

EMPIRICAL BAYES TESTS BASED ON KERNEL SEQUENCE ESTIMATION

Jianjun Li and Shanti S. Gupta

Purdue Univeristy

Abstract: In this paper, we consider the hypothesis-testing problem in the continuous one-parameter exponential family using the nonparametric empirical Bayes approach. In order to estimate an unknown marginal density and its derivative, a kernel sequence method is introduced. This method uses a sequence of kernel functions and allows the kernel index and window bandwidth to vary simultaneously. Thus improved estimates are obtained. Then we construct a monotone empirical Bayes test based on these estimates and show that the rule has a rate of convergence of $(\ln n)^{3+\epsilon}/n$ for any $\epsilon > 0$. This rate substantially improves the previous results and is much closer to the lower bound rate $1/n$. Since the rule is monotone, it also has good performance for small samples.

Key words and phrases: Empirical bayes, kernel sequence method, rate of convergence, regret bayes risk.

1. Introduction

Assume that X is an observation from the distribution with density

$$f(x|\theta) = c(\theta) \exp\{\theta x\}h(x), \quad -\infty \leq a < x < b \leq +\infty, \quad (1.1)$$

where $h(x)$ is continuous, positive for $x \in (a, b)$, θ is a parameter distributed according to an unknown prior G on the parameter space Ω , a subset of the natural parameter space $\{\theta : c(\theta) > 0\}$.

We consider the problem of testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is known. The loss function is $l(\theta, 0) = \max\{\theta - \theta_0, 0\}$ for accepting H_0 and $l(\theta, 1) = \max\{\theta_0 - \theta, 0\}$ for accepting H_1 . A test $\delta(x)$ is a measurable mapping from (a, b) into $[0, 1]$ so that $\delta(x) = P\{\text{accepting } H_1 | X = x\}$, i.e., $\delta(x)$ is the probability of accepting H_1 when $X = x$ is observed. Let $R(G, \delta)$ denote the Bayes risk of a test δ when G is a prior distribution. Let $\phi_G(x) = E[\theta | X = x]$. Given that $E[|\theta|] < \infty$, a Bayes test δ_G is found as

$$\delta_G(x) = \begin{cases} 1 & \text{if } \phi_G(x) \geq \theta_0, \\ 0 & \text{if } \phi_G(x) < \theta_0. \end{cases} \quad (1.2)$$

Because $\phi_G(x)$ involves G , the above solution works only if the prior G is given. If G is unknown, this testing problem is formed as a compound decision problem and the empirical Bayes approach is used. Let X_1, \dots, X_n be the observations from n independent past experiences and let X be the present observation. Based on $\tilde{X}_n = (X_1, \dots, X_n)$ and X , an empirical Bayes rule $\delta_n(X, \tilde{X}_n) (\equiv \delta_n(X) \equiv \delta_n)$ can be constructed. The performance of δ_n is measured by $R(G, \delta_n) - R(G, \delta_G)$, where $R(G, \delta_n)$ is the overall risk of δ_n and $R(G, \delta_n) = E[(1 - \delta_n)(\theta - \theta_0)I_{[\theta > \theta_0]} + \delta_n(\theta_0 - \theta)I_{[\theta \leq \theta_0]}]$. The quantity $R(G, \delta_n) - R(G, \delta_G)$ is referred to as the regret Bayes risk (or regret) in the literature.

Let $\alpha_G(x) = \int c(\theta) \exp(\theta x) dG(\theta)$ and $\psi_G(x) = \int \theta c(\theta) \exp(\theta x) dG(\theta)$. It is clear that $\phi_G(x) = \psi_G(x) / \alpha_G(x)$ and $\phi_G(x) \geq \theta_0 \iff w(x) \equiv \theta_0 \alpha_G(x) - \psi_G(x) \leq 0$. So the construction of δ_n involves estimation of $\alpha_G(x)$ and $\phi_G(x)$. This is usually done using a kernel method. In this paper, we introduce a kernel sequence method and apply it to obtain the estimates of $\alpha_G(x)$ and $\phi_G(x)$. The idea of the kernel sequence method is to use a sequence of kernel functions and allow both the kernel index and window bandwidth to vary simultaneously.

Based on the estimates of $\alpha_G(x)$ and $\phi_G(x)$, we construct an empirical Bayes rule δ_n for the testing problem mentioned above. Then we show that δ_n has a rate of convergence of $(\ln n)^{3+\epsilon}/n$ ($\epsilon > 0$) with the assumption $E[|\theta|] < \infty$, a substantial improvement over previous results.

The readers interested in empirical Bayes approach may refer to two introductory papers of Robbins (1956, 1964). For the above empirical Bayes testing problem, Johns and Van Ryzin (1972) made an early contribution. Van Houwelingen (1976) constructed monotone empirical Bayes tests which achieve the rate $O(n^{-2r/(2r+1)}(\ln n)^2)$ if $E[|\theta|^{r+1}] < \infty$. Van Houwelingen also showed that his rules have good performance for small samples since they are monotone. Karunamuni and Yang (1995) studied monotone rules and their asymptotic behavior. With the assumption $c_G \in [-A, A]$, they obtain the rate $O(n^{-2r/(2r+1)})$. Karunamuni (1996) made an attempt to find the optimal rate of convergence of the monotone empirical Bayes rule. Recently Liang (2000) investigated a special case of (1.1), namely a $N(\theta, 1)$ model, and obtained improved results. See also Gupta and Li (2001a). Another related work is Stijnen (1985), in which the asymptotic behavior of both monotone empirical Bayes rules and non-monotone rules was studied.

