ON EMPIRICAL BAYES PROCEDURES FOR SELECTING GOOD POPULATIONS IN POSITIVE EXPONENTIAL FAMILY

by

Shanti S. Gupta and Jianjun Li
Purdue University

Technical Report # 01-10

Department of Statistics
Purdue University
West Lafayette, IN USA

August 2001

*

ON EMPIRICAL BAYES PROCEDURES FOR SELECTING GOOD POPULATIONS IN POSITIVE EXPONENTIAL FAMILY [1]

Shanti S. Gupta *and* Jianjun Li

Department of Statistics, Purdue University, W. Lafayette, IN 47907

Department of Statistics, Purdue University, W. Lafayette, IN 47907

Abstract:

The problem of selecting good ones compared with a control from $k(\geq 2)$ positive exponential family populations is considered in this paper. A nonparametric empirical Bayes approach is used to construct the selection procedures. It has been shown that the risks of the empirical Bayes procedures converge to the (minimum) Bayes risk with a rate of $O(1/n)$, where n is the number of accumulated past observations at hand. Simulations were carried out to study the performance of the procedures for small to moderate values of n. The results of this study are provided in the paper.

AMS 1991 Subject classification: Primary 62F07; secondary 62C12.

Key words and phrases: Empirical Bayes; selection procedures, positive exponential family; regret Bayes risk; rate of convergence.

[1]This research was supported in part by US Army Research Office, Grant DAAH 04-95-1-0165 and Grant DAAD 19-00-1-0502 at Purdue University. The authors are very thankful to Professor Friedrich Liese and Professor Tachen Liang for their helpful comments which led to improvements in the paper.

1. Introduction and Formulation.

In this paper, we are interested in the problem of simultaneous inference and selection from among $k(\geq 2)$ populations in comparison with a standard or control. The populations are denoted by $\pi_1, \cdots, \pi_k$. The random variable X_i associated with π_i has the density $f(x_i|\theta_i) = c(\theta_i)e^{-x_i/\theta_i}h(x_i)$ with $h(x) > 0$ in $(0, \infty)$, where the unknown parameter θ_i is the characterization of population π_i.

Let θ_0 denote a standard or a control. In practical situations, we desire to differentiate between *good* and *bad* populations and select *good* ones and exclude *bad* ones. Here a population π_i is said to be *good* if $\theta_i \geq \theta_0$ and *bad* otherwise. This type of decision problem has been considered by many authors. For example, see early papers: Paulson (1952), Dunnett (1955), Gupta and Sobel (1958), Lehmann (1961), and later: Gupta and Hsiao (1983), and more recently: Gupta, Liang and Rau (1994), Gupta and Liang (1999), among others.

Let $\Omega = \{\tilde{\theta} = \{\theta_1, \cdots, \theta_k\} : \theta_i > 0, i = 1, 2, \cdots, k\}$ be the parameter space. Let $A = \{\tilde{a} = \{a_1, \cdots, a_k\} : a_i = 0 \text{ or } 1, i = 1, \cdots, k\}$ be the action space, where $a_i = 1$ means that population π_i is selected as good, $a_i = 0$ means population π_i is excluded as bad.

The loss function we use is

$$L(\tilde{\theta}, \tilde{a}) = \sum_{i=1}^{k} l(\theta_i, a_i) \tag{1.1}$$

with

$$l(\theta_i, a_i) = a_i\theta_i(\theta_0 - \theta_i)I_{[\theta_i < \theta_0]} + (1 - a_i)\theta_i(\theta_i - \theta_0)I_{[\theta_i \geq \theta_0]}.$$

We also assume that θ_i is a realization of a random variable Θ_i, and $\Theta_1, \cdots, \Theta_k$ are independently distributed with priors $G_1, \cdots, G_k$ respectively. Let $G = \prod_{i=1}^{k} G_i(\theta_i)$.

Let $\widetilde{X} = (X_1, \cdots, X_k)$ and $\mathcal{X}$ be the sample space of $\widetilde{X}$. Here X_i may be thought of as a sufficient statistic based on several i.i.d. samples.

The selection procedure $\tilde{\delta} = (\delta_1, \cdots, \delta_k)$, where $\delta_i(\tilde{x})$ is the probability of selecting population π_i as good when $\widetilde{X} = \tilde{x}$ is observed. To ensure that the Bayes rule exists, we assume $\int_0^\infty \theta^2 dG_i(\theta) < \infty$ for $i = 1, \cdots, k$.

Based on previous assumptions, a straightforward computation shows that

$$R(G, \tilde{\delta}) = \sum_{i=1}^{k} R_i(G, \delta_i) \tag{1.2}$$

and

$$R_i(G, \delta_i) = \int_{\mathcal{X}} \delta_i(\tilde{x})[\prod_{j \neq i} f_j(x_j)]w_i(x_i)h(x_i)d\tilde{x} + T_i \tag{1.3}$$

where

$$\begin{aligned} f_i(x_i) &= \int_0^\infty c(\theta)e^{-x_i/\theta}h(x_i)dG_i(\theta), \\ w_i(x_i) &= \int_0^\infty \theta(\theta_0-\theta)c(\theta)e^{-x_i/\theta}dG_i(\theta), \\ T_i &= \int_0^\infty \theta(\theta-\theta_0)I_{[\theta>\theta_0]}dG_i(\theta). \end{aligned}$$

Here $f_i(x_i)$ is the marginal density of X_i and T_i is independent of the selection rule $\tilde{\delta}$. Clearly, a Bayes selection procedure $\tilde{\delta}_G = (\delta_{Gi}, \cdots, \delta_{Gk})$ is given by

$$\delta_{Gi} = \begin{cases} 1 & \text{if } w_i(x_i) \le 0, \\ 0 & \text{if } w_i(x_i) > 0. \end{cases} \tag{1.4}$$

Let $\alpha_i(x_i) = \int_0^\infty \theta c(\theta)e^{-x_i/\theta}dG_i(\theta)$ and $\psi_i(x_i) = \int_0^\infty \theta^2 c(\theta)e^{-x_i/\theta}dG_i(\theta)$. Denote $\phi_i(x_i) = \psi_i(x_i)/\alpha_i(x_i)$. Then δ_{Gi} can be expressed as

$$\delta_{Gi} = \begin{cases} 1 & \text{if } \phi_i(x_i) \ge \theta_0, \\ 0 & \text{if } \phi_i(x_i) < \theta_0. \end{cases} \tag{1.5}$$

If G_i is unknown, the Bayes rule cannot be applied and the selection cannot be made. The empirical Bayes approach is a way to help one to make the decision when past data are available. Since Robbins (1956, 1964) introduced the empirical Bayes approach, it has become a powerful tool in decision-making. Empirical Bayes approach in statistical inferences has been used recently by Singh and Wei (1992), van Houwelingen and Stijnen (1993), Pensky (1998), Pensky and Singh (1999), and Liang (2000a, 2000b).

For each $i = 1, \ldots, k$, let $(X_{ij}, \Theta_{ij}), j = 1, 2, \ldots$ be random vectors associated with population π_i and stage j, where X_{ij} is observable while Θ_{ij} is unobservable. It is assumed that Θ_{ij} has a prior distribution G_i, for all $j = 1, 2, \ldots$, and conditioning on $\Theta_{ij} = \theta_{ij}, X_{ij}$ follows a distribution with density $f(x_{ij}|\theta_{ij})$ and $(X_{ij}, \Theta_{ij}), i = 1, \ldots, k, j = 1, 2, \ldots$ are mutually independent. At the present stage, say, stage $n+1$, we have observed $\tilde{X} = \tilde{x}$. The past accumulated observations are denoted by $(\tilde{X}_1, \cdots, \tilde{X}_n) = \tilde{\tilde{X}}_n$, where $\tilde{X}_j = (X_{1j}, \cdots, X_{kj})$ is the observation at stage j. Based on $\tilde{\tilde{X}}_n$ and $\tilde{x}$, we wish to construct an empirical Bayes rule to select all good populations and to exclude all bad populations. Such an empirical Bayes rule can be expressed as

$$\tilde{\delta}_n(\tilde{x}, \tilde{\tilde{X}}_n) = (\delta_{n1}(\tilde{x}, \tilde{\tilde{X}}_n), \cdots, \delta_{nk}(\tilde{x}, \tilde{\tilde{X}}_n))$$

where $\delta_{ni}(\tilde{x}, \tilde{\tilde{X}}_n)$ is the probability of selecting π_i as good if $\tilde{\tilde{X}}_n$ and $\tilde{x}$ are observed. Let

$R(G, \widetilde{\delta}_n)$ denote the overall Bayes risk of $\widetilde{\delta}_n$. Then

$$(1.6) \qquad R(G, \widetilde{\delta}_n) = \sum_{i=1}^{k} R_i(G, \delta_{ni}),$$

where

$$(1.7) \qquad R_i(G, \delta_{ni}) = \int_{\mathcal{X}} E[\delta_{ni}(\widetilde{x}, \widetilde{\widetilde{X}})] \cdot [\prod_{j \neq i} f_j(x_j)] \cdot w_i(x_i) h(x_i) d\widetilde{x} + T_i.$$

The regret Bayes risk is defined as $r_n = R(G, \widetilde{\delta}_n) - R(G, \widetilde{\delta}_G)$, which is used to measure the performance of empirical Bayes rule $\widetilde{\delta}_n$. If $r_n = o(1)$, we say that $\widetilde{\delta}_n$ is asymptotically optimal (a.o.). If $r_n = O(\beta_n)$ for some positive β_n such that $\lim_{n \to \infty} \beta_n = 0$, we say that $\widetilde{\delta}_n$ is asymptotically optimal at a rate of $O(\beta_n)$.

