ON SIMULTANEOUS SELECTION OF GOOD POPULATIONS

by

Shanti S. Gupta and TaChen Liang

Purdue University | Wayne State University

Technical Report #95-42C

Department of Statistics

Purdue University

September 1995

Revised October 1996

Revised December 1997

Revised January 1998

◇

ON SIMULTANEOUS SELECTION OF GOOD POPULATIONS*

Shanti S. Gupta and TaChen Liang

Received: revised version:

Abstract

This paper considers the problem of simultaneous selection of k populations in comparison with a known standard. The empirical Bayes method is employed to incorporate information from among the k populations to improve the decision for each of the k component problems. Under the assumption of the linearity of the posterior mean, an empirical Bayes procedure for simultaneous selection of good populations is derived. The asymptotic optimality of the empirical Bayes selection procedure is investigated. Three statistical distribution models are studied in detail. For each of the three models, it is shown that the relative regret risk of the empirical Bayes selection procedure has a rate of convergence of order $O(k^{-1})$.

1. Introduction

Consider k independent populations $\pi_1, \ldots, \pi_k$. For each i, the population π_i is characterized by a parameter θ_i. Let θ_0 denote a known standard or an unknown control value. A population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. In this paper, we study the problem of selecting all good populations as compared with a known standard θ_0. This selection problem may arise, for example, in the industrial applications in which one may be interested in finding all the manufacturing processes with product quality satisfying a specified standard. Also, in the developing stage of clinical trials, an experimenter may be interested in seeking a potential drug (treatment) which would achieve a certain requirement of efficacy as compared with the standard. In the literature, this selection problem

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165.

AMS 1991 Subject classifications: 62C12, 62F07.

Keywords and phrases: Asymptotically optimal, empirical Bayes, rate of convergence, posterior linearity, simultaneous selection.

has been studied extensively. Bechhofer [1] has made multiple comparisons with a control for multifactor experiments involving variances of normal populations. Bechhofer [2], Bechhofer and Nocturne [3] and Bechhofer and Turnbull [5] have considered the problem of optimally allocating the observations when comparing several treatments with a control. Dunnett [9] and Gupta and Sobel [22] have considered problems of selecting a subset containing all populations better than a control. Lehmann [23] and Spjøtvoll [30] have treated the problem using methods from the theory of testing hypothesis. Randles and Hollander [27], Miescke [26] and Gupta and Miescke [18] have derived optimal procedures via the minimax or Γ-minimax approaches. Gupta and Liang [15] have developed empirical Bayes simultaneous testing procedures for Poisson populations.

It should be noted that the problem of comparing populations with a control under different types of formulation has been investigated in the literature. We mention a few here: Bechhofer and Turnbull [6], Dunnett [10], Wilcox [31], Dudewicz and Taneja [8] and Gupta, Liang and Rau [17] have discussed problems of selecting the best population provided it is better than a control. Gupta and Singh [21] and Gupta and Hsiao [14] have derived Bayes, Γ-minimax and minimax procedures for selecting populations close to a control. Mee, Shah and Lefante [25] have developed multiple testing procedures to compare the means of k normal populations with respect to a control. Liang [24] has studied simultaneous selection procedures for selecting normal populations close to a control based on Kullback-Leibler discrimination information. Also, Chapter 5 of Bechhofer, Santner and Goldsman [4], Chapter 20 of Gupta and Panchapakesan [19], and Gupta and Panchapakesan [20] have provided comprehensive overviews and good references on this research area.

In this article, we study the problem of selecting good populations versus a control using an empirical Bayes approach. A simultaneous empirical Bayes selection procedure is constructed under the assumption that the posterior expectation of the parameter θ_i is a linear function of the sample observations. Such a property is, thus, referred to as posterior linearity. This assumption holds in many statistical models, for example, when the sampling distribution is normal distribution with mean θ, and θ also follows a normal prior distribution; or, the random observation follows a binomial $B(n, \theta)$ distribution and θ has a Beta prior distribution. For more general statistical models with posterior linearity see Goldstein [13] and Diaconis and Ylvisaker [7]. Also Ghosh and Lahiri [12] have studied some empirical Bayes estimation of means from stratified samples under the posterior linearity assumption.

This paper is organized as follows. We formulate the selection problem in Section 2. A Bayes selection procedure to the selection problem is derived. Since the Bayes selection procedure depends on certain unknown parameters, in Section 3, by incorporating information from each of the k populations, we first present estimates for the unknown parameters. Then, by mimicking the behavior of the Bayes selection procedure, we propose an empirical Bayes procedure. The asymptotic optimality of the empirical Bayes selection procedure is investigated. In Section 4, we study the rate of convergence of the relative regret risk of the empirical Bayes selection procedure for three statistical distribution models. For each of these models, it is shown that the empirical Bayes selection procedure has a rate of convergence of order $O(k^{-1})$.

2. Formulation of the Selection Problem and a Bayes Procedure

Let $\pi_1, \ldots, \pi_k$ be k independent populations with unknown means $\theta_1, \ldots, \theta_k$ respectively. For a given standard θ_0, population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. Our goal is to derive selection procedures to select all good populations and to exclude all bad populations.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | \theta_i \in \mathcal{I},\ i = 1, \ldots, k\}$ be the parameter space, where $\mathcal{I}$ is an interval in R. Let $\underset{\sim}{a} = (a_1, \ldots, a_k)$ be an action, where $a_i = 0,\ 1,\ i = 1, \ldots, k$. When $\underset{\sim}{a}$ is taken, $a_i = 1$ means that population π_i is selected as good while $a_i = 0$ means that π_i is excluded as bad. For the parameter $\underset{\sim}{\theta}$ and action $\underset{\sim}{a}$, the loss function is given by

$$L(\underset{\sim}{\theta}, \underset{\sim}{a}) = \sum_{i=1}^{k} L_i(\theta_i, a_i) \tag{2.1}$$

and

$$L_i(\theta_i, a_i) = a_i(\theta_0 - \theta_i)J(\theta_0 - \theta_i) + (1 - a_i)(\theta_i - \theta_0)J(\theta_i - \theta_0), \tag{2.2}$$

where $J(x) = 1(0)$ if $x \geq 0$ (otherwise).

In (2.2), the first term is the loss due to selecting π_i as good when π_i is bad, and the second term is the loss due to excluding π_i as bad while it is good.

For each $i = 1, \ldots, k$, let $X_{i1}, \ldots, X_{im}$ be a sample of size m taken from population π_i. Denote $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im})$ and $\underset{\sim}{X} = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k)$. It is assumed that $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k$ satisfy the following assumptions:

A1. For each $i = 1, \ldots, k$, given θ_i, $X_{i1}, \ldots, X_{im}$ are iid with

$$E[X_{ij}|\theta_i] = \theta_i \quad \text{and} \quad \text{Var}\,(X_{ij}|\theta_i) = V(\theta_i).$$

A2. For each $i = 1, \ldots, k$, θ_i is a realization of a random variable Θ_i; $\Theta_1, \ldots, \Theta_k$ are iid with the mean μ and a finite variance σ^2, and $0 < \tau^2 \equiv E[V(\Theta_i)] < \infty$.

A3. $E[\Theta_i|\underset{\sim}{X}_i] = a\overline{X}_i + b$ for some constants a and b, where $\overline{X}_i = \frac{1}{m} \sum_{j=1}^{m} X_{ij}$.

A4. $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k$ are marginally iid.

Note that these assumptions are satisfied, for example, when X_{ij} follows a normal distribution with mean θ_i and Θ_i also has a normal prior. One may also see that the assumptions hold when X_{ij} belongs to an exponential family with Θ_i having a conjugate prior distribution.

Under the assumptions A1–A4, it follows from Ericson [11] that $a = \sigma^2/(\sigma^2 + \frac{\tau^2}{m})$ and $b = (1-a)\mu$. Hence $E[\Theta_i|\underset{\sim}{X}_i] = a\overline{X}_i + (1-a)\mu$. Thus the posterior mean of Θ_i is a linear function of the sample mean $\overline{X}_i$. This property is referred to as posterior linearity. Note that the posterior mean is also a weighted average of the sample mean $\overline{X}_i$ and the mean μ.