This paper is organized as follows. In Section 2 we introduce a few preliminary results and Section 3, the idea of kernel sequence method. In Section 4, we construct the monotone empirical Bayes test δ_n and obtain its rate of convergence. Section 5 contains proofs of the main results in Section 4. In the appendix, we provide proofs of a few lemmas used in Section 5.

2. Preliminary

We assume that G satisfies $\int |\theta|dG(\theta) < \infty$ throughout this paper. Note that $\alpha_G(x)$ and $\phi_G(x)$ exist for all $x \in (a, b)$ under the assumption $\int |\theta|dG(\theta) < \infty$ and are infinitely differentiable for $x \in (a, b)$. Furthermore $\phi'_G(x) \geq 0$. If $\lim_{x \downarrow a} \phi_G(x) \geq \theta_0$, then $\phi_G(x) \geq \theta_0$ and $\delta_G(x) \equiv 1$ for all $x \in (a, b)$; if $\lim_{x \uparrow b} \phi_G(x) \leq \theta_0$, then $\phi_G(x) \leq \theta_0$ and $\delta_G(x) \equiv 0$ for all $x \in (a, b)$. In both cases, we say $\delta_G(x)$ is degenerate. We assume $\delta_G(x)$ is non-degenerate with $\lim_{x \downarrow a} \phi_G(x) < \theta_0 < \lim_{x \uparrow b} \phi_G(x)$. Then G is non-degenerate and $\phi'_G(x) > 0$. Therefore there exists a unique point $c_G \in (a, b)$ such that $\phi_G(x) > \theta_0$ for $x > c_G$, $\phi_G(x) = \theta_0$ for $x = c_G$ and $\phi_G(x) < \theta_0$ for $x < c_G$ (see Van Houwelingen (1976) and others). Note that $w(x) = \theta_0\alpha_G(x) - \psi_G(x)$. Then c_G is the unique root of $w(x)$.

Now the Bayes rule defined by (1.2) can be represented as

$$\delta_G(x) = \begin{cases} 1 & \text{if } \phi_G(x) \geq \theta_0 \iff w(x) \leq 0 \iff x \geq c_G, \\ 0 & \text{if } \phi_G(x) < \theta_0 \iff w(x) > 0 \iff x < c_G. \end{cases} \tag{2.1}$$

The Bayes rule δ_G defined by (2.1) is characterized by a single number c_G , so a monotone empirical Bayes test (MEBT) can be constructed through estimating c_G by $c_n(X_1, \dots, X_n)$, say, and defining

$$\delta_n = \begin{cases} 1 & \text{if } x \geq c_n, \\ 0 & \text{if } x < c_n. \end{cases} \tag{2.2}$$

The regret of δ_n is

$$R(G, \delta_n) - R(G, \delta_G) = E \int_{c_n}^{c_G} w(x)h(x)dx. \tag{2.3}$$

3. Kernel Sequence Method

We introduce a kernel sequence method which uses a sequence of kernel functions instead of a single one. As the number of observations increases, the kernel function and kernel bandwidth are set to vary simultaneously.

For $i = 0, 1$ and $m = 1, 2, \dots$, let $K_{im}(y)$ be a Borel-measurable function such that $K_{im}(y)$ vanishes outside the interval $[A_{im}, B_{im}]$. Suppose

$$\int y^j K_{0m}(y)dy \begin{cases} = 1 & \text{if } j = 0, \\ = 0 & \text{if } j = 1, \dots, k_{0m} - 1, \\ \neq 0 & \text{if } j = k_{0m}, \end{cases} \tag{3.1}$$

$$\int y^j K_{1m}(y)dy \begin{cases} = 0 & \text{if } j = 0, 2, 3, \dots, k_{1m} - 1, \\ = 1 & \text{if } j = 1, \\ \neq 0 & \text{if } j = k_{1m}. \end{cases} \tag{3.2}$$

Let $u = u_n$ be a sequence of positive numbers and $v = v_n$ be a sequence of positive integers. For any $x \in (a, b)$, define

$$\alpha_n(x) = \frac{1}{nu} \sum_{j=1}^n K_{0v} \left(\frac{X_j - x}{u} \right) / h(X_j), \quad \psi_n(x) = \frac{1}{nu^2} \sum_{j=1}^n K_{1v} \left(\frac{X_j - x}{u} \right) / h(X_j).$$

For appropriate u and v , $\alpha_n(x)$ and $\psi_n(x)$ are the estimates of $\alpha_G(x)$ and $\psi_G(x)$, respectively. In these kernel estimates, u is called the kernel bandwidth and v is called the kernel index.

Note that the kernel index v of functions K_{0v} and K_{1v} depends on n . As n increases, v changes, and so K_{0v} and K_{1v} change. Both kernel indices and window bandwidths vary in the estimates and this is a little different from the traditional fixed index kernel method.

4. MEBT For General Exponential Family

We use the idea of the kernel sequence method to find estimators of $\alpha_G(x)$ and $\psi_G(x)$. Then we construct c_n based on them.