The aim of this paper is to construct an empirical Bayes rule δ_n for the selection problem described above. For most distributions in the family $f(x_i|\theta_i)$, under the above general setting or, in some cases, with one additional condition $\int_0^\infty \theta^3 dG(\theta) < \infty$, we show that $nr_n \to m$, where m is a computable constant depending on G.

It should be pointed out that Gupta and Liang (1999) studied the selection problem for $gamma(x|\theta, s)$ populations, a special case of above problem, firstly through an empirical Bayes approach. They constructed an empirical Bayes rule $\widetilde{\delta}_n^*$ and established its convergence rate $O(n^{-1})$ under some regularity conditions. A rate of $O(n^{-1} \log n)$ was obtained there under the condition that $\Theta_i' s$ are bounded. Gupta and Liese (2000) showed that the limiting distribution of nR_n is a linear combination of independent χ^2 random variables, where R_n is the conditional regret of a modified version of rule $\widetilde{\delta}_n^*$.

The remaining part of this paper is organized as follows. In Section 2, an empirical Bayes selection rule $\widetilde{\delta}_n$ is constructed. The asymptotic behavior of $\widetilde{\delta}_n$ is investigated in Section 3. In Section 4, we provide a few typical examples as applications of our results. The proofs of our results are given in Section 5.

2. Construction of Empirical Bayes Selection Procedure $\widetilde{\delta}_n$.

The construction of $\widetilde{\delta}_n$ can be divided into three steps. First, we construct an estimator of $w_i(x)$. Second, we localize the Bayes rule. And then we complete the construction by mimicking the Bayes rule using the estimator of $w_i(x)$.

The construction of an estimator of $w_i(x)$ follows the idea of Gupta and Liang (1999). For the loss function (1.1), an unbiased and consistent estimator of $w_i(x)$ can be obtained.

For each $i = 1, \cdots, k$, $j = 1, \cdots, n$, and $x > 0$, define

$$V_{ij}(x) = \frac{\theta_0 + x - X_{ij}}{h(X_{ij})} I_{[X_{ij} \geq x]}. \tag{2.1}$$

Through a standard calculation, we have $E[V_{ij}(x)] = w_i(x)$. Based on this nice property, an unbiased and consistent estimator of $w_i(x)$ can be constructed as:

$$W_{ni}(x) = \frac{1}{n} \sum_{j=1}^{n} V_{ij}(x), \tag{2.2}$$

for each $i = 1, \cdots, k$, and $x \in (0, \infty)$.

We call the next step as a localization of the Bayes test. Examining the Bayes selection rule $\widetilde{\delta}_G$, we see that δ_{Gi} depends on $\widetilde{x}$ only through x_i. Also $\phi_i(x)$ is increasing for $i = 1, \cdots, k$. If x_i is large so that $\phi_i(x_i) \geq \theta_0$, we have $\delta_{Gi} = 1$; If x_i is small so that $\phi_i(x_i) < \theta_0$, we have $\delta_{Gi} = 0$. Since G is unknown, we do not know at which point we should accept H_0 or reject it. But, one will be more likely to take action $a_i = 1$ if the observation of $X_i = x_i$ is quite large and take action $a_i = 0$ if it is quite small. By knowing this, we want to find two numbers B_n and L_n such that we select π_i as good if we observe $x_i > L_n$ and exclude it as bad if $x_i < B_n$. Here both B_n and L_n depend on n. This could be understood as follows. As n increases, we have more information from the accumulated data, and we should adapt new B_n and L_n so that our decision can be made more precisely.

Certainly, the exact form of $f(x|\theta)$ and the distribution G affect the choice of B_n and L_n. Since we have no knowledge about G except that $\int_0^\infty \theta dG(\theta) < \infty$, we rely on $f(x|\theta)$ itself.

If $\lim_{x \downarrow 0} h(x) > 0$, let $B_n = 1/L_n$ and $L_n = (\theta_0 \log n/4) \vee 10$. If $\lim_{x \downarrow 0} h(x) = 0$, let H_n and L_n be the two sequences of positive numbers such that $H_n e^{L_n/\theta_0} = n^{1/4}$ and $H_n \to \infty$, $L_n \to \infty$ as $n \to \infty$. For example, $L_n = (\theta_0 \log n/12) \vee 10$ and $H_n = n^{1/4} e^{-L_n/\theta_0}$. Then define $B_n = [\inf\{x < 1 : h(x) \leq 1/H_n\} \vee (1/L_n)] \wedge 0.1$. It follows that $B_n \to 0$ since $H_n \to \infty$ as $n \to \infty$.

According to what we mentioned at the beginning of this section, we propose the following empirical Bayes procedure: For each $i = 1, \cdots, k$, and x_i,

$$\delta_{ni}(x_i) = \begin{cases} 1 & \text{if} \quad (x_i > L_n) \text{ or } (B_n \leq x_i \leq L_n \text{ and } W_{ni}(x_i) \leq 0), \\ 0 & \text{if} \quad (x_i < B_n) \text{ or } (B_n \leq x_i \leq L_n \text{ and } W_{ni}(x_i) > 0). \end{cases} \tag{2.3}$$

This empirical Bayes procedure says that, at stage $n+1$, if the present observation x_i from π_i is relatively big or small, a decision will be made based on x_i only. If it is not too small or too big, we have to resort to past data information and use $W_{ni}(x)$, the estimator of $w_i(x)$, to make the decision.

3. Asymptotic Optimality of $\widetilde{\delta}_n$.

In this section, the asymptotic behavior of $\widetilde{\delta}_n$ is investigated. We derive the regret Bayes risk first. From (1.2) and (1.3), the Bayes risk of $\widetilde{\delta}_G$ is $R(G, \widetilde{\delta}_G) = \sum_{i=1}^k R_i(G, \delta_{Gi})$ with

$$R_i(G, \delta_{Gi}) = \int_0^\infty \delta_{Gi}(\widetilde{x}) w_i(x_i) h(x_i) dx_i + T_i.$$

From (1.6) and (1.7), the Bayes risk of $\widetilde{\delta}_n(\widetilde{x})$ is $R(G, \widetilde{\delta}_n) = \sum_{i=1}^k R_i(G, \delta_{ni})$ with

$$R_i(G, \delta_{ni}) = \int_0^\infty E[\widetilde{\delta}_{ni}(\widetilde{x})] w_i(x_i) h(x_i) dx_i + T_i.$$

Thus the regret Bayes risk of $\widetilde{\delta}_n$, the difference between $R(G, \widetilde{\delta}_n)$ and $R(G, \widetilde{\delta}_G)$, is

$$r_n = \sum_{i=1}^k r_{ni}, \tag{3.1}$$

where

$$r_{ni} = \int_{B_n}^{L_n} P(W_{ni}(x) \le 0) w_i(x) I_{[w_i(x)>0]} h(x) dx + \int_{B_n}^{L_n} P(W_{ni}(x) > 0) w_i(x) I_{[w_i(x)<0]} h(x) dx. \tag{3.2}$$

Under the assumption $\int_0^\infty \theta^2 dG_i(\theta) < \infty$, we have $\int_0^\infty |w_i(x)| h(x) dx < \infty$ from the inequality

$$\int_0^\infty |w_i(x)| h(x) dx \le \theta_0 \int_0^\infty \alpha_i(x) h(x) dx + \int_0^\infty \psi_i(x) h(x) dx \le \theta_0 \int_0^\infty \theta dG_i(\theta) + \int_0^\infty \theta^2 dG_i(\theta).$$

Since $W_n(x)$ is a consistent estimator of $w_i(x)$, $P(W_{ni}(x) \le 0) \to 0$ if $w_i(x) > 0$, and $P(W_{ni}(x) > 0) \to 0$ if $w_i(x) < 0$. Applying the dominated convergence theorem, we have $r_{ni} = o(1)$. Thus we have the following theorem.

