EMPIRICAL BAYES TESTS IN SOME CONTINUOUS EXPONENTIAL FAMILY

by

Shanti S. Gupta and Jianjun Li
Purdue University Purdue University

Technical Report #99-21

Department of Statistics
Purdue University
West Lafayette, IN USA

September 1999

*

Empirical Bayes Tests
In Some Continuous Exponential Family

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

Empirical Bayes tests for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ in some continuous exponential family with density $c(\theta)exp\{\theta x\}h(x)$, $\infty \leq \alpha < x < \beta \leq \infty$, where, $c(\theta)$ satisfies certain condition, are investigated under the linear loss. Using a mild assumption that $\int_\Omega |\theta| dG(\theta) < \infty$, we construct the empirical Bayes test and show that its regret goes to zero with a convergence rate of order $o(n^{-1+\frac{1}{\log\log n}})$. The applications of our result to $N(\theta, 1)$ and the general exponential family distributions are given as corollaries.

AMS 1991 Subject classification: 62c12.

Keywords: empirical Bayes; exponential family; rate of convergence; regret.

---

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

§ 1 Introduction

Since the empirical Bayes approach was introduced by Robbins (1956, 1964), it has become a powerful tool when situations involve sequences of similar and independent statistical problems. In these situations, the independent component problems in the sequence are formulated as the Bayes statistical decision problems involving a common, unknown prior distribution over the parameter space. The accumulated observations and present observations are used to construct the decision rule at each stage and so a sequence of empirical Bayes decision rule is formed as observations continue. If the empirical Bayes rule is constructed such that the risk for the nth decision problem converges to the Bayes risk which would have been obtained if the prior distribution were known and the Bayes rule with respect to this prior distribution were used, the empirical Bayes rule is said to be asymptotically optimal (a.o.). Among a.o. empirical Bayes rules, the convergence rate is used to measure their performances.

The detailed structure we consider in this paper is as follows: An observation X is obtained from a distribution with density

$$f(x|\theta) = c(\theta)exp\{\theta x\}h(x), \qquad -\infty \leq \alpha < x < \beta \leq +\infty, \tag{1.1}$$

where $h(x)$ is continuous, positive for $x \in (\alpha, \beta)$, θ is the value of the parameter, which is regarded as the value of a random variable Θ having unknown distribution G on the parameter space $\Omega \subset \{\theta : c(\theta) > 0\}$. The observation X may be thought of as the value of a sufficient statistic based on several i.i.d observations.

The hypotheses $H_0 : \theta \leq \theta_0$ is to be tested against $H_1 : \theta > \theta_0$, where $\theta_0 \in \Omega$, under the linear loss

$$l(\theta, a) = a(\theta_0 - \theta)I_{[\theta \leq \theta_0]} + (1 - a)(\theta - \theta_0)I_{[\theta > \theta_0]}, \tag{1.2}$$

where $a = 0$ or 1 according to taking action in favor of H_0 or H_1 respectively.

A decision rule $\delta(x)$ is the probability of accepting H_1 when $X = x$ is observed. That is $\delta(x) = P\{$ accepting $H_1|X = x\}$. $R(G, \delta)$ is used to denote the Bayes risk of the test δ when G is the prior distribution. Then $R(G, \delta) = \int_\Omega \int_\alpha^\beta l(\theta, \delta(x))f(x|\theta)dxdG(\theta)$.

We consider only prior G such that $E|\Theta| < \infty$ to ensure that the risk is always finite. To find Bayes test, rewrite $R(G, \delta)$ as

$$\begin{aligned} R(G, \delta) &= C_G + \int_\Omega \int_\alpha^\beta \delta(x)(\theta_0 - \theta)c(\theta)e^{\theta x}h(x)dxdG(\theta) \qquad (1.3) \\ &= C_G + \int_\alpha^\beta \delta(x)[\int_\Omega (\theta_0 - \theta)c(\theta)e^{\theta x}dG(\theta)]h(x)dx \\ &= C_G + \int_\alpha^\beta \delta(x)[\theta_0\alpha_G(x) - \psi_G(x)]h(x)dx \\ &= C_G + \int_\alpha^\beta \delta(x)W(x)h(x)dx \\ &= C_G + \int_\alpha^\beta \delta(x)[\theta_0 - \phi_G(x)]\alpha_G(x)h(x)dx, \end{aligned}$$

where

$$\begin{cases} C_G = \int_\Omega (\theta - \theta_0) I_{[\theta > \theta_0]} dG(\theta), \\ \alpha_G(x) = \int_\Omega c(\theta) e^{\theta x} dG(\theta), \\ \psi_G(x) = \int_\Omega \theta c(\theta) e^{\theta x} dG(\theta), \\ w(x) = \alpha_G(x) - \psi_G(x), \\ \phi_G(x) = \psi_G(x)/\alpha_G(x). \end{cases}$$

Here, $\alpha_G(x)$ and $\psi_G(x)$ are well defined since it is easy to verify that $\alpha_G(x) < \infty$ and $\psi_G(x) < \infty$ for $x \in (\alpha, \beta)$. From (1.3), a Bayes test δ_G is clearly given by

$$\delta_G(x) = \begin{cases} 1 & \text{if} \quad w(x) \leq 0 \\ 0 & \text{if} \quad w(x) > 0 \end{cases}$$

or, equivalently

$$\delta_G(x) = \begin{cases} 1 & \text{if} \quad \phi_G(x) \geq \theta_0 \\ 0 & \text{if} \quad \phi_G(x) < \theta_0. \end{cases}$$

and the Bayes risk is

$$R(G, \delta_G) = C_G + \int_\alpha^\beta \delta_G(x) w(x) h(x) dx.$$

The trivial cases are expected to be excluded, so we assume that

$$\lim_{x \downarrow \alpha} \phi_G(x) < \theta_0 < \lim_{x \uparrow \beta} \phi_G(x). \tag{1.4}$$

A consequence of (1.4) is that $\phi_G(x)$ is strictly increasing and there exists the unique point b_0 (critical value) such that $\phi_G(b_0) = \theta_0$, $\phi_G(x) < \theta_0$ for $x < b_0$, and $\phi_G(x) > \theta_0$ for $x > b_0$. Therefore, the Bayes test δ_G can be simply represented as

$$\delta_G(x) = \begin{cases} 1 & \text{if} \quad x \geq b_0, \\ 0 & \text{if} \quad x < b_0. \end{cases}$$

In the empirical Bayes context, a sequence of problems having the above structure occurs but $G(\theta)$ is not known. The only thing we can obtain are the past n independent observations $X_1, X_2, \cdots, X_n$ and the present observation X. We need to make the decision at present based on $(X, \widetilde{X}_n)$ where $\widetilde{X}_n = (X_1, X_2, \cdots, X_n)$. Let $\delta_n(X, \widetilde{X}_n)$ be a empirical Bayes test, i.e., $\delta_n(X, \widetilde{X}_n)$ is the probability of accepting H_1 when X and $\widetilde{X}_n$ are observed. The risk related to δ_n is $E[R(G, \delta | \widetilde{X}_n)]$, where $R(G, \delta | \widetilde{X}_n)$ is the risk of δ_n given $\widetilde{X}_n$. $R(G, \delta_n) - R(G, \delta_G)$ is often called the regret of δ_n, and is used as a measure of performance of the empirical Bayes test of δ_n.

In the literature, Johns and Van Ryzin (1972) constructed empirical Bayes test for the above problem and studied the rate of convergence for its associated regret. Van Houwelingen (1976) improved Johns and Van Ryzin's result and showed that the empirical Bayes test there has a convergence rate of order $O(n^{-2r/(2r+3)} \log^2 n)$, where $r \geq 1$

is an integer, associated with the moment requirement that $E[|\Theta|^{r+1}] < \infty$. Note that if r is small, this convergence rate is still slow.

Karunamuni and Yang (1995) considered this problem under a new setting. They assumed that there exist two known constants C_1, C_2, $\alpha < C_1 < C_2 < \beta$ such that

$$C_1 \leq b_0 \leq C_2. \tag{1.5}$$

Then they constructed the empirical Bayes test such that it achieves a rate of convergence of order $O(n^{-2r/(2r+3)})$ if $E[|\Theta|^{r+1}] < \infty$ for some $r \geq 1$. With (1.5) only, Liang (1999) constructed an empirical Bayes test for the positive exponential family and proved that the regret of the empirical Bayes test there converges to zero with a rate of order $O(n^{-s/(s+1)})$, where s is an arbitrarily prespecified positive integer. Later, Gupta and Li (1999) generalized Liang's result to the general continuous exponential family. They showed that, under (1.5) and without the restriction $E[|\Theta|^{r+1}] < \infty$, the empirical Bayes test can be constructed such that its regret has a rate of convergence of order $o(n^{-1+\epsilon})$, where $\epsilon > 0$ is any prespecified number.