We now present two sequences of kernel functions K_{0v} and K_{1v} . For odd v , let $K_{0v}(y) = K_{0(v+1)}(y)$; for even v ,

$$K_{0v}(y) = \begin{cases} p_v y^v + p_{v-1} y^{v-1} + \dots + p_0, & \text{if } -1 \leq y \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

$$p_i = \begin{cases} 0, & \text{if } i \text{ is odd,} \\ \frac{(-1)^{i/2} v!(v+i)!v(v-i)}{i!(i+1)2^{2v+1}[(\frac{v}{2})!]^2(\frac{v+i}{2})!(\frac{v-i}{2})!}, & \text{if } i \text{ is even.} \end{cases} \tag{4.1}$$

For even v , $K_{1v}(y) = K_{1(v+1)}(y)$; for odd v ,

$$K_{1v}(y) = \begin{cases} q_v y^v + q_{v-1} y^{v-1} + \dots + q_0, & \text{if } -1 \leq y \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

$$q_i = \begin{cases} 0, & \text{if } i \text{ is even,} \\ \frac{(-1)^{(i+1)/2} (v+1)!(v+i)!(v-1)(v-i)}{i!(i+2)2^{2v+1}(\frac{v-1}{2})!(\frac{v+1}{2})!(\frac{v+i}{2})!(\frac{v-i}{2})!}, & \text{if } i \text{ is odd.} \end{cases} \tag{4.2}$$

The kernel functions (4.1) and (4.2) are in Gasser, Muller and Mammitzsch (1985). They showed that $K_{0v}(y)$ satisfies (3.1) with $A_{0v} = -1$, $B_{0v} = 1$, $k_{0v} = v$ if v is even and $k_{0v} = v + 1$ if v is odd; $K_{1v}(y)$ satisfies (3.2) with $A_{1v} = -1$, $B_{1v} = 1$, $k_{1v} = v$ if v is odd and $k_{1v} = v + 1$ if v is even.

For $\epsilon > 0$, let $\epsilon_n = (\ln n)^{-\epsilon} \wedge 1$. Denote $u = u_n = \epsilon_n^{1/3}$ and $v = v_n = [-\frac{\ln n}{\ln u}] \vee 0 + 1$, where $[x]$ denote the integer part of x . Then $u^v \leq n^{-1}$. For any $x \in (a, b)$, define

$$\alpha_n(x) = \frac{1}{nu} \sum_{j=1}^n K_{0v} \left(\frac{X_j - x}{u} \right) / h(X_j), \quad \psi_n(x) = \frac{1}{nu^2} \sum_{j=1}^n K_{1v} \left(\frac{X_j - x}{u} \right) / h(X_j). \tag{4.3}$$

It is shown later that $\alpha_n(x)$ and $\psi_n(x)$ are consistent estimators of $\alpha_G(x)$ and $\phi_G(x)$, respectively. Define $W_n(x) = \theta_0 \alpha_n(x) - \psi_n(x)$. Then $W_n(x)$ is a consistent estimator of $w(x)$.

Since c_G is the unique root of $w(x)$, we are going to use $W_n(x)$ to construct c_n . Before doing this, let us examine δ_G . Note that δ_G is a monotone rule. If x is larger than c_G , we accept H_1 ; if x is smaller than c_G , we accept H_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 accept H_1 if the present observation x is quite large and accept H_0 if it is quite small. Knowing this, we want numbers c_{1n} and c_{2n} depending on n such that we accept H_1 if we observe $x > c_{2n}$ and accept H_0 if we observe $x < c_{1n}$. Once proper c_{1n} and c_{2n} are found, we can concentrate our effort on $x \in [c_{1n}, c_{2n}]$.

The idea of splitting (a, b) into (a, c_{1n}) , $[c_{1n}, c_{2n}]$ and (c_{2n}, b) is called the localization technique. To implement it, we need the following lemma.

Lemma 4.1. *Four sequences of numbers $\{a_n, \bar{a}_n, b_n, \bar{b}_n\}$ can be found such that $a_n \downarrow a$, $b_n \uparrow b$, and for large n*

- (i) $-[(\ln \ln n) \wedge u^{-1}] \leq a_n < b_n \leq [(\ln \ln n) \wedge u^{-1}]$;
- (ii) $\min_{a_n < x < b_n} h(x) \geq u$;
- (iii) $\int_{\bar{a}_n}^{a_n} h(t) dt \geq 2u$, $\int_{b_n}^{\bar{b}_n} h(t) dt \geq 2u$.

The proof is obvious, hence omitted.

