ON EMPIRICAL BAYES SIMULTANEOUS SELECTION PROCEDURES FOR COMPARING NORMAL POPULATIONS WITH A STANDARD

by

Shanti S. Gupta and TaChen Liang

Purdue University / Wayne State University

Technical Report #96-25C

Department of Statistics
Purdue University
West Lafayette, IN USA

May 1996
Revised June 1998

*

ON EMPIRICAL BAYES SIMULTANEOUS SELECTION PROCEDURES FOR COMPARING NORMAL POPULATIONS WITH A STANDARD

by

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

and

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

Abstract

In this paper, we derive statistical selection procedures to partition k normal populations into "good" or "bad" ones, respectively, using the nonparametric empirical Bayes approach. The relative regret risk of a selection procedure is used as a measure of its performance. We establish the asymptotic optimality of the proposed empirical Bayes selection procedures and investigate the associated rates of convergence. Under a very mild condition, the proposed empirical Bayes selection procedures are shown to have rates of convergence of order close to $O(k^{-\frac{1}{2}})$ where k is the number of populations involved in the selection problem. With further strong assumptions, the empirical Bayes selection procedures have rates of convergence of order $O(k^{-\frac{\alpha(r-1)}{2r+1}})$, where $1 < \alpha < 2$ and r is an integer greater than 2.

AMS 1991 Subject Classification: 62F07, 62C12.

Keywords and Phrases: Asymptotically optimal, comparison with a standard, empirical Bayes, rate of convergence, simultaneous selection, relative regret risk.

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

1. Introduction

Consider k independent normal populations $\pi_i \equiv N(\theta_i, \sigma^2)$, $i = 1, \ldots, k$, with unknown means $\theta, \ldots, \theta_k$, respectively, and a common variance σ^2. Let θ_0 denote a standard or a control. A population π_i is said to be good if $\theta_i \geq \theta_0$, and bad otherwise. In certain practical situations, one may be interested in the selection of all good populations while excluding all bad populations. For example, let θ_i denote the quality level of a newly developed manufacturing process $\pi_i, i = 1, \ldots, k$, and θ_0 be a specified standard level. Then, one may be interested in finding out all the potential manufacturing processes for further investigation. The preceding described selection goal can also be viewed as a first step of a selection problem in which the selection goal is to select the best from among k populations provided that the best is at least as good as the specified standard level.

In the literature, the problem of comparing normal populations with a control has been extensively studied by many authors. To mention a few, for example, Dunnett (1955), Gupta and Sobel (1958) and Tong (1969) have proposed and studied some natural selection procedures. Lehmann (1961) and Spjøtvoll (1972) have treated the problem using methods from the theory of testing hypothesis. Randles and Hollander (1971), Miescke (1981) and Gupta and Miescke (1985) have derived optimal procedures via minimax or Γ-minimax approaches.

The purpose of this paper is to derive statistical procedures which partition the k normal populations into "good" and "bad" ones, respectively, using the empirical Bayes approach. It is assumed that the parameter θ_0 is the value of a specified standard level, and therefore is assumed to be known.

The paper is organized as follows. In Section 2, the statistical model of the selection problem is introduced and a Bayes selection procedure for the selection problem is also derived. As seen in the later part of the paper, the Bayes selection procedure depends on the prior distribution. When the prior distribution is unknown, the Bayes selection procedure cannot be implemented. In such a situation, using the empirical Bayes approach and by mimicking the behavior of the Bayes selection procedure, we have developed empirical Bayes selection procedures in Section 3. The relative regret risk of an empirical Bayes selection procedure is used as a measure of the performance of this empirical Bayes selection procedure. We establish the asymptotic optimality of the proposed empirical Bayes selection procedures in Section 4. The rate of convergence of the relative regret risks is studied in Section 5. Under a very mild condition, the proposed empirical Bayes selection procedures have rates of convergence of order close to $O(k^{-1/2})$. With some further assumptions, the empirical Bayes selection procedures have rates of convergence of order $O(k^{-\alpha(r-1)/[2r+1]})$ where $1 < \alpha < 2$ and $r > 2$ is an integer.

2. The Selection Problem and A Bayes Selection Procedure

Let $X_{i1}, \ldots, X_{im}$ be a sample of size m taken from a normal population $\pi_i \equiv N(\theta_i, \sigma^2)$, $i = 1, \ldots, k$. All samples are assumed to be mutually independent. Let $X_i = \frac{1}{m} \sum_{j=1}^{m} X_{ij}$ and when $m \geq 2$, let $W_k = \sum_{i=1}^{k} \sum_{j=1}^{m} (X_{ij} - X_i)^2/[k(m-1)]$. Note that given $\theta_i, X_i \sim N(\theta_i, \frac{\sigma^2}{m})$, $k(m-1)\ W_k/\sigma^2 \sim \chi^2(k(m-1))$, and $X_1, \ldots, X_k$ and W_k are mutually independent.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) | -\infty < \theta_i < \infty, i = 1, \ldots, k\}$ be the parameter space, and let $\mathcal{A} = \{\underset{\sim}{a} = (a_1, \ldots, a_k) | a_i = 0,\ 1, i = 1, \ldots, k\}$ be the action space. When action $\underset{\sim}{a}$ is taken, it means that population π_i is selected as good if $a_i = 1$, and excluded as bad if $a_i = 0$. For each $\underset{\sim}{\theta} \in \Omega$ and $\underset{\sim}{a} \in \mathcal{A}$, the loss function $L(\underset{\sim}{\theta}, \underset{\sim}{a})$ is defined to be

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

and

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

where $I(x) = 1(0)$ if $x \geq 0$ (otherwise). Note that in (2.2), the first term is the loss due to selecting π_i as good when $\theta_i < \theta_0$, and the second term is the loss due to wrongly excluding π_i as bad when $\theta_i \geq \theta_0$.

It is assumed that for each i, the parameter θ_i is a realization of a random variable Θ_i; and $\Theta_1, \ldots, \Theta_k$ are independently distributed with a common but unknown prior distribution G. Under the preceding assumption, $X_1, \ldots, X_k$ and W_k are mutually independent, and $X_1, \ldots, X_k$ are identically distributed, having a marginal pdf $f(x) = \int f(x|\theta, \sigma^2) dG(\theta)$, where $f(x|\theta, \sigma^2)$ denotes the pdf of a normal $N(\theta, \frac{\sigma^2}{m})$ distribution.