Let $f_i(\underset{\sim}{x}_i|\theta_i)$ denote the probability density function (or probability function in discrete case) of $\underset{\sim}{X}_i$ given θ_i and let $f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{k} f_i(\underset{\sim}{x}_i|\theta_i)$. Also, let $G_i(\theta_i)$ denote the prior distribution of Θ_i, $i = 1, \ldots, k$, and let $G(\underset{\sim}{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$. Note that $G_1, \ldots, G_k$ are identical since $\Theta_1, \ldots, \Theta_k$ are marginally iid. Finally, let $f_i(\underset{\sim}{x}_i) = \int f_i(\underset{\sim}{x}_i|\theta_i) dG_i(\theta_i)$ and $f(\underset{\sim}{x}) = \prod_{i=1}^{k} f_i(\underset{\sim}{x}_i)$.

Let $\mathcal{X}$ be the sample space generated by $\underset{\sim}{X}$. A selection procedure $\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k)$ is a mapping defined on $\mathcal{X}$ into $[0,1]^k$, such that for each i and $\underset{\sim}{x} \in \mathcal{X}$, $\delta_i(\underset{\sim}{x})$ is the probability of selecting population π_i as good when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. Denote the Bayes risk of the selection procedure $\underset{\sim}{\delta}$ by $R_G(\underset{\sim}{\delta}, k)$. Then, with the loss $L(\underset{\sim}{\theta}, \underset{\sim}{a})$ given in (2.1)–(2.2), we have

$$R_G(\underset{\sim}{\delta}, k) = \sum_{i=1}^{k} R_i(\delta_i), \tag{2.3}$$

and

$$R_i(\delta_i) = \int_{\mathcal{X}} \delta_i(\underset{\sim}{x})\{\theta_0 - E[\Theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]\}f(\underset{\sim}{x})d\underset{\sim}{x} + C, \tag{2.4}$$

where $C = E[(\Theta_i - \theta_0)J(\Theta_i - \theta_0)]$.

From (2.4), a Bayes selection procedure $\underset{\sim}{\delta}_B = (\delta_{B1}, \ldots, \delta_{Bk})$ which minimizes the Bayes risks $R_G(\underset{\sim}{\delta}, k)$ among the class of all selection procedures is given by: For each $i = 1, \ldots, k$ and $\underset{\sim}{x} \in \mathcal{X}$,

$$\begin{aligned} \delta_{Bi}(\underset{\sim}{x}) &= \begin{cases} 1 & \text{if } E[\Theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i] \geq \theta_0 \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} 1 & \text{if } \bar{x}_i \geq \frac{(m\sigma^2+\tau^2)\theta_0 - \tau^2\mu}{m\sigma^2} \equiv C_0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2.5}$$

From (2.5), we see that for each i, $\delta_{Bi}(\underset{\sim}{x})$ depends on $\underset{\sim}{x}$ only through the sample mean value $\bar{x}_i$. Let $h_i(\bar{x}_i)$ denote the marginal pdf of $\overline{X}_i$. Note that $h_1 = \cdots = h_k$ since $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k$ are marginally iid. Then,

$$R_G(\underset{\sim}{\delta}_B, k) = \sum_{i=1}^{k} R_i(\delta_{Bi}) \tag{2.6}$$

and

$$R_i(\delta_{Bi}) = \int \delta_{Bi}(\bar{x}_i)\{\theta_0 - \varphi_i(\bar{x}_i)\}h_i(\bar{x}_i)d\bar{x}_i + C. \tag{2.7}$$

where $\varphi_i(\bar{x}_i) = a\bar{x}_i + (1-a)\mu$.

3. Empirical Bayes Approach

Note that, from (2.5), the Bayes selection procedure $\underset{\sim}{\delta}_B$ depends on the parameters τ^2, σ^2 and μ. When the values of these parameters are unknown, it is not possible to implement the Bayes procedure $\underset{\sim}{\delta}_B$ for the selection problem. In such a situation, we incorporate the information from among the k populations and construct estimates for these parameters. Then, by mimicking the behavior of the Bayes procedure $\underset{\sim}{\delta}_B$, we propose an empirical Bayes selection procedure.

For each i, the data $\underset{\sim}{X} = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k)$ is partitioned into two parts, namely, $\underset{\sim}{X}_i$ and $\underset{\sim}{X}(i) = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_{i-1}, \ \underset{\sim}{X}_{i+1}, \ldots, \underset{\sim}{X}_k)$. Let

$$\hat{\mu}(i) = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} \overline{X}_j,$$

$$SSB(i) = \sum_{\substack{j=1 \\ j \neq i}}^{k} m(\overline{X}_j - \hat{\mu}(i))^2, \quad MSB(i) = SSB(i)/(k-1),$$

$$SSW(i) = \sum_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\ell=1}^{m} (X_{j\ell} - \overline{X}_j)^2, \quad MSW(i) = SSW(i)/[(k-1)(m-1)].$$

A straightforward computation shows that $E[MSW(i)] = \tau^2$, $E[MSB(i)] = \tau^2 + m\sigma^2$ and $E[\hat{\mu}(i)] = \mu$. Therefore, we use $\hat{\tau}^2(i) = MSW(i)$, $MSB(i)$ and $\hat{\mu}(i)$ to estimate τ^2, $\tau^2 + m\sigma^2$ and μ, respectively, and use $m\hat{\sigma}^2(i) = \max(0, \ MSB(i) - MSW(i))$ to estimate $m\sigma^2$.

Empirical Bayes Selection Procedure $\underset{\sim}{\delta}^* = (\delta_1^*, \ldots, \delta_k^*)$

By mimicking the Bayes selection procedure $\underset{\sim}{\delta}_B$ of (2.5), we propose an empirical Bayes selection procedure $\underset{\sim}{\delta}^* = (\delta_1^*, \ldots, \delta_k^*)$ as follows: For each $i = 1, \ldots, k$ and $\underset{\sim}{x} \in \mathcal{X}$, define

$$\delta_i^*(\underset{\sim}{x}) = \delta_i^*(\bar{x}_i, \underset{\sim}{x}(i)) = \begin{cases} 1 & \text{if } m\hat{\sigma}^2(i)\bar{x}_i + \hat{\tau}^2(i)\hat{\mu}(i) - MSB(i)\theta_0 \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

When $m\hat{\sigma}^2(i) > 0$, we let $C_i = \frac{MSB(i)\theta_0 - \hat{\tau}^2(i)\hat{\mu}(i)}{m\hat{\sigma}^2(i)}$. Then, the empirical Bayes selection procedure δ_i^* can be expressed as:

$$\delta_i^*(\bar{x}_i, \underset{\sim}{x}(i)) = \begin{cases} 1 & \text{if } \bar{x}_i \geq C_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1$'$}$$

Asymptotic Optimality of the Selection Procedure $\underset{\sim}{\delta}^*$

We denote the Bayes risk of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ by $R_G(\underset{\sim}{\delta}^*, k)$. Then,

$$R_G(\underset{\sim}{\delta}^*, k) = \sum_{i=1}^{k} R_i(\delta_i^*) \tag{3.2}$$

where $R_i(\delta_i^*) = E_i[R_i(\delta_i^* | \underset{\sim}{X}(i))]$ and

$$R_i(\delta_i^* | \underset{\sim}{X}(i)) = \int \delta_i^*(\bar{x}_i, \underset{\sim}{X}(i))[\theta_0 - \varphi_i(\bar{x}_i)]h_i(\bar{x}_i)d\bar{x}_i + C. \tag{3.3}$$

Note that $R_i(\delta_i^* | \underset{\sim}{X}(i))$ is the conditional Bayes risk of the ith component selection procedure δ_i^* conditioning on $\underset{\sim}{X}(i)$ and the expectation E_i is taken with respect to the probability measure generated by $\underset{\sim}{X}(i)$.

Since $\underset{\sim}{\delta}_B$ is the Bayes selection procedure, $R_i(\delta_i^* | \underset{\sim}{X}(i)) \geq R_i(\delta_{Bi})$ for all $\underset{\sim}{X}(i)$ and all $i = 1, \ldots, k$. Therefore, $R_i(\delta_i^*) \geq R_i(\delta_{Bi})$, $i = 1, \ldots, k$. Hence, $R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k) = \sum_{i=1}^{k} [R_i(\delta_i^*) - R_i(\delta_{Bi})] \geq 0$. $R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)$ is called the regret risk of the selection procedure $\underset{\sim}{\delta}^*$. Define $\rho_G(\underset{\sim}{\delta}^*, k) = [R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)]/R_G(\underset{\sim}{\delta}_B, k)$. Then, $\rho_G(\underset{\sim}{\delta}^*, k)$ is called the relative regret risk of the procedure $\underset{\sim}{\delta}^*$. In the following, $\rho_G(\underset{\sim}{\delta}^*, k)$ is used as a measure of performance of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$.

