqslant O((\ln N)^2/N)$. Therefore,

$$I_{11} \leqslant O((\ln N)^2/N).$$

Also, using the Lipschitz condition, it can be shown that I_{12} and I_{13} all converge to 0 with rate of order $O((\ln N)^2/N)$. Thus,

$$I_1 \leqslant O((\ln N)^2/N). \tag{4.2}$$

By using similar arguments and the results in Gupta et al. (1994), we also have

$$I_2 \leqslant O((\ln N)^2/N). \tag{4.3}$$

Now, we evaluate the rate of convergence of I_3, I_4, I_5, I_6, and I_7. Huang and Lai (1999) have proved that

$$P_{\boldsymbol{n}}(\hat{\sigma}_i^2 \leqslant \sigma_0^2;\, \sigma_i^2 > \sigma_0^2) = O(\exp(-c_2 n_i)) \tag{4.4}$$

and

$$P_{\boldsymbol{n}}(\hat{\sigma}_i^2 > \sigma_0^2;\, \sigma_i^2 < \sigma_0^2) = O(\exp(-c_2 n_i)), \tag{4.5}$$

where $c_2 = \max_{1 \leqslant i \leqslant k}(M_i - 1)/2|(\sigma_0^2 - \sigma_i^2)/\sigma_i^2 - \ln(\sigma_0^2/\sigma_i^2)| > 0$ for $i = 1, \ldots, k$.

Recall that $\psi_i(x_i) = g_1(\sigma_i^2)(x_i\tau_i^2 + (\sigma_i^2/M_i)\mu_i)/v_i^2 + g_2(\sigma_i^2)$ and X_i is marginally $\mathcal{N}(\mu_i, v_i^2)$ distributed. Therefore, $\psi_i(X_i)$ follows $\mathcal{N}(g_1(\sigma_i^2)\mu_i + g_2(\sigma_i^2),\, g_1^2(\sigma_i^2)\tau_i^4/v_i^2)$. Hence,

$$\begin{aligned}
&\int_R |\psi_i(x_i) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\quad \leqslant \int_R |g_1(\sigma_i^2)(\phi(x_i) - \mu_i)| f_i(x_i)\,\mathrm{d}x_i + \int_R |g_1(\sigma_i^2)\mu_i + g_2(\sigma_i^2) - \eta_0| f_i(x_i)\,\mathrm{d}x_i \\
&\quad = |g_1(\sigma_i^2)| \frac{2\tau_i^2}{\sqrt{2\pi} v_i} + |g_1(\sigma_i^2)\mu_i + g_2(\sigma_i^2) - \eta_0| < \infty.
\end{aligned} \tag{4.6}$$

Also, X_i and X_j are mutually independent for all $i \neq j$. Similarly as in case of (4.8), we conclude that

$$\int\!\!\int_{R^2} |\psi_i(x_i) - \psi_j(x_j)| f_i(x_i) f_j(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j < \infty. \tag{4.7}$$

Noting that $|\sigma_i^2 - \sigma_0^2|$ is finite and combining (4.4)–(4.7), it is easy to see that I_3, I_4, I_5, I_6, and I_7, all converge to 0 with rate of order $1/N$. Finally, by combining (4.1), (4.2), and (4.3), we complete the proof. □

5. Examples

In following examples, we consider some special functions of g_1 and g_2. For η_i associated with such g_1 and g_2, the empirical Bayes selection rule is asymptotically optimal of order $O((\ln N)^2)/N)$.

Example 5.1. Take $g_1(\sigma_i^2) = 1$ and $g_2(\sigma_i^2) = \Phi^{-1}(p)\sigma_i$, then $\eta_i = g_1(\sigma_i^2)\theta_i + g_2(\sigma_i^2) = \theta_i + \Phi^{-1}(p)\sigma_i$ is the pth quantile. It is obvious that both g_1 and g_2 are Lipschitz continuous.