Let $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{im}), i = 1 \ldots k$ and $\underset{\sim}{X} = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_k)$. Let $\mathcal{X}$ be the sample space of $\underset{\sim}{X}$. A selection procedure $\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k)$ is defined to be a mapping from the sample space $\mathcal{X}$ into the product space $[0,1]^k$, such that for each $i = 1, \ldots, k, \delta_i(\underset{\sim}{x})$ is the probability of selecting population π_i as a good population when $\underset{\sim}{X} = \underset{\sim}{x}$ is observed. Let $\mathcal{D}$ be the class of all selection procedures. Also let $R(G, \underset{\sim}{\delta})$ denote the Bayes risk associated with the selection procedure $\underset{\sim}{\delta}$. It is assumed that $E[|\Theta_i|] < \infty$ so that the Bayes risk $R(G, \underset{\sim}{\delta})$ is finite. By Fubini's theorem, a straightforward computation yields that the Bayes risk $R(G, \underset{\sim}{\delta})$ can be expressed as:

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

and

$$R_i(G, \delta_i) = \int_{\mathcal{X}} [\theta_0 - \varphi_i(\underset{\sim}{x}_i)]\delta_i(\underset{\sim}{x}) \prod_{j=1}^{k} f_j\ (\underset{\sim}{x}_j) d\underset{\sim}{x} + C \tag{2.4}$$

where $\varphi_i(\underset{\sim}{x}_i) = E[\Theta_i|\underset{\sim}{X}_i = \underset{\sim}{x}_i]$: the posterior mean of Θ_i given $\underset{\sim}{X}_i = \underset{\sim}{x}_i$; $f_j(\underset{\sim}{x}_j)$: the marginal joint pdf of $\underset{\sim}{X}_j$, and $C = \int_{\theta_0}^{\infty} (\theta - \theta_0) dG(\theta)$. Thus, a Bayes selection procedure $\underset{\sim}{\delta}_B = (\delta_{B1}, \ldots, \delta_{Bk})$, which minimizes the Bayes risks among all selection procedures in $\mathcal{D}$, is clearly given by: For each $i = 1, \ldots, k$ and $\underset{\sim}{x} \in \mathcal{X}$,

$$\delta_{Bi}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \varphi_i(\underset{\sim}{x}_i) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Algebraic computation yields that

$$\varphi_i(\underset{\sim}{x}_i) \equiv x_i + \frac{\sigma^2}{m} f^{(1)}(x_i)/f(x_i) \equiv \psi_i(x_i), \tag{2.6}$$

where $f(x_i)$ is the marginal pdf of the sample mean X_i and $f^{(1)}(x_i)$ denotes its corresponding derivative. That is, the posterior mean $\varphi_i(\underset{\sim}{x}_i)$ depends on $\underset{\sim}{x}_i$ only through the sample mean value x_i. From (2.5) and (2.6), the i-th component Bayes selection procedure δ_{Bi} depends on $\underset{\sim}{x}$ only through x_i. Therefore, (2.5) can be expressed as

$$\delta_{Bi}(x_i) = \begin{cases} 1 & \text{if } \psi_i(x_i) \geq \theta_0 \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 1 & \text{if } T_i(x_i) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.5$'$}$$

where $T_i(x_i) = \frac{\sigma^2}{m} f^{(1)}(x_i) + (x_i - \theta_0)f(x_i)$. Note that $\psi_1(\cdot) = \ldots = \psi_k(\cdot)$ and $T_1(\cdot) = \ldots = T_k(\cdot)$ since $X_1, \ldots, X_k$ are identically distributed.

One can see that the posterior mean $\psi_i(x_i)$ is a continuous function in x_i; also $\psi_i(x_i)$ is strictly increasing in x_i if the prior distribution G is non-degenerate. Let $A_1 = \{x|\psi_i(x) < \theta_0\}$ and $A_2 = \{x|\psi_i(x) \geq \theta_0\}$. Define

$$a^* = \begin{cases} \sup A_1 & \text{if } A_1 \neq \phi, \\ -\infty & \text{if } A_1 = \phi. \end{cases} \tag{2.7}$$

Note that if $A_2 \neq \phi$ and $A_1 \neq \phi$, then $-\infty < a^* < \infty$; and if $A_2 = \phi, a^* = \infty$. In terms of a^*, the Bayes selection procedure $\underset{\sim}{\delta}_B$ can be written as:

$$\delta_{Bi}(x_i) = \begin{cases} 1 & \text{if } x_i \geq a^* \\ 0 & \text{otherwise.} \end{cases} \tag{2.5$''$}$$

Finally, the minimum Bayes risk is:

$$R(G, \underset{\sim}{\delta}_B) = \sum_{i=1}^{k} R_i(G, \delta_{Bi}), \tag{2.8}$$

and

$$R_i(G, \delta_{Bi}) = \int [\theta_0 - \psi_i(x_i)]\delta_{Bi}(x_i)f(x_i)dx_i + C. \tag{2.9}$$

In the following analysis, we consider those prior distributions G such that $\lim_{x_i \to -\infty} \psi_i(x_i) < \theta_0 < \lim_{x_i \to +\infty} \psi_i(x_i)$. Hence $A_1 \neq \phi$ and $A_2 \neq \phi$. Therefore, $-\infty < a^* < \infty$.

3. Empirical Bayes Selection Procedures

Since the prior distribution G is unknown, it is not possible to implement the Bayes selection procedure $\underset{\sim}{\delta}_B$ for the selection problem at hand. However, according to the statistical model described previously, the k components share certain similarity. Therefore, the empirical Bayes approach is employed to incorporate information from among the k populations to provide robust selection procedures for each of the k component selection problems.

The proposed empirical Bayes selection procedures mimick the behavior of the Bayes selection procedure $\underset{\sim}{\delta}_B$. For this, the forms of (2.5′) and (2.5″) provide important motivation for the construction of the empirical Bayes selection procedures. To construct the empirical Bayes selection procedures, first, we need to have estimates for $f(x)$ and $f^{(1)}(x)$.

For an integer $r > 2$ and for each $i = 0, 1$, let $\mathcal{K}_i^r$ be the class of all Borel-measurable bounded functions vanishing outside the interval $(0, 1)$, such that for $k_0 \in \mathcal{K}_0^r$,

$$\int_0^1 y^j k_0(y)dy = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1, \dots, r-1, \\ B_1 & \text{if } j = r; \end{cases} \tag{3.1}$$

and for $k_1 \in \mathcal{K}_1^r$,

$$\int_0^1 y^j k_1(y)dy = \begin{cases} 1 & \text{if } j = 1, \\ 0 & \text{if } j = 0, 2, \dots, r-1, \\ B_2 & \text{if } j = r. \end{cases} \tag{3.2}$$

We may let B_3 be a positive value such that $|k_i(y)| \leq B_3$ for all $y \in (0, 1)$ and $i = 0, 1$. Also, let $h = h(k)$ be a decreasing function of k such that $h(k) \to 0$ as $k \to \infty$.