It should be pointed out that the improved rate is because of assumption (1.5). To apply the result in Gupta and Li (1999), one needs to know more about the prior distribution G. For example, even if we know that θ is bounded, C_1 and C_2 required in (1.5) may not be obtainable. So to find a better empirical Bayes test in the asymptotic sense without assumption (1.5) is still an interesting problem.

In this paper we deal with those exponential distributions in which, for some constants $\eta > 0$, $B_0 \geq 0$ and $B_1 > 0$,

$$c(\theta) \leq \begin{cases} e^{-|\theta|^{1+\eta}} & \text{if} \quad \Omega[|\theta| > B_0], \\ B_1 & \text{if} \quad \Omega[|\theta| \leq B_0]. \end{cases} \tag{1.6}$$

For these distributions, we are able to construct the empirical Bayes test such that the convergence rate of its regret is of order $o(n^{-1+\frac{1}{\log\log n}})$, which is clearly faster than $o(n^{-1+\epsilon})$. This result not only gets rid of the restriction (1.5), but also improves the previous convergence rates.

One application of our result is for $X \sim N(\theta, 1)$ (see Corollary 3.6). For $N(\theta, 1)$, the only necessary condition we need is that $E[|\Theta|] < \infty$. This condition guarantees that the risk of Bayes decision rules is finite and is quite generally used in Bayes analysis. So our result can be applied in most situations. As for the convergence rate, our result is quite strong, since $o(n^{-1+\frac{1}{\log\log n}})$ is much closer to $O(n^{-1})$, which has been thought of as the best possible convergence rate for the empirical Bayes rules (see Singh (1979)).

For the general exponential family, we know that $c(\theta)$ is a continuous function on the natural space $\{\theta : c(\theta) > 0\}$. So if Ω is an inner closed subset of $\{\theta : c(\theta) > 0\}$, then (1.6) is satisfied and our result holds (see Corollary 3.7). If we know $\theta \in [\theta_{01}, \theta_{02}] \subset \{\theta : c(\theta) > 0\}$, where θ_{01} and θ_{02} are known or unknown constants, then (1.6) is met. So our result can be applied. But in both cases of Corollary 3.6 and Corollary 3.7, we may not be able to find C_1 and C_2 required in (1.5).

Karunamuni (1996) has claimed that he found the optimal convergence rates of the empirical Bayes tests for $N(\theta, 1)$ and $exp(\theta)$. From our result, one can see that the result for $N(\theta, 1)$ in Karunamuni (1996) is not correct, since we can have a much faster rate than his optimal rate. In fact, his result for $exp(\theta)$ is also not correct according to our investigation.

The paper is organized as follows: §1 gives the introduction; §2 constructs the empirical Bayes test δ_n; §3 proves that the empirical Bayes test has a convergence rate of order $o(n^{-1+\frac{1}{\log\log n}})$. Two specific results are given as corollaries; §4 proves the lemmas stated in §3; §5 gives a few comments about possible improvements of our result. Last, we attach an appendix about the existence of the kernel functions used in §2.

§ 2 Construction of Empirical Bayes Tests

We use the kernel method to construct the empirical Bayes test. The kernel method has been used by many authors over the years. The method here is a little different from the previous ones. We use a sequence of kernel functions instead of the single one.

For each $i = 0, 1$ and $m = 1, 2, \cdots$, let $K_{im}(y)$ be a Borel-measurable function such that $K_{im}(y)$ vanishes outside the internal $[0, 1]$, and for $K_{0m}(y)$

$$\int_0^1 y^j K_{0m}(y)dy = \begin{cases} 1 & \text{if} \quad j = 0, \\ 0 & \text{if} \quad j = 1, 2, \cdots, m-1, \end{cases} \tag{2.1}$$

and for $K_{1m}(y)$

$$\int_0^1 y^j K_{1m}(y)dy = \begin{cases} 0 & \text{if} \quad j = 0, 2, 3, \cdots, m, \\ 1 & \text{if} \quad j = 1. \end{cases} \tag{2.2}$$

For fixed m, the construction of bounded $K_{im}(y)$ is well-known (see Stijnen (1985)). When m varies, the bound of $K_{im}(y)$ is a function of m. We denote this function as $B(m)$. The construction of empirical Bayes test and the convergence rate of its regret depend clearly on $B(m)$. The best result we obtained about $B(m)$ is that $K_{im}(y)$ can be constructed such that $|K_{im}(y)| \leq (m+1)^2 8^{m+1}$ (see Appendix).

For any $x \in (\alpha, \beta)$, define

$$\alpha_n(x) = \frac{1}{nu} \sum_{j=1}^{n} K_{0v}(\frac{X_j - x}{u})/h(X_j), \tag{2.3}$$

and

$$\psi_n(x) = \frac{1}{nu^2} \sum_{j=1}^{n} K_{1v}(\frac{X_j - x}{u})/h(X_j), \tag{2.4}$$

where $u = u(n) = n^{-1/8\log\log n}$ and $v = v_n = 16[\frac{\log\log n}{2}] - 1$, here $[x]$ means the integer part of x. Let $W_n(x) = \theta_0 \alpha_n(x) - \psi_n(x)$. $W_n(x)$ is an asymptotically unbiased estimator of $w(x)$(see Lemma 3.2).

Next we take some steps to localize $h(x)$. Let $L_n = \log\log n$, $h(\alpha+) = \lim_{x\downarrow\alpha} h(x)$ and $h(\beta-) = \lim_{x\uparrow\beta} h(x)$. Choose any $\gamma \in (\alpha, \beta)$. Define

$$C_{1n} = \begin{cases} \max\{\alpha < x < \gamma : h(x) \le \frac{1}{\log n}\} \vee (\alpha + \frac{1}{n}) \vee (-L_n) & \text{if } h(\alpha+) = 0, \\ (\alpha + \frac{1}{n}) \vee (-L_n) & \text{if } 0 < h(\alpha+) < \infty, \\ \max\{\alpha < x < \gamma : h(x) \ge L_n\} \vee (\alpha + \frac{1}{n}) \vee (-L_n) & \text{if } h(\alpha+) = \infty, \end{cases} \tag{2.5}$$

and

$$C_{2n} = \begin{cases} (\min\{\gamma < x < \beta : h(x) \le \frac{1}{\log n}\} - u) \wedge (\beta - \frac{1}{n}) \wedge L_n & \text{if } h(\beta-) = 0, \\ (\beta - \frac{1}{n}) \wedge L_n & \text{if } 0 < h(\beta-) < \infty, \\ \min\{\gamma < x < \beta : h(x) \ge L_n\} \wedge (\beta - \frac{1}{n}) \wedge L_n & \text{if } h(\beta-) = \infty. \end{cases} \tag{2.6}$$

Since $L_n \to \infty$ as $n \to \infty$, $C_{1n} \to \alpha$, and $C_{2n} \to \beta$. And so b_0 will fall in $[C_{1n}, C_{2n}]$ for $n > N_0$, there N_0 is some integer.

The direct consequence of (2.5) and (2.6) is

$$C_{2n} - C_{1n} \le 2L_n. \tag{2.7}$$

Other consequences are

$$\min_{C_{1n} \le x \le C_{2n}+u} h(x) \ge \frac{1}{\log n} \tag{2.8}$$

and

$$\max_{C_{1n} \le x \le C_{2n}} h(x) \le L_n, \tag{2.9}$$

for large value of n. Without loss of generality, assume that (2.8) and (2.9) are true for $n > N_0$.