Theorem 3.1. *Assume that $\int_0^\infty \theta^2 dG_i(\theta) < \infty$ for each $i = 1, 2, \cdots, k$. Then $\widetilde{\delta}_n$, as defined by (2.3), is asymptotically optimal.*

Besides the asymptotic optimality, the convergence rate of an empirical Bayes procedure is also an important factor to be considered when the procedure is applied. The following discussion shows that the procedure $\widetilde{\delta}_n$ achieves the rate of convergence of order $O(n^{-1})$.

From now on, we consider only those members of the family $f(x|\theta)$ in which $\lim_{x\uparrow\infty} h(x) > 0$ and $h(x)$ is bounded from below for any inner closed subset of $(0, \infty)$. These members belong to one of the following cases:

Case 1. $\lim_{x\uparrow\infty} \frac{h(x)}{x} > 0$ and $\lim_{x\downarrow 0} h(x) > 0$.

Case 2. $\lim_{x\uparrow\infty} \frac{h(x)}{x} > 0$ and $\lim_{x\downarrow 0} h(x) = 0$.

Case 3. $\lim_{x\uparrow\infty} \frac{h(x)}{x} = 0$ and $\lim_{x\downarrow 0} h(x) > 0$.

Case 4. $\lim_{x\uparrow\infty}\frac{h(x)}{x}=0$ and $\lim_{x\downarrow 0}h(x)=0$.

Before presenting the main results, we introduce the following definition. If $\delta_{Gi}=1$ for all $x\in(0,\infty)$ or $\delta_{Gi}=0$ for all $x\in(0,\infty)$, we say that δ_{Gi} is degenerate; otherwise we say that δ_{Gi} is non-degenerate.

If δ_{Gi} is non-degenerate, i.e., $\lim_{x\downarrow 0}\phi_i(x)<\theta_0<\lim_{x\uparrow\infty}\phi_i(x)$, then $\phi_i(x)$ is strictly increasing. Therefore there exists a point $c_i\in(0,\infty)$ such that $\phi_i(c_i)=\theta_0$, $\phi_i(x)>\theta_0$ for $x>c_i$ and $\phi_i(x)<\theta_0$ for $x<c_i$.

Theorem 3.2. *Assume that* $\int_0^\infty\theta^2dG_i(\theta)<\infty$ *for* $i=1,\cdots,k$. *In Case 3 and Case 4, we also assume that* $\int_1^\infty\theta^4c(\theta)dG_i(\theta)<\infty$ *for* $i=1,\cdots,k$. *Then*

$$\lim_{n\to\infty}nr_n=\sum_{i=1}^k m_i, \tag{3.3}$$

where

$$m_i=\begin{cases}0 & \text{if } \delta_{Gi} \text{ is degenerate,}\\ \frac{h(c_i)Var([V_{i1}(c_i)])}{2|w_i'(c_i)|} & \text{if } \delta_{Gi} \text{ is non-degenerate.}\end{cases} \tag{3.4}$$

and $Var([V_{i1}(c_i)])$ *is the variance of* $V_{i1}(c_i)$, $w_i'(c_i)$ *is the derivative of* $w_i(x)$ *at* c_i, $w_i'(c_i)\neq 0$ *if* δ_{Gi} *is non-degenerate.*

Proof. The proof is given in Section 5.

In Case 3 and Case 4, the assumptions $\int_0^\infty\theta^2dG_i(\theta)<\infty$ and $\int_1^\infty\theta^4c(\theta)dG_i(\theta)<\infty$ can be simplified into $\int_0^\infty\theta^3dG_i(\theta)<\infty$. So we have the following corollary.

Corollary 3.3. *In Case 3 and Case 4, if* $\int_0^\infty\theta^3dG_i(\theta)<\infty$ *for* $i=1,\cdots,k$, *then (3.3) as well as (3.4) holds.*

Proof. Note that $\theta c(\theta)=\theta[\int_0^\infty exp(-x/\theta)h(x)dx]^{-1}$ and for $\theta>1$,

$$\theta^{-1}\int_0^\infty e^{-x/\theta}h(x)dx=\int_0^\infty e^{-y}h(y\theta)dy\geq e^{-2}\int_1^2 h(y\theta)dy>e^{-2}[\min_{t\geq 1}h(t)]. \tag{3.5}$$

It follows that $\theta c(\theta)$ is bounded for $\theta>1$. Thus from $\int_0^\infty\theta^3dG_i(\theta)<\infty$ we have both $\int_0^\infty\theta^2dG_i(\theta)<\infty$ and $\int_1^\infty\theta^4c(\theta)dG_i(\theta)<\infty$. Then Corollary 3.3 follows Theorem 3.2.

From Theorem 3.2, one sees a rate of order $O(n^{-1})$ is obtained under a (quite) weak condition. We only require $\int_0^\infty\theta^2dG(\theta)<\infty$ in Case 1 and Case 2. The assumption

$\int_0^\infty \theta^2 dG(\theta) < \infty$ guarantees the existence of the Bayes rule. This assumption is natural and not very stringent. In Case 3 and Case 4, we require one moment condition, $\int_0^\infty \theta^3 dG(\theta) < \infty$.

The applications of our results to a few typical distributions are presented in the following section. It includes the construction of $\widetilde{\delta}_n$ and the statement of convergence rate for each distribution there.

4. Examples and Results.

We select a few distributions as examples.

Example 4.1 (*exp* (θ)*-family*). Consider the exponential populations having density

$$f(x_i|\theta_i) = \frac{1}{\theta_i} e^{-x_i/\theta_i}, \qquad x_i > 0, \;\; \theta_i > 0, \;\; i = 1, \cdots, k. \tag{4.1}$$

Here $h(x) \equiv 1$. This family belongs to Case 3. Take $B_n = 1/L_n$, $L_n = (\theta_0 \log n/4) \vee 10$ and construct $\widetilde{\delta}_n$ as

$$\delta_{ni}(x_i) = \begin{cases} 1 & \text{if} \quad (x_i > L_n) \text{ or } (B_n \le x_i \le L_n \text{ and } W_{ni}(x_i) \le 0), \\ 0 & \text{if} \quad (x_i < B_n) \text{ or } (B_n \le x_i \le L_n \text{ and } W_{ni}(x_i) > 0). \end{cases} \tag{4.2}$$

Then applying Corollary 3.3, we have the following result.

Result 4.1. *If X_i has density $f(x_i|\theta_i)$ given in (4.1) and $\int_0^\infty \theta^3 dG_i(\theta) < \infty$ for all $i = 1$, $\cdots$, k, then $\widetilde{\delta}_n$, as constructed in (4.2), has a rate of convergence of order $O(n^{-1})$.*

Example 4.2 (*Gamma* (θ, s)*-family with known* $s > 1$). Consider the gamma populations having density

$$f(x_i|\theta_i) = \frac{x_i^{s-1}}{\Gamma(s)\theta_i^s} e^{-x_i/\theta_i}, \qquad x_i > 0, \;\; \theta_i > 0, \;\; i = 1, \cdots, k. \tag{4.3}$$

Here $h(x) = x^{s-1}$. This family belongs to Case 2. Let $L_n = (\theta_0 \log n/12) \vee 10$ and $B_n = [n^{-1/[6(s-1)]} \vee L_n^{-1}] \wedge 0.1$. Construct $\widetilde{\delta}_n$ as:

$$\delta_{ni}(x_i) = \begin{cases} 1 & \text{if} \quad (x_i > L_n) \text{ or } (B_n \le x_i \le L_n \text{ and } W_{ni}(x_i) \le 0), \\ 0 & \text{if} \quad (x_i < B_n) \text{ or } (B_n \le x_i \le L_n \text{ and } W_{ni}(x_i) > 0). \end{cases} \tag{4.4}$$

Then applying Theorem 3.2, we have the following result.