Example 5.2. If we have a prior information that there exists a positive small real number σ^* such that $\sigma_i > \sigma^*$ for all i. Take $g_1(\sigma_i^2)=1/\sigma_i$ and $g_2(\sigma_i^2)=0$, then $\eta_i=g_1(\sigma_i^2)\theta_i+g_2(\sigma_i^2)=\theta_i/\sigma_i$ is the signal-to-noise ratio (or standardized mean). Again, it is obvious that both g_1 and g_2 are Lipschitz continuous.

Example 5.3. Let$X_{01}, \ldots, X_{0M_i}$ be an independent random sample of size M_i from a control normal population π_0 with known mean θ_0 and variance σ_0^2. The reliability parameter can be defined by $P(X_{ij} > X_{0j}) = \Phi((\theta_i - \theta_0)(\sigma_i^2 + \sigma_0^2)^{-1/2})$ for $i \neq 0$. If $g_1(\sigma_i^2) = (\sigma_i^2 + \sigma_0^2)^{-1/2}$ and $g_2(\sigma_i^2) = -\theta_0(\sigma_i^2 + \sigma_0^2)^{-1/2}$, then $\eta_i = g_1(\sigma_i^2)\theta_i + g_2(\sigma_i^2) = (\theta_i - \theta_0)(\sigma_i^2 + \sigma_0^2)^{-1/2}$ and $\Phi(\eta_i)$ is the reliability parameter. Since $\Phi(\cdot)$ is strictly increasing, the parameter for selection criterion can be defined by η_i. Here both g_1 and g_2 are Lipschitz continuous.

Example 5.4. Suppose that $Y_{ij}=\exp(X_{ij})$, where X_{ij} are independently distributed as $\mathcal{N}(\theta_i, \sigma_i^2)$ $(j = 1, \ldots, M_i, i = 1, \ldots, k)$. Then Y_{ij} has a lognormal distribution with mean $\exp(\theta_i + \sigma_i^2/2)$ and coefficient of variation $\mathrm{CV} = \{\exp(\sigma_i^2) - 1\}^{1/2}$. If $g_1(\sigma_i^2) = 1$ and $g_2(\sigma_i^2) = \sigma_i^2/2$, then $\eta_i = g_1(\sigma_i^2)\theta_i + g_2(\sigma_i^2) = \theta_i + \sigma_i^2/2$ and $\exp(\eta_i)$ is equal to the mean of Y_{ij}. Since $\exp(\cdot)$ is strictly increasing, the parameter for selection criterion can be defined by η_i. It is obvious that both g_1 and g_2 are Lipschitz continuous. That is, Y_{ij} are independent random samples having lognormal distributions. For given control values ϑ_0 and CV_0, we want to select some population whose associated mean is larger than ϑ_0 and whose associated coefficient of variation is less than or equal to CV_0. If we define $\eta_0 = \ln(\vartheta_0)$ and $\sigma_0^2 = \ln(\mathrm{CV}_0^2 + 1)$, then the problem is just equivalent to selecting some population whose associated η_i is larger than η_0 and whose associated σ_i^2 is less than or equal to σ_0^2.

6. Simulation study

In order to investigate the performance of proposed empirical Bayes selection rule $\boldsymbol{d}^{*\boldsymbol{n}}$ defined in Section 3, we have carried out a simulation study and which is summarized in this section. The quality $E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{*\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}})$, mentioned in Definition 4.1, is used as a measure of performance of the empirical Bayes selection rule $\boldsymbol{d}^{*\boldsymbol{n}}$. For a given current observations $\boldsymbol{x}$ and given past observation x_{ijt}, let

$$D^{\boldsymbol{n}}(\boldsymbol{x}) = \alpha \sum_{i=0}^{k} \{d_i^{\mathrm{B}}(\boldsymbol{x}) - d_i^{*\boldsymbol{n}}(\boldsymbol{x})\}\psi_i'(\boldsymbol{x}) + (1-\alpha) \sum_{i=0}^{k} d_i^{*\boldsymbol{n}}(\boldsymbol{x}) \left(\frac{\sigma_i}{\sigma_0} - 1\right) I(\sigma_i > \sigma_0)$$
$$= \alpha\{\psi_{i^*}'(\boldsymbol{x}) - \psi_{i_{\boldsymbol{n}}^*}'(\boldsymbol{x})\} + (1-\alpha)\left(\frac{\sigma_{i_{\boldsymbol{n}}^*}}{\sigma_0} - 1\right) I(\sigma_{i_{\boldsymbol{n}}^*} > \sigma_0).$$