Now we propose an empirical Bayes test $\delta_n(x, \widetilde{X}_n)$ by

$$\delta_n = \begin{cases} 1 & \text{if} \quad (x > C_{2n}) \text{ or } (C_{1n} \le x \le C_{2n} \text{ and } W_n(x) \le 0), \\ 0 & \text{if} \quad (x < C_{1n}) \text{ or } (C_{1n} \le x \le C_{2n} \text{ and } W_n(x) > 0). \end{cases}$$

The conditional Bayes risk of the empirical Bayes test δ_n is:

$$R(G, \delta_n | \widetilde{X}_n) = C_G + \int_\alpha^\beta \delta_n(x) w(x) h(x) dx.$$

In the following discussion of this paper, we consider the case $n > N_0$ without further mention. Note that $w(x) \ge 0$ if $x \in [C_{1n}, b_0]$; $w(x) \le 0$ if $x \in [b_0, C_{2n}]$. Then the conditional regret can be expressed as,

$$\begin{aligned} R(G, \delta_n | \widetilde{X}_n) - R(G, \delta) &= \int_\alpha^\beta (\delta_n - \delta) w(x) h(x) dx \\ &= \int_{C_{1n}}^{b_0} I_{[W_n(x) \le 0]} w(x) h(x) dx \\ &\quad + \int_{b_0}^{C_{2n}} I_{[W_n(x) > 0]} |w(x)| h(x) dx \end{aligned}$$

and the regret becomes

$$R(G,\delta_n) - R(G,\delta) = \int_{C_{1n}}^{b_0} P(W_n(x) \le 0) w(x) h(x) dx + \int_{b_0}^{C_{2n}} P(W_n(x) > 0) |w(x)| h(x) dx. \tag{2.10}$$

§3 Asymptotic Optimality of $\delta_n(x)$

In this section, we shall investigate the asymptotic optimality of $\delta_n(x)$. Clearly, the convergence rate of $R(G,\delta_n) - R(G,\delta_G)$ depends on the properties of $w(x)$ and $W_n(x)$, and depends on the inequality used to estimate $P(W_n(x) \le 0)$ and $P(W_n(x) > 0)$. The more information about $w(x)$ and $W_n(x)$ is obtained, the more accurate rate we can get. So we will dig out a few properties of $w(x)$ and $W_n(x)$. That is a few lemmas, whose proofs are left to §4. Then we state a well-known inequality about the non-uniform estimation of the distance between the normal distribution and the distribution of a sum of i.i.d random variables. After that, our main result and two corollaries are given. The final part in this section is the proof of the main result.

Note that $\alpha_G(x) < \infty$ for $x \in (\alpha, \beta)$. So $\alpha_G(x)$ is infinitely differentiable and so is $\psi_G(x)$ since $\psi_G(x) = \alpha'_G(x)$. Then we know that $w(x)$ is infinitely differentiable and the analytic properties of $w(x)$ can be investigated. The following property of $w(x)$ plays an important role in the estimation of $\int_{C_{1n}}^{b_0} P(W_n(x) \le 0) w(x) h(x) dx$ and $\int_{b_0}^{C_{2n}} P(W_n(x) > 0) w(x) h(x) dx$.

Lemma 3.1 *There exist $A_1 > 0$, $A_2 > 0$, $B > 0$, c_1, c_2 such that $\alpha < c_1 < b_0 < c_2 < \beta$, and for all $x \in [c_1, c_2]$,*

$$-w'(x) \ge A_1^{-1}, \tag{3.1}$$

for all $x \in [C_{1n}, c_1] \cup [c_2, C_{2n}]$ and $n > N_1$, and where $N_1 (> N_0)$ is some integer,

$$|w(x)| \ge A_2 (\log n)^{-B}. \tag{3.2}$$

Now we consider $W_n(x)$. Note that

$$W_n(x) = \theta_0 \alpha_n(x) - \psi_n(x) = \frac{1}{n} \sum_{j=1}^{n} V_n(X_j, x),$$

where

$$V_n(X_j, x) = \frac{\theta_0}{u} \times \frac{K_{0v}(\frac{X_j - x}{u})}{h(X_j)} - \frac{1}{u^2} \times \frac{K_{1v}(\frac{X_j - x}{u})}{h(X_j)}.$$

For fixed n and x, $V_n(X_j, x)$ are i.i.d. random variables. So the classical results about the sum of i.i.d random variables can be used. But a little moment information about $W_n(x)$ is necessarily to investigate first. We have two lemmas for that purpose. Both are direct results of computations.

The first lemma also shows that $W_n(x)$ is an asymptotically unbiased and consistent estimator of $w(x)$. That is the basis of the construction of the Bayes rule $\delta_n(x)$.

Lemma 3.2 *Let $w_n(x) = E[V_n(X_j, x)]$. Then we have*

$$w_n(x) = w(x) + u^v d_n(x), \tag{3.3}$$

where $d_n(x)$ is some function of x such that $|u^v d_n(x)| \le \frac{1}{2\sqrt{n}}$ for all $x \in [C_{1n}, C_{2n}]$ and $n > N_2$, and where $N_2(> N_1)$ is some integer.

Let $Z_{jn}(x) = V_n(X_j, x) - w_n(x)$. Then $EZ_{jn}(x) = 0$. And

Lemma 3.3 *There is an integer $N_3(> N_2)$ such that for all $n > N_3$ and all $x \in [C_{1n}, C_{2n}]$,*

$$\sigma_n^2(x) \equiv E[Z_{jn}(x)]^2 \le \frac{1}{u^6} \quad \text{and} \quad \frac{E|Z_{jn}(x)|^3}{\sigma_n^2(x)} \le \frac{1}{u^3}. \tag{3.4}$$

We see that $W_n(x)$ is not an unbiased estimator of $w(x)$. $w(x) > 0$ cannot guarantee that $w_n(x) > 0$ and $w(x) < 0$ cannot guarantee that $w_n(x) < 0$. (To see this, look at the expression of $d_n(x)$ in (4.10).) This sign difference between $w_n(x)$ and $w(x)$ causes some trouble for us later. So we introduce the following lemma.

Lemma 3.4 *For all n> N_2 and $x \in [C_{1n}, C_{2n}]$,*

$$w(x) > \frac{1}{\sqrt{n}} \Longrightarrow w_n(x) \ge \frac{1}{2} w(x), \tag{3.5}$$

and

$$w(x) < -\frac{1}{\sqrt{n}} \Longrightarrow w_n(x) \le \frac{1}{2} w(x). \tag{3.6}$$

Note that $Z_{jn}(x)$ are i.i.d random variables for fixed n and x. For large n, the central limit theorem tells us that $\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} Z_{jn}$ 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 (1975, pp125) or Michel (1981). Michel proved $A < 30.54$ in his paper.

Result *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*

$$|F_n(x) - \Psi(x)| \leq A \frac{\rho}{\sqrt{n}(1+|x|)^3}. \tag{3.7}$$

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

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

Now, we are ready to introduce our main result:

Theorem 3.5 *Suppose $\int_\Omega |\theta| dG(\theta) < \infty$ and (1.4) is assumed. If (1.6) is true, then we have, as $n \to \infty$,*

$$\lim_{n\to\infty} (n^{1-\frac{1}{\log\log n}})[R(G, \delta_n) - R(G, \delta)] = 0. \tag{3.8}$$

Before proving it, we state two corollaries of Theorem 3.5. For $N(\theta, 1)$, (1.6) holds naturally. So we have the following lemma.

Corollary 3.6 *For $X \sim N(\theta, 1)$, if $\int_\Omega |\theta| dG(\theta) < \infty$ and (1.4) is assumed, then δ_n has a rate of convergence of order $o(n^{-1+\frac{1}{\log\log n}})$.*

For the general exponential family distributions, we know $c(\theta)$ is a continuous function. So we have the following lemma.

Corollary 3.7 *Suppose X has density (1.1) and (1.4) is true. If Ω is an inner closed subset of the natural parameter space $\{\theta : c(\theta) > 0\}$, then δ_n has a rate of convergence of order $o(n^{-1+\frac{1}{\log\log n}})$.*

For $N(\theta, 1)$, the only necessary condition we need is that $\int_\Omega |\theta| dG(\theta) < \infty$. This condition guarantees that the risk of Bayes decision rules is finite and is quite generally used in Bayes analysis. So our result can be applied in most situations. As for the convergence rate, our result is quite strong, since $o(n^{-1+\frac{1}{\log\log n}})$ is much closer to $O(n^{-1})$, which has been thought of as the best possible convergence rate for the empirical Bayes rules. Karunamuni (1996) claimed that the optimal minimax rate of empirical Bayes rule for $N(\theta, 1)$ is $O(n^{-2(r-1)/(2r+1)})$, $r > 1$. From our result, one can see that the regret of δ_n has the convergence rate of order $o(n^{-1+\frac{1}{\log\log n}})$ even in his minimax setting. So his minimax rate result for $N(\theta, 1)$ is not correct. For the general exponential family distributions, $c(\theta)$ is a continuous function. If we know $\theta \in [\theta_{01}, \theta_{02}] \subset \{\theta : c(\theta) > 0\}$, where θ_{01} and θ_{02} are known or unknown constants, then we can apply Corollary 3.7 and construct an empirical Bayes test such that its regret converges to zero with a rate of order $o(n^{-1+\frac{1}{\log\log n}})$.