Result 4.2. *If X_i has density $f(x_i|\theta_i)$ given in (4.3) and $\int_0^\infty \theta^2 dG_i(\theta) < \infty$ for all $i = 1, \cdots, k$, then $\widetilde{\delta}_n$, as constructed in (4.4), has a rate of convergence of order $O(n^{-1})$.*

Example 4.3 (*A population having the density with infinite many discontinuities*). Consider the exponential populations having density

$$f(x_i|\theta_i) = c(\theta_i)e^{-x_i/\theta_i}\sum_{l=0}^{\infty}(l+1)I_{[l<x_i\le l+1]}, \qquad x_i > 0,\ \theta_i > 0,\ i = 1, \cdots, k. \tag{4.5}$$

Here $h(x) = \sum_{l=0}^{\infty}(l+1)I_{[l<x\le l+1]}$. This family belongs to Case 1. Take $B_n = 1/L_n$, $L_n = (\theta_0 \log n/4) \vee 10$ and construct $\widetilde{\delta}_n$ as

$$\delta_{ni}(x_i) = \begin{cases} 1 & \text{if} \quad (x_i > L_n) \text{ or } (B_n \le x_i \le L_n \text{ and } W_{ni}(x_i) \le 0), \\ 0 & \text{if} \quad (x_i < B_n) \text{ or } (B_n \le x_i \le L_n \text{ and } W_{ni}(x_i) > 0). \end{cases} \tag{4.6}$$

Then applying Theorem 3.2, we have the following result.

Result 4.3. *If X_i has density $f(x_i|\theta_i)$ given in (4.5) and $\int_0^\infty \theta^2 dG_i(\theta) < \infty$ for all $i = 1, \cdots, k$, then $\widetilde{\delta}_n$ as constructed in (4.6), has a rate of convergence of order $O(n^{-1})$.*

Remark. Gupta and Liang (1999) considered the same selection problem for the gamma population (4.3). In that paper, an empirical Bayes rule was constructed as

$$\delta^*_{ni}(x_i) = \begin{cases} 1 & \text{if } W_{ni}(x_i) \le 0, \\ 0 & \text{if } W_{ni}(x_i) > 0. \end{cases}$$

The convergence rate of $\widetilde{\delta}^*_n = (\delta^*_{n1}, \cdots, \delta^*_{nk})$ is affected by the tail probability of the underlying distributions. In our paper, we split the interval $(0, \infty)$ into three parts $(0, B_n)$, $[B_n, L_n]$ and (L_n, ∞) by localizing the Bayes test. Then we construct the empirical Bayes rule as (4.4). So the influence of the tail probability of the underlying distributions is controlled and a rate of $O(n^{-1})$ is obtained under quite weak conditions as shown in Result 4.2.

5. Proof of Theorem 3.2.

The main idea of the proof is to use a classic result about the non-uniform estimation of the difference between the normal distribution and the distribution of the sum of i.i.d. random variables. We shall prove it in the following two subsections according to whether all δ_{Gi} are non-degenerate or not.

5.1. All δ_{Gi} are non-degenerate. We shall prove Theorem 3.2 assuming that all δ_{Gi} are non-degenerate in this subsection. Then there exists a point $c_i \in (0, \infty)$ such that $\phi_i(c_i) = \theta_0$, $\phi_i(x) > \theta_0$ for $x > c_i$ and $\phi_i(x) < \theta_0$ for $x < c_i$. Since we consider the asymptotic behavior of $\widetilde{\delta}_n$, we assume $c_i \in (B_n, L_n)$ for all n without loss of generality.

Lemma 5.1. *For each $i = 1, \cdots, k$, $w_i'(c_i) < 0$ and further there is a neighborhood of c_i, denoted by $N(c_i, \epsilon_i)$, such that $N(c_i, \epsilon_i) \subset (B_1, L_1)$ and*

$$A_i = \min_{x \in N(c_i, \epsilon_i)} |w_i'(x)| > 0. \tag{5.1}$$

Denote $c_{i1} = c_i - \epsilon_i$, $c_{i2} = c_i + \epsilon_i$. Then for all $x \in [B_n, c_{i1}] \cup [c_{i2}, L_n]$, there exists an $M_i > 0$ such that

$$|w_i(x)| \geq M_i e^{-L_n/\theta_0}. \tag{5.2}$$

Proof. For $x > 0$, the derivative of $w_i(x)$ exists and can be expressed as

$$w_i'(x) = -\theta_0 \int_0^\infty e^{-x/\theta} c(\theta) dG_i(\theta) + \int_0^\infty \theta e^{-x/\theta} c(\theta) dG_i(\theta).$$

From Jensen's inequality, we see that for $x > 0$

$$\frac{\int_0^\infty \theta e^{-x/\theta} c(\theta) dG_i(\theta)}{\int_0^\infty e^{-x/\theta} c(\theta) dG_i(\theta)} < \frac{\int_0^\infty \theta^2 e^{-x/\theta} c(\theta) dG_i(\theta)}{\int_0^\infty \theta e^{-x/\theta} c(\theta) dG_i(\theta)}.$$

Plugging c_i for x in the above inequality, we have

$$\frac{\int_0^\infty \theta e^{-c_i/\theta} c(\theta) dG_i(\theta)}{\int_0^\infty e^{-c_i/\theta} c(\theta) dG_i(\theta)} < \theta_0.$$

This implies that $w_i'(c_i) < 0$.

Note that $w_i'(x)$ is continuous in $(0, \infty)$. So an $N(c_i, \epsilon_i)$ can be found such that $N(c_i, \epsilon_i) \subset (B_1, L_1)$ and

$$A_i = \min_{x \in N(c_i, \epsilon_i)} |w_i'(x)| > 0.$$

Then (5.1) is proved. On the other hand, rewrite $w_i(x)$ as $w_i(x) = \alpha_i(x)[\theta_0 - \phi_i(x)]$. Noting that $\phi_i(x)$ is strictly increasing in x and $\phi_i(c_i) = \theta_0$, then for $x \in [B_n, c_{i1}] \cup [c_{i2}, L_n]$,

$$|\theta_0 - \phi_i(x)| \geq (\theta_0 - \phi_i(c_{i1})) \wedge (\phi_i(c_{i2}) - \theta_0).$$

For $x \leq L_n$,

$$\alpha_i(x) \geq \int_{\theta_0}^\infty \theta c(\theta) e^{-x/\theta} dG_i(\theta) \geq e^{-L_n/\theta_0} \int_{\theta_0}^\infty \theta c(\theta) dG_i(\theta).$$

Thus

$$|w_i(x)| \geq [(\theta_0 - \phi_i(c_{i1})) \wedge (\phi_i(c_{i2}) - \theta_0)]e^{-L_n/\theta_0} \int_{\theta_0}^{\infty} \theta c(\theta) dG_i(\theta).$$

This completes the proof of Lemma 5.1.

Next lemma deals with the bounds of the moments of $W_{ni}(x)$.

In Case 1 and Case 3, $\min_{0<x<\infty} h(x) > 0$. Let $S_n \equiv 1/[\min_{0<x<\infty} h(x)]$. In Case 2 and Case 4, Let $S_n = H_n \vee [1/\min_{1\leq x<\infty} h(x)]$. Then $h(x) \geq S_n^{-1}$ for $x > B_n$ in all four cases. Recall $L_n = \theta_0 \log n/4$ in Case 1 ans Case 3 and $H_n e^{L_n/\theta_0} = n^{1/4}$ in Case 2 and Case 4. Then we have $S_n e^{L_n/\theta_0} \sim n^{1/4}$.

In Case 3 and Case 4, we know $\int_1^\infty \theta^4 c(\theta) dG_i(\theta) < \infty$ and let $C_i = \int_1^\infty \theta^4 c(\theta) dG_i(\theta)$.

Without loss of generality, we assume $h(x) \geq x$ for $x > 1$ in Case 1 and Case 2.