Definition 3.1 (a) A selection procedure $\underset{\sim}{\delta}$ is said to be asymptotically optimal relative to the prior distribution G if $\rho_G(\underset{\sim}{\delta}, k) \to 0$ as $k \to \infty$.

(b) A selection procedure $\underset{\sim}{\delta}$ is said to be asymptotically optimal relative to the prior distribution G with a rate of convergence of order $\{\varepsilon_k\}$ if $\rho_G(\underset{\sim}{\delta}, k) = O(\varepsilon_k)$ where $\{\varepsilon_k\}$ is a sequence of positive numbers such that $\lim_{k\to\infty} \varepsilon_k = 0$.

According to the statistical model described before, one can see that $R_i(\delta_i^* | \underset{\sim}{X}(i)), i = 1, \ldots, k$, are identically distributed. Therefore, $R_1(\delta_1^*) = \cdots = R_k(\delta_k^*)$. Also, for the Bayes selection procedure $\underset{\sim}{\delta}_B$, $R_1(\delta_{B1}) = \cdots = R_k(\delta_{Bk})$. Hence, $R_G(\underset{\sim}{\delta}^*, k) = kR_1(\delta_1^*)$. $R_G(\underset{\sim}{\delta}_B, k) = kR_1(\delta_{B1})$, and $\rho_G(\underset{\sim}{\delta}^*, k) = [R_1(\delta_1^*) - R_1(\delta_{B1})]/R_1(\delta_{B1})$, where $R_1(\delta_{B1})$ is a fixed positive value. That is, the relative regret risk of the selection procedure $\underset{\sim}{\delta}^*$ and the regret risk of the component procedure δ_1^* have the same asymptotic behavior. Therefore, it suffices to investigate the asymptotic behavior of the difference $R_1(\delta_1^*) - R_1(\delta_{B1})$ for sufficiently large k.

From (2.5), (2.7), (3.1) and (3.3),

$$\begin{aligned}
0 &\leq R_1(\delta_1^*) - R_1(\delta_{B1}) \\
&= \int E_1\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) - \delta_{B1}(\bar{x}_1)\}[\theta_0 - \varphi_1(\bar{x}_1)]h_1(\bar{x}_1)d\bar{x}_1 \\
&= \int_{-\infty}^{c_0} P\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) = 1, \delta_{B1}(\bar{x}_1) = 0\}[\theta_0 - \varphi_1(\bar{x}_1)]h_1(\bar{x}_1)d\bar{x}_1 \\
&\quad + \int_{c_0}^{\infty} P\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) = 0, \delta_{B1}(\bar{x}_1) = 1\}[\varphi_1(\bar{x}_1) - \theta_0]h_1(\bar{x}_1)d\bar{x}_1
\end{aligned} \tag{3.4}$$

where

$$\begin{aligned}
&P\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) = 1,\ \delta_{B1}(\bar{x}_1) = 0\} \\
&= P\{m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1) = \hat{\mu}(1) - MSB(1)\theta_0 \geq 0, \\
&\qquad m\sigma^2\bar{x}_1 + \tau^2\mu - (m\sigma^2 + \tau^2)\theta_0 < 0\},
\end{aligned} \tag{3.5}$$

and

$$\begin{aligned}
&P\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) = 0,\ \delta_{B1}(\bar{x}_1) = 1\} \\
&= P\{m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0, \\
&\qquad m\sigma^2\bar{x}_1 + \tau^2\mu - (m\sigma^2 + \tau^2)\theta_0 \geq 0\}.
\end{aligned} \tag{3.6}$$

To show the asymptotic optimality of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ we proceed as follows. By Corollary 1 of Robbins [28], it suffices to prove that $P\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) = 1,\ \delta_{B1}(\bar{x}_1) = 0\} \to 0$ and $P\{\delta_1^*(\bar{x}_1, \underset{\sim}{X}(1)) = 0,\ \delta_{B1}(\bar{x}_1) = 1\} \to 0$ as $k \to \infty$. For this purpose, one only need to show that for each $\bar{x}_1$, $[m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1)\hat{\mu}(1))] - MSB(1)\theta_0$ converges to $(m\sigma^2\bar{x}_1 + \tau^2\mu) - (m\sigma^2 + \tau^2)\theta_0$ in probability, which is true, since $m\hat{\sigma}^2(1)$, $\hat{\tau}^2(1)$, $\hat{\mu}(1)$ and $MSB(1)$ are consistent estimates of $m\sigma^2$, τ^2, μ and $m\sigma^2 + \tau^2$, respectively. Hence, we have the following theorem.

Theorem 3.1. Let $\underset{\sim}{\delta}^*$ be the empirical Bayes selection procedure constructed previously. Then, under the assumptions A1–A4, $\underset{\sim}{\delta}^*$ is asymptotically optimal. That is, $\rho_G(\underset{\sim}{\delta}^*, k) \to 0$ as $k \to \infty$.

Comparison with a Natural Selection Procedure

Consider a natural selection procedure $\underset{\sim}{\delta}^N = (\delta_1^N, \ldots, \delta_k^N)$ which is based on the sample means and defined as follows:

$$d_i^N(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \bar{x}_i \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

The selection criterion of the procedure $\underset{\sim}{\delta}^N$ is simple and natural. It just compares the sample means with the control θ_0, and selects those populations with sample mean values at least equal to θ_0 as good populations. The Bayes risk associated with the selection procedure $\underset{\sim}{\delta}^N$ is:

$$R_G(\underset{\sim}{\delta}^N, k) = \sum_{i=1}^{k} R_i(\delta_i^N), \tag{3.8}$$

and

$$R_i(\delta_i^N) = \int \delta_i^N(\bar{x}_i)[\theta_0 - \varphi_i(\bar{x}_i)]h_i(\bar{x}_i)d\bar{x}_i + C. \tag{3.9}$$

Note that $R_1(\delta_1^N) = \cdots = R_k(\delta_k^N)$. Hence the regret risk and relative regret risk of $\underset{\sim}{\delta}^N$ with respect to the Bayes procedure $\underset{\sim}{\delta}_B$ are $R_G(\underset{\sim}{\delta}^N, k) - R_G(\underset{\sim}{\delta}_B, k) = k[R_1(\delta_1^N) - R_1(\delta_{B1})]$ and $\rho_G(\underset{\sim}{\delta}^N, k) = [R_1(\delta_1^N) - R_1(\delta_{B1})]/R_1(\delta_{B1})$, respectively. Since $R_1(\delta_1^N) - R_1(\delta_{B1})$ is a positive constant and independent of k, the natural selection procedure $\underset{\sim}{\delta}^N$ does not possess the asymptotic optimality. We define the relative performance of $\underset{\sim}{\delta}^*$ with respect to $\underset{\sim}{\delta}^N$ as:

$$\rho_G(\underset{\sim}{\delta}^*, \underset{\sim}{\delta}^N, k) = \frac{\rho_G(\underset{\sim}{\delta}^N, k)}{\rho_G(\underset{\sim}{\delta}^*, k)} = \frac{R_1(\delta_1^N) - R_1(\delta_{B1})}{R_1(\delta_1^*) - R_1(\delta_{B1})}.$$

Since $\underset{\sim}{\delta}^*$ is asymptotically optimal, while $R_1(\delta_1^N) - R_1(\delta_{B1})$ is a fixed positive value independent of k, it can be seen that $\rho_G(\underset{\sim}{\delta}^*, \underset{\sim}{\delta}^N, k) \to \infty$ as $k \to \infty$.

Small Sample Performance: A Simulation Study

A simulation study was carried out to investigate the small sample performance of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$. For each i, let $D_i(\underset{\sim}{X}) = [\delta_i^*(\overline{X}_i, \underset{\sim}{X}(i)) - \delta_{Bi}(\overline{X}_i)][\theta_0 - \varphi_i(\overline{X}_i)]$, and $D(\underset{\sim}{X}, k) = \sum_{i=1}^{k} D_i(\underset{\sim}{X})$. Note that

$$\begin{aligned} &ED_i(\underset{\sim}{X}) \\ &= E_i[E_{(i)}\{[\delta_i^*(\overline{X}_i, \underset{\sim}{X}(i)) - \delta_{Bi}(\overline{X}_i)][\theta_0 - \varphi_i(\overline{X}_i)]\}] \\ &= E_i[R_i(\delta_i^*|\underset{\sim}{X}(i)) - R_i(\delta_{Bi})] \\ &= R_i(\delta_i^*) - R_i(\delta_{Bi}). \end{aligned} \tag{3.10}$$

In (3.10), the expectation $E_{(i)}$ is taken with respect to the probability measure generated by $\underset{\sim}{X}_i$, and E_i denotes the expectation taken with respect to the probability measure generated by $\underset{\sim}{X}(i)$. Therefore, from (3.10), one can see that $ED(\underset{\sim}{X}, k) = R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)$.