Proof of Theorem 3.5. From (2.10),

$$\begin{aligned} R(G,\delta_n) - R(G,\delta) &= \int_{C_{1n}}^{b_0} P(W_n(x) \le 0) w(x) h(x) I_{[0<w(x)\le\frac{1}{\sqrt{n}}]} dx \\ &\quad + \int_{b_0}^{C_{2n}} P(W_n(x) > 0) |w(x)| h(x) I_{[-\frac{1}{\sqrt{n}}\le w(x)<0]} dx \\ &\quad + \int_{C_{1n}}^{b_0} P(W_n(x) \le 0) w(x) h(x) I_{[w(x)>\frac{1}{\sqrt{n}}]} dx \\ &\quad + \int_{b_0}^{C_{2n}} P(W_n(x) > 0) |w(x)| h(x) I_{[w(x)<-\frac{1}{\sqrt{n}}]} dx \\ &\equiv I + II + III + IV. \end{aligned}$$

We decompose $R(G,\delta_n) - R(G,\delta)$ into four parts and estimate each part separately. In case of $0 < w(x) \le \frac{1}{\sqrt{n}}$, $P(W_n(x) \le 0)$ converges quite slowly. We will use the property of $w(x)$ to estimate the bound of Part I. We use a similar method in Part II. In case of $w(x) > \frac{1}{\sqrt{n}}$, we depend mainly on the estimation of $P(W_n(x) \le 0)$ to obtain a bound of Part III. Similarly, we estimate the bound of Part IV.

Part I. Since $h(x) \le L_n$ for $x \in [C_{1n}, C_{2n}]$ and $0 < w(x) \le \frac{1}{\sqrt{n}}$, we have

$$I \le \frac{L_n}{\sqrt{n}} \int_{C_{1n}}^{b_0} I_{[0<w(x)\le\frac{1}{\sqrt{n}}]} dx.$$

From (3.1), $\min_{x\in[C_{1n},c_1]\cup[c_1,C_{2n}]} |w(x)| \ge A_2(\log n)^{-B}$. So there exists an integer $N_4(> N_3)$ such that as $n > N_4$ $\sqrt{n}\min_{x\in[C_{1n},c_1]\cup[c_1,C_{2n}]} |w(x)| \ge 1$. It follows that $\{C_{1n} \le x \le C_{2n} : |w(x)| \le \frac{1}{\sqrt{n}}\} \subset [c_1, c_2]$. Thus $\int_{C_{1n}}^{b_0} I_{[0<w(x)\le\frac{1}{\sqrt{n}}]} dx = \int_{c_1}^{b_0} I_{[0<w(x)\le\frac{1}{\sqrt{n}}]} dx$. Using (3.2),

$$\int_{c_1}^{b_0} I_{[0<w(x)\le\frac{1}{\sqrt{n}}]} dx \le -A_1 \int_{c_1}^{b_0} I_{[0<w(x)\le\frac{1}{\sqrt{n}}]} w'(x) dx \le A_1 \int_0^{w(c_1)} I_{[0<y\le\frac{1}{\sqrt{n}}]} dy \le \frac{A_1}{\sqrt{n}}.$$

Thus

$$I \le \frac{A_1 L_n}{n}. \tag{3.9}$$

Part II. Similar to Part I, we can have, as $n > N_4$,

$$II \le \frac{A_1 L_n}{n}. \tag{3.10}$$

Part III. We use (3.5), (3.7) and (3.4) to estimate $P(W_n(x) \le 0) I_{[w(x)>\frac{1}{\sqrt{n}}]}$ first.

$$P(W_n(x) \le 0) I_{[w(x)>\frac{1}{\sqrt{n}}]}$$

$$
\begin{aligned}
&= P\Big(\frac{1}{\sqrt{n\sigma_n^2(x)}}\sum_{j=1}^{n} Z_{jn} \le -\frac{\sqrt{n}w_n(x)}{\sigma_n(x)}\Big)I_{[w(x)>\frac{1}{\sqrt{n}}]}\\
&\le P\Big(\frac{1}{\sqrt{n\sigma_n^2(x)}}\sum_{j=1}^{n} Z_{jn} \le -\frac{\sqrt{n}w(x)}{2\sigma_n(x)}\Big)I_{[w(x)>\frac{1}{\sqrt{n}}]}\\
&\le \Phi\Big(-\frac{\sqrt{n}w(x)}{2\sigma_n(x)}\Big) + \frac{A\frac{E|Z_{jn}|^3}{\sigma_n^3(x)}}{\sqrt{n}(1+|-\frac{\sqrt{n}w(x)}{2\sigma_n(x)}|)^3}I_{[w(x)>\frac{1}{\sqrt{n}}]}\\
&\le \Phi\Big(-\frac{1}{2}\sqrt{nu^6}w(x)\Big) + \frac{2A}{nw(x)} \times \frac{E|Z_{jn}|^3}{\sigma_n^2(x)}I_{[w(x)>\frac{1}{\sqrt{n}}]}\\
&\le \Phi\Big(-\frac{1}{2}\sqrt{nu^6}w(x)\Big) + \frac{2A}{nu^3w(x)}I_{[w(x)>\frac{1}{\sqrt{n}}]}.
\end{aligned}
$$

One can see why we need Lemma 3.4 and apply it in the first equality above. If we applied (3.7) to $P(W_n(x) \le 0)$ without substituting $w_n(x)$ with $w(x)$, then we would have $\frac{A}{nu^3 w_n(x)}$ in the second part of the last inequality above. Since $w_n(x)$ is 0 for some $x \in \{x : w(x) > 0\}$, the bound of $P(W_n(x) \le 0)$ would be infinite. Another benefit of Lemma 3.4 is that we can handle $w(x)$ easier than $w_n(x)$ after the substitution. Applying the inequality above, we have

$$
\begin{aligned}
III &\le \int_{C_{1n}}^{b_0} P(W_n(x) \le 0)I_{[w(x)>\frac{1}{\sqrt{n}}]}w(x)h(x)dx\\
&\le \int_{C_{1n}}^{b_0} \Phi\Big(-\frac{1}{2}\sqrt{nu^6}w(x)\Big)w(x)h(x)dx + \int_{C_{1n}}^{b_0} \frac{2A}{nu^3w(x)}I_{[w(x)>\frac{1}{\sqrt{n}}]}w(x)h(x)dx\\
&= \int_{C_{1n}}^{c_1} \Phi\Big(-\frac{1}{2}\sqrt{nu^6}w(x)\Big)w(x)h(x)dx + \int_{c_1}^{b_0} \Phi\Big(-\frac{1}{2}\sqrt{nu^6}w(x)\Big)w(x)h(x)dx\\
&\quad + \int_{C_{1n}}^{b_0} \frac{2A}{nu^3}h(x)dx\\
&\equiv III_1 + III_2 + III_3.
\end{aligned}
$$

Here $\int_{C_{1n}}^{b_0} \Phi(-\frac{1}{2}\sqrt{nu^6}w(x))w(x)h(x)dx$ is decomposed into two parts. The reason is that $\Phi(-\frac{1}{2}\sqrt{nu^6}w(x))$ goes to zero very slowly if $w(x)$ is relatively small. So we have to separate it into $\int_{C_{1n}}^{c_1} \Phi(-\frac{1}{2}\sqrt{nu^6}w(x))w(x)h(x)dx$ and $\int_{c_1}^{b_0} \Phi(-\frac{1}{2}\sqrt{nu^6}w(x))w(x)h(x)dx$. If $x \in [C_{1n}, c_1]$, $w(x) \ge A_2(\log n)^{-B}$ from (3.2). Then

$$
\begin{aligned}
III_1 &\le \int_{C_{1n}}^{c_1} \Phi\Big(-\frac{A_2}{2}\sqrt{nu^6}(\log n)^{-B}\Big)w(x)h(x)dx\\
&= \Phi\Big(-\frac{A_2}{2}\sqrt{nu^6}(\log n)^{-B}\Big)\int_{C_{1n}}^{c_1} w(x)h(x)dx\\
&\le \Phi\Big(-\frac{A_2}{2}\sqrt{nu^6}(\log n)^{-B}\Big)\int_{\alpha}^{\beta} |w(x)|h(x)dx.
\end{aligned}
$$

Since $\sqrt{nu^6}(\log n)^{-B} = (nn^{-\frac{6}{8\log\log n}})^{\frac{1}{2}}(\log n)^{-B}$, there exists an integer $N_5(> N_4)$ such that for $n > N_5$, $\Phi(-\frac{A_2}{2}\sqrt{nu^6}(\log n)^{-B}) \le \frac{1}{n}$ and

$$V \le \frac{A_0}{n}, \tag{3.11}$$

where $A_0 = \int_\alpha^\beta |w(x)|h(x)dx \le \int_\alpha^\beta \alpha_G(x)h(x)dx + \int_\alpha^\beta |\psi_G(x)|h(x)dx \le 1 + E|\Theta| < \infty$. As for Part III$_2$, noting $h(x) \le L_n$ for $x \in [c_1, c_2]$ and making change of variable $y = \frac{1}{2}\sqrt{nu^6}w(x)$,

$$\begin{aligned} III_2 &\le -A_1 L_n \int_{c_1}^{b_0} \Phi(-\frac{1}{2}\sqrt{nu^6}w(x))w(x)w'(x)dx \\ &= \frac{4A_1L_n}{nu^6}\int_0^{\frac{1}{2}\sqrt{nu^6}w(c_1)} \Phi(-y)ydy \\ &\le \frac{4A_1L_n}{nu^6}\int_0^{\infty} \Phi(-y)ydy \\ &= \frac{2A_1L_n}{nu^6}. \end{aligned} \tag{3.12}$$