Lemma 5.2. *Let $\sigma_i^2(x) = E[(V_{ij}(x) - w_i(x))^2]$ and $\gamma_i(x) = E[|V_{ij}(x) - w_i(x)|^3]$. Then for $x \in [B_n, L_n]$,*

$$\sigma_i^2(x) \leq \begin{cases} [2S_n(\theta_0+1)+1]^2 & \text{for Case 1 and Case 2 ,} \\ S_n[(\theta_0+1)^2\alpha_i(x) + 2(\theta_0+1)C_i] & \text{for Case 3 and Case 4,} \end{cases} \tag{5.3}$$

and

$$\gamma_i(x) \leq \begin{cases} 4[2S_n(\theta_0+1)+1]^3 + 4|w_i(x)|^3 & \text{for Case 1 and Case 2 ,} \\ 16S_n^2[(\theta_0^3+6)\alpha_i(x) + 6C_i] + 4|w_i(x)|^3 & \text{for Case 3 and Case 4 .} \end{cases} \tag{5.4}$$

For $x \in [c_{i1}, c_{i2}]$, there exist a constants $C_{i\gamma} > 0$ such that

$$\gamma_i \leq C_{i\gamma}. \tag{5.5}$$

For $x \in [B_n, c_{i1}] \cup [c_{i2}, L_n]$ and large n ,

$$n^{3/8}|w_i(x)|/|\sigma_i(x)| \geq 1. \tag{5.6}$$

Proof. Consider $x \in [B_n, L_n]$. Note that $h(x) \geq S_n^{-1}$. In Case 1 and Case 2, if $x \geq 1$, $h(x) \geq x$, and

$$|V_{ij}(x)| \leq I_{[X_j \geq x]}\theta_0/h(X_j) + I_{[X_j\geq x]}(X_j - x)/h(X_j) \leq \theta_0 S_n + 1.$$

If $B_n \leq x < 1$, it can be shown that $|V_{ij}(x)| \leq 2S_n(\theta_0+1)+1$. Thus

$$\sigma_i^2(x) \leq E[|V_{ij}(x)|^2] \leq [2S_n(\theta_0+1)+1]^2.$$

For $\gamma_i(x)$, using $|a+b|^3 \le 4|a|^3 + 4|b|^3$, we have

$$\gamma_i(x) \le 4E[|V_{ij}(x)|^3] + 4|w_i(x)|^3 \le 4[2S_n(\theta_0+1)+1]^3 + 4|w_i(x)|^3.$$

Then (5.3) and (5.4) are proved for Case 1 and Case 2. In Case 3 and Case 4, a simple calculation shows that

$$\sigma_i^2(x) \le S_n[\theta_0^2\alpha_i(x) + 2\theta_0\psi_i(x) + 2\int_0^\infty \theta^3 c(\theta)e^{-x/\theta}dG_i(\theta)].$$

By breaking the interval $(0,\infty)$ into $(0,1)$ and $[1,\infty)$, we have $\psi_i(x) \le C_i + \alpha_i(x)$ and $\int_0^\infty \theta^3 c(\theta)e^{-x/\theta}dG_i(\theta) \le C_i + \alpha_i(x)$. Thus

$$\sigma_i^2(x) \le S_n[(\theta_0+1)^2\alpha_i(x) + 2(\theta_0+1)C_i].$$

Similarly,

$$\gamma_i(x) \le 16S_n^2[(\theta_0^3+6)\alpha_i(x) + 6C_i] + 4|w_i(x)|^3.$$

Now consider $x \in [c_{i1}, c_{i2}]$. It is easy to see that

$$\gamma_i(x) \le \begin{cases} 4[\frac{1+2(\theta_0+1)}{[\min_{x>c_{i1}} h(x)]}]^3 + 4\max_{b_{i1}\le x\le c_{i2}} |w_i(x)|^3 \equiv C_{i\gamma} & \text{in Case 1 and Case 2} \\ 16\frac{(\theta_0^3+6)\alpha_i(c_{i1})+6C_i}{[\min_{x\ge c_{i1}} h(x)]^2} + 4\max_{c_{i1}\le x\le c_{i2}} |w_i(x)|^3 \equiv C_{i\gamma} & \text{in Case 3 and Case 4} \end{cases}$$

Then (5.5) holds. Next we prove (5.6). From (5.2), $|w_i(x)| \ge M_i e^{-L_n/\theta_0}$ for $x \in [B_n, c_{i1}] \cup [c_{i2}, L_n]$. In Case 1 and Case 2,

$$|\frac{w_i(x)}{\sigma_i(x)}| \ge \frac{M_i e^{-L_n/\theta_0}}{2S_n(\theta_0+1)+1} \sim S_n^{-1}e^{-L_n/\theta_0} \sim n^{-1/4}.$$

In Case 3 and Case 4

$$|\frac{w_i(x)}{\sigma_i(x)}| \ge \frac{|\theta_0 - \phi_i(x)|}{S_n^{1/2}[(\theta_0+1)^2/\alpha_i(x) + 2(\theta_0+1)C_i/[\alpha_i(x)]^2]^{1/2}}. \tag{5.7}$$

It is easy to see that $|\theta_0 - \phi_i(x)| \ge \min\{|\theta_0 - \phi_i(c_{i1})|, |\theta_0 - \phi_i(c_{i2})|\}$. We know from the proof of Lemma 5.1 that $\alpha_i(x) \ge e^{-L_n/\theta_0}\int_{\theta_0}^\infty \theta c(\theta)dG_i(\theta)$. Then

$$S_n^{1/2}[(\theta_0+1)^2/\alpha_i(x) + 2(\theta_0+1)C_i/[\alpha_i(x)]^2]^{1/2} \sim S_n^{1/2}e^{L_n/\theta}.$$

Thus $|w_i(x)/\sigma_i(x)| = O(S_n^{-1/2}e^{-L_n/\theta_0}) = O(S_n^{1/2}n^{-1/4})$. This completes the proof of Lemma 5.2.

Note that $V_{ij}(x)$ are i.i.d random variables for fixed x. For large n, the central limit theorem tells us that $\sum_{j=1}^n[V_{ij}(x) - w_i(x)]/[\sigma_i(x)\sqrt{n}]$ is close to $N(0,1)$ in distribution. Furthermore, we have the following non-uniform estimation of the difference between the normal distribution and the distribution of the sum of i.i.d random variables. This result can be found in Petrov (1995, P168).

Fact. *Let $X_1, X_2, \cdots, X_n$ be i.i.d random variables, $EX_1 = 0$, $EX_1^2 = \sigma^2 > 0$, $E|X_1|^3 < \infty$. Then for all x*

$$|F_n(x) - \Phi(x)| \le A\frac{\rho}{\sqrt{n}(1+|x|)^3}. \tag{5.8}$$

Here $\Phi(x)$ is the c.d.f. of $N(0,1)$, $F_n(x)$ and ρ are given by

$$F_n(x) = P(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n} X_j \le x), \qquad \rho = \frac{E|X_1|^3}{\sigma^3}.$$

Now, we are ready to prove our main result.

Proof of Theorem 3.2. Rewrite $P(W_{ni}(x) < 0)$ as

$$P(\frac{1}{\sqrt{n\sigma_i^2(x)}}\sum_{j=1}^{n}[V_{ij}(x) - w_i(x)] \le -\frac{\sqrt{n}w_i(x)}{\sigma_i(x)}).$$

Then applying (5.8), we have

$$P(W_{ni}(x) < 0) \le \Phi(-\frac{\sqrt{n}|w_i(x)|}{\sigma_i(x)}) + \frac{A\gamma_i(x)}{\sqrt{n}(\sigma_i(x) + \sqrt{n}|w_i(x)|)^3}.$$

Similarly,

$$P(W_{ni}(x) > 0) \le 1 - \Phi(\frac{\sqrt{n}|w_i(x)|}{\sigma_i(x)}) + \frac{A\gamma_i(x)}{\sqrt{n}(\sigma_i(x) + \sqrt{n}|w_i(x)|)^3}.$$

Plugging above two inequalities in (3.2), we obtain

$$\begin{aligned} r_{ni} &\le \int_{B_n}^{c_i}[\Phi(-\frac{\sqrt{n}w_i(x)}{\sigma_i(x)}) + \frac{A\gamma_i(x)}{\sqrt{n}(\sigma_i(x) + \sqrt{n}|w_i(x)|)^3}]w_i(x)h(x)dx \\ &+ \int_{c_i}^{L_n}[1 - \Phi(\frac{\sqrt{n}w_i(x)}{\sigma_i(x)}) + \frac{A\gamma_i(x)}{\sqrt{n}(\sigma_i(x) + \sqrt{n}|w_i(x)|)^3}]|w_i(x)|h(x)dx \\ &\equiv I + II. \end{aligned}$$

From (5.3), (5.4), (5.5) and (5.6), we see that $w_i(x)$, $\sigma_i^2(x)$ and $\gamma_i(x)$ have different behavior for different x. So we decompose I into four parts.