Then,

$$E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{*\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}}) = E[E_{\boldsymbol{n}}\{D^{\boldsymbol{n}}(\mathbf{X})\}].$$

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

The simulation scheme is similar to that of Gupta et al. (1994) and Huang and Lai (1999). We briefly explain the scheme as follows:

Table 1
Behavior of empirical Bayes selection rules when $\eta_i^{(1)} = \theta_i + \Phi^{-1}(0.9)\sigma_i$

n	f_n	$\bar{D}_n$	$n\bar{D}_n$	$SE(\bar{D}_n)$
20	0.8493	1.3396×10^{-2}	2.6792×10^{-1}	5.3299×10^{-4}
40	0.9228	4.7644×10^{-3}	1.9058×10^{-1}	2.7052×10^{-4}
60	0.9553	2.4116×10^{-3}	1.4469×10^{-1}	1.8885×10^{-4}
80	0.9645	1.4147×10^{-3}	1.1317×10^{-1}	1.1664×10^{-4}
100	0.9764	8.6692×10^{-4}	8.6692×10^{-2}	1.0258×10^{-4}
200	0.9873	1.9440×10^{-4}	3.8879×10^{-2}	2.3420×10^{-5}
300	0.9901	1.0970×10^{-4}	3.2909×10^{-2}	1.4088×10^{-5}
400	0.9901	9.8765×10^{-5}	3.9506×10^{-2}	1.3060×10^{-5}
500	0.9922	8.2909×10^{-5}	4.1455×10^{-2}	1.2275×10^{-5}
600	0.9928	6.4165×10^{-5}	3.8499×10^{-2}	1.0051×10^{-5}
700	0.9925	6.5966×10^{-5}	4.6176×10^{-2}	9.4488×10^{-6}
800	0.9921	5.3319×10^{-5}	4.2655×10^{-2}	7.7244×10^{-6}
900	0.9925	4.7624×10^{-5}	4.2862×10^{-2}	7.2539×10^{-6}
1000	0.9944	4.4789×10^{-5}	4.4789×10^{-2}	8.0138×10^{-6}

(1) For each $t = 1, \ldots, n_i$ and for each population π_i, $i = 1, \ldots, k$, generate observations $x_{i1t}, \ldots, X_{iM_it}$ by the following way:

   a. Take a value θ_{it} according to distribution $\mathcal{N}(\mu_i, \tau_i^2)$.
   b. For given θ_{it} and σ_i^2, generate random samples $x_{i1t}, \ldots, x_{iM_it}$ according to distribution $\mathcal{N}(\theta_{it}, \sigma_i^2)$.

(2) Based on the samples x_{ijt}, $i = 1, \ldots, k$, $j = 1, \ldots, M_i$, $t = 1, \ldots, n_i$, estimate the unknown parameters $\mu_i, \sigma_i^2, \tau_i^2$ according to (3.1) and they are denoted by $\hat{\mu}_i, \hat{\sigma}_i^2, \hat{\tau}_i^2$, respectively.
(3) For population π_i, $i = 1, \ldots, k$, repeat step (1) with $t = n_i + 1$ and take its sample mean as our current sample x_i. Thus the current sample vector is given by $\boldsymbol{x} = (x_1, \ldots, x_k)$.
(4) For given value of α and control values σ_0^2 and η_0, based on the current sample vector, determine the Bayes selection rule $\boldsymbol{d}^{\mathrm{B}}$ and the empirical Bayes selection rule $\boldsymbol{d}^{*\boldsymbol{n}}$ according (2.6) and (3.7). Then, compute $D_{\boldsymbol{n}}(\boldsymbol{x})$.
(5) Repeat step (1) through step (4) 10 000 times, and then take its average denoted by $\bar{D}_{\boldsymbol{n}}$ which is used as an estimate of $E_{\boldsymbol{n}}\{r(\boldsymbol{d}^{*\boldsymbol{n}})\} - r(\boldsymbol{d}^{\mathrm{B}})$. In addition, $SE(\bar{D}_{\boldsymbol{n}})$,the estimated standard error and $N\bar{D}_{\boldsymbol{n}}$ are also computed.