Consider III_3 now. Since $b_0 - C_{1n} \le 2L_n$ and $h(x) \le L_n$ for $x \in [C_{1n}, C_{2n}]$, we have

$$III_3 = \frac{2A}{nu^3}\int_{C_{1n}}^{b_0} h(x)dx \le \frac{4AL_n^2}{nu^3}. \tag{3.13}$$

Combining (3.11), (3.12) and (3.13), we obtain that when $n \ge N_5$,

$$III \le \frac{A_0}{n} + \frac{2A_1L_n}{nu^6} + \frac{4AL_n^2}{nu^3}.$$

Since $u^3L_n^2 = n^{-\frac{3}{8\log\log n}}(\log\log n)^2 \to 0$, we can find some integer $N_6(> N_5)$ such that, for $n > N_6$,

$$III \le \frac{4A_1L_n}{nu^6}. \tag{3.14}$$

Part IV. Similar to $P(W_n(x) \le 0)I_{[w(x)>\frac{1}{\sqrt{n}}]}$, we have

$$\begin{aligned} &P(W_n(x) > 0)I_{[w(x)<-\frac{1}{\sqrt{n}}]} \\ \le\ & 1 - \Phi(\frac{1}{2}\sqrt{nu^6}|w(x)|) + \frac{2A}{nu^3|w(x)|}I_{[w(x)<-\frac{1}{\sqrt{n}}]}. \end{aligned}$$

Then as in Part III,

$$\begin{aligned} IV &\le \int_{b_0}^{C_{2n}} P(W_n(x) > 0)I_{[w(x)<-\frac{1}{\sqrt{n}}]}|w(x)|h(x)dx \\ &\le \int_{c_2}^{C_{2n}} [1 - \Phi(\frac{1}{2}\sqrt{nu^6}|w(x)|)]|w(x)|h(x)dx \end{aligned}$$

$$
\begin{aligned}
&+\int_{b_0}^{c_2}[1-\Phi(\frac{1}{2}\sqrt{nu^6}|w(x)|)]|w(x)|h(x)dx+\int_{b_0}^{C_{2n}}\frac{2A}{nu^3}h(x)dx\\
\le\ & [1-\Phi(\frac{1}{2}\sqrt{nu^6}|w(x)|)]A_0+\frac{4A_1L_n}{nu^6}\int_0^{\frac{1}{2}\sqrt{nu^6}|w(c_2)|}[1-\Phi(y)]ydy+\frac{4AL_n^2}{nu^3}\\
\le\ & \frac{A_0}{n}+\frac{2A_1L_n}{nu^6}+\frac{4AL_n^2}{nu^3},
\end{aligned}
$$

where the last inequality holds for $n > N_5$. And as $n > N_6$,

$$
IV \le \frac{4A_1L_n}{nu^6}. \tag{3.15}
$$

From (3.9), (3.10), (3.14) and (3.15), we get

$$
R(G,\delta_n)-R(G,\delta) \le \frac{2A_1L_n}{n}+\frac{8A_1L_n}{nu^6}.
$$

As $n\to\infty$, $n^{1-\frac{1}{\log\log n}}\times\frac{L_n}{nu^6}=n^{1-\frac{1}{\log\log n}}\times\frac{\log\log n}{nn^{-\frac{6}{8\log\log n}}}\to 0$. Thus

$$
\lim_{n\to\infty} n^{1-\frac{1}{\log\log n}}[R(G,\delta_n)-R(G,\delta)]=0.
$$

The proof is completed.

§ 4 Proofs of Lemmas

Proof of Lemma 3.1 We know that $w(x)$ is infinite differentiable. And

$$
w'(x)=\theta_0\int_\Omega \theta c(\theta)e^{\theta x}dG(\theta)-\int_\Omega \theta^2 c(\theta)e^{\theta x}dG(\theta). \tag{4.1}
$$

If $\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)=0$, then $w'(b_0)=-\int_\Omega \theta^2 c(\theta)e^{\theta b_0}dG(\theta)<0$. If $\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)>0$, then

$$
\frac{\int_\Omega \theta^2 c(\theta)e^{\theta b_0}dG(\theta)}{\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)} > \frac{\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)}{\int_\Omega c(\theta)e^{\theta b_0}dG(\theta)}=\theta_0.
$$

Thus $w'(b_0)=\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)[\theta_0-\frac{\int_\Omega \theta^2 c(\theta)e^{\theta b_0}dG(\theta)}{\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)}]<0$. If $\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)<0$, then

$$
\frac{\int_\Omega \theta^2 c(\theta)e^{\theta b_0}dG(\theta)}{\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)} < \frac{\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)}{\int_\Omega c(\theta)e^{\theta b_0}dG(\theta)}=\theta_0.
$$

Thus $w'(b_0)=\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)[\theta_0-\frac{\int_\Omega \theta^2 c(\theta)e^{\theta b_0}dG(\theta)}{\int_\Omega \theta c(\theta)e^{\theta b_0}dG(\theta)}]<0$. Now we have proved that

$$
w'(b_0)<0. \tag{4.2}
$$

Next, we prove that there exist c_3 and c_4 such that $\alpha < c_3 < b_0 < c_4 < \beta$ and $w(x)$ is strictly monotone in (α, c_3) and (c_4, β). Taking the derivative of $w'(x)$, we have

$$w''(x) = \theta_0 \int_\Omega \theta^2 c(\theta) e^{\theta x} dG(\theta) - \int_\Omega \theta^3 c(\theta) e^{\theta x} dG(\theta).$$

And $w''(x)$ can be written as

$$w''(x) = \int_\Omega \theta^2 c(\theta) e^{\theta x} dG(\theta)[\theta_0 - \frac{\int_\Omega \theta^3 c(\theta) e^{\theta x} dG(\theta)}{\int_\Omega \theta^2 c(\theta) e^{\theta x} dG(\theta)}].$$

Note that, under (1.4),

$$[\frac{\int_\Omega \theta^3 c(\theta) e^{\theta x} dG(\theta)}{\int_\Omega \theta^2 c(\theta) e^{\theta x} dG(\theta)}]' = \frac{\int_\Omega \theta^4 c(\theta) e^{\theta x} dG(\theta) \int_\Omega \theta^2 c(\theta) e^{\theta x} dG(\theta) - [\int_\Omega \theta^3 c(\theta) e^{\theta x} dG(\theta)]^2}{[\int_\Omega \theta^2 c(\theta) e^{\theta x} dG(\theta)]^2} > 0.$$

Thus $w''(x)$ is strictly decreasing and has at most one sign change in (α, β).

If $w''(x)$ has no sign change, we have that $w'(x) > 0$ for all $x \in (\alpha, \beta)$ or $w'(x) < 0$ for all $x \in (\alpha, \beta)$. Then we can choose any point from (α, b_0) as c_3 and any point from (b_0, β) as c_4.

If $w''(x)$ has one sign change, let b_2 be the change point. Since $w''(x)$ is strictly decreasing, $w''(x) > 0$ for $x \in (\alpha, b_2)$ and $w''(x) < 0$ for $x \in (b_2, \beta)$. Then $w'(x)$ is strictly increasing in (α, b_2) and strictly decreasing in (b_2, β). So $w'(x)$ has at most one sign change in (α, b_2) and at most one sign change in (b_2, β).

Consider the case of (α, b_2). If $w'(x)$ has no sign change in (α, b_2), we can choose any point in $(\alpha, b_2 \wedge b_0)$ as c_3. If $w'(x)$ has one sign change in (α, b_2), let b_{11} be the change point. Then we can choose any point in $(\alpha, b_{11} \wedge b_0)$ as c_3. So $\alpha < c_3 < b_0$ and $w'(x)$ has no sign change in (α, c_3).

Consider the case of (b_2, β). If $w'(x)$ has no sign change in (b_2, β), we can choose any point in $(b_2 \vee b_0, \beta)$ as c_4. If $w'(x)$ has one sign change in (b_2, β), let b_{12} be the change point. Then we can choose any point in $(b_0 \vee b_{12}, \beta)$ as c_4. Thus $b_0 < c_4 < \beta$ and $w'(x)$ has no sign change in (c_4, β).