$$\begin{aligned} I &= \int_{B_n}^{c_{i1}} \Phi(-\frac{\sqrt{n}w_i(x)}{\sigma_i(x)})w_i(x)h(x)dx + \int_{c_{i1}}^{c_i} \Phi(-\frac{\sqrt{n}w_i(x)}{\sigma_i(x)})w_i(x)h(x)dx \\ &+ \int_{B_n}^{c_{i1}} \frac{A\gamma_i(x)w_i(x)h(x)}{\sqrt{n}(\sigma_i(x) + \sqrt{n}|w_i(x)|)^3}dx + \int_{c_{i1}}^{c_i} \frac{A\gamma_i(x)w_i(x)h(x)}{\sqrt{n}(\sigma_i(x) + \sqrt{n}|w_i(x)|)^3}dx \\ &\equiv I_1 + I_2 + I_3 + I_4. \end{aligned} \tag{5.9}$$

Consider I_1 first. According to (5.6), as n is large, $w_i(x)/\sigma_i(x) \geq n^{-3/8}$ for $x \in [B_n, c_{i1}]$, It follows that $\sqrt{n}w_i(x)/\sigma_i(x) \geq n^{1/8}$. Then applying it to I_1, we have

$$I_1 \leq \Phi(-n^{1/8}) \int_{B_n}^{c_{i1}} w_i(x)h(x)dx = o(n^{-1}). \tag{5.10}$$

For I_2, letting $\bar{h}_i$ and $\bar{\sigma}_i$ be the maximum values of $h(x)$ and $\sigma_i(x)$ on $[c_{i1}, c_{i2}]$. Thus

$$I_2 \leq \bar{h}_i \int_{c_{i1}}^{c_i} \Phi(-\frac{\sqrt{n}w_i(x)}{\bar{\sigma}_i})w_i(x)dx.$$

Using (5.1) and letting $y = \sqrt{n}w_i(x)/\bar{\sigma}_i$,

$$\int_{c_{i1}}^{c_i} \Phi(-\sqrt{n}w_i(x)/\bar{\sigma}_i)w_i(x)dx \leq \frac{\bar{\sigma}_i^2}{A_i n}\int_0^\infty \Phi(-y)ydy = \frac{\bar{\sigma}_i^2}{4A_i n}.$$

Then

$$\limsup_{n\to\infty} nI_2 \leq \frac{\bar{h}_i\bar{\sigma}_i^2}{4A_i}. \tag{5.11}$$

Next we consider I_3. From (5.2), $|w_i(x)| \geq M_i e^{-L_n/\theta_0}$ for $x \in [B_n, c_{i1}]$. In Case 1 and Case 2, applying (5.4), we have

$$\begin{aligned} I_3 &\leq \int_{B_n}^{c_{i1}} \frac{4[2S_n(\theta_0+1)+1]^3 + 4|w_i(x)|^3}{\sqrt{n}(\sigma_i(x)+\sqrt{n}|w_i(x)|)^3} w_i(x)h(x)dx \\ &\leq \frac{4[2S_n(\theta_0+1)+1]^3}{n^2M_i^3e^{-3L_n/\theta_0}}\int_{B_n}^{c_{i1}} w_i(x)h(x)dx + \frac{4}{n^2}\int_{B_n}^{c_{i1}} w_i(x)h(x)dx \\ &= o(n^{-1}). \end{aligned} \tag{5.12}$$

In Case 3 and Case 4, using (5.4) again,

$$\begin{aligned} I_3 &\leq \int_{B_n}^{c_{i1}} \frac{16S_n^2[(\theta_0^3+6)\alpha_i(x)+6C_i] + 4|w_i(x)|^3}{\sqrt{n}(\sigma_i(x)+\sqrt{n}|w_i(x)|)^3} w_i(x)h(x)dx \\ &\leq \frac{16S_n^2(\theta_0^3+6)}{n^2M_i^2e^{-2L_n/\theta_0}}\int_{B_n}^{c_{i1}} \alpha_i(x)h(x)dx + \frac{96S_n^2C_i}{n^2M_i^3e^{-3L_n/\theta_0}}\int_{B_n}^{c_{i1}} w_i(x)h(x)dx + \\ &\quad \frac{4}{n^2}\int_{B_n}^{c_{i1}} w_i(x)h(x)dx \\ &= o(n^{-1}). \end{aligned} \tag{5.13}$$

For $x \in [c_i, c_{i1}]$, $\gamma_i(x) \leq C_{ir}$. Let $\underline{\sigma}_i$ be the minimum value of $\sigma_i(x)$ on $[c_{i1}, c_{i2}]$. It is easy to see that $\sigma_i(x) > 0$ for each $x \in [c_{i1}, c_{i2}] \subset [B_1, L_1]$. Noting that $\sigma_i(x)$ is a continuous function of x, then $\underline{\sigma}_i > 0$. It follows that

$$I_4 \leq AC_{i\gamma}\bar{h}_i \int_{c_{i1}}^{c_i} \frac{w_i(x)dx}{\sqrt{n}(\underline{\sigma}_i+\sqrt{n}|w_i(x)|)^3} \leq \frac{AC_{i\gamma}\bar{h}_i}{n^{3/2}A_i}\int_0^{\sqrt{n}w(c_i)} \frac{y}{(\underline{\sigma}_i+y)^3}dy = o(\frac{1}{n}). \tag{5.14}$$

Combining (5.9) to (5.14), we have $\limsup_{n\to\infty} nI \leq \frac{\bar{h}_i\bar{\sigma}_i^2}{4A_i}$. Note that I is independent of ϵ_i, $\bar{h}_i \to h(c_i)$, $\bar{\sigma}_i^2 \to Var([V_{i1}(c_i)])$ and $A_i \to |w_i'(c_i)|$ as $\epsilon_i \to 0$. Then

$$\limsup_{n\to\infty} nI \leq \frac{h(c_i)Var([V_{i1}(c_i)])}{4|w_i'(c_i)|} = \frac{m_i}{2}.$$

Similarly $\limsup_{n\to\infty} nII \le m_i/2$. Therefore

$$\limsup_{n\to\infty} nr_{ni} \le m_i.$$

Note that

$$\begin{aligned} r_{ni} \ \ge \ & \int_{B_n}^{c_i}[\Phi(-\frac{\sqrt{n}w_i(x)}{\sigma_i(x)}) - \frac{A\gamma_i(x)}{\sqrt{n}(\sigma_i(x)+\sqrt{n}|w_i(x)|)^3}]w_i(x)h(x)dx \\ & + \int_{c_i}^{L_n}[1-\Phi(\frac{\sqrt{n}w_i(x)}{\sigma_i(x)}) - \frac{A\gamma_i(x)}{\sqrt{n}(\sigma_i(x)+\sqrt{n}|w_i(x)|)^3}]|w_i(x)|h(x)dx. \end{aligned}$$

We have proved that $\int_{B_n}^{L_n} \frac{A\gamma_i(x)w_i(x)h(x)}{\sqrt{n}(\sigma_i(x)+\sqrt{n}|w_i(x)|)^3}dx = o(n^{-1})$, $\int_{B_n}^{c_{i1}} \Phi(-\frac{\sqrt{n}w_i(x)}{\sigma_i(x)})w_i(x)h(x)dx = o(n^{-1})$ and $\int_{c_{i2}}^{L_n}[1-\Phi(\frac{\sqrt{n}w_i(x)}{\sigma_i(x)})]w_i(x)h(x)dx = o(n^{-1})$. Then

$$\liminf_{n\to\infty} nr_{ni} \ge \liminf_{n\to\infty} n[\int_{c_{i1}}^{c_i}[\Phi(-\frac{\sqrt{n}w_i(x)}{\sigma_i(x)})w_i(x)h(x)dx + \int_{c_i}^{c_{i2}}[1-\Phi(\frac{\sqrt{n}w_i(x)}{\sigma_i(x)})w_i(x)h(x)dx].$$

Using the same idea applied to I_2, it is easy to prove the left-hand-side of above inequality is not smaller than m_i. So the proof of (3.3) and (3.4) is complete.

5.2. Some components of $\tilde{\delta}_G$ are degenerate. We shall prove Theorem 3.2 assuming that some of components of $\tilde{\delta}_G$ are degenerate.

For simplicity, we assume that only δ_{G1} is degenerate without loss of generality. We need to show $r_{n1} = o(n^{-1})$. If $P(\theta_1 = \theta_0) = 1$, $w_1(x) = 0$ and $r_{n1} = 0$. Assume $P(\theta_1 = \theta_0) < 1$ in the following. From the proof in Subsection 5.1, we only need to prove that (5.2)-(5.4) and (5.6) hold for all $x \in [B_n, L_n]$. Notice that (5.3) and (5.4) do not depend on the assumption of non-degeneracy of δ_{G1}. Then we only need to show (5.2) and (5.6). If G_1 is degenerate, then (5.2) and (5.6) are obvious. Next we assume that G_1 is non-degenerate and $\lim_{x\downarrow 0}\phi_1(x) \ge \theta_0$ or $\lim_{x\uparrow\infty}\phi_1(x) \le \theta_0$. Denote $\phi_1(0+) = \lim_{x\downarrow 0}\phi_1(x)$ and $\phi_1(\infty-) = \lim_{x\uparrow\infty}\phi_1(x)$. We shall show (5.2) first.