Let $c_{1n} = a_n + u + u^{1/3}$ and $c_{2n} = b_n - u - u^{1/3}$. From Lemma 4.1, we know that $c_{1n} \downarrow a$ and $c_{2n} \uparrow b$. So c_G will fall in $[c_{1n}, c_{2n}]$ for large values of n and therefore can be expressed as $c_G = \int_{c_{1n}}^{c_{2n}} I_{[w(x)>0]} dx + c_{1n}$. Recall $W_n(x)$, defined as $W_n(x) = \theta_0 \alpha_n(x) - \psi_n(x)$, is a consistent estimator of $w(x)$. Then we define c_n as:

$$c_n = \int_{c_{1n}}^{c_{2n}} I_{[W_n(x)>0]} dx + c_{1n}. \tag{4.4}$$

A monotone empirical Bayes test $\delta_n(x)$ is now proposed as follows:

$$\delta_n = \begin{cases} 1 & \text{if } x \geq c_n, \\ 0 & \text{if } x < c_n. \end{cases} \tag{4.5}$$

It is obvious that $c_n \in [c_{1n}, c_{2n}]$ so if $x > c_{2n}$, we will accept H_1 , if $x < c_{1n}$ we will accept H_0 . If $x \in [c_{1n}, c_{2n}]$, we calculate c_n and compare x with c_n to make the decision.

The use of the localization technique helps us avoid the boundary effect of kernel estimates. It provides bounds on the moments of $W_n(x)$ for $x \in [c_{1n}, c_{2n}]$ (see Lemma 5.3 below). Also it results in a nice lower bound on $|w(x)|$ for $x \in [c_{1n}, c_G - \epsilon_G] \cup [c_G + \epsilon_G, c_{2n}]$ and $\epsilon_G > 0$ (see Lemma 5.2 below). This is crucial to get the desired rate of convergence in Section 5. The localization technique has also been used in Gupta and Li (1999), and Gupta and Li (2001a, b).

Note that since $W_n(x)$ is an estimate of $w(x)$, a natural construction of the empirical Bayes rule should be $\delta_n = 1$ if $W_n(x) \leq 0$ and $\delta_n = 0$ if $W_n(x) > 0$. Unfortunately this construction will lead to a non-monotone rule. We use the integration of $I_{[W_n(x) > 0]}$ in (4.4) instead. This technique is borrowed from Brown, Cohen, and Strawderman (1976), Van Houwelingen (1976) and Stijnen (1985).

Now we study the large sample behavior of δ_n . The next two lemmas enable us to express the regret of δ_n through $c_n - c_G$.

Lemma 4.2. $w'(c_G) < 0$.

Since $w'(x)$ is continuous in (a, b) , we can find $N_{\epsilon_G}(c_G)$, a neighborhood of c_G , such that $N_{\epsilon_G}(c_G) \subset (c_{1n}, c_{2n}) \subset (a, b)$ (for large n), and $A_\epsilon = \min_{x \in N_{\epsilon_G}(c_G)} [-w'(x)] > 0$. Denote $\eta_1 = c_G - \epsilon_G$ and $\eta_2 = c_G + \epsilon_G$ in the following.

Lemma 4.3. Let $\bar{h} = \sup\{h(x) : x \in [\eta_1, \eta_2]\}$ and $\bar{w} = \sup\{-w'(x) : x \in [\eta_1, \eta_2]\}$. Then

$$R(G, \delta_n) - R(G, \delta_G) \leq 2^{-1} \bar{h} \bar{w} E(c_n - c_G)^2 + (\theta_0 + E[|\theta|]) \epsilon_G^{-4} E(c_n - c_G)^4.$$

Following (4.4) and $c_G \in [c_{1n}, c_{2n}]$, we have

$$c_n - c_G = - \int_{c_{1n}}^{c_G} I_{[W_n(x) \leq 0]} dx + \int_{c_G}^{c_{2n}} I_{[W_n(x) > 0]} dx. \quad (4.6)$$

So an upper bound for $c_n - c_G$ is easy to obtain through the properties of $W_n(x)$ and $w(x)$. Note that $W_n(x)$ can be written as

$$W_n(x) = \frac{1}{n} \sum_{j=1}^n V_n(X_j, x), \quad \text{where } V_n(X_j, x) = \frac{\theta_0}{u} \cdot \frac{K_{0v}\left(\frac{X_j - x}{u}\right)}{h(X_j)} - \frac{1}{u^2} \cdot \frac{K_{1v}\left(\frac{X_j - x}{u}\right)}{h(X_j)}.$$

For fixed n and x , $V_n(X_j, x)$ are i.i.d. random variables, so $W_n(x)$ is the average of the i.i.d. random variables. After applying the results in Petrov (1995), we have the following result.

Lemma 4.4. $\lim_{n \rightarrow \infty} [n \epsilon_n (\ln n)^{-3} E(c_n - c_G)^2] = 0$, $\lim_{n \rightarrow \infty} [n \epsilon_n (\ln n)^{-3} E(c_n - c_G)^4] = 0$.

The proofs of Lemmas 4.2–4.4 are given in Section 5. Note that $\epsilon_n \leq (\ln n)^{-\epsilon}$. Then as a result of Lemma 4.3 and Lemma 4.4, we have the following.

Theorem 4.1. *Assume that $\int |\theta| dG(\theta) < \infty$ and the Bayes rule δ_G is non-degenerate. Then $R(G, \delta_n) - R(G, \delta_G) = o((\ln n)^{3+\epsilon}/n)$.*

Remark 4.1. To apply the kernel sequence method, a key question is how to construct the sequence of kernel functions. Here we use results of Gasser, Muller and Mammitzsch (1985), but expect that the rate can be improved with a “better” kernel sequence.

Remark 4.2. The rule δ_n is monotone and is weakly admissible as discussed by Van Houwelingen (1976). It therefore has good performance for small samples.