Note that $\rho_G(\underset{\sim}{\delta}^*, k) = [R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)]/[kR_1(\delta_{B1})]$, where $R_1(\delta_{B1})$ is a fixed value, independent of the value k. Hence, $\rho_G(\underset{\sim}{\delta}^*, k)$ depends on k only through the part $[R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)]/k = E[D(\underset{\sim}{X}, k)/k]$.

Let $\overline{D}(k)$ be the sample mean based on n independent repetitions of $D(\underset{\sim}{X}, k)$'s. By the law of large numbers, $\overline{D}(k)$ is a consistent estimator of $R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)$. Therefore, we use $\overline{D}(k)/k$ to estimate $[R_G(\underset{\sim}{\delta}^*, k) - R_G(\underset{\sim}{\delta}_B, k)]/k$.

Two statistical distribution models were used for the simulation study. They are normal-normal model and binomial-beta model, respectively. In the simulation study, the random variables are generated by using the subroutines DRNNOA, RNBET and DRNUN from IMSL STAT/LIBRARY. The simulation scheme used in this paper is described as follows.

(1) For each $i = 1, \ldots, k$, generate $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im})$ by the following:

(a) Normal-normal model:

(a^1) Generate Θ_i from a $N(\mu, \sigma^2)$ distribution;

(a^2) Given $\Theta_i = \theta_i$, generate random sample $X_{i1}, \ldots, X_{im}$ from a $N(\theta_i, \tau^2)$ distribution.

(b) Binomial-beta model:

(b^1) Generate Θ_i/N from a Beta (p, q) distribution,

(b^2) Given $\Theta_i = \theta_i$, generate random sample $X_{i1}, \ldots, X_{im}$ from a $B(N, \theta_i/N)$ distribution.

(2) Based on the data $\underset{\sim}{X} = (\underset{\sim}{X}, \ldots, \underset{\sim}{X}_k)$, construct the Bayes and the empirical Bayes selection procedures $\underset{\sim}{\delta}_B$ and $\underset{\sim}{\delta}^*$ and compute $D(\underset{\sim}{X}, k)$.

(3) For each k, steps (1) and (2) were repeated n=1000 times. Then $\overline{D}(k)$ is the sample mean based on n=1000 independent repetitions of $D(\underset{\sim}{X}, k)$'s

The results of the simulated small sample performance of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ based on 1000 repetitions are reported in Tables 1 and 2. In the Tables, S.D. denotes the estimated standard deviation of $\overline{D}(k)/k$. In each of the two tables, the $\overline{D}(k)$

decreases in k quickly for $k \le 20$. For the normal-normal model case, for moderate values of k, $\overline{D}(k)$ seems oscillating about some value, say 1.5000×10^{-3}. For the binomial-beta model case, it seems that $\overline{D}(k)$ tends to decrease in k for moderate values of k. These results indicate that the relative regret risk of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ has a rate of convergence of order $O(k^{-1})$.

Note: In Table 2, for binomial-beta case, it is assumed that for each $i = 1, \ldots, k, X_{ij} \sim B(N, \frac{\theta_i}{N}), j = 1, \ldots, m, \frac{\theta_i}{N} \sim \text{Beta}(p, q)$. Therefore, $\mu = \frac{Np}{p+q}, \sigma^2 = N^2pq/[(p+q)^2(p+q+1)]$ and $\tau^2 = Npq/[(p+q)(p+q+1)]$.

4. Rate of Convergence

In this section, we study the rate of convergence of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$. Three statistical distribution models are investigated: (1) normal-normal; (2) Poisson-Gamma, and (3) Case where the random variables are bounded, $|X_{ij}| \le B$, and $0 \le h_i(\overline{x}_i|\theta_i) \le M$ for all $\overline{x}_i$ and θ_i, but both the sampling density function $h_i(\overline{x}_i|\theta_i)$ and the prior distribution $G_i(\theta_i)$ are unknown.

Case 1. Normal-Normal Case

It is assumed that conditioning on θ_i, X_{ij}, $j = 1, \ldots, m$, is a sample of size m arising from a normal population π_i with mean θ_i and variance τ^2. Also, the parameter θ_i is a realization of a normal random variable Θ_i with mean μ and variance σ^2. Hence, given θ_i, $\overline{X}_i = \frac{1}{m}\sum_{j=1}^{m} X_{ij} \sim N(\theta_i, \frac{\tau^2}{m})$, and marginally, $\overline{X}_i \sim N(\mu, \sigma^2 + \frac{\tau^2}{m})$. Hence, $h_i(\overline{x}_i)$ is the probability density function of a $N(\mu, \sigma^2 + \frac{\tau^2}{m})$ distribution. In the following analysis, we assume that $\mu < \theta_0$ and $\theta_0 > 0$, so that $c_0 = \frac{\tau^2(\theta_0-\mu)}{m\sigma^2} + \theta_0 > \theta_0$. From (3.4)–(3.6),

$$
\begin{aligned}
&R_1(\delta_1^*) - R_1(\delta_{B1}) \\
&\le \int_{-\infty}^{\mu} [\theta_0 - \varphi_1(\bar{x}_1)] h_1(\bar{x}_1) P\{m\hat{\sigma}^2(1) > 0,\ m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1)\hat{\mu}(1) \\
&\qquad\qquad - MSB(1)\theta_0 \ge 0\} d\bar{x}_1 \\
&+ \int_{\mu}^{\theta_0} [\theta_0 - \varphi_1(\bar{x}_1)] h_1(\bar{x}_1) P\{m\hat{\sigma}^2(1) > 0, m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1)\hat{\mu}(1) \\
&\qquad\qquad - MSB(1)\theta_0 \ge 0\} d\bar{x}_1
\end{aligned}
$$

Table 1. Simulated Small Sample Performance of $\underset{\sim}{\delta}^*$ for Normal-Normal Case with $\mu = 10,\ \sigma^2 = 2,\ \tau^2 = 1.5,\ \theta_0 = 10$ and $m = 10$.

k	$\overline{D}(k)/k$	$\overline{D}(k)$	S.D.
10	12.5432×10^{-4}	12.5432×10^{-3}	92.8273×10^{-4}
20	1.7004×10^{-4}	3.4008×10^{-3}	7.4388×10^{-4}
30	0.7050×10^{-4}	2.1151×10^{-3}	2.4761×10^{-4}
40	0.5324×10^{-4}	2.1294×10^{-3}	1.7029×10^{-4}
60	0.3153×10^{-4}	1.8917×10^{-3}	1.0336×10^{-4}
80	0.1856×10^{-4}	1.4848×10^{-3}	0.7394×10^{-4}
100	0.1599×10^{-4}	1.5993×10^{-3}	0.4729×10^{-4}
120	0.1522×10^{-4}	1.8258×10^{-3}	0.4178×10^{-4}
140	0.0991×10^{-4}	1.3871×10^{-3}	0.3215×10^{-4}
160	0.1076×10^{-4}	1.7208×10^{-3}	0.2710×10^{-4}
180	0.0808×10^{-4}	1.4552×10^{-3}	0.2004×10^{-4}
200	0.0751×10^{-4}	1.5009×10^{-3}	0.2108×10^{-4}
240	0.0537×10^{-4}	1.2892×10^{-3}	0.1473×10^{-4}
280	0.0544×10^{-4}	1.5225×10^{-3}	0.1346×10^{-4}
320	0.0472×10^{-4}	1.5096×10^{-3}	0.1107×10^{-4}
360	0.0389×10^{-4}	1.4005×10^{-3}	0.0996×10^{-4}
400	0.0356×10^{-4}	1.4240×10^{-3}	0.0878×10^{-4}

Table 2. Simulated Small Sample Performance of $\underset{\sim}{\delta}^*$ for Binomial (N)-Beta(p, q) Case with $N = 10,\ p = 2,\ q = 1,\ \theta_0 = 5.5$ and $m = 5$.