If $\phi_1(0+) \ge \theta_0$ and $G_1(\theta_0-) = 0$, for $x \in [B_n, L_n]$, $|w_1(x)| \ge e^{-L_n/\theta_0}\int_{\theta_0}^{\infty}(\theta-\theta_0)\theta c(\theta)dG_1(\theta)$ and $\alpha_1(x) \ge e^{-L_n/\theta_0}\int_{\theta_0}^{\infty}\theta c(\theta)dG_1(\theta)$. Then (5.2) holds.

If $\phi_1(0+) \ge \theta_0$ and $G_1(\theta_0-) > 0$, then $\alpha_1(0) < \infty$ and $\psi_1(0) < \infty$. The reason is in the following. Since $G_1(\theta_0-) > 0$, we can find $\epsilon > 0$ such that $G_1(\theta_0 - \epsilon) > 0$. From (3.3), we know $\theta c(\theta)$ is bounded on $[\theta_0 - \epsilon, \infty)$. Then $\int_{\theta-\epsilon}^{\infty}\theta c(\theta)dG_1(\theta) < \infty$. Therefore, if $\alpha_1(0) = \infty$, then $\int_0^{\theta_0-\epsilon}\theta c(\theta)e^{-x/\theta}dG_1(\theta) \to \infty$ as $x \to 0$. And since

$$\phi_1(x) \le \frac{(\theta_0-\epsilon)\int_0^{\theta_0-\epsilon}\theta c(\theta)e^{-x/\theta}dG_1(\theta) + \int_{\theta_0-\epsilon}^{\infty}\theta^2 c(\theta)e^{-x/\theta}dG_1(\theta)}{\int_0^{\theta_0-\epsilon}\theta c(\theta)e^{-x/\theta}dG_1(\theta) + \int_{\theta_0-\epsilon}^{\infty}\theta c(\theta)e^{-x/\theta}dG_1(\theta)},$$

we have $\phi_1(0+) \leq \theta_0 - \epsilon$. This contradicts $\phi_1(0+) \geq \theta_0$. Thus $\alpha_1(0) < \infty$ and $\psi_1(0) < \infty$. Then we have $w_1'(0+) < 0$. Let $c_{11} > 0$ such that $-w_1'(x) > d_1 > 0$ for $x \in (0, c_{11})$. As n is large, $B_n < c_{11}$. Then for $x \in [B_n, c_{11}]$, $x > 1/L_n$ and $|w(x)| = |w(x) - w(0)| \geq x|w_1'(x^*)| \geq d_1/L_n$, where $x^* \in (0, x)$. It is easy to see that (5.2) is true for $x \in [c_{11}, L_n]$. Then (5.2) holds for all $x \in [B_n, L_n]$ in this case.

If $\phi_1(\infty-) \leq \theta_0$, we must have $G_1(\theta_0) = 1$. The reason is in the following. If $G_1(\theta_0) < 1$, let $\epsilon > 0$ satisfy $G(\theta_0 + \epsilon) \neq 1$. Then

$$\phi_1(x) > \frac{\int_0^{\theta_0+\epsilon} \theta^2 c(\theta) e^{-x[\theta^{-1}-(\theta_0+\epsilon)^{-1}]} dG_1(\theta) + (\theta_0+\epsilon)\int_{\theta_0+\epsilon}^{\infty} \theta c(\theta) e^{-x[\theta^{-1}-(\theta_0+\epsilon)^{-1}]} dG_1(\theta)}{\int_0^{\theta_0+\epsilon} \theta c(\theta) e^{-x[\theta^{-1}-(\theta_0+\epsilon)^{-1}]} dG_1(\theta) + \int_{\theta_0+\epsilon}^{\infty} \theta c(\theta) e^{-x[\theta^{-1}-(\theta_0+\epsilon)^{-1}]} dG_1(\theta)}.$$

Therefore $\lim_{x\uparrow\infty} \phi_1(x) > \theta_0 + \epsilon$. This contradicts $\phi_1(\infty-) \leq \theta_0$. Since $G_1(\theta_0) = 1$, $w_1(x) = \int_0^{\theta_0} \theta(\theta_0 - \theta) c(\theta) e^{-x/\theta} dG_1(\theta) \geq e^{-L_n/\theta} \int_0^{\theta_0} \theta(\theta_0 - \theta) c(\theta) dG_1(\theta)$ for $x \in [B_n, L_n]$.

Next we shall prove (5.6). It is obvious for Case 1 and Case 2 from (5.2) and (5.3). We only prove (5.6) for Case 3 and Case 4.

If $\phi_1(0+) \geq \theta_0$ and $G_1(\theta_0-) = 0$, $\alpha_1(x) \leq \int_{\theta_0}^{\infty} \theta c(\theta) dG_1(\theta) < \infty$ for all $x > 0$. Then (5.6) follows from (5.2) and (5.3).

If $\phi_1(0+) \geq \theta_0$ and $G_1(\theta_0-) > 0$, $\alpha_1(0) < \infty$ from previous result. Then (5.6) follows from (5.2) and (5.3).

If $\phi_1(\infty-) \leq \theta_0$, $G_1(\theta_0) = 1$ and $\phi_1(x) < \theta_0$ for all $x > 0$. We know $B_n < 1$ for large n. Then for $x \in [1, L_n]$, $\alpha_1(x)$ is bounded. Then (5.6) follows from (5.2) and (5.3). For $x \in [B_n, 1]$, $\alpha_1(x)$ and $\theta_0 - \phi_1(x)$ are bounded from below. Then (5.6) follows from (5.7).

Now the proof of Theorem 3.2 is complete.

6. Simulation Study.

A simulation study was carried out to investigate the performance of the proposed empirical Bayes selection procedure (2.3) for small to moderate values of n.

We consider $\exp(\theta)$ family in this simulation study, i.e.,

$$\pi_i \sim f(x_i|\theta_i) = \frac{1}{\theta_i} e^{-\frac{x_i}{\theta_i}}.$$

This is also Example 4.1. We consider the case in which $k = 3$. That is, there are 3 exponential populations (treatments) π_1, π_2 and π_3. We shall make a selection using the procedure (4.2). Assume that the prior distributions G_1, G_2 and G_3 are i.i.d. having a density

$$g_i(\theta) = \frac{1}{(s-2)!\theta^s} e^{-\frac{1}{\theta}}, \qquad \theta > 0,$$

where $s > 4$ so that $\int \theta^3 g_i(\theta)d\theta < \infty$ and the requirement of Corollary 3.3 is satisfied. It is easy to compute that

$$\phi_i(x) = \frac{x+1}{s-2}, \qquad x > 0$$

and the marginal density of X

$$f_i(x) = \frac{s-1}{(x+1)^s}, \qquad x > 0$$

for $i = 1, 2, 3$. As $(s-2)\theta_0 > 1$, δ_G is nondegenerate and $c_i = (s-2)\theta_0 - 1 \in (0, \infty)$. Then it can be computed that

$$m \equiv \sum_{i=1}^{3} m_i = \frac{3(s-1)(s-2)\theta_0^2}{2(s-3)}.$$

According Corollary 3.3, the regret of the empirical Bayes selection (4.2) is close to $\frac{m}{n}$ as n is large. Note that

$$\frac{w_i(x)}{f_i(x)} = \frac{[\theta_0(s-2) - (x+1)](x+1)}{(s-1)(s-2)}.$$

Following (3.1) and (3.2), the regret of the empirical Bayes selection rule can be expressed as

$$r_n = E\Big\{\sum_{i=1}^{3}[I_{[W_{ni}(X_i)\le 0, X_i<(s-2)\theta_0-1, B_n<X_i<L_n]}\frac{[\theta_0(s-2)-(X_i+1)](X_i+1)}{(s-1)(s-2)}]\Big\}$$

where the expectation is taken over the probability generated by $(\widetilde{X}_1, \widetilde{X}_2, \cdots, \widetilde{X}_n, \widetilde{X})$. Denote D as

$$D = \sum_{i=1}^{3}[I_{[W_{ni}(X_i)\le 0, X_i<(s-2)\theta_0-1, B_n<X_i<L_n]}\frac{[\theta_0(s-2)-(X_i+1)](X_i+1)}{(s-1)(s-2)}]$$

The scheme of the simulation is described as follows:

(1) For each i and n, generate independent random variables as follows:

$$\begin{cases} \text{for} \quad j = 1, \ldots, n, \\ \text{(a) first generate } \Theta_{ij} \text{ from an inverse gamma distribution with density } g_i(\theta), \\ \text{(b) then generate } X_{ij} \text{ from an exponential } exp(\theta_{ij}) \text{ distribution.} \end{cases}$$

Likewise, generate Θ_i from an inverse gamma distribution with density $g_i(\theta)$, X_i from $exp\ (\theta_i)$.