Remark 4.3. What is the best possible rate of MEBT’s? In Gupta and Li (2001b), it is proved that $1/n$ is a (natural) minimax lower bound rate for MEBT’s in the exponential family (1.1). That is,

$$\inf_{\delta_n^* \in \mathcal{D}} \sup_{G \in \mathcal{G}} [R(G, \delta_n^*) - R(G, \delta_G)] \geq l/n \quad \text{for some } l > 0,$$

where \mathcal{D} is the set of monotone empirical Bayes rules of type (2.2) and \mathcal{G} is the set of prior distributions with supports inside $[\theta_{01}, \theta_{02}] \ni \theta_0$. However, we do not know whether the rate $1/n$ is achievable or not.

5. Proofs

We prove the results in the previous sections. First we state some useful lemmas, proofs of which are in the appendix.

Lemma 5.1. *Let $\bar{\alpha}_n = \max\{\alpha_G(x) : x \in [a_n, b_n]\}$. For sufficiently large n , $\bar{\alpha}_n \leq (2u)^{-1}$.*

Lemma 5.2. *For sufficiently large n , if $x \in [c_{1n}, c_{2n}]$, $|w(x)| \leq 2u^{-2}$, and if $x \in [c_{1n}, \eta_1] \cup [\eta_2, c_{2n}]$, $|w(x)| \geq M \cdot u(\ln n)^{-B}$ for some positive constants M and B .*

Lemma 5.3. *Let $w_n(x) = E[V_n(X_j, x)]$, $Z_{jn} = V_n(X_j, x) - w_n(x)$, $\sigma_n^2(x) = E[|Z_{jn}|^2]$ and $\gamma_n(x) = E[|Z_{jn}|^3]$. For sufficiently large n , there exist positive constants l_1, l_2, l_3 and l_4 such that*

- (i) for $x \in [c_{1n}, c_{2n}]$, $|w_n(x) - w(x)| \leq 1/\sqrt{n}$;
- (ii) for $x \in [c_{1n}, c_{2n}]$, $\sigma_n(x) \leq l_1 v^{3/2} u^{-5/2}$;
- (iii) for $x \in [\eta_1, \eta_2]$, $l_2 \leq \sigma_n(x) \leq l_3 (v/u)^{3/2}$;
- (iv) for $x \in [c_{1n}, c_{2n}]$, $\gamma_n(x) \leq l_4 v^{13} 36^v u^{-6}$.

Lemma 5.4. *Let $d_n = \sqrt{v^3/n u^3}$. For $x \in [c_{1n}, c_{2n}]$, n sufficiently large, $w(x) > d_n \implies w_n(x) \geq w(x)/2$, $w(x) < -d_n \implies w_n(x) \leq w(x)/2$.*

Proof of lemma 4.2. Note that $\alpha_G(x)$ is infinitely differentiable, $\alpha'_G(x) = \psi_G(x)$ and $w'(x) = \theta_0\psi_G(x) - \psi'_G(x)$. If $\psi_G(c_G) = 0$, then $w'(c_G) = -\int \theta^2 c(\theta) e^{\theta c_G} dG(\theta) < 0$. If $\psi_G(c_G) > 0$, by Jensen's inequality $\psi'_G(c_G)/\psi_G(c_G) > \psi_G(c_G)/\alpha_G(c_G) = \theta_0$. Thus $w'(c_G) < 0$. Similarly if $\psi_G(c_G) < 0$, $w'(c_G) < 0$. The proof of Lemma 4.2 is complete.

Proof of lemma 4.3. From (2.3),

$$\begin{aligned} R(G, \delta_n) - R(G, \delta_G) &\leq E \left[I_{[|c_n - c_G| > \epsilon_G]} \int_{c_n}^{c_G} w(x) h(x) dx \right] + \bar{h} E \left[I_{[|c_n - c_G| \leq \epsilon_G]} \int_{c_n}^{c_G} w(x) dx \right] \\ &\leq (\theta_0 + \mu_G) \epsilon_G^{-4} E (c_n - c_G)^4 + 1/2 \bar{h} \bar{w} E (c_n - c_G)^2, \end{aligned}$$

where $\int_{c_n}^{c_G} w(x) h(x) dx \leq (\theta_0 + \mu_G)$ and, by a Taylor expansion,

$$I_{[|c_n - c_G| \leq \epsilon_G]} \int_{c_n}^{c_G} w(x) dx = -1/2 \times w'(\hat{c}_n) (c_n - c_G)^2 I_{[|c_n - c_G| \leq \epsilon_G]} \leq 1/2 \bar{w} (c_n - c_G)^2.$$

Proof of lemma 4.4. From (4.6),

$$E (c_n - c_G)^2 \leq E \left[\int_{c_{1n}}^{c_G} I_{[W_n(x) \leq 0]} dx \right]^2 + E \left[\int_{c_G}^{c_{2n}} I_{[W_n(x) > 0]} dx \right]^2 \equiv r_{1n} + r_{2n}. \quad (5.1)$$