So we have proved that $w(x)$ is strictly monotone in $x \in (\alpha, c_3)$ and $x \in (c_4, \beta)$, where $\alpha < c_3 < b_0 < c_4 < \beta$. Without loss of generality, assume $C_{1n} < c_3 < c_4 < C_{2n}$ for all $n > N_0$. Then

$$\min_{x \in [C_{1n}, c_3] \cup [c_4, C_{2n}]} |w(x)| = \min\{w(C_{1n}), |w(C_{2n})|, w(c_3), |w(c_4)|\}. \tag{4.3}$$

From (4.1), we see that $w'(x)$ is continuous. Then we can find c_1 and c_2 such that $c_3 < c_1 < b_0 < c_2 < c_4$, and for any $x \in [c_1, c_2]$,

$$-w'(x) \geq \frac{1}{2}[-w'(b_0)] \equiv A_1^{-1}. \tag{4.4}$$

Note that $w(x) > 0$ for $x \in [c_3, b_0)$; $w(x) < 0$ for $x \in (b_0, c_4]$. Then

$$\zeta_0 \equiv [\min_{c_3 \leq x \leq c_1} w(x)] \wedge [\min_{c_2 \leq x \leq c_4} |w(x)|] > 0.$$

Obviously,

$$\min_{x\in[C_{1n},c_1]\cup[c_2,C_{2n}]} |w(x)| = \min\{w(C_{1n}), |w(C_{2n})|\} \wedge \zeta_0. \tag{4.5}$$

Let B be a positive constant such that $\int_{\Omega[|\theta|\le B]} dG(\theta) \neq 0$. From (2.5), we know $|C_{1n}| \le \log\log n$. Then

$$\begin{aligned} w(C_{1n}) &\ge [\theta_0 - \phi_G(c_1)] \int_\Omega c(\theta) e^{-|\theta||C_{1n}|} dG(\theta) \\ &\ge [\theta_0 - \phi_G(c_1)] \int_\Omega c(\theta) e^{-|\theta|\log\log n} dG(\theta) \\ &\ge [\theta_0 - \phi_G(c_1)] \int_{\Omega[|\theta|\le B]} c(\theta)(\log n)^{-|\theta|} dG(\theta) \\ &\ge [\theta_0 - \phi_G(c_1)](\log n)^{-B} \int_{\Omega[|\theta|\le B]} c(\theta) dG(\theta). \end{aligned}$$

Similarly, $|w(C_{2n})| \ge [\phi_G(c_2) - \theta_0](\log n)^{-B} \int_{\Omega[|\theta|\le B]} c(\theta)dG(\theta)$. Let $A_2 = \min\{[\theta_0 - \phi_G(c_1)], [\phi_G(c_2) - \theta_0]\} \times \int_{\Omega[|\theta|\le B]} c(\theta)dG(\theta) > 0$. Thus, there exists an integer $N_1(> N_0)$ such that for $n \ge N_1$, $A_2(\log n)^{-B} \le \zeta_0$ and

$$\min_{x\in[C_{1n},c_1]\cup[c_1,C_{2n}]} |w(x)| \ge A_2(\log n)^{-B}. \tag{4.6}$$

This completes the proof of Lemma 3.1.

Proof of Lemma 3.2 Using Taylor expansion, (2.1) and (2.2), a straight-forward computation shows that

$$\begin{aligned} & E[\frac{K_{0v}(\frac{X_j-x}{u})}{uh(X_j)}] \qquad\qquad (4.7)\\ = & \int_\Omega \int_\alpha^\beta \frac{K_{0v}(\frac{y-x}{u})}{uh(y)} c(\theta) e^{\theta y} h(y) dy dG(\theta) \\ = & \int_\Omega \int_0^1 K_{0v}(t) c(\theta) e^{\theta x} e^{\theta u t} dt dG(\theta) \\ = & \int_\Omega c(\theta) e^{\theta x} [\int_0^1 K_{0v}(t) e^{\theta u t} dt] dG(\theta) \\ = & \int_\Omega c(\theta) e^{\theta x} [1 + \int_0^1 \frac{K_{0v}(t) t^v u^v \theta^v e^{\theta u t^*}}{v!} dt] dG(\theta) \\ = & \int_\Omega c(\theta) e^{\theta x} dG(\theta) + u^v \int_\Omega \theta^v c(\theta) e^{\theta x} [\int_0^1 \frac{K_{0v}(t) t^v e^{\theta u t^*}}{v!} dt] dG(\theta), \end{aligned}$$

where $0 \le t^* = t^*(\theta, u, t, v) \le 1$. Also,

$$E[\frac{K_{1v}(\frac{X_j-x}{u})}{u^2 h(X_j)}] \tag{4.8}$$

$$
\begin{aligned}
&= \int_\Omega \int_\alpha^\beta \frac{K_{1v}(\frac{y-x}{u})}{u^2 h(y)} c(\theta) e^{\theta y} h(y) dy dG(\theta) \\
&= \frac{1}{u} \int_\Omega \int_0^1 K_{1v}(t) c(\theta) e^{\theta x} e^{\theta u t} dt dG(\theta) \\
&= \frac{1}{u} \int_\Omega c(\theta) e^{\theta x} [\int_0^1 K_{1v}(t) e^{\theta u t} dt] dG(\theta) \\
&= \frac{1}{u} \int_\Omega c(\theta) e^{\theta x} [u\theta + \int_0^1 \frac{K_1(t) t^{v+1} u^{v+1} \theta^{v+1} e^{\theta u t^{**}}}{(v+1)!} dt] dG(\theta) \\
&= \int_\Omega \theta c(\theta) e^{\theta x} dG(\theta) + u^v \int_\Omega \theta^{v+1} c(\theta) e^{\theta x} [\int_0^1 \frac{K_1(t) t^{v+1} e^{\theta u t^{**}}}{(v+1)!} dt] dG(\theta),
\end{aligned}
$$

where $0 \le t^{**} = t^{**}(\theta, u, t, v) \le 1$. From (4.7) and (4.8), we get that

$$
E[V_n(X_j, x)] = w(x) + u^v d_n(x), \tag{4.9}
$$

where

$$
\begin{aligned}
d_n(x) &= \theta_0 \int_\Omega \theta^v c(\theta) e^{\theta x} [\int_0^1 \frac{K_{0v}(t) t^v e^{\theta u t^*}}{v!} dt] dG(\theta) \\
&\quad - \int_\Omega \theta^{v+1} c(\theta) e^{\theta x} [\int_0^1 \frac{K_{1v}(t) t^{v+1} e^{\theta u t^{**}}}{(v+1)!} dt] dG(\theta).
\end{aligned} \tag{4.10}
$$

Recalling that $|K_{iv}(t)| \le (v+1)^2 8^{v+1}$ for $i = 0, 1$,

$$
\int_0^1 K_{0v}(t) t^v e^{\theta u t^*} dt \le \frac{(v+1)^2 8^{v+1}}{(v+1)} e^{|\theta|} \le (v+1) 8^{v+1} e^{|\theta|},
$$

and

$$
\int_0^1 K_{1v}(t) t^{v+1} e^{\theta u t^{**}} dt \le \frac{(v+1)^2 8^{v+1}}{v+2} e^{|\theta|} \le (v+1) 8^{v+1} e^{|\theta|}.
$$

It follows that

$$
\begin{aligned}
|d_n(x)| &\le |\theta_0| \int_\Omega \frac{(v+1)|\theta|^v 8^{v+1} c(\theta) e^{|\theta|(|x|+1)}}{v!} dG(\theta) \\
&\quad + \int_\Omega \frac{(v+1)|\theta|^{v+1} 8^{v+1} c(\theta) e^{|\theta|(|x|+1)}}{(v+1)!} dG(\theta).
\end{aligned}
$$

Since $\frac{|\theta|^v 8^v}{v!} \le e^{8|\theta|}$ and $\frac{|\theta|^{v+1} 8^{v+1}}{(v+1)!} \le e^{8|\theta|}$, we have

$$
|d_n(x)| \le (8|\theta_0| + 1)(v+1) \int_\Omega c(\theta) e^{|\theta|(|x|+9)} dG(\theta).
$$

For $x \in [C_{1n}, C_{2n}]$, $|x| \le L_n$. Now, without loss of generality, we assume $|x| + 9 \le 2L_n$. And note that (1.6) is assumed. Then, for $x \in [C_{1n}, C_{2n}]$,

$$
\int_\Omega c(\theta) e^{|\theta|(|x|+9)} dG(\theta)
$$

$$\begin{aligned}
&\leq \int_\Omega c(\theta)e^{2L_n|\theta|}dG(\theta) \\
&\leq \int_\Omega e^{-|\theta|^{1+\eta}}e^{2L_n|\theta|}dG(\theta) + B_1 e^{2B_0 L_n} \\
&\leq \int_{\Omega[|\theta|\leq(2L_n)^{1/\eta}]} e^{(2L_n)^{(1+\eta)/\eta}}dG(\theta) + \int_{\Omega[|\theta|>(2L_n)^{1/\eta}]} e^{|\theta|(2L_n-|\theta|^\eta)}dG(\theta) + B_1 e^{2L_n B_0} \\
&\leq e^{(2\log\log n)^{(\eta+1)/\eta}} + 1 + B_1(\log n)^{2B_0}.
\end{aligned}$$

Since $u = n^{-\frac{1}{8\log\log n}}$ and $v = 16[\frac{\log\log n}{2}] - 1$, we have $v + 1 \leq 8\log\log n$ and

$$u^v = n^{-\frac{16[\frac{\log\log n}{2}]-1}{8\log\log n}} \leq n^{-1} n^{\frac{17}{8\log\log n}}.$$

As $n \to \infty$, $\sqrt{n}u^v(v+1)e^{(2\log\log n)^{(\eta+1)/\eta}} \to 0$ and $\sqrt{n}u^v(v+1)(\log n)^{2B_0} \to 0$. Thus

$$\sqrt{n}u^v(v+1)\int_\Omega c(\theta)e^{|\theta|(|x|+9)}dG(\theta) \to 0.$$

So we can find an $N_2(> N_1)$ such that for $n > N_2$, $u^v|d_n(x)| \leq \frac{1}{2\sqrt{n}}$.