Table 1

Performance of the selection rule

when $s = 6$ and $\theta_0 = 0.6$

n	D_n	$SE(D_n)$	$\frac{m}{n}$
10	0.13577026	0.003197456	0.16000000
20	0.06920616	0.003120035	0.08000000
30	0.04476671	0.003669832	0.05333333
40	0.03566289	0.002416336	0.04000000
50	0.02786863	0.003738685	0.03200000
60	0.02122455	0.003135153	0.02666667
70	0.01999013	0.003188093	0.02285714
80	0.01898793	0.002517613	0.02000000
90	0.01945126	0.002356803	0.01777777
100	0.01715247	0.002609822	0.01600000
150	0.01190080	0.002147165	0.01066667
200	0.01011384	0.002052432	0.00800000

Likewise, generate Θ_i from an inverse gamma distribution with density $g_i(\theta)$, X_i from *exp* (θ_i).

(2) Based on the past observations $(\widetilde{X}_j, j = 1, \ldots, n)$ and the present observations $\widetilde{X} = (X_1, \ldots, X_k)$, we compute D.

(3) Repeat steps (1) and (2) 10000 times. The average of the D's from the 10000 repetitions, denoted by D_n, is used as an estimator of the difference r_n. The standard error, denoted by $SE(D_n)$, is also computed.

Table 1 gives the results of this simulation study on the performance of the proposed empirical selection procedure. For a specific n, three columns give D_n, $SE(D_n)$ and $\frac{m}{n}$. In this case we choose $s = 6$ and $\theta_0 = 0.4$.

Figure 1 gives the plots of (n, D_n) and $(n, \frac{m}{n})$. The dotted line gives the values of $\frac{m}{n}$; the solid line denotes the values of D_n.

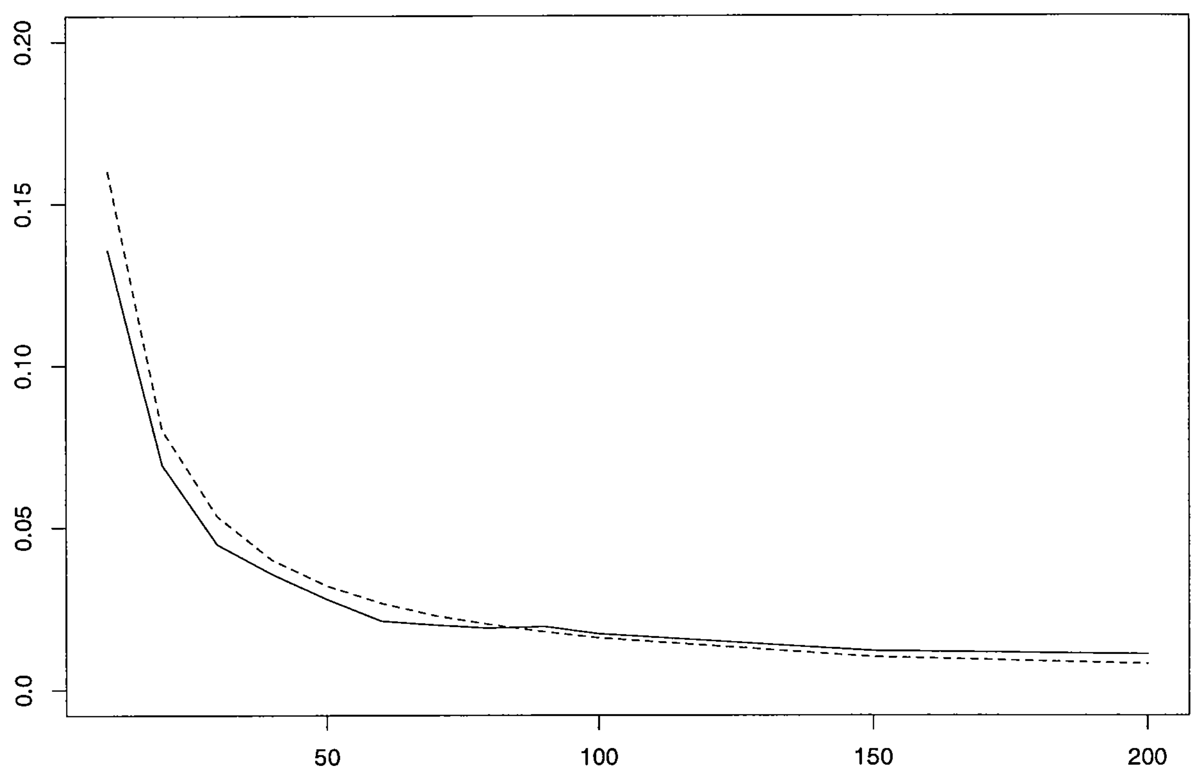

Figure 1: Graph for Table 1.

References

[1] Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *J. Amer. Statist. Asso.* **50**, 1096-1121.

[2] Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Inference.* **8**, 87-101.

[3] Gupta, S. S. and Liang, T. (1999). Selecting good exponential populations compared with a control : A nonparametric Bayes approach. *Sankhya, Ser. B.* **61**, 289-304.

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

[5] Gupta, S. S. and Liese, F. (2000). Asymptotic distribution of the conditional regret risk for selecting good exponential populations. *Kybernetika.* **36**, 571-588.

[6] 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.

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

[8] Liang, T. (2000a). On empirical Bayes tests in a positive exponential family. *J. Statist. Plann. Inference* **83**, 169-181.

[9] Liang, T. (2000b). On an empirical Bayes test for a normal mean. *Ann. Statist.* **28**, 648-655.

[10] Paulson, E. (1952). An Optimum Solution to the k-Sample Slippage Problem for the Normal Distribution. *Ann. Math. Statist.* **23**, 610-616.

[11] Petrov, V. V. (1995). *Limit Theorems of Probability Theory.* Clarendon Press · Oxford.

[12] Pensky, M. (1998). Empirical Bayes estimation based on wavelets. *Sankhya Ser. A* **60**, 214-231.

[13] Pensky, M. and Singh, R. S. (1999). Empirical Bayes estimation of reliability characteristics for an exponential family. *Canad. J. Statist.* **27**, 127-136.

[14] Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third. Berkeley Symp. Math. Statist, Probab.* **1**, 157-163, Univ. California Press, Berkeley.

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

[16] Singh, R. S. and Wei, L. S. (1992). Empirical Bayes with rates and best rates of convergence in $u(x)C(\theta)\exp(-x/\theta)$-family: estimation case. *Ann. Inst. Statist. Math.* **44**, 435-449.

[17] van Houwelingen, H. and Stijnen, T. (1993). Monotone empirical Bayes estimators based on more informative samples. *J. Amer. Statist. Assoc.* **88**, 1438–1443.

SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
August 2001

3. REPORT TYPE AND DATES COVERED
Technical Report August 2001

4. TITLE AND SUBTITLE
On Empirical Bayes Procedures for Selecting Good Populations in Positive Exponential Family

5. FUNDING NUMBERS
DAAH 04-95-1-0165
DAAD 19-00-1-0502

6. AUTHOR(S)
Shanti S. Gupta and Jianjun Li

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

8. PERFORMING ORGANIZATION REPORT NUMBER
Technical Report #01-10

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The problem of selecting good ones compared with a control from $k(\geq 2)$ positive exponential family populations is considered in this paper. A nonparametric empirical Bayes approach is used to construct the selection procedures. It has been shown that the risks of the empirical Bayes procedures converge to the (minimum) Bayes risk with a rate of $O(1/n)$, where n is the number of accumulated past observations at hand. Simulations were carried out to study the performance of the procedures for small to moderate values of n. The results of this study are provided in the paper.

14. SUBJECT TERMS
Empirical Bayes; selection procedures; positive exponential family; regret Bayes risk; rate of convergence

15. NUMBER IF PAGES
20

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500 Enclosure 1 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18 298-102