Define

$$\begin{cases} f_{ik}(X_i) = \frac{1}{(k-1)h} \sum\limits_{\substack{j=1 \\ j \neq i}}^{k} k_0\left(\frac{X_i - X_j}{h}\right), \\ f_{ik}^{(1)}(X_i) = \frac{1}{(k-1)h^2} \sum\limits_{\substack{j=1 \\ j \neq i}}^{k} k_1\left(\frac{X_i - X_j}{h}\right). \end{cases} \tag{3.3}$$

Note that for each fixed $X_i = x_i$, $f_{ik}(x_i)$ and $f_{ik}^{(1)}(x_i)$ are consistent estimators of $f(x_i)$ and $f^{(1)}(x_i)$, respectively; see Singh (1977, 1979).

When the variance σ^2 is known, for each $i = 1, \dots, k$, let

$$T_{ik}^*(X_i) = (X_i - \theta_0) f_{ik}(X_i) + \frac{\sigma^2}{m} f_{ik}^{(1)}(X_i).$$

Also, let $\{C_k^*\}$ be a sequence of positive numbers such that C_k^* is increasing in k and $C_k^* \to \infty$ as $k \to \infty$. We consider an empirical Bayes selection procedures $\underset{\sim}{\delta}^* = (\delta_1^*, \ldots, \delta_k^*)$ defined as follows,

$$\delta_i^*(\underset{\sim}{X}) = \delta_i(X_i, \underset{\sim}{X}(i)) = \begin{cases} 0 & \text{if either } (X_i < -C_k^*) \\ & \text{or } (|X_i| \leq C_k^* \text{ and } T_{ik}^*(X_i) < 0), \\ 1 & \text{if either } (X_i > C_k^*) \\ & \text{or } (|X_i| \leq C_k^* \text{ and } T_{ik}^*(X_i) \geq 0), \end{cases} \tag{3.4}$$

where $\underset{\sim}{X}(i) = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$.

When the variance σ^2 is unknown, we estimate σ^2 by W_k. For each $i = 1, \ldots, k$, let

$$\tilde{T}_{ik}(X_i) = (X_i - \theta_0) f_{ik}(X_i) + \frac{W_k}{m} f_{ik}^{(1)}(X_i).$$

We then consider an empirical Bayes selection procedure $\hat{\underset{\sim}{\delta}} = (\hat{\delta}_1, \ldots, \hat{\delta}_k)$ defined as follows,

$$\hat{\delta}_i(\underset{\sim}{X}) = \hat{\delta}_i(X_i, \underset{\sim}{X}(i), W_k) = \begin{cases} 0 & \text{if either } (X_i < -C_k^*) \\ & \text{or } (|X_i| \leq C_k^* \text{ and } \hat{T}_{ik}(X_i) < 0), \\ 1 & \text{if either } (X_i > C_k^*) \\ & \text{or } (|X_i| \leq C_k^* \text{ and } \hat{T}_{ik}(X_i) \geq 0). \end{cases} \tag{3.5}$$

The Bayes risk of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ is:

$$R(G, \underset{\sim}{\delta}^*) = \sum_{i=1}^{k} R_i(G, \delta_i^*) \tag{3.6}$$

and

$$\begin{aligned} R_i(G, \delta_i^*) &= E_i^*[\int [\theta_0 - \psi_i(x_i)] \delta_i^*(x_i, \underset{\sim}{X}(i)) f(x_i) dx_i] + C \\ &= \int [\theta_0 - \psi_i(x_i)] E_i^*[\delta_i^*(x_i, \underset{\sim}{X}(i))] f(x_i) dx_i + C, \end{aligned} \tag{3.7}$$

where the expectation E_i^* is taken with respect to the probability measure generated by $\underset{\sim}{X}(i)$.

The Bayes risk of the empirical Bayes selection procedure $\hat{\underset{\sim}{\delta}}$ is:

$$R(G, \hat{\underset{\sim}{\delta}}) = \sum_{i=1}^{k} R_i(G, \hat{\delta}_i) \tag{3.8}$$

and

$$R_i(G,\hat{\delta}_i) = \int [\theta_0 - \psi_i(x_i)]\hat{E}_i[\hat{\delta}_i(x_i, \underset{\sim}{X}(i), W_k)]f(x_i)dx_i + C. \tag{3.9}$$

where the expectation $\hat{E}_i$ is taken with respect to the probability measure generated by $(\underset{\sim}{X}(i), W_k)$.

Since $\underset{\sim}{\delta}_B$ is the Bayes selection procedure, for any selection procedure $\underset{\sim}{\delta} = (\delta_1, \ldots, \delta_k)$, $R_i(G,\delta_i) - R_i(G,\delta_{Bi}) \geq 0$, $i = 1, \ldots, k$ and $R(G,\underset{\sim}{\delta}) - R(G,\underset{\sim}{\delta}_B) \geq 0$. Define

$$\rho(G,\underset{\sim}{\delta}) = [R(G,\underset{\sim}{\delta}) - R(G,\underset{\sim}{\delta}_B)]/R(G,\underset{\sim}{\delta}_B). \tag{3.9}$$

$\rho(G,\underset{\sim}{\delta})$ is called as the relative regret risk of the selection procedure $\underset{\sim}{\delta}$ compared with the Bayes selection procedure $\underset{\sim}{\delta}_B$. The relative regret risk $\rho(G,\underset{\sim}{\delta})$ is used to measure the performance of the selection procedure $\underset{\sim}{\delta}$.

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

(b) A selection procedure $\underset{\sim}{\delta}$ is said to be asymptotically optimal of order $\{\alpha_k\}$ if $\rho(G,\underset{\sim}{\delta}) = O(\alpha_k)$ where $\{\alpha_k\}$ is a sequence of decreasing positive numbers such that $\alpha_k \to 0$ as $k \to \infty$.

The asymptotic optimality of the empirical Bayes selection procedures $\underset{\sim}{\delta}^*$ and $\hat{\underset{\sim}{\delta}}$ will be investigated in the next two sections.

4. Asymptotic Optimality of the Empirical Bayes Selection Procedures

Under the preceding described statistical model and the loss function, one can see that for the Bayes selection procedure $\underset{\sim}{\delta}_B$, $R_1(G,\delta_{B1}) = \ldots = R_k(G,\delta_{Bk})$ and $R(G,\underset{\sim}{\delta}_B) = k\ R_1(G,\delta_{B1})$.