Simulation results are summarized in Tables 1–4 in which the parameters of selection criterion are respectively given by $\eta_i^{(1)} = \theta_i + \Phi^{-1}(0.9)\sigma_i$, $\eta_i^{(2)} = \theta_i/\sigma_i$, $\eta_i^{(3)} = (\theta_i - \theta_0)(\sigma_i^2 + \sigma_0^2)^{-1/2}$, and $\eta_i^{(4)} = \theta_i + \sigma_i^2/2$, which have been considered in Section 5. In Tables 1–4, we take $k = 4$, $M_i = M = 5$, $n_i = n$ (i.e. $N = n$), $\mu_i = \sigma_i^2 = i$, $\tau_i^2 = 1$, $i = 1, \ldots, 4$, $\alpha = 0.5$, $\theta_0 = \sigma_0^2 = 2.5$, and $\eta_0 = g_1(\sigma_0^2)\theta_0 + g_2(\sigma_0^2)$. The relative frequency that the population selected according to the proposed empirical Bayes selection rule coincides with that selected by the Bayes selection rule is computed and denoted by $f_{\boldsymbol{n}}$. It can be seen from Tables 1–4 that values of $\bar{D}_{\boldsymbol{n}}$ decrease quite rapidly as n increases. The performance of the proposed empirical Bayes selection rules behave satisfactorily when $n \geqslant 40$. Also, the value of $n\bar{D}_{\boldsymbol{n}}$ oscillates locally while it decreases globally as n increases. This supports Theorem 4.1 that the rate of convergence is at least of order

Table 2
Behavior of empirical Bayes selection rules when $\eta_i^{(2)} = \theta_i/\sigma_i$

n	$f_{\boldsymbol{n}}$	$\bar{D}_{\boldsymbol{n}}$	$n\bar{D}_{\boldsymbol{n}}$	$SE(\bar{D}_{\boldsymbol{n}})$
20	0.8041	2.0550×10^{-2}	4.1101×10^{-1}	6.6991×10^{-4}
40	0.8905	7.7007×10^{-3}	3.0803×10^{-1}	3.5467×10^{-4}
60	0.9260	4.0101×10^{-3}	2.4061×10^{-1}	2.5312×10^{-4}
80	0.9401	2.3060×10^{-3}	1.8448×10^{-1}	1.4981×10^{-4}
100	0.9505	1.4891×10^{-3}	1.4891×10^{-1}	1.0143×10^{-4}
200	0.9668	5.9964×10^{-4}	1.1993×10^{-1}	4.3995×10^{-5}
300	0.9712	4.2832×10^{-4}	1.2850×10^{-1}	3.5850×10^{-5}
400	0.9807	2.3013×10^{-4}	9.2052×10^{-2}	2.2665×10^{-5}
500	0.9789	2.4893×10^{-4}	1.2446×10^{-1}	2.4819×10^{-5}
600	0.9815	1.9500×10^{-4}	1.1700×10^{-1}	2.1062×10^{-5}
700	0.9805	1.5893×10^{-4}	1.1125×10^{-1}	1.5587×10^{-5}
800	0.9867	1.0963×10^{-4}	8.7707×10^{-2}	1.3204×10^{-5}
900	0.9844	1.1864×10^{-4}	1.0677×10^{-1}	1.2894×10^{-5}
1000	0.9858	1.1367×10^{-4}	1.1367×10^{-1}	1.3306×10^{-5}