The Hölder inequality and a little algebra shows that

$$r_{1n} \leq 2(c_{2n} - c_{1n})I_1 + 2I_2 + 2I_3, \quad (5.2)$$

where $I_1 = \int_{c_{1n}}^{\eta_1} P(W_n(x) \leq 0) dx$, $I_2 = \left(\int_{\eta_1}^{c_G} I_{[w(x) \leq d_n]} dx \right)^2$, $I_3 = E \left[\int_{\eta_1}^{c_G} I_{[W_n(x) \leq 0, w(x) > d_n]} dx \right]^2$. For $w(x) > d_n$, $w_n(x) \geq w(x)/2$ from Lemma 5.4. Then we have

$$P(W_n(x) \leq 0) = P \left(\frac{1}{\sqrt{n\sigma_n^2}} \sum_{j=1}^n Z_{jn} \leq \frac{-\sqrt{n}w_n(x)}{\sigma_n} \right) \leq P \left(\frac{1}{\sqrt{n\sigma_n^2}} \sum_{j=1}^n Z_{jn} \leq \frac{-\sqrt{n}w(x)}{2\sigma_n} \right).$$

Applying Theorem 5.16 on page 168 in Petrov (1995) to the RHS of the above inequality,

$$P(W_n(x) \leq 0) \leq \Phi \left(-\frac{\sqrt{n}w(x)}{2\sigma_n} \right) + \frac{8A\gamma_n(x)}{\sqrt{n}[2\sigma_n + \sqrt{n}w(x)]^3} \equiv S_n(x) + T_n(x), \quad (5.3)$$

where A is a constant and $\Phi(\cdot)$ is the cdf of $N(0, 1)$. For $x \in [c_{1n}, \eta_1]$, $w(x) \geq Mu(\ln n)^{-B}$ and certainly $w(x) > d_n$ for large n . Also note that $\sigma_n \leq l_1 u^{-5/2} v^{3/2}$ and $\gamma_n(x) \leq l_4 v^{13} 36^v u^{-6}$. It follows that $S_n(x) \leq \Phi(-n^{1/4})$ and $T_n(x) \leq n^{-3/2}$ for large n . Thus

$$(c_{2n} - c_{1n})I_1 = (c_{2n} - c_{1n}) \int_{c_{1n}}^{\eta_1 \vee c_{1n}} P(W_n(x) \leq 0) dx = o(n^{-1}). \quad (5.4)$$

For $x \in [\eta_1, c_G]$, $|w'(x)| \geq A_\epsilon$. Thus $I_2 \leq A_\epsilon^{-2} [\int_{\eta_1}^{c_G} I_{[w(x) \leq d_n]} w'(x) dx]^2$. Letting $y = w(x)/d_n$, $I_2 \leq A_\epsilon^{-2} d_n^2 \int_0^\infty I_{[y \leq 1]} dy = A_\epsilon^{-2} d_n^2$. Therefore

$$I_2 = O(d_n^2) = o((\ln n)^3 / (n\epsilon_n)). \tag{5.5}$$

By the Hölder inequality again,

$$I_3 \leq \int_{\eta_1}^{c_G} P(W_n(x) \leq 0) [w(x)]^{3/2} I_{[w(x) > d_n]} dx \times \int_{\eta_1}^{c_G} [w(x)]^{-3/2} I_{[w(x) > d_n]} dx.$$

Letting $y = w(x)/d_n$, $\int_{\eta_1}^{c_G} [w(x)]^{-3/2} I_{[w(x) > d_n]} dx \leq 2/[A_\epsilon \sqrt{d_n}]$. Using the previous two inequalities and (5.3), we have

$$I_3 \leq 2/(A_\epsilon d_n^{1/2}) \left\{ \int_{\eta_1}^{c_G} S_n(x) [w(x)]^{3/2} dx + \int_{\eta_1}^{c_G} T_n(x) [w(x)]^{3/2} dx \right\}. \tag{5.6}$$

For $x \in [\eta_1, c_G]$, $l_2 \leq \sigma_n \leq l_3 \sqrt{v^3/u^3}$ and $\gamma_n(x) \leq l_4 v^{13} 36^v u^{-6}$. Therefore

$$\begin{aligned} \int_{\eta_1}^{c_G} S_n(x) w^{\frac{3}{2}}(x) dx &\leq \frac{1}{A_\epsilon} \int_{\eta_1}^{c_G} \Phi\left(-\frac{\sqrt{nu^3 w(x)}}{2l_3 \sqrt{v^3}}\right) [w(x)]^{\frac{3}{2}} dw(x) \\ &\leq \frac{(2l_3 d_n)^{5/2}}{A_\epsilon} \int_0^\infty \Phi(-y) y^{\frac{3}{2}} dy, \end{aligned} \tag{5.7}$$

$$\int_{\eta_1}^{c_G} T_n(x) [w(x)]^{3/2} dx \leq \frac{8Al_4 v^{13} 36^v}{A_\epsilon n^{7/4} u^6} \int_0^\infty \frac{y^{3/2}}{[2l_2 + y]^3} dy. \tag{5.8}$$

Combining (5.6)–(5.8), we have $I_3 = o((\ln n)^3 / (n\epsilon_n))$. This together with (5.4) and (5.5) yields $r_{1n} = o((\ln n)^3 / (n\epsilon_n))$. Similarly, $r_{2n} = o((\ln n)^3 / (n\epsilon_n))$. Then $E(c_n - c_G)^2 = o((\ln n)^3 / (n\epsilon_n))$. Observe that $(c_n - c_G)^4 \leq (\int_{c_{1n}}^{c_G} I_{[W_n(x) \leq 0]} dx)^4 + (\int_{c_G}^{c_{2n}} I_{[W_n(x) > 0]} dx)^4$, and

$$E\left(\int_{c_{1n}}^{c_G} I_{[W_n(x) \leq 0]} dx\right)^4 \leq 8(c_{2n} - c_{1n})^3 I_1 + 4\epsilon_G^2 (I_2 + I_3) = o((\ln n)^3 / (n\epsilon_n)).$$

Then $E(c_n - c_G)^4 = o((\ln n)^3 / (n\epsilon_n))$. This completes the proof of Lemma 4.4.