k	$\overline{D}(k)/k$	$\overline{D}(k)$	S.D.
10	24.4245×10^{-4}	24.4245×10^{-3}	184.9388×10^{-4}
20	3.0943×10^{-4}	6.1887×10^{-3}	13.8346×10^{-4}
30	2.0912×10^{-4}	6.2736×10^{-3}	11.1489×10^{-4}
40	1.4222×10^{-4}	5.6886×10^{-3}	4.0207×10^{-4}
60	0.8726×10^{-4}	5.2358×10^{-3}	2.7469×10^{-4}
80	0.7535×10^{-4}	6.0283×10^{-3}	2.5808×10^{-4}
100	0.6369×10^{-4}	6.3679×10^{-3}	2.1940×10^{-4}
120	0.4717×10^{-4}	5.6603×10^{-3}	2.0161×10^{-4}
140	0.4124×10^{-4}	5.7735×10^{-3}	1.7238×10^{-4}
160	0.4175×10^{-4}	6.6792×10^{-3}	1.6754×10^{-4}
180	0.2500×10^{-4}	4.5000×10^{-3}	1.2816×10^{-4}
200	0.2561×10^{-4}	5.1226×10^{-3}	1.4074×10^{-4}
240	0.1675×10^{-4}	4.0188×10^{-3}	1.1077×10^{-4}
280	0.1071×10^{-4}	3.0000×10^{-3}	0.8258×10^{-4}
320	0.0531×10^{-4}	1.6981×10^{-3}	0.5400×10^{-4}
360	0.0613×10^{-4}	2.2075×10^{-3}	0.6409×10^{-4}
400	0.0340×10^{-4}	1.3584×10^{-3}	0.4304×10^{-4}

$$
\begin{aligned}
&+\int_{\theta_0}^{c_0}[\theta_0-\varphi_1(\bar{x}_1)]h_1(\bar{x}_1)P\{m\hat{\sigma}^2(1)>0, m\hat{\sigma}^2(1)\bar{x}_1+\hat{\tau}^2(1)\hat{\mu}(1)\\
&\qquad\qquad -MSB(1)\theta_0\geq 0\}d\bar{x}_1\\
&+\int_{c_0}^{\infty}[\varphi_1(\bar{x}_1)-\theta_0]h_1(\bar{x}_1)P\{m\hat{\sigma}^2(1)>0, m\hat{\sigma}^2(1)\bar{x}_1+\hat{\tau}^2(1)\hat{\mu}(1)\\
&\qquad\qquad -MSB(1)\theta_0< 0\}d\bar{x}_1\\
&+\int_{-\infty}^{\infty}|\theta_0-\varphi_1(\bar{x}_1)|h_1(\bar{x}_1)d\bar{x}_1P\{m\hat{\sigma}^2(1)=0\}.\\
&=A_1+A_2+A_3+A_4+A_5.
\end{aligned}
\tag{4.1}
$$

Under the normality assumption, $\hat{\mu}(1)$, $MSB(1)$ and $MSW(1)$ are mutually independent, and $\hat{\mu}(1)\sim N(\mu,(\sigma^2+\frac{\tau^2}{m})/(k-1))$, $\frac{(k-1)MSB(1)}{m\sigma^2+\tau^2}\sim\chi^2(k-1)$ and $\frac{(k-1)(m-1)MSW(1)}{\tau^2}\sim\chi^2((k-1)(m-1))$. Then, following an argument similar to that of Gupta and Liang [16], one can obtain that, for each term on the right hand side of (4.1), $A_i=O(k^{-1})$. We summarize the result as a theorem as follows.

Theorem 4.1 For the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$, under the normal-normal statistical model, we have $\rho_G(\underset{\sim}{\delta}^*,k)=O(k^{-1})$.

Case 2. Poisson-Gamma Case

It is assumed that conditioning on θ_i, $X_{ij}, j=1,\ldots,m$, is a sample of size m arising from a Poisson distribution $\mathcal{P}(\theta_i)$ and the parameter θ_i is a realization of a random variable Θ_i having a $\Gamma(\alpha,\beta)$ prior distribution with pdf $g(\theta_i)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta_i^{\alpha-1}e^{-\beta\theta_i}$. Then, conditioning on θ_i, $X_i=\sum_{j=1}^{m}X_{ij}\sim\mathcal{P}(m\theta_i)$ and X_i has a marginal probability function $h(x_i)=\frac{\beta^\alpha m^{x_i}}{x_i!\Gamma(\alpha)}\times\frac{\Gamma(\alpha+x_i)}{(m+\beta)^{\alpha+x_i}}$. Furthermore, $E[\overline{X}_i|\theta_i]=\theta_i$, Var $(X_{ij}|\theta_i)=\theta_i$, $\mu=E[\Theta_i]=\frac{\alpha}{\beta}$, $\sigma^2=$ Var $(\Theta_i)=\frac{\alpha}{\beta^2}$, $\tau^2=E[\mathrm{Var}(X_{ij}|\Theta_i)]=E[\Theta_i]=\mu=\frac{\alpha}{\beta}$, and $E[\Theta_i|\overline{X}_i]=\frac{m}{m+\beta}\overline{X}_i+\frac{\beta}{m+\beta}\mu=\frac{1}{m+\beta}X_i+\frac{\beta}{m+\beta}\mu=\psi(X_i)$. Let $A=\{x|\psi(x)<\theta_0\}$ and $B=\{x|\psi(x)>\theta_0\}$. Define

$$
a^*=\begin{cases}\sup A & \text{if } A\neq\phi,\\ -1 & \text{if } A=\phi,\end{cases}
$$

and

$$
b^*=\begin{cases}\inf B & \text{if } B\neq\phi,\\ \infty & \text{if } B=\phi.\end{cases}
$$

Note that $\lim_{x\to\infty} \psi(x) = \infty$. So $B \neq \phi$ and therefore $b^* < \infty$. Also, $A = \phi$ iff $\psi(0) \geq \theta_0$.

Let $I(E)$ denote the indicator function of the event E. Now, according to the definitions of a^*, b^* and (2.5) and (3.1),

$$\begin{aligned} R_1(\delta_1^*|\underset{\sim}{X}(1)) - R_1(\delta_{B1}) = &\sum_{x=0}^{a^*}[\theta_0 - \psi(x)]I(\delta_1^*(x|\underset{\sim}{X}(1)) = 1)h(x) \\ &+ \sum_{x=b^*}^{\infty}[\psi(x) - \theta_0]I(\delta_1^*(x|\underset{\sim}{X}(1)) = 0)h(x) \end{aligned} \tag{4.2}$$

where $\sum_{x=0}^{a*} = 0$ if $a^* = -1$. For $0 \leq x \leq a^*$,

$$\begin{aligned} &I(\delta_1^*(x|\underset{\sim}{X}(1)) = 1) \\ &= I(m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)x + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0) \\ &\quad + I(m\hat{\sigma}^2(1) = 0,\ \hat{\sigma}^2(1)x + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0) \\ &\leq I(m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)a^* + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0) + I(m\hat{\sigma}^2(1) = 0), \end{aligned} \tag{4.3}$$

and for $b^* \leq x < \infty$,

$$\begin{aligned} &I(\delta_1^*(x|\underset{\sim}{X}(1)) = 0) \\ &= I(m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)x + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0) \\ &\quad + I(m\hat{\sigma}^2(1) = 0,\ \hat{\sigma}^2(1)x + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0) \\ &\leq I(m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)b^* + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0) + I(m\hat{\sigma}^2(1) = 0). \end{aligned} \tag{4.4}$$

Plugging (4.3) and (4.4) into (4.2), we obtain

$$\begin{aligned} &R_1(\delta_1^*|\underset{\sim}{X}(1)) - R_1(\delta_{B1}) \\ &\leq b_1 I(m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)a^* + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0) \\ &\quad + b_2 I(m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)b^* + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0) \\ &\quad + b_3 I(m\hat{\sigma}^2(1) = 0), \end{aligned} \tag{4.5}$$

where $b_1 = \sum_{x=0}^{a^*}[\theta_0 - \psi(x)]h(x)$, $b_2 = \sum_{x=b^*}^{\infty}[\psi(x) - \theta_0]h(x)$ and $b_3 = b_1 + b_2$. Note that $0 \leq b_i < \infty$, $i = 1, 2, 3$. Therefore,

$$\begin{aligned} &R_1(\delta_1^*) - R_1(\delta_{B1}) \\ &= E[R_1(\delta_1^*|\underset{\sim}{X}(1))] - R_1(\delta_{B1}) \\ &\leq b_1 P\{m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)a^* + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0\} \\ &\quad + b_2 P\{m\hat{\sigma}^2(1) > 0,\ \hat{\sigma}^2(1)b^* + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0\} \\ &\quad + b_3 P\{m\hat{\sigma}^2(1) = 0\}. \end{aligned} \tag{4.6}$$

Thus, it suffices to study the asymptotic behaviors of the three terms on the right hand side of (4.6). For this, we first present certain useful preliminary results.