Table 3
Behavior of empirical Bayes selection rules when $\eta_i^{(3)} = (\theta_i - \theta_0)(\sigma_i^2 + \sigma_0^2)^{-1/2}$

n	$f_{\boldsymbol{n}}$	$\bar{D}_{\boldsymbol{n}}$	$n\bar{D}_{\boldsymbol{n}}$	$SE(\bar{D}_{\boldsymbol{n}})$
20	0.7864	2.4795×10^{-2}	4.9591×10^{-1}	6.9934×10^{-4}
40	0.8723	9.6728×10^{-3}	3.8691×10^{-1}	3.9064×10^{-4}
60	0.8993	5.9760×10^{-3}	3.5856×10^{-1}	2.8090×10^{-4}
80	0.9224	3.4700×10^{-3}	2.7760×10^{-1}	1.8737×10^{-4}
100	0.9278	2.9127×10^{-3}	2.9127×10^{-1}	1.6209×10^{-4}
200	0.9515	1.1360×10^{-3}	2.2720×10^{-1}	6.9261×10^{-5}
300	0.9637	7.5687×10^{-4}	2.2706×10^{-1}	5.1834×10^{-5}
400	0.9675	5.4517×10^{-4}	2.1807×10^{-1}	4.0938×10^{-5}
500	0.9685	4.4958×10^{-4}	2.2479×10^{-1}	3.4530×10^{-5}
600	0.9696	3.8113×10^{-4}	2.2868×10^{-1}	2.8698×10^{-5}
700	0.9731	3.0783×10^{-4}	2.1548×10^{-1}	2.5120×10^{-5}
800	0.9751	2.8380×10^{-4}	2.2704×10^{-1}	2.3748×10^{-5}
900	0.9765	2.5210×10^{-4}	2.2689×10^{-1}	2.1380×10^{-5}
1000	0.9761	2.3985×10^{-4}	2.3985×10^{-1}	2.0894×10^{-5}

$O((\ln N)^2/N)$. These results may also indicate that the best obtainable rate of convergence is of order $O(1/N)$. We also consider a quantity $f_{\boldsymbol{n}}/(n\bar{D}_{\boldsymbol{n}})$ which combines both correct selection frequency and difference of the empirical Bayes risk from the Bayes risk. The larger the value of $f_{\boldsymbol{n}}/(n\bar{D}_{\boldsymbol{n}})$, the higher the efficiency of the empirical Bayes rule. Finally, in Figs. 1–3, we respectively draw the quantities of $f_{\boldsymbol{n}}$, $\bar{D}_{\boldsymbol{n}}/SE(\bar{D}_{\boldsymbol{n}})$ and $f_{\boldsymbol{n}}/(n\bar{D}_{\boldsymbol{n}})$ with respect to n for each case of $\eta_i^{(1)}$, $\eta_i^{(2)}$, $\eta_i^{(3)}$, and $\eta_i^{(4)}$ with equal $M_i = 5$ and equal n_i.

Table 4
Behavior of empirical Bayes selection rules when $\eta_i^{(4)} = \theta_i + \sigma_i^2/2$