Also, by the symmetric properties of the empirical Bayes selection procedures $\underset{\sim}{\delta}^*$ and $\hat{\underset{\sim}{\delta}}$, we have $R_1(G,\delta_1^*) = \ldots = R_k(G,\delta_k^*)$ and $R(G,\underset{\sim}{\delta}^*) = k\ R_1(G,\delta_1^*)$, $R_1(G,\hat{\delta}_1) = \ldots = R_k(G,\hat{\delta}_k)$ and $R(G,\hat{\underset{\sim}{\delta}}) = k\ R_1(G,\hat{\delta}_1)$.

Therefore, $\rho(G,\underset{\sim}{\delta}^*) = [R_1(G,\delta_1^*) - R_1(G,\delta_{B1})]/R_1(G,\delta_{B1})$ and $\rho(G,\hat{\underset{\sim}{\delta}}) = [R_1(G,\hat{\delta}_1) - R_1(G,\delta_{B1})]/R_1(G,\delta_{B1})$. Since $R_1(G,\delta_{B1})$ is a fixed positive value, to study the asymptotic optimality of the empirical Bayes selection procedures, it suffices to investigate the asymptotic behavior of $R_1(G, \delta_1^*) - R_1(G, \delta_{B1})$ and $R_1(G, \hat{\delta}_1) - R_1(G, \delta_{B1})$ for sufficiently large k.

Theorem 4.1 For the statistical model previously described in Section 2, assume that $E[|\Theta_i|] < \infty$ and a^* is a finite number. Then, the empirical Bayes selection procedures $\underset{\sim}{\delta}^*$ and $\hat{\underset{\sim}{\delta}}$ are asymptotically optimal in the sense that $\rho(G, \underset{\sim}{\delta}^*) \to 0$ and $\rho(G, \hat{\underset{\sim}{\delta}}) \to 0$ as $k \to \infty$.

Proof: It is assumed that k is sufficiently large so that $a^* \in (-C_k^*, C_k^*)$, where $\{C_k^*\}$ is a sequence of increasing positive numbers such that $\lim_{k\to\infty} C_k^* = \infty$ as described in Section 3. By the finiteness of a^* and from (2.5″), (2.9), (3.4) and (3.7),

$$
\begin{aligned}
&R_1(G, \delta_1^*) - R_1(G, \delta_{B1}) \\
&= \int [\theta_0 - \psi_1(x_1)] E_1^* \{\delta_1^*(x_1, \underset{\sim}{X}(1)) - \delta_{B1}(x_1)\} f(x_1) dx_1 \\
&= \int_{-C_k^*}^{a^*} [\theta_0 - \psi_1(x_1)] P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 1, \delta_{B1}(x_1) = 0\} f(x_1) dx_1 \\
&\quad + \int_{a^*}^{C_k^*} [\psi_1(x_1) - \theta_0] P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 0, \delta_{B1}(x_1) = 1\} f(x_1) dx_1.
\end{aligned}
\tag{4.1}
$$

Also, from (2.5″), (2.9), (3.5) and (3.9),

$$
\begin{aligned}
&R_1(G, \hat{\delta}_1) - R_1(G, \delta_{B1}) \\
&= \int_{-C_k^*}^{a^*} [\theta_0 - \psi_1(x_1)] P\{\hat{\delta}_1(x_1, \underset{\sim}{X}(1), W_k) = 1, \delta_{B1}(x_1) = 0\} f(x_1) dx_1 \\
&\quad + \int_{a^*}^{C_k^*} [\psi_1(x_1) - \theta_0] P\{\hat{\delta}_1(x_1, \underset{\sim}{X}(1), W_k) = 0, \delta_{B1}(x_1) = 1\} f(x_1) dx_1.
\end{aligned}
\tag{4.2}
$$

From a corollary of Robbins (1964), to prove the asymptotic optimality of $\underset{\sim}{\delta}^*$ and $\hat{\underset{\sim}{\delta}}$, it suffices to show that for each x_1, and $\epsilon = 0, 1$,

$$P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = \epsilon, \delta_{B1}(x_1) = 1 - \epsilon\} \to 0$$

and $$P\{\hat{\delta}_1(x_1, \underset{\sim}{X}(1), W_k) = \epsilon,\ \delta_{B1}(x_1) = 1-\epsilon\} \to 0$$
as $k \to \infty$. Note that for each fixed $x_1, f_{1k}(x_1)$ and $f_{1k}^{(1)}(x_1)$ are consistent estimates of $f(x_1)$ and $f^{(1)}(x_1)$, respectively. Also, W_k is a consistent estimator of σ^2. Hence, $T_{1k}^*(x_1)$ is a consistent estimator of $T_1(x_1)$ for the σ^2 known case, and $\hat{T}_{1k}(x_1)$ is a consistent estimator of $T_1(x_1)$ for the σ^2 unknown case. Therefore,

$$P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 1, \delta_{B1}(x_1) = 0\}$$

$$= P\{T_{1k}^*(x_1) \geq 0,\ T_1(x_1) < 0\} \to 0 \text{ as } k \to \infty, \text{ and}$$

$$P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 0, \delta_{B1}(x_1) = 1\}$$

$$= P\{T_{1k}^*(x_1) < 0, T_1(x_1) \geq 0\} \to 0 \text{ as } k \to \infty.$$

Similarly,

$$P\{\hat{\delta}_1(x_1, \underset{\sim}{X}(1), W_k) = \epsilon,\ \delta_{B1}(x_1) = 1-\epsilon\} \to 0 \text{ as } k \to \infty.$$

Hence the proof of the theorem is complete. □

5. Rates of Convergence

The following theorem gives the main results of the paper concerning rates of convergence of the empirical Bayes selection procedures.

Theorem 5.1 For the statistical model described in Section 2, assume that $E[|\Theta_1|] < \infty$ and a^* is a finite number.

(a) If we take $h = c_1(k-1)^{-1/(2r+1)}$ and $C_k^* = c_2 h^{-1}$ for some fixed positive values c_1 and c_2, then, $\rho(G, \underset{\sim}{\delta}^*) = O(k^{-\frac{r-2}{2r+1}})$ and $\rho(G, \hat{\underset{\sim}{\delta}}) = O(k^{-\frac{r-2}{2r+1}})$.