Note that $SSB(1) = \sum_{j=2}^k m(\overline{X}_j - \hat{\mu}(1))^2 = m\sum_{j=2}^k \overline{X}_j^2 - m(k-1)\hat{\mu}^2(1)$; $SSW(1) = \sum_{j=2}^k \sum_{\ell=1}^m (X_{j\ell} - \overline{X}_j)^2 = \sum_{j=2}^k \sum_{\ell=1}^m X_{j\ell}^2 - m\sum_{j=2}^k \overline{X}_j^2$; $\overline{X}_j, j = 1, \ldots, k$, are iid with $E[\overline{X}_j] = \mu$ and Var $(\overline{X}_j) \equiv V_1 < \infty$; $\overline{X}_j^2$, $j = 1, \ldots, k$, are iid with $E[\overline{X}_j^2] = \frac{\tau^2}{m} + \mu^2 + \sigma^2$ and Var $(\overline{X}_j^2) \equiv V_2 < \infty$; $\sum_{\ell=1}^m X_{j\ell}^2$, $j = 1, \ldots, k$, are iid with $E[\sum_{\ell=1}^m X_{j\ell}^2] = m(\tau^2 + \mu^2 + \sigma^2)$ and Var $(\sum_{\ell=1}^m X_{j\ell}^2) \equiv V_3 < \infty$. Note that the finiteness of the variances V_i, $i = 1, 2, 3$, is guaranteed under the Poisson-Gamma model.

Then from Tchebychev inequality, we have the following results.

Lemma 4.1 For $c > 0$,

(a) $P\{|\sum_{j=2}^k \sum_{\ell=1}^m X_{j\ell}^2/(k-1) - m(\tau^2 + \mu^2 + \sigma^2)| \geq c\} \leq \frac{V_3}{(k-1)c^2}$,

(b) $P\{|\sum_{j=2}^k \overline{X}_j^2/(k-1) - (\frac{\tau^2}{m} + \mu^2 + \sigma^2)| \geq c\} \leq \frac{V_2}{(k-1)c^2}$,

(c) $P\{|\hat{\mu}(1) - \mu| > c\} \leq \frac{V_1}{(k-1)c^2}$,

(d) $P\{\hat{\mu}^2(1) - \mu^2 > c\} \leq \frac{V_1}{(k-1)(\sqrt{\mu^2+c}-\mu)^2}$,

$P\{\hat{\mu}^2(1) - \mu^2 < -c\} \leq \frac{V_1 J(\mu^2-c)}{(k-1)(\mu-\sqrt{\mu^2-c})^2}$

where $J(x) = 1(0)$ if $x \geq 0$ (otherwise).

Lemma 4.2 For $c > 0$,

(a) $P\{MSW(1) - \tau^2 > c\} \leq \frac{4V_3}{(k-1)(m-1)^2c^2} + \frac{4m^2V_2}{(k-1)(m-1)^2c^2}$,

(b) $P\{MSW(1) - \tau^2 < -c\} \leq \frac{4V_3}{(k-1)(m-1)^2c^2} + \frac{4m^2V_2}{(k-1)(m-1)^2c^2}$.

(c) $P\{MSB(1) - (\tau^2 + m\sigma^2) > c\} \leq \frac{4m^2V_2}{(k-1)c^2} + \frac{V_1 J(\mu^2 - \frac{c}{2m})}{(k-1)(\mu-\sqrt{\mu^2-\frac{c}{2m}})^2}$,

(d) $P\{MSB(1) - (\tau^2 + m\sigma^2) < -c\} \leq \frac{4m^2V_2}{(k-1)c^2} + \frac{V_1}{(k-1)(\sqrt{\mu^2+\frac{c}{2m}}-\mu)^2}$.

Proof: We prove (a) and (c) only. (b) and (d) can be obtained in a similar way.

(a) $\quad P\{MSW(1) - \tau^2 > c\}$

$$= P\{\frac{SSW(1)}{(k-1)(m-1)} - \tau^2 > c\}$$

$$= P\{\frac{\sum_{j=2}^{k}\sum_{\ell=1}^{m} X_{j\ell}^2}{(k-1)(m-1)} - \frac{m}{m-1} \times \frac{1}{k-1}\sum_{j=2}^{k}\overline{X}_j^2 - \frac{m}{m-1}(\tau^2+\mu^2+\sigma^2)$$

$$+ \frac{m}{m-1}(\frac{\tau^2}{m} + \mu^2 + \sigma^2) > c\}$$

$$\leq P\{\frac{\sum_{j=2}^{k}\sum_{\ell=1}^{m} X_{j\ell}^2}{k-1} - m(\tau^2+\mu^2+\sigma^2) > \frac{(m-1)c}{2}\}$$

$$+ P\{\frac{1}{k-1}\sum_{j=2}^{k}\overline{X}_j^2 - (\frac{\tau^2}{m} + \mu^2 + \sigma^2) < -\frac{(m-1)c}{2m}\}$$

$$\leq \frac{4V_3}{(k-1)(m-1)^2c^2} + \frac{4m^2V_2}{(k-1)(m-1)^2c^2}.$$

(c) $$P\{MSB(1) - (\tau^2 + m\sigma^2) > c\}$$

$$= P\{\frac{SSB(1)}{k-1} - (\tau^2 + m\sigma^2) > c\}$$

$$= P\{\frac{m\sum_{j=2}^{k} X_j^2}{k-1} - m\hat{\mu}^2(1) - (\tau^2 + m\sigma^2 + m\mu^2) + m\mu^2 > c\}$$

$$\leq P\{\frac{1}{k-1}\sum_{j=2}^{k}\overline{X}_j^2 - (\frac{\tau^2}{m} + \sigma^2 + \mu^2) > \frac{c}{2m}\}$$

$$+ P\{\hat{\mu}^2(1) - \mu^2 < -\frac{c}{2m}\}$$

$$\leq \frac{4m^2V_2}{(k-1)c^2} + \frac{V_1 J(\mu^2 - \frac{c}{2m})}{(k-1)(\mu - \sqrt{\mu^2 - \frac{c}{2m}})^2}. \qquad \square$$

Let $d(a^*) = m\sigma^2\frac{a^*}{m} + \tau^2\mu - (\tau^2 + m\sigma^2)\theta_0$ and $d(b^*) = m\sigma^2\frac{b^*}{m} + \tau^2\mu - (\tau^2 + m\sigma^2)\theta_0$. Then $d(a^*) < 0 < d(b^*)$, by the definitions of a^* and b^*. Then,

$$P\{m\hat{\sigma}^2(1) > 0,\ m\hat{\sigma}^2(1)\frac{a^*}{m} + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0\}$$

$$\leq P\{[MSB(1) - MSW(1)]\frac{a^*}{m} + MSW(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0\}$$

$$
\begin{aligned}
&= P\left\{\begin{array}{ll} [MSB(1)-(\tau^2+m\sigma^2)](\frac{a^*}{m}-\theta_0) & \\ +MSW(1)(\hat{\mu}(1)-\mu) & > -d(a^*) \\ +(MSW(1)-\tau^2)(\mu-\frac{a^*}{m}) & \end{array}\right\} \\
&\le P\{[MSB(1)-(\tau^2+m\sigma^2)](\frac{a^*}{m}-\theta_0) > -\frac{d(a^*)}{3}\} \\
&\quad + P\{MSW(1)(\hat{\mu}(1)-\mu) > -\frac{d(a^*)}{3}\} \\
&\quad + P\{(MSW(1)-\tau^2)(\mu-\frac{a^*}{m}) > \frac{d(a^*)}{3}\}.
\end{aligned} \tag{4.7}
$$