Appendix

Lemma A.1. *The following hold.*

- (i) $|K_{iv}(y)| \leq kv^{10} 36^v$ for some constant k , $i = 0, 1$.
- (ii) $v^{-1} \int |K_{0v}(y)|^2 dy \rightarrow \pi^{-1}$ as $v \rightarrow \infty$.
- (iii) $v^{-3} \int |K_{1v}(y)|^2 dy \rightarrow (3\pi)^{-1}$ as $v \rightarrow \infty$.

Proof. (i) is obtained by simple calculations. By definition and Theorem 1 of Gasser, Muller and Mammitzsch (1985), for an even v ,

$$\int_{-1}^1 K_{0v}^2(y) dy = \frac{v^2[(v-1)!!]^2}{2[v!!]^2}, \quad \int_{-1}^1 K_{1v}^2(y) dy = \frac{v^2[(v+1)!!]^2}{6[v!!]^2},$$

where $v!! = v \times (v - 2) \times \dots \times 1$ if v is odd and $v!! = v \times (v - 2) \times \dots \times 2$ if v is even. Since $s[(2s - 1)!!]^2 / [(2s)!!]^2 \rightarrow \pi^{-1}$ as $s \rightarrow \infty$, (ii) and (iii) are obvious. The case of odd v can be proved similarly.

Proof of Lemma 5.1. Note that $\alpha_G''(x) = \int \theta^2 c(\theta) e^{\theta x} dG(\theta) > 0$ for $x \in (a, b)$. Then $\alpha_G(x)$ is a convex function and $\bar{\alpha}_n = \alpha_G(a_n) \vee \alpha_G(b_n)$. We prove $\alpha_G(a_n) \leq (2u)^{-1}$, the proof of $\alpha_G(b_n) \leq (2u)^{-1}$ is similar. Since $c(\theta) = 1 / \{\int_a^b h(x) e^{\theta x} dx\}$ and $\alpha_G(a_n) = \int c(\theta) e^{\theta a_n} dG(\theta)$, it follows

$$\alpha_G(a_n) \leq \int_{[\theta \geq 0]} \frac{1}{\int_{a_n}^{\bar{b}_n} h(x) \exp(\theta(x - a_n)) dx} dG(\theta) + \int_{[\theta < 0]} \frac{1}{\int_{\bar{a}_n}^{a_n} h(x) \exp(\theta(x - a_n)) dx} dG(\theta).$$

Note that $\int_{a_n}^{\bar{b}_n} \exp(\theta(x - a_n)) h(x) dx \geq 2u$ as $\theta \geq 0$ and $\int_{\bar{a}_n}^{a_n} \exp(\theta(x - a_n)) h(x) dx \geq 2u$ as $\theta < 0$, from Lemma 4.1. Then Lemma 5.1 holds.

Proof of Lemma 5.2. Since $\psi_G(x) = \int \theta c(\theta) \exp(\theta x) dG(\theta)$ and $u|\theta| \leq \exp(u|\theta|)$,

$$|\psi_G(x)| \leq u^{-1} \left[\int_{[\theta \geq 0]} c(\theta) \exp(\theta(x + u)) dG(\theta) + \int_{[\theta < 0]} c(\theta) \exp(\theta(x - u)) dG(\theta) \right].$$

From Lemma 5.1, for $x \in [c_{1n}, c_{2n}]$, $\alpha_G(x) \leq 1/(2u)$. Then $|\psi_G(x)| \leq 1/u^2$ and $|w(x)| \leq 2/u^2$ for large n . Assume that $B > 0$ such that $\int_{\{|\theta| \leq B\}} dG(\theta) > 0$. Let $\Omega_B = \{\theta : \theta \in \Omega, |\theta| \leq B\}$. Since $1/c(\theta)$ is a convex function of θ on Ω , $c(\theta)$ is bounded on Ω_B . Thus $\int_{\Omega_B} c(\theta) dG(\theta)$ is finite.

Recall that $w(x) = \alpha_G(x)[\theta_0 - \phi_G(x)]$. Since $\phi_G(x)$ is increasing and $\phi_G(c_G) = 0$, then for $x \in [c_{1n}, \eta_1]$, $\theta_0 - \phi_G(x) \geq \theta_0 - \phi_G(\eta_1) > 0$; for $x \in [\eta_2, c_{2n}]$, $\phi_G(x) - \theta_0 \geq \phi_G(\eta_2) - \theta_0 > 0$. For $x \in [c_{1n}, c_{2n}]$, $|x| \leq \ln \ln n$ and

$$\alpha_G(x) \geq \int_{\Omega_B} c(\theta) \exp(-\theta |\ln \ln n|) dG(\theta) \geq (\ln n)^{-B} \int_{\Omega_B} c(\theta) dG(\theta).$$

Let $M = \{[\theta_0 - \phi_G(\eta_1)] \wedge [\phi_G(\eta_2) - \theta_0]\} \cdot \int_{\Omega_B} c(\theta) dG(\theta)$. Then Lemma 5.2 is proved.