(b) Furthermore, suppose that for some $1 < \alpha < 2$, and for some $h > 0$,

$$\text{(A1)} \int \frac{|x_1|^\alpha M^{\alpha/2}(x_1, h)}{f^{\alpha-1}(x_1)} dx_1 < \infty \text{ and (A2)} \int \frac{|x_1|^\alpha N^\alpha(x_1, h)}{f^{\alpha-1}(x_1)} dx_1 < \infty$$

where $M(x_1, h)$ and $N(x_1, h)$ are defined in Lemma 6.3. If we take $h = (k-1)^{-1/(2r+1)}$, then, $\rho(G, \underset{\sim}{\delta}^*) = O(k^{-\frac{\alpha(r-1)}{2r+1}})$ and $\rho(G, \hat{\underset{\sim}{\delta}}) = O(k^{-\frac{\alpha(r-1)}{2r+1}})$.

Proof: We provide proof of the theorem for the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ only. The proof for the empirical Bayes selection procedure $\hat{\underset{\sim}{\delta}}$ is analogous to that of $\underset{\sim}{\delta}^*$ and hence is omitted here. Also, assume that k is large enough so that $a^* \in (-C_k^*, C_k^*)$.

From (4.1) and Lemma 6.1 - 6.3, by noting that $[\psi_1(x_1)-\theta_0]f(x_1) = T_1(x_1)$, we have,

$$
\begin{aligned}
& R_1(G,\delta_1^*) - R_1(G,\delta_{B1}) \\
\leq & \int_{-C_k^*}^{C_k^*} E[|T_{1k}^*(x_1) - T_1(x_1)|^\alpha] \; / \; |T_1(x_1)|^{\alpha-1} dx_1 \\
\leq & \frac{b_1(\alpha)}{[(k-1)h^3]^{\alpha/2}} \int_{-C_k^*}^{C_k^*} [M(x_1,h)]^{\alpha/2} \; / \; |T_1(x_1)|^{\alpha-1} dx_1 \\
& + \; b_1(\alpha)h^{\alpha(r-1)} \int_{-C_k^*}^{C_k^*} N^\alpha(x_1,b) \; / \; |T_1(x_1)|^{\alpha-1} dx_1 \\
& + \; \frac{b_2(\alpha)}{[(k-1)h]^{\alpha/2}} \int_{-C_k^*}^{C_k^*} |x_1-\theta_0|^\alpha \; M^{\alpha/2}(x_1,h) \; / \; |T_1(x_1)|^{\alpha-1} dx_1 \\
& + \; b_2(\alpha)h^{\alpha r} \int_{-C_k^*}^{C_k^*} |x_1-\theta_0|^\alpha N^\alpha(x_1,h) \; / \; |T_1(x_1)|^{\alpha-1} dx_1.
\end{aligned}
\tag{5.1}
$$

Since $f(x_1) = \int \frac{\sqrt{m}}{\sqrt{2\pi}\sigma} \, e^{-\frac{m(x_1-\theta)^2}{2\sigma^2}} dG(\theta) \leq \frac{\sqrt{m}}{\sqrt{2\pi}\sigma} = B_4$ for all x_1,

$$
0 \leq M(x_1,h) = \int_0^1 f(x_1-vh)dv \leq B_4 \tag{5.2}
$$

for all x_1 and h. Also,

$$
\left.\frac{\partial^r f(t)}{\partial \, t^r}\right|_{t \, = \, x_1-hw} = \int \frac{\sqrt{m}}{\sqrt{2\pi}\sigma} \; e^{-\frac{m(x_1-hw-\theta)^2}{2\sigma^2}} \left[\sum_{j=0}^{r} a_j(x_1-hw-\theta)^j\right] dG(\theta)
$$

where $a_j, j=0,\ldots,r$, are finite numbers. Therefore, $\left| \left[\left.\frac{\partial^r f(t)}{\partial \, t^r}\right|_{t \, = \, x_1-hw} \right] \right| \leq B_5$ for some positive number B_5 for all x_1, w and h and hence, for all x_1 and h

$$
N(x_1,h) \leq B_5. \tag{5.3}
$$

For $\alpha = 1$, substituting the inequalities of (5.2) and (5.3) into (5.1) we obtain,

$$
\begin{aligned}
& R_1(G, \delta_1^*) - R_1(G, \delta_{B1}) \\
\leq\ & 2b_1(1)B_4^{1/2}\ C_k^*\ /\ [(k-1)h^3]^{1/2} + 2b_1(1)B_5\ C_k^*\ h^{r-1} \\
+\ & b_2(1)B_4^{1/2} \int_{-C_k^*}^{C_k^*} |x_1 - \theta_0| dx_1\ /\ [(k-1)h]^{1/2} \\
+\ & b_2(1)B_5\ h^r \int_{-C_k^*}^{C_k^*} |x_1 - \theta_0| dx_1 \\
\leq\ & 2b_1(1)B_4^{1/2}\ C_k^*\ /\ [(k-1)h^3]^{1/2} + 2b_1(1)B_5\ C_k^*\ h^{r-1} \\
+\ & b_2(1)B_4^{1/2} \left\{ 2|\theta_0|C_k^* + \frac{C_k^{*2}}{2} \right\}\ /\ [(k-1)h]^{1/2} \\
+\ & b_2(1)B_5\ h^r \left\{ 2|\theta_0|C_k^* + \frac{C_k^{*2}}{2} \right\} \\
=\ & O(C_k^*\ /\ [(k-1)h^3]^{1/2}) + O(C_k^*\ h^{r-1}) \\
+\ & O(C_k^{*2}\ /\ [(k-1)h]^{1/2}) + O(C_k^{*2}\ h^r).
\end{aligned}
\tag{5.4}
$$

Thus, if we let $h = c_1(k-1)^{-1/(2r+1)}$ and $C_k^* = c_2 h^{-1}$ where c_1 and c_2 are positive constants, then, from (5.4), $R_1(G, \delta_1^*) - R_1(G, \delta_{B1}) = O(k^{-\frac{r-2}{2r+1}})$.

This completes the proof of part (a).