Note that $P\{[MSB(1)-(\tau^2+m\sigma^2)]\ (\frac{a^*}{m}-\theta_0) > -\frac{d(a^*)}{3}\} = 0$ if $\frac{a^*}{m}-\theta_0 = 0$ and $P\{(MSW(1)-\tau^2)(\mu-\frac{a^*}{m}) > -\frac{d(a^*)}{3}\} = 0$ if $\mu-\frac{a^*}{m} = 0$. Hence, without loss of generality, we assume $\mu < \frac{a^*}{m} < \theta_0$. Therefore, from Lemma 4.2,

$$
\begin{aligned}
&P\{(MSB(1)-(\tau^2+m\sigma^2)][\frac{a^*}{m}-\theta_0] > -\frac{d(a^*)}{3}\} \\
&= P\{MSB(1)-(\tau^2+m\sigma^2) < \frac{d(a^*)}{3(\theta_0-\frac{a^*}{m})}\} \\
&\le \frac{36m^2V_2(\theta_0-\frac{a^*}{m})^2}{(k-1)d^2(a^*)} + \frac{V_1}{(k-1)(\sqrt{\mu^2+\frac{-d(a^*)}{6m(\theta_0-\frac{a^*}{m})}}-\mu)^2} \\
&= O(k^{-1}).
\end{aligned} \tag{4.8}
$$

$$
\begin{aligned}
&P\{(MSW(1)-\tau^2)(\mu-\frac{a^*}{m}) > \frac{d(a^*)}{3}\} \\
&= P\{MSW(1)-\tau^2 < \frac{d(a^*)}{3(\frac{a^*}{m}-\mu)}\} \\
&\le \frac{36V_3(\frac{a^*}{m}-\mu)^2}{(k-1)(m-1)^2d^2(a^*)} + \frac{36m^2V_2(\frac{a^*}{m}-\mu)^2}{(k-1)(m-1)^2d^2(a^*)} \\
&= O(k^{-1}).
\end{aligned} \tag{4.9}
$$

Also,

$$
\begin{aligned}
&P\{MSW(1)(\hat{\mu}(1)-\mu) > -\frac{d(a^*)}{3}\} \\
&= P\{MSW(1)(\hat{\mu}(1)-\mu) > -\frac{d(a^*)}{3}, |MSW(1)-\tau^2| \le \frac{\tau^2}{2}\} \\
&\quad + P\{MSW(1)(\hat{\mu}(1)-\mu) > -\frac{d(a^*)}{3}, |MSW(1)-\tau^2| > \frac{\tau^2}{2}\} \\
&\le P\{\hat{\mu}(1)-\mu > \frac{1}{\frac{3}{2}\tau^2} \times \frac{-d(a^*)}{3}\} + P\{|MSW(1)-\tau^2| > \frac{\tau^2}{2}\} \\
&\le \frac{V_1(\frac{9}{2}\tau^2)^2}{(k-1)d^2(a^*)} + \frac{4V_3}{(k-1)(m-1)^2\frac{\tau^4}{4}} + \frac{4m^2V_2}{(k-1)(m-1)^2\frac{\tau^4}{4}} \\
&= O(k^{-1}).
\end{aligned} \tag{4.10}
$$

Combining (4.7)–(4.10) yields that

$$P\{m\hat{\sigma}^2(1) > 0,\ m\hat{\sigma}^2(1)\frac{a^*}{m} + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0\} = O(k^{-1}). \tag{4.11}$$

Similarly, one can see that

$$P\{m\hat{\sigma}^2(1) > 0,\ m\hat{\sigma}^2(1)\frac{b^*}{m} + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0\} = O(k^{-1}). \tag{4.12}$$

Also

$$\begin{aligned}
&P\{m\hat{\sigma}^2(1) = 0\}\\
&= P\{MSB(1) - MSW(1) \leq 0\}\\
&= P\{[MSB(1) - (\tau^2 + m\sigma^2)] - [MSW(1) - \tau^2] \leq -m\sigma^2\}\\
&\leq P\{MSB(1) - (\tau^2 + m\sigma^2) \leq -\frac{m}{2}\sigma^2\}\\
&\qquad + P\{MSW(1) - \tau^2 \geq \frac{m\sigma^2}{2}\}\\
&= O(k^{-1}) \text{ by Lemma 4.2.}
\end{aligned} \tag{4.13}$$

Substituting the results of (4.11)–(4.13) into (4.6), we obtain that $R_1(\delta_1^*) - R_1(\delta_{B1}) = O(k^{-1})$. We summarize this result as a theorem as follows.

Theorem 4.2. For the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$, under the Poisson-Gamma statistical model, we have $\rho_G(\underset{\sim}{\delta}^*, k) = O(k^{-1})$.

Case 3. Bounded Random Variables with Bounded Sampling Probability Density

It is assumed that for each $i = 1, \ldots, k$, $|X_{ij}| \leq B$ and $0 \leq h_i(\bar{x}_i|\theta_i) \leq M$ for all θ_i and $\bar{x}_i$, with no further assumption on the model. Under the assumptions,

$$|\overline{X}_i| \leq B \text{ and } h_i(\bar{x}_i) = \int h_i(\bar{x}_i|\theta_i)dG_i(\theta_i) \leq \int MdG_i(\theta_i) = M < \infty.$$

In this case, it is also assumed that $|\theta_0| < B$.

Let D denote the event that $|m\hat{\sigma}^2(1) - m\sigma^2| \leq \frac{1}{2}m\sigma^2$, and D^c its complement. Also, let I_D be the indicator function of the event D.

Then, from (2.5), (2.7), (3.1) and (3.3), we have

$$
\begin{aligned}
&R_1(\delta_1^*|\underset{\sim}{X}(1)) - R_1(\delta_{B1}) \\
&= \int_{c_0}^{C_1} [\varphi(\bar{x}_1) - \theta_0] h_1(\bar{x}_1) d\bar{x}_1 I_D \\
&\quad + \int_{-B}^{c_0} [\theta_0 - \varphi(\bar{x}_1)] h_1(\bar{x}_1) I_{D^c} I(m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 \geq 0) d\bar{x}_1 \\
&\quad + \int_{c_0}^{B} [\varphi(\bar{x}_1) - \theta_0] h_1(\bar{x}_1) I_{D^c} I(m\hat{\sigma}^2(1)\bar{x}_1 + \hat{\tau}^2(1)\hat{\mu}(1) - MSB(1)\theta_0 < 0) d\bar{x}_1 \\
&\leq \int_{c_0}^{C_1} [\varphi(\bar{x}_1) - \theta_0] h_1(\bar{x}_1) d\bar{x}_1 I_D + \int_{-B}^{B} |\varphi(\bar{x}_1) - \theta_0| h_1(\bar{x}_1) d\bar{x}_1 I_{D^c}
\end{aligned}
\tag{4.14}
$$

where $C_1 = [MSB(1)\theta_0 - \hat{\tau}^2(1)\hat{\mu}(1)]/m\hat{\sigma}^2(1)$.

Note that $\varphi(\bar{x}_1) = a\bar{x}_1 + (1-a)\mu$, $\varphi(c_0) = ac_0 + (1-a)\mu = \theta_0$ and $0 \leq h(\bar{x}_1) \leq M$ for all $\bar{x}_1$. By the mean-value theorem, there exists a C_1^* between C_1 and c_0 such that

$$
\begin{aligned}
\int_{c_0}^{C_1} [\varphi(\bar{x}_1) - \theta_0] h_1(\bar{x}_1) d\bar{x}_1 &= [\varphi(C_1^*) - \theta_0](C_1 - c_0) h_1(C_1^*) \\
&= a(C_1^* - c_0)(C_1 - c_0) h_1(C_1^*) \\
&\leq a(C_1 - c_0)^2 M.
\end{aligned}
\tag{4.15}
$$

Also

$$
\int_{-B}^{B} |\varphi(\bar{x}_1) - \theta_0| h_1(\bar{x}_1) d\bar{x}_1 \leq B. \tag{4.16}
$$

Therefore, by combining (4.14)–(4.16), we obtain

$$
\begin{aligned}
R_1(\delta_1^*) - R_1(\delta_{B1}) &= E_1[R_1(\delta_1^*|\underset{\sim}{X}(1))] - R_1(\delta_{B1}) \\
&\leq aME_1[(C_1 - c_0)^2 I_D] + BE_1[I_{D^c}].
\end{aligned}
\tag{4.17}
$$

Now, by the definition of $m\hat{\sigma}^2(1)$, we have

$$
\begin{aligned}
&E_1[I_{D^c}] = P\{|m\hat{\sigma}^2(1) - m\sigma^2| > \frac{1}{2} m\sigma^2\} \\
&= P\left\{\begin{array}{l} [MSB(1) - (\tau^2 + m\sigma^2)] - [MSW(1) - \tau^2] > \frac{1}{2} m\sigma^2 \\ \text{or } [MSB(1) - (\tau^2 + m\sigma^2)] - [MSW(1) - \tau^2] < -\frac{1}{2} m\sigma^2 \end{array}\right\} \\
&\leq P\{|MSB(1) - (\tau^2 + m\sigma^2)| > \frac{1}{4} m\sigma^2\} \\
&\qquad + P\{|MSW(1) - \tau^2| > \frac{1}{4} m\sigma^2\} \\
&= O(k^{-1}).
\end{aligned}
\tag{4.18}
$$

In (4.18), the last equality can be obtained by an argument similar to that of Lemma 4.2 by noting that $|X_{ij}| \leq B$.