Proof of Lemma 3.3 Note that for $i = 0$ or 1, $|K_{iv}(t)| \leq (v+1)^2 8^{v+1}$ if $0 \leq t \leq 1$ and $|K_{iv}(t)| = 0$ if $t < 0$ or $t > 1$. Then $\frac{K_{iv}(\frac{y-x}{u})}{h(y)} \leq \frac{(v+1)^2 8^{v+1}}{h(y)} I_{[x\leq y\leq x+u]}$. For $x \in [C_{1n}, C_{2n}]$, $[x, x+u] \subset [C_{1n}, C_{2n}+u]$. From (2.8), we have $h(y) \geq (\log n)^{-1}$ for $y \in [C_{1n}, C_{2n}+u]$. Thus, for any $y \in (\alpha,\beta)$, $|\frac{K_{iv}(\frac{y-x}{u})}{h(y)}| \leq (v+1)^2 8^{v+1}\log n$ if $x \in [C_{1n}, C_{2n}]$. Since $u(v+1)^2 8^{v+1}\log n \leq n^{-\frac{1}{\log\log n}}(\log n)^{8\log 8+1}(8\log\log n)^2$, $u(v+1)^2 8^{v+1}\log n \to 0$ as $n \to \infty$. Then

$$|\frac{uK_{iv}(\frac{y-x}{u})}{h(y)}| I_{[C_{1n}\leq x\leq C_{2n}]} \to 0 \tag{4.11}$$

uniformly for any $y \in (\alpha,\beta)$. On the other hand,

$$\begin{aligned}
u\int_\Omega c(\theta)e^{\theta x}dG(\theta) &\leq u\int_\Omega c(\theta)e^{|\theta x|}dG(\theta) \\
&\leq u\int_\Omega e^{-|\theta|^{1+\eta}}e^{|\theta|L_n}dG(\theta) + B_1 e^{B_0 L_n} \\
&\leq n^{-\frac{1}{\log\log n}}[e^{(\log\log n)^{(\eta+1)/\eta}} + 1 + B_1(\log n)^{B_0}], \\
&\to 0.
\end{aligned}$$

as $n \to \infty$. Similarly,

$$u\int_\Omega \theta c(\theta)e^{\theta x}dG(\theta) \to 0.$$

It follows that, as $n \to \infty$,

$$u[w(x) + u^v d_n(x)] I_{[C_{1n}\leq x\leq C_{2n}]} \to 0. \tag{4.12}$$

From (4.11) and (4.12), we see that, as $n \to \infty$,

$$|\theta_0 u^2 \frac{K_{0v}(\frac{y-x}{u})}{h(y)} - \frac{uK_{1v}(\frac{y-x}{u})}{h(y)} - u^3 w_n(x)| I_{[C_{1n} \le x \le C_{2n}]} \to 0$$

uniformly for $y \in (\alpha, \beta)$. So there exists an $N_4 (> N_3)$ such that for all $n > N_4$, $y \in (\alpha, \beta)$,

$$|\theta_0 u^2 \frac{K_{0v}(\frac{y-x}{u})}{h(y)} - \frac{uK_{1v}(\frac{y-x}{u})}{h(y)} - u^3 w_n(x)| I_{[C_{1n} \le x \le C_{2n}]} \le 1. \tag{4.13}$$

Recall $Z_{jn}(x) = V_n(X_j, x) - w_n(x) = \theta_0 \frac{K_{0v}(\frac{X_j - x}{u})}{uh(X_j)} - \frac{K_{1v}(\frac{X_j - x}{u})}{u^2 h(X_j)} - w_n(x)$.

$$\begin{aligned} EZ_{jn}^2(x) &= \int_\Omega \int_\alpha^\beta [\theta_0 \frac{K_{0v}(\frac{y-x}{u})}{uh(y)} - \frac{K_{1v}(\frac{y-x}{u})}{u^2 h(y)} - w_n(x)]^2 c(\theta) e^{\theta y} h(y) dy dG(\theta) \\ &= \frac{1}{u^6} \int_\Omega \int_\alpha^\beta [\theta_0 u^2 \frac{K_{0v}(\frac{y-x}{u})}{h(y)} - \frac{uK_{1v}(\frac{y-x}{u})}{h(y)} - u^3 w_n(x)]^2 c(\theta) e^{\theta y} h(y) dy dG(\theta) \\ &\le \frac{1}{u^6}. \end{aligned}$$

Also,

$$\begin{aligned} E|Z_{jn}(x)|^3 &= \int_\Omega \int_\alpha^\beta |\theta_0 \frac{K_{0v}(\frac{y-x}{u})}{uh(y)} - \frac{K_{1v}(\frac{y-x}{u})}{u^2 h(y)} - w_n(x)|^3 c(\theta) e^{\theta y} h(y) dy dG(\theta) \\ &= \frac{1}{u^9} \int_\Omega \int_\alpha^\beta |\theta_0 u^2 \frac{K_{0v}(\frac{y-x}{u})}{h(y)} - \frac{uK_{1v}(\frac{y-x}{u})}{h(y)} - u^3 w_n(x)|^3 c(\theta) e^{\theta y} h(y) dy dG(\theta) \\ &\le \frac{1}{u^9} \int_\Omega \int_\alpha^\beta [\theta_0 u^2 \frac{K_{0v}(\frac{y-x}{u})}{h(y)} - \frac{uK_{1v}(\frac{y-x}{u})}{h(y)} - u^3 w_n(x)]^2 c(\theta) e^{\theta y} h(y) dy dG(\theta), \\ &= \frac{1}{u^3} \sigma_n^2(x). \end{aligned}$$

The proof is completed.

Proof of Lemma 3.4 From lemma 3.2, we have that $|u^v d_n(x)| \le \frac{1}{2\sqrt{n}}$ for all $x \in [C_{1n}, C_{2n}]$ and $n > N_2$. If $w(x) > \frac{1}{\sqrt{n}}$,

$$\sqrt{n} w_n(x) = \sqrt{n}[w(x) + u^v d_n(x)] = \sqrt{n} w(x) + \sqrt{n} u^v d_n(x) > 1 - \frac{1}{2} = \frac{1}{2} > 0$$

and

$$\begin{aligned} \frac{w(x)}{d_n(x)} &= \frac{\sqrt{n} w(x)}{\sqrt{n} w(x) + \sqrt{n} u^v d_n(x)} \\ &\le \frac{\sqrt{n} w(x) - 1 + 1}{\sqrt{n} w(x) - 1 + \frac{1}{2}} \\ &\le 2. \end{aligned}$$

Then (3.5) is proved. (3.6) can be proved in a similar way.

§ 5 A Few Comments

From §4, we see that Lemma 3.2 and Lemma 3.3 depend strongly on $B(m)$. Now we use the result that $B(m) = (m+1)^2 8^{m+1}$. In the appendix, we use the polynomial functions as kernels. For us, the only things we need for those kernels are their common compact supports, (2,1) and (2.2). So if we can find some kernels such that $B(m)$ is smaller than $(m+1)^2 8^{m+1}$ significantly, then our result can be much improved. If $K_{im}(y)$ can be constructed such that $B(m) \leq C$, for some constant C, then our result can be improved to

$$n\epsilon_n[R(G,\delta_n) - R(G,\delta)] \to 0, \tag{5.1}$$

where ϵ_n can be any prespecified positive sequence such that $\epsilon_n \to 0$ as $n \to \infty$. The proof of this is almost similar to the proof of Theorem 3.5. Then (5.1) tells us that we can construct the empirical Bayes test such that it has the convergence rate as close to $O(n^{-1})$ as possible. So the problem to find the empirical Bayes test such that its regret converges to zero with a rate as close to $O(n^{-1})$ as possible is reduced to find the kernels such that $B(m)$ is as small as possible.