For each k, let $I_k^c = [-C_k^*, a^* - c_0) \cup (a^* + c_0, C_k^*]$. Since $T_1(x_1) = [\psi_1(x_1) - \theta_0] f(x_1)$, by Lemma 6.4, for $x_1 \in I, |T_1(x_1)| \geq |x_1 - a^*| f(x_1) b_3$, and for $x_1 \in I_k^c$, $|T_1(x_1)| \geq f(x_1) b_4$. Combining these inequalities with (5.1) together, we obtain

$$R_1(G,\delta_1^*) - R_1(G,\delta_{B1})$$

$$\begin{aligned}
&\le \frac{b_1(\alpha)E_1}{b_3^{\alpha-1}((k-1)h^3)^{\alpha/2}} + \frac{b_1(\alpha)E_2}{b_4^{\alpha-1}((k-1)h^3)^{\alpha/2}} + \frac{b_1(\alpha)h^{\alpha(r-1)}E_3}{b_3^{\alpha-1}} \\
&+ \frac{b_1(\alpha)h^{\alpha(r-1)}E_4}{b_4^{\alpha-1}} + \frac{b_2(\alpha)E_5}{((k-1)h)^{\alpha/2}\, b_3^{\alpha-1}} + \frac{b_2(\alpha)E_6}{((k-1)h)^{\alpha/2}\, b_4^{\alpha-1}} \qquad (5.5)\\
&+ \frac{b_2(\alpha)h^{\alpha r}E_7}{b_3^{\alpha-1}} + \frac{b_2(\alpha)h^{\alpha r}E_8}{b_4^{\alpha-1}},
\end{aligned}$$

where

$$E_1 = \int_I \frac{M^{\alpha/2}(x_1,h)}{[|x_1-a^*|f(x_1)]^{\alpha-1}}dx_1,$$

$$E_2 = \int_{I_k^c} \frac{M^{\alpha/2}(x_1,h)}{f^{\alpha-1}(x_1)}dx_1,$$

$$E_3 = \int_I \frac{N^{\alpha}(x_1,h)}{[|x_1-\alpha^*|f(x_1)]^{\alpha-1}}dx_1,$$

$$E_4 = \int_{I_k^c} \frac{N^{\alpha}(x_1,h)}{f^{\alpha-1}(x_1)}dx_1,$$

$$E_5 = \int_I \frac{|x_1-\theta_0|^\alpha\ M^{\alpha/2}(x_1,h)}{[|x_1-a^*|f(x_1)]^{\alpha-1}}dx_1,$$

$$E_6 = \int_{I_k^c} \frac{|x_1-\theta_0|^\alpha\ M^{\alpha/2}(x_1,h)}{f^{\alpha-1}(x_1)}dx_1,$$

$$E_7 = \int_I \frac{|x_1-\theta_0|^\alpha\ N^{\alpha}(x_1,h)}{[|x_1-a^*|f(x_1)]^{\alpha-1}}dx_1,$$

and

$$E_8 = \int_{I_k^c} \frac{|x_1-\theta_0|^\alpha\ N^{\alpha}(x_1,h)}{f^{\alpha-1}(x_1)}dx_1.$$

Since I is a bounded interval and $0 \le M(x_1,h) \le B_4$, $0 \le N(x_1,h) \le B_5$ for all x_1 and h, one can see that for $1 < \alpha < 2$, $0 \le E_i < \infty, i = 1,3,5,7$. Also, under the assumptions (A1) and (A2), $0 \leq E_i < \infty, i = 2, 4, 6, 8$. Therefore,

$$R_1(G, \delta_1^*) - R_1(G, \delta_{B1}) = O(((k-1)h^3)^{-\alpha/2}) + O(h^{\alpha(r-1)}).$$

If we take $h = (k-1)^{-\frac{1}{2r+1}}$, then $R_1(G, \delta_1^*) - R_1(G, \delta_{B1}) = O(k^{-\frac{\alpha(r-1)}{2r+1}})$.

Hence, this completes the proof of part (b). □

Remarks

(a) In Theorem 5.1(a), we have demonstrated that under the very mild conditions that $E[|\Theta_1|] < \infty$ and a^* is finite, the relative regret risk converges to zero with a rate of order $O(k^{-\frac{r-2}{2r+1}})$ where $r > 2$ is a positive integer, involving in the choice of the kernels k_0 and k_1, see (3.1) and (3.2). According to (3.1) and (3.2), there is no restriction about the choice of the value r. Therefore, we may choose r as large as possible, so that the rate of convergence of the relative regret risk may have an order close to $O(k^{-1/2})$.

(b) Though under the assumptions A1 and A2, the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ may have a better convergence rate, it is possible that the assumptions A1 and/or A2 may not hold for any $1 < \alpha < 2$ and any $0 < h < 1$. For example, consider a prior distribution with density $g(\theta)$ where

$$g(\theta) = \begin{cases} \frac{1}{4} & \text{if } |\theta| \leq 1, \\ \frac{1}{4\theta^2} & \text{if } |\theta| > 1. \end{cases}$$

Then, one can see that $f(x)$ is symmetric about the point 0 and $f(x)$ is decreasing in x for $x \geq 0$. Hence, for $x > vh > 0$, $M(x,h) = \int_0^1 f(x - vh)dx \geq f(x)$. Also, as $x > 2$, $f(x) \geq \frac{c}{x^2}$ for some $c > 0$. Therefore,

$$\int_{-\infty}^{\infty} \frac{|x|^\alpha M^{\alpha/2}(x,h)}{f^{\alpha-1}(x)} dx \geq \int_2^{\infty} \frac{x^\alpha M^{\alpha/2}(x,h)}{f^{\alpha-1}(x)} dx$$

$$\geq \int_2^{\infty} x^\alpha f^{1-\alpha/2}(x) dx$$

$$\geq \int_2^\infty x^\alpha \left(\frac{c}{x^2}\right)^{1-\alpha/2} dx$$

$$= c^{1-\alpha/2} \int_2^\infty x^{2\alpha-2} dx$$

$$= \infty \text{ since } 1 < \alpha < 2.$$

Therefore A1 does not hold in this case.

(c) We provide an example in which both the assumptions A1 and A2 hold for any $1 < \alpha < 2$.

Assume that $\Theta_1 \sim N(0, \tau^2)$. Then, marginally $X_1 \sim N(0, q)$ where $q = \frac{\sigma^2}{m} + \tau^2$. We need to verify that

$$\int \frac{|x|^\alpha \ M^{\alpha/2}(x,h)}{f^{\alpha-1}(x)} dx < \infty \text{ and } \int \frac{|x|^\alpha \ N^\alpha(x,h)}{f^{\alpha-1}(x)} dx < \infty.$$

Since $f(x) = \frac{1}{\sqrt{2\pi q}} \ e^{-\frac{x^2}{2q}}$,

$$\frac{M(x,h)}{f(x)} = \int_{v=0}^1 e^{\frac{2vhx - v^2h^2}{2q}} \ dv \leq \begin{cases} 1 & \text{if } x < 0, \\ e^{\frac{hx}{q}} & \text{if } x > 0. \end{cases}$$

Hence,

$$\int \frac{|x|^\alpha \ M^{\alpha/2}(x,h)}{f^{\alpha-1}(x)} dx \leq \int_{-\infty}^0 |x|^\alpha f^{1-\alpha/2}(x) dx + \int_0^\infty x^\alpha \ f^{1-\alpha/2} \ (x) e^{\frac{hx}{q}} \ dx < \infty.$$