Next, on the event D, $m\hat{\sigma}^2(1) = MSB(1) - MSW(1) \geq \frac{1}{2}m\sigma^2$ and $-B \leq C_1 \equiv \frac{MSB(1)\theta_0 - MSW(1)\hat{\mu}(1)}{MSB(1)-MSW(1)} \leq B$. Also, $-B \leq c_0 \equiv \frac{(m\sigma^2+\tau^2)\theta_0 - \tau^2\mu}{m\sigma^2} \leq B$. Hence, on D, $|C_1 - c_0| \leq 2B$. Then, using Lemma of Singh [29],

$$\begin{aligned}
&E_1[(C_1-c_0)^2 I_D] \\
&\leq E_1\left[|\frac{MSB(1)\theta_0 - MSW(1)\hat{\mu}(1)}{MSB(1)-MSW(1)} - \frac{(m\sigma^2+\tau^2)\theta_0 - \tau^2\mu}{m\sigma^2}| \wedge 2B\right]^2 \\
&\leq \frac{8}{(m\sigma^2)^2}E_1[(MSB(1)\theta_0 - MSW(1)\hat{\mu}(1)) - ((m\sigma^2+\tau^2)\theta_0 - \tau^2\mu)]^2 \\
&\quad + \frac{8}{(m\sigma^2)^2}E_1[MSB(1)-MSW(1)-m\sigma^2]^2[(\frac{(m\sigma^2+\tau^2)\theta_0 - \tau^2\mu}{m\sigma^2})^2 + 2B^2] \\
&= O(k^{-1}),
\end{aligned} \tag{4.19}$$

since

$$\begin{aligned}
&E_1[(MSB(1)\theta_0 - MSW(1)\hat{\mu}(1)) - ((m\sigma^2+\tau^2)\theta_0 - \tau^2\mu)]^2 \\
&\leq 3E_1[MSB(1)\theta_0 - (m\sigma^2+\tau^2)\theta_0]^2 + 3E_1[MSW(1)(\hat{\mu}(1)-\mu)]^2 \\
&\qquad + 3E_1[(MSW(1)-\tau^2)\mu]^2 \\
&\leq 3\theta_0^2 E_1[MSB(1) - (M\sigma^2+\tau^2)]^2 + 3B^4 E_1[\hat{\mu}(1)-\mu]^2 \\
&\qquad + 3B^2 E_1[MSW(1)-\tau^2]^2 \\
&= O(k^{-1})
\end{aligned} \tag{4.20}$$

and

$$\begin{aligned}
&E_1[MSB(1)-MSW(1)-m\sigma^2]^2 \\
&\leq 2E_1[MSB(1)-(m\sigma^2+\tau^2)]^2 + 2E_1[MSW(1)-\tau^2]^2 \\
&= O(k^{-1}).
\end{aligned} \tag{4.21}$$

Therefore, we obtain that

$$R_1(\delta_1^*) - R_1(\delta_{B1}) = O(k^{-1}). \tag{4.22}$$

Thus, we have the following theorem.

Theorem 4.3. Assume that $|X_{ij}| \leq B$ and $0 \leq h_i(\bar{x}_i|\theta_i) \leq M$ for all $\bar{x}_i$ and θ_i. Also assume A1–A4 hold. Then, the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ is asymptotically optimal and $\rho_G(\underset{\sim}{\delta}^*, k) = O(k^{-1})$.

Remark: We have studied the problem of simultaneous selection of k populations in comparison with a known standard θ_0. When θ_0 is an unknown control, supposed to be the mean of the control population, we may first estimate θ_0 by $\tilde{\theta}_0$ based on a sample of size ℓ taken from the control population. We then construct a selection procedure, say $\tilde{\underset{\sim}{\delta}}$, by replacing θ_0 in $\underset{\sim}{\delta}^*$ by the estimator $\tilde{\theta}_0$. In order to possess the asymptotic optimality, the sample size ℓ should be large enough. However, one may or may not achieve the same rate of convergence as claimed in Theorem 4.3. Actually the rate of convergence for the θ_0 unknown case depends on the sample size l, the property of the marginal *pdf* $h_i(\bar{x}_i), i = 1, \ldots, k$ and the properties related to the estimate $\tilde{\theta}_0$.

References

[1] Bechhofer, R. E. (1968). Multiple comparisons with a control for multiply-classified variances of normal populations. *Technometrics*, **10**, 715-718.

[2] Bechhofer, R. E. (1969). Optimal allocation of observations when comparing several treatments with a control. *Multivariate Analysis*-II (Ed. P. R. Krishnaiah), Academic, New York, 463-473.

[3] Bechhofer, R. E. and Nocturne, D. J.-M. (1972). Optimal allocation of observations when comparing several treatments with a control, II: 2-sided comparisons. *Technometrics*, **14**, 423-436.

[4] Bechhofer, R. E. Santner, T. J. and Goldsman, D. M. (1995). *Design and Analysis of Experiments for Statitsitcal Selection, Screening, and Multiple Comparisons*, Wiley, New York.

[5] Bechhofer, R. E. and Turnbull, B. W. (1971). Optimal allocation of observations when comparing several treatments with a control, III. Globally best one-sided intervals for unequal variances. *Statistical Design Theory and Related Topics*, (Eds. S. S. Gupta and J. Yackel), Academic, New York, 41-78.

[6] 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.

[7] Diaconis, P. and Yluisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.*, **7**, 269-281.

[8] Dudewicz, E. J. and Taneja, B. K. (1992). Selection of the best experimental category provided it is better than a standard: the heteroscedastic method solution. *The Frontiers of Modern Statistical Inference Procedures*, II (Eds. E. Bofinger, E. J. Dudewicz, G. J. Lewis and K. Mengersen). American Sciences Press, Inc., Columbus, Ohio, 1992, 47-94.

[9] Dunnett, C. W. (1995). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096-1121.

[10] Dunnett, C. 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.

[11] Ericson, W. A. (1969). A note on the posterior mean of a population mean. *J. Royal Statist. Soc.*, **31B**, 332–334.

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

[13] Goldstein, M. (1975). A note on some Bayesian nonparametric estimates. *Ann. Statist.*, **3**, 736–740.

[14] Gupta, S. S. and Hsiao, P. (1981). On Γ-mimimax, and Bayes procedures for selecting populations close to a control. *Sankhya, B.*, **43**, 291-318.

[15] Gupta, S. S. and Liang, T. (1991). On the asymptotic optimality of certain empirical Bayes simultaneous testing procedures for Poisson populations. *Statistics and Decisions*, **9**, 263–283.

[16] Gupta, S. S. and Liang, T. (1996-97). Selecting good normal regression models: an empirical Bayes approach. *J. India Agricultural Statist.*, **49**, 335-352.

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

[18] Gupta, S. S. and Miescke, K. J. (1985). Minimax multiple t-test for comparing k normal populations with a control. *J. Statist. Plann. Inference*, **12**, 161–169.

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

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

[21] Gupta, S. S. and Singh, A. K. (1979). On selection rules for treatments versus control problems. *Proc. 42nd Session of the International Statistical Institute*, 229-232.

[22] Gupta, S. S. and Sobel, M. (1958). On selecting a subset which containing all populations better than a control. *Ann. Math. Statist.*, **29**, 235–244.

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

[24] Liang, T. (1997). Simultaneously selecting normal populations close to a control. *J. Statist. Plann. Inference*, **61**, 297-316.

[25] Mee, R. W., A. K. Shah and J. J. Lefante (1987). Comparing k independent sample means with a known standard. *J. Quality Technol.* **19**, 75-81.

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

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

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

[29] Singh, R. S. (1977). Applications of estimates of estimators of a density and its derivatives to certain statistical problems. *J. Royal Statist. Soc.*, **39B**, 357–363.

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

[31] 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.

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

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