n	f_n	$\bar{D}_n$	$n\bar{D}_n$	$SE(\bar{D}_n)$
20	0.8495	1.2978×10^{-2}	2.5957×10^{-1}	5.1314×10^{-4}
40	0.9242	4.6423×10^{-3}	1.8570×10^{-1}	2.6054×10^{-4}
60	0.9525	2.5035×10^{-3}	1.5021×10^{-1}	1.8546×10^{-4}
80	0.9667	1.3497×10^{-3}	1.0798×10^{-1}	1.1241×10^{-4}
100	0.9753	8.8269×10^{-4}	8.8269×10^{-2}	1.0002×10^{-4}
200	0.9878	2.2617×10^{-4}	4.5235×10^{-2}	2.5858×10^{-5}
300	0.9893	1.2700×10^{-4}	3.8100×10^{-2}	1.7801×10^{-5}
400	0.9892	1.2399×10^{-4}	4.9595×10^{-2}	1.5311×10^{-5}
500	0.9910	9.1090×10^{-5}	4.5545×10^{-2}	1.2259×10^{-5}
600	0.9914	1.0104×10^{-4}	6.0624×10^{-2}	1.3772×10^{-5}
700	0.9932	6.9747×10^{-5}	4.8823×10^{-2}	1.0832×10^{-5}
800	0.9931	6.0182×10^{-5}	4.8146×10^{-2}	9.3833×10^{-6}
900	0.9920	6.2221×10^{-5}	5.6000×10^{-2}	8.7588×10^{-6}
1000	0.9939	4.4816×10^{-6}	4.4816×10^{-2}	7.3154×10^{-6}

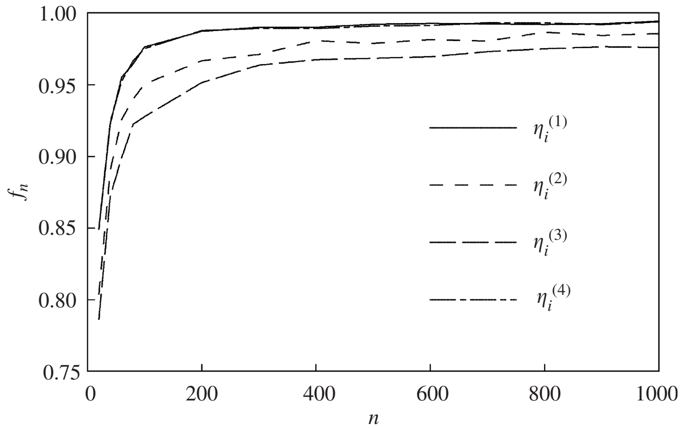

Fig. 1. Plots of f_n with respect to n.

7. Conclusions and discussion

Most literature in decision theory focus only on single criterion. In this paper we consider a selection problem with three criteria in which both a parameter of interest and the dispersion parameter are involved. An empirical Bayes selection rule is proposed and which has been shown to be asymptotically optimal with convergence rate of order $O((\ln N)^2/N)$. A simulation study has been carried out and it shows that the performance of the proposed empirical Bayes selection rule is rather acceptable.

Furthermore, for given control values η_0, θ_0, and σ_0^2, we are interested in selecting some population whose parameter of interest η_i is the largest in the qualified subset in which each parameter of interest is larger than η_0 and whose mean and variance should be no smaller than

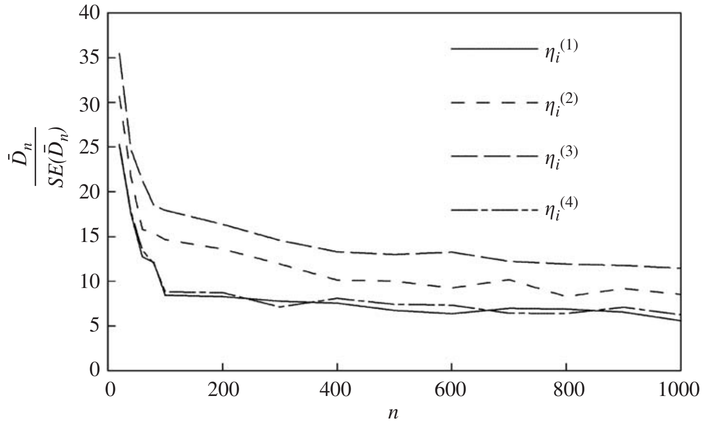

Fig. 2. Plots of $\bar{D}_{\boldsymbol{n}}/SE(\bar{D}_{\boldsymbol{n}})$ with respect to n.