Also,

$$\left.\frac{\partial^r f(t)}{\partial \ t^r}\right|_{t=x-hw} = \frac{1}{\sqrt{2\pi q}} \ e^{-\frac{(x-hw)^2}{2q}} \ \left[\sum_{j=0}^r a_j (x-hw)^j\right]$$

for some real values $a_j, j = 0, 1, \ldots, r$. Then,

$$\frac{\left.\frac{\partial^r f(t)}{\partial \ t^r}\right|_{t=x-hw}}{f(x)} = e^{\frac{2hwx - h^2w^2}{2q}} \ \left[\sum_{j=0}^r a_j (x-hw)^j\right].$$

Hence,

$$\frac{N(x,h)}{f(x)} \leq e^{\frac{h|x|}{q}} \sup_{0 \leq w \leq 1} \sum_{j=0}^r |a_j| \ |x - hw|^j.$$

Therefore,

$$\int \frac{|x|^\alpha \ N^\alpha(x,h)}{f^{\alpha-1}(x)} dx \le \sum_{j=0}^{r} \sup_{0\le w\le 1} \int |x|^\alpha f(x)|a_j| \ |x-hw|^j \ e^{\frac{h|x|}{q}} \ dx < \infty.$$

That is, both the assumptions A1 and A2 hold.

6. Useful Lemmas

The following lemmas are useful for presenting a concise proof of Theorem 5.1.

Lemma 6.1 (a) For $x_1 < a^*$ and $\alpha > 0$,

$$P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 1\} \le E[|T_{1k}^*(x_1) - T_1(x_1)|^\alpha] \ / \ |T_1(x_1)|^\alpha.$$

(b) For $x_1 > a^*$ and $\alpha > 0$.

$$P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 0\} \le E[|T_{1k}^*(x_1) - T_1(x_1)|^\alpha] \ / \ |T_1(x_1)|^\alpha.$$

Proof: For $x_1 < a^*, T_1(x_1) = \frac{\sigma^2}{m} f^{(1)}(x_1) + (x_1 - \theta_0) f(x_1) < 0$. Then by the definition of the empirical Bayes selection procedure $\underset{\sim}{\delta}^*$ and by Markov inequality,

$$\begin{aligned} P\{\delta_1^*(x_1, \underset{\sim}{X}(1)) = 1\} &= P\{T_{1k}^*(x_1) \ge 0\} \\ &= P\{T_{1k}^*(x_1) - T_1(x_1) \ge -T_1(x_1)\} \\ &\le E[|T_{1k}^*(x_1) - T_1(x_1)|^\alpha] \ / \ |T_1(x_1)|^\alpha. \end{aligned}$$

Part (b) can also be obtained in a similar way. □

Lemma 6.2 For $0 < \alpha \le 2$, for each fixed x_1,

$$\begin{aligned} & E[|T_{1k}^*(x_1) - T_1(x_1)|^\alpha] \\ \le \ & C_\alpha^2 \left(\frac{\sigma^2}{m}\right)^\alpha \{\mathrm{Var}^{\alpha/2}(f_{1k}^{(1)}(x_1)) + |Ef_{1k}^{(1)}(x_1) - f^{(1)}(x_1)|^\alpha\} \\ & + \ C_\alpha^2 |x_1 - \theta_0|^\alpha \{ \ \mathrm{Var}^{\alpha/2}(f_{1k}(x_1)) + |Ef_{1k}(x_1) - f(x_1)|^\alpha\}, \end{aligned}$$

where

$$C_\alpha = \begin{cases} 1 & \text{if } 0 < \alpha \le 1, \\ 2^{\alpha-1} & \text{if } 1 < \alpha \le 2. \end{cases}$$

Proof: By the definitions of $T^*_{1k}(x_1)$ and $T_1(x_1)$ and by C_r-inequality,

$$\begin{aligned} & E[|T^*_{1k}(x_1) - T_1(x_1)|^\alpha] \\ = \ & E\Big[\Big|\Big[\frac{\sigma^2}{m}\ f^{(1)}_{1k}(x_1) + (x_1-\theta_0)f_{1k}(x_1)\Big] - \Big[\frac{\sigma^2}{m} f^{(1)}(x_1) + (x_1-\theta_0)f(x_1)\Big]\Big|^\alpha\Big] \\ \le \ & C_\alpha\left(\frac{\sigma^2}{m}\right)^\alpha E\big[|f^{(1)}_{1k}(x_1) - f^{(1)}(x_1)|^\alpha\big] \\ & + \ C_\alpha|x_1-\theta_0|^\alpha\ E\big[|f_{1k}(x_1) - f(x_1)|^\alpha\big]. \end{aligned} \tag{6.1}$$

Again, by C_r-inequality, for $0 < \alpha \le 2$,

$$\begin{aligned} & E[|f^{(1)}_{1k}\ (x_1) - f^{(1)}(x_1)|^\alpha] \\ = \ & E[|f^{(1)}_{1k}\ (x_1) - Ef^{(1)}_{1k}(x_1) + Ef^{(1)}_{1k}(x_1) - f^{(1)}(x_1)|^\alpha] \\ \le \ & C_\alpha E[|f^{(1)}_{1k}(x_1) - Ef^{(1)}_{1k}(x_1)|^\alpha] + C_\alpha|Ef^{(1)}_{1k}(x_1) - f^{(1)}(x_1)|^\alpha \\ \le \ & C_\alpha\ \mathrm{Var}^{\alpha/2}(f^{(1)}_{1k}(x_1)) + C_\alpha|Ef^{(1)}_{1k}(x_1) - f^{(1)}(x_1)|^\alpha, \end{aligned} \tag{6.2}$$

and

$$\begin{aligned} & E[|f_{1k}(x_1) - f(x_1)|^\alpha] \\ \le \ & C_\alpha E[|f_{1k}(x_1) - Ef_{1k}(x_1)|^\alpha] + C_\alpha|Ef_{1k}(x_1) - f(x_1)|^\alpha \\ \le \ & C_\alpha\ \mathrm{Var}^{\alpha/2}(f_{1k}(x_1)) + C_\alpha|Ef_{1k}(x_1) - f(x_1)|^\alpha. \end{aligned} \tag{6.3}$$