Appendix

We will prove that $K_{im}(y)$, used in §2, can be chosen properly such that $B(m) = (m+1)^2 8^{m+1}$. Fix m for a moment. For $i = 0$, let

$$K_{0m}(y) = \begin{cases} a_m y^m + a_{m-1} y^{m-1} + \cdots + a_0, & \text{if } 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

For $i = 1$, let

$$K_{1m}(y) = \begin{cases} b_m y^m + b_{m-1} y^{m-1} + \cdots + b_0, & \text{if } 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

We treat $K_{0m}(y)$ first. (2.1) implies

$$\begin{cases} \frac{a_m}{m+1} + \frac{a_{m-1}}{m} + \cdots + a_0 & = 1 \\ \frac{a_m}{m+2} + \frac{a_{m-1}}{m+1} + \cdots + \frac{a_0}{2} & = 0 \\ \cdots \quad \cdots \quad \cdots & \\ \frac{a_m}{2m+1} + \frac{a_{m-1}}{2m} + \cdots + \frac{a_0}{m+1} & = 0. \end{cases}$$

Using Cramer's rule, we have, for $0 \le s \le m$,

$$a_s = \frac{\begin{vmatrix} \frac{1}{m+1} & \cdots & 1 & \cdots & 1 \\ \frac{1}{m+2} & \cdots & 0 & \cdots & \frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots & \\ \frac{1}{2m+1} & \cdots & 0 & \cdots & \frac{1}{m+1} \end{vmatrix}}{\begin{vmatrix} \frac{1}{m+1} & \cdots & \frac{1}{s+1} & \cdots & 1 \\ \frac{1}{m+2} & \cdots & \frac{1}{s+2} & \cdots & \frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots & \\ \frac{1}{2m+1} & \cdots & \frac{1}{s+m+1} & \cdots & \frac{1}{m+1} \end{vmatrix}} = \frac{det_2}{det_1},$$

where det_2 is the numerator of a_s and det_1 is the denominator of a_s. A simple calculation shows that

$$det_1 = \frac{[m!(m-1)!\cdots 2!]^3}{(m+1)!(m+2)!\cdots(2m+1)!}$$

and

$$det_2 = \frac{(-1)^{1+(m-s+1)}(m!)^2[(m-1)!(m-2)!\cdots 2!]^3(s+m+1)!}{(m+2)!(m+3)!\cdots(2m+1)!(s+1)!s!(m-s)!}.$$

Thus

$$a_s = \frac{(-1)^{1+(m-s+1)}(m+1)(m+s+1)!}{(s+1)!s!(m-s)!}.$$

And

$$|a_s| = (m+1)\binom{m+s+1}{s+1}\binom{m}{s} \le (m+1)\binom{2m+1}{m+1}\binom{m}{s}.$$

If $m = 2l$,

$$\begin{aligned} |a_s| &\le (2l+1)\binom{4l+1}{2l+1}\binom{2l}{l} \\ &= \frac{(4l+1)!}{(2l)!l!l!}. \end{aligned}$$

Using Stirling's formula $(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}e^{-n} < n! < 2(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}e^{-n}$ in above inequality and then simplifying it, we have

$$|a_s| \le 2^{6l+3} \le 8^{m+1}.$$

If $m = 2l+1$,

$$\begin{aligned} |a_s| &\le (2l+2)\binom{4l+3}{2l+2}\binom{2l+1}{l+1} \\ &= \frac{(4l+3)!}{(2l+1)!(l+1)!l!l!}. \end{aligned}$$

By Stirling's formula, we can get

$$|a_s| \le \frac{1}{\pi} 2^{6l+6} \le 8^{m+1}.$$

As for $K_{1m}(y)$, (2.2) implies

$$\begin{cases} \frac{b_m}{m+1} + \frac{b_{m-1}}{m} + \cdots + b_0 & = 0 \\ \frac{b_m}{m+2} + \frac{b_{m-1}}{m+1} + \cdots + \frac{b_0}{2} & = 1 \\ \cdots \quad \cdots \quad \cdots & \\ \frac{b_m}{2m+1} + \frac{b_{m-1}}{2m} + \cdots + \frac{b_0}{m+1} & = 0. \end{cases}$$

Then, for $0 \le s \le m$,

$$b_s = \frac{\begin{vmatrix} \frac{1}{m+1} & \cdots & 0 & \cdots & 1 \\ \frac{1}{m+2} & \cdots & 1 & \cdots & \frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots & \\ \frac{1}{2m+1} & \cdots & 0 & \cdots & \frac{1}{m+1} \end{vmatrix}}{\begin{vmatrix} \frac{1}{m+1} & \cdots & \frac{1}{s+1} & \cdots & 1 \\ \frac{1}{m+2} & \cdots & \frac{1}{s+2} & \cdots & \frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots & \\ \frac{1}{2m+1} & \cdots & \frac{1}{s+m+1} & \cdots & \frac{1}{m+1} \end{vmatrix}} = \frac{(-1)^{2+(m-s+1)}(m+2)!(s+m+1)!}{(m-1)!s!(s+2)s!(m-s)!}.$$

Thus

$$|\frac{b_s}{a_s}| = \frac{m(m+2)(s+1)}{s+2} \le (m+1)^2.$$

So we prove that, for $i = 0, 1$,

$$|K_{im}(y)| \le (m+1)^2 8^{m+1}.$$

Acknowledgment We would like to thank Prof. TaChen Liang for his careful reading of this paper and very helpful comments. We also thank Prof. Herman Rubin for a helpful suggestion on the proof in the appendix.

References

[1] Gupta, S.S. and Li, J. (1999). Empirical Bayes tests with $n^{-1+\epsilon}$ convergence rate in continuous one-parameter exponential family. **Department of Statistics, Purdue University, Technical Report #99-09C**.

[2] Johns, M.V., Jr. and Van Ryzin, J.R. (1972). Convergence rates for empirical Bayes two-action problems, II. Continuous Case. **Ann. Math. Statist. 43**, 934-947.

[3] Karunamuni, R. J. (1996). Optimal rates of convergence of empirical Bayes tests for the continuous one-parameter exponential family. **Ann. Statist. 24**, 212-231.

[4] Karunamuni, R. J. and Yang, H. (1995). On convergence rates of monotone empirical Bayes tests for the continuous one-parameter exponential family. **Statistics and Decisions, 13**, 181-192.
[5] Liang, T. (1999). On empirical Bayes tests in a positive exponential family. To appear in **J. Statist. Plann. Inference.**
[6] Michel, R. (1981). On the constant in the nonuniform version of the Berry-Esseen Theorem. **Z.W., 55** 109-117.
[7] Petrov, V.V. (1975). **Sums of Independent Random Variables**. Springer-verlag, Berlin Heidelberg New York.
[8] Robbins, H.(1956). An empirical Bayes approach to statistics. **Proc. Third. Berkeley Symp. Math. Statist, Probab. 1**, 157-163, Univ. California Press, Berkeley.
[9] Robbins, H.(1964). The empirical Bayes approach to statistical decision problems. **Ann. Math. Statist. 35** 1-20.
[10] Singh, R.S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with convergence rate near the best possible rate. **Ann. Statist. 7**, 890-902.
[11] Stijnen, T.(1985). On the asymptotic behavior of empirical Bayes tests for the continuous one-parameter exponential family. **Ann. Statist. 13**, 403-412.
[12] Van Houwelingen, J.C.(1976). Monotone empirical Bayes tests for the continuous one-parameter exponential family. **Ann. Statist. 4**, 981-989.

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	3. REPORT TYPE AND DATES COVERED
	September 1999	Technical Report, September 1999

4. TITLE AND SUBTITLE
Empirical Bayes Tests In Some Continuous Exponential Family

5. FUNDING NUMBERS
DAAH04-95-1-0165

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

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

8. PERFORMING ORGANIZATION REPORT NUMBER
Technical Report 99-21

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)*

Empirical Bayes tests for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ in some continuous exponential family with density $c(\theta)exp\{\theta x\}h(x)$, $\infty \leq \alpha < x < \beta \leq \infty$, where, $c(\theta)$ satisfies certain condition, are investigated under the linear loss. Using a mild assumption that $\int_\Omega |\theta| dG(\theta) < \infty$, we construct the empirical Bayes test and show that its regret goes to zero with a convergence rate of order $o(n^{-1+\frac{1}{\log\log n}})$. The applications of our result to $N(\theta, 1)$ and the general exponential family distributions are given as corollaries.

14. SUBJECT TERMS
empirical Bayes; exponential family, rate of convergence, regret

15. NUMBER IF PAGES
22

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