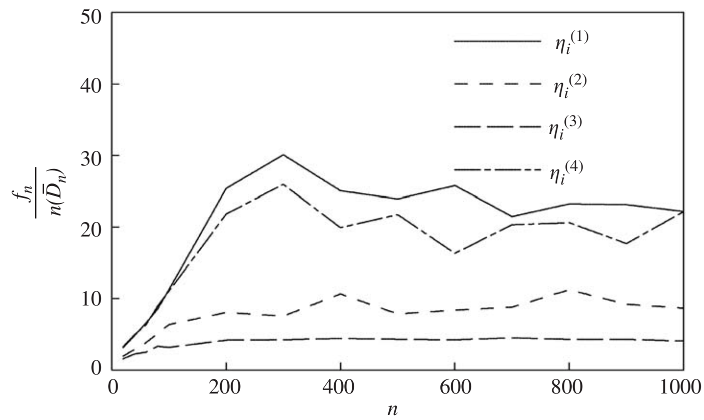

Fig. 3. Plots of $f_{\boldsymbol{n}}/(n\bar{D}_{\boldsymbol{n}})$ with respect to n.

θ_0 and no larger than σ_0^2, respectively. This is a selection problem with four criteria in which parameter of interest η_i, location parameter θ_i, and dispersion parameter σ_i^2 are all involved. Under same formulation, the loss function can be defined by

$$\begin{aligned} L(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\sigma}; \boldsymbol{a}) &= L(\boldsymbol{\eta}', \boldsymbol{\theta}, \boldsymbol{\sigma}; \boldsymbol{a}) \\ &= \alpha_1 \left\{ \max(\eta'_{[k]}, \eta_0) - \sum_{i=0}^{k} a_i \eta'_i \right\} + \alpha_2 \sum_{i=0}^{k} a_i (\theta_0 - \theta_i) I(\theta_i < \theta_0) \\ &\quad + (1 - \alpha_1 - \alpha_2) \sum_{i=0}^{k} a_i \left(\frac{\sigma_i}{\sigma_0} - 1 \right) I(\sigma_i > \sigma_0) \end{aligned}$$

for prefixed α_1 and α_2 such that $\alpha_1, \alpha_2 \geqslant 0$, $\alpha_1 + \alpha_2 \leqslant 1$, where

$$\eta'_i = \begin{cases} \eta_0 & \text{if } i = 0 \text{ or } \theta_i < \theta_0 \text{ or } \sigma_i > \sigma_0, \\ \eta_i & \text{otherwise} \end{cases}$$

and $\eta'_{[k]} = \max_{1 \leqslant i \leqslant k} \eta'_i$. The Bayes selection rule and the empirical Bayes selection rule can be obtained by using a similar approach in this paper. However, the corresponding convergence rate of the empirical Bayes selection rule may not be obtained analogously.

On the other hand, it is worthwhile to consider a general location-scale model. More study is needed for this model since it covers a large family and important in practical applications.

Acknowledgements

The authors would like to thank two referees for helpful comments and suggestions which have improved presentation of this paper. This research of the first author was partially supported by NSC 92-2213-E-032-016, and the second author was partially supported by NSC 89-2118-M-031-002 from the National Science Council of Taiwan, R.O.C.

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.

Ghosh, M., Lahiri, P., 1987. Robust empirical estimation of means from stratified sample. J. Amer. Statist. Assoc. 82, 1153–1162.

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

Gupta, S.S., 1956. On a decision rule for a problem in ranking means. Mimeograph Series No. 150. Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Gupta, S.S., 1965. On some multiple decision (selection and ranking) rules. Technometrics 7, 225–245.

Gupta, S.S., Panchapakesan, S., 1979. Multiple Decision Procedures. Wiley, New York.

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

Huang, W.T., Lai, Y.T., 1999. Empirical Bayes procedures for selecting the best population with multiple criteria. Ann. Inst. Statist. Math. 51, 281–299.

Huang, W.T., Wu, J.S., Lai, Y.T., 1998. Note on empirical Bayes procedures for selecting the best population with conjugate prior. Internat. J. Inform. Manage. Sci. 9, 1–10.