Substituting (6.2) and (6.3) into (6.1) yields the result of the lemma. □

Lemma 6.3 For each fixed x_1,

(a) $\mathrm{Var}\ (f_{1k}(x_1)) \le B_3^2\ M(x_1,h)\ /\ [(k-1)h]$,

(b) $\mathrm{Var}\ (f^{(1)}_{1k}(x_1))\ \le\ B_3^2\ M(x_1,h)\ /\ [(k-1)h^3]$,

(c) $|Ef_{1k}(x_1) - f(x_1)| \leq B_3\ h^r\ N(x_1, h)$,

(d) $|Ef_{1k}^{(1)}(x_1) - f^{(1)}(x_1)| \leq B_3 h^{r-1} N(x_1, h)$,

(e) $E[|T_{1k}^*(x_1) - T_1(x_1)|^\alpha]$

$$\leq\ b_1(\alpha)\{[M(x_1,h)\ /\ ((k-1)h^3)]^{\alpha/2} + [h^{r-1}N(x_1,h)]^\alpha\}$$

$$+\ b_2(\alpha)|x_1 - \theta_0|^\alpha\{[M(x_1,h)\ /\ ((k-1)h)]^{\alpha/2} + [h^r\ N(x_1,h)]^\alpha\},$$

where $b_1(\alpha) = C_\alpha^2\ B_3^\alpha(\frac{\sigma^2}{m})^\alpha$, $b_2(\alpha) = C_\alpha^2\ B_3^\alpha$, $M(x_1,h) = \int_0^1 f(x_1 - vh)dv$, and $N(x_1,h) = \sup_{0\leq w\leq 1}\left|\left[\frac{\partial^r f(t)}{\partial\ t^r}\Big|_{t=x_1-hw}\right]\right|$.

Proof (a)

$$\begin{aligned}
\text{Var}\ (f_{1k}(x_1)) &= \frac{1}{(k-1)h^2}\ \text{Var}(k_0(\frac{x_1 - X_2}{h})) \\
&\leq \frac{1}{(k-1)h^2}\ E[k_0^2(\frac{x_1 - X_2}{h})] \\
&= \frac{1}{(k-1)h}\int_0^1 k_0^2(v) f(x_1 - vh)dv \\
&\leq \frac{B_3^2}{(k-1)h}\int_0^1 f(x_1 - vh)dv \\
&= \frac{B_3^2}{(k-1)h}\ M(x_1,h).
\end{aligned}$$

Part (b) can be obtained in a similar way.

(c) $$Ef_{1k}(x_1) = \frac{1}{h}\ E\ k_0(\frac{x_1 - X_2}{h}) = \int_0^1 k_0\ (w) f(x_1 - hw)dw.$$

Since $\int_0^1 k_0(x)dw = 1$, hence, by the property of k_0,

$$|Ef_{1k}(x_1) - f(x_1)| = |\int_0^1 k_0(w)[f(x_1 - hw) - f(x_1)]dw|$$

$$= |\int_0^1 k_0(w)\left[\frac{\partial^r f(t)}{\partial\, t^r}\Big|_{t=x_1-hw^*}\right](-hw)^r dw|$$

where $0 < w^* < w$

$$\leq B_3\, h^r \sup_{0\leq w\leq 1}\left|\left[\frac{\partial^r f(t)}{\partial\, t^r}\Big|_{t=x_1-hw}\right]\right|$$

$$= B_3\, h^r N(x_1, h).$$

The proof of part (d) can be obtained in a way similar to part (c).

Finally, part (e) is obtained by plugging the results of parts (a)-(d) into the inequality in Lemma 6.2. □

For a positive constant c_0, define interval $I = [a^* - c_0, a^* + c_0]$. Note that $\psi_1(x_1) = \int \theta\, e^{\frac{m\theta x_1}{\sigma^2} - \frac{m\theta^2}{2\sigma^2}} dG(\theta) \;/\; \int e^{\frac{m\theta x_1}{\sigma^2} - \frac{m\theta^2}{2\sigma^2}} dG(\theta)$. Hence,

$$\psi_1^{(1)}(x_1) = \frac{\frac{m}{\sigma^2}\{\int \theta^2 e^{\frac{m\theta x_1}{\sigma^2} - \frac{m\theta^2}{2\sigma^2}} dG(\theta) \int e^{\frac{m\theta x_1}{\sigma^2} - \frac{m\theta^2}{2\sigma^2}} dG(\theta) - [\int \theta\, e^{\frac{m\theta x_1}{\sigma^2} - \frac{m\theta^2}{2\sigma^2}} dG(\theta)]^2\}}{[\int e^{\frac{m\theta x_1}{\sigma^2} - \frac{m\theta^2}{2\sigma^2}} dG(\theta)]^2},$$

which is continuous in x_1 and positive for all x_1 since G is nondegenerate.

Lemma 6.4

(a) Let $b_3 = \inf\{\psi_1^{(1)}(x_1) | x_1 \in I\}$. Then $b_3 > 0$.

(b) For any $x_1 \in I, |\psi_1(x_1) - \theta_0| \geq |x_1 - a^*| b_3$.

(c) For any $x_1 \notin I$, $|\psi_1(x_1) - \theta_0| \geq b_4$ for some $b_4 > 0$.

Proof: (a) Since $\psi_1^{(1)}(x_1)$ is continuous in x_1 on the compact interval I, there exists an $x^* \in I$ such that $\psi_1^{(1)}(x^*) = b_3$. Also, note that $\psi_1^{(1)}(x^*)$ can also be viewed as the variance of some non-degenerate distribution. Hence $b_3 > 0$.

(b) By the definition of $a^*, \psi_1(a^*) = \theta_0$. By mean-value theorem, for $x_1 \in I$.

$$\begin{aligned} |\psi_1(x_1) - \theta_0| &= |\psi_1(x_1) - \psi_1(a^*)| \\ &= |(x_1 - a^*)\psi_1^{(1)}(x_1^*)| \\ &\geq |x_1 - a^*| b_3, \end{aligned}$$

where x_1^* is a point between x_1 and a^*.

(c) Note that $\psi_1(x_1) - \theta_0$ is strictly increasing in x_1 and $\psi_1(a^*) - \theta_0 = 0$. So, $|\psi_1(x_1) - \theta_0| \geq \min(|\psi_1(a^* - c_0) - \theta_0|, |\psi_1(a^* + c_0) - \theta_0|) \equiv b_4 > 0$ for all $x_1 \notin I$. □

References

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.*, **50**, 1096-1121.

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.

Gupta, S. S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.*, **29**, 235-244.

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

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

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

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

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

Singh, R. S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate. *Ann. Statist.*, **7**, 890-902.

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

Tong, Y. L. (1969). On partitioning a set of normal populations by their locations with respect to a control. *Ann. Math. Statist.*, **40**, 1300-1324.