Proof of Lemma 5.3. We prove (i) for even v only, odd v follows similarly. Using a Taylor expansion of $e^{\theta u x}$, simple calculations show that

$$E \left[\frac{K_{0v}(\frac{X_j - x}{u})}{uh(X_j)} \right] = \int c(\theta) e^{\theta x} dG(\theta) + u^v \int \theta^v c(\theta) e^{\theta x} \left[\int_{-1}^1 \frac{K_{0v}(t) t^v e^{\theta u t}}{v!} dt \right] dG(\theta),$$

$$E \left[\frac{K_{1v}(\frac{X_j - x}{u})}{u^2 h(X_j)} \right] = \int \theta c(\theta) e^{\theta x} dG(\theta) + u^v \int \theta^{v+1} c(\theta) e^{\theta x} \left[\int_{-1}^1 \frac{K_1(t) t^{v+1} e^{\theta u t}}{(v+1)!} dt \right] dG(\theta),$$

where $|t^*|, |t^{**}| < 1$. Then $E[V_n(X_j, x)] = w(x) + u^{v/2}d_n(x)$ and

$$d_n(x) = \theta_0 u^{v/2} \int \frac{\theta^v}{v!} c(\theta) e^{\theta x} \left[\int_{-1}^1 K_{0v}(t) t^v e^{\theta u t^*} dt \right] dG(\theta) - u^{v/2} \int \frac{\theta^{v+1}}{(v+1)!} c(\theta) e^{\theta x} \left[\int_{-1}^1 K_{1v}(t) t^{v+1} e^{\theta u t^{**}} dt \right] dG(\theta).$$

Since $(u^{1/3}\theta)^v/v! \leq \exp(|\theta|u^{1/3})$ and $(u^{1/3}\theta)^{v+1}/(v+1)! \leq \exp(|\theta|u^{1/3})$, for $x \in [c_{1n}, c_{2n}]$, $|d_n(x)| \leq u^{v/6-1} \int c(\theta) e^{\theta x + |\theta|u + |\theta|u^{1/3}} dG(\theta) \cdot [|\theta_0| \int_{-1}^1 |K_{0v}(t)| dt + \int_{-1}^1 |K_{1v}(t)| dt] \leq u^{v/6-1} \bar{\alpha}_n \{ |\theta_0| [2 \int_{-1}^1 |K_{0v}(y)|^2 dy]^{1/2} + [2 \int_{-1}^1 |K_{1v}(y)|^2 dy]^{1/2} \}$. From Lemma A.1 and Lemma 5.1, $|d_n(x)| \rightarrow 0$ uniformly for $x \in [c_{1n}, c_{2n}]$. Then (i) is proved. For $x \in [c_{1n}, c_{2n}]$, $h(x+u) \geq u$ from Lemma 4.1 and

$$\begin{aligned} \sigma_n^2(x) &\leq E \left[\theta_0 \frac{K_{0v}(\frac{X_j-x}{u})}{uh(X_j)} - \frac{K_{1v}(\frac{X_j-x}{u})}{u^2h(X_j)} \right]^2 \\ &= u^{-3} \int \int_{-1}^1 [\theta_0 u K_{0v}(t) - K_{1v}(t)]^2 c(\theta) e^{\theta x} e^{\theta u t} [h(x+ut)]^{-1} dt dG(\theta) \\ &\leq l_1^2 u^{-4} v^3 \int c(\theta) e^{\theta x} e^{|\theta|u} dG(\theta) \leq l_1^2 u^{-5} v^3. \end{aligned}$$

Especially, for $x \in [\eta_1, \eta_2]$, letting $\underline{h} = \min\{h(x+ut) : x \in [\eta_1, \eta_2], |t| \leq 1\}$, $\sigma_n^2(x) \leq l_5 u^{-3} \bar{h}^{-1} v^3 \int c(\theta) e^{\theta x} e^{|\theta|u} dG(\theta) \leq l_3^2 u^{-3} v^3$. It is easy to see that $\sigma_n^2(x) > l_2^2$. We prove (iii) next. From Lemma A.1, for $i = 0$ or 1 , $|K_{iv}(t)| \leq kv^{10}36^v$. Also note that $|K_{iv}(t)| = 0$ if $|t| > 1$. Then

$$|K_{iv}((y-x)/u)/h(y)| I_{[c_{1n} \leq x \leq c_{2n}]} \leq kv^{10}36^v/h(y) I_{[c_{1n} \leq y \leq c_{2n}+u]} \leq kv^{10}36^v u^{-1}.$$

For $x \in [c_{1n}, c_{2n}]$, $E[|Z_{jn}(x)|^3] \leq 2kv^{10}36^v u^{-1} E[Z_{jn}^2(x)] \leq l_4 v^{13} 36^v u^{-6}$. The proof of Lemma 5.3 is complete.

Proof of Lemma 5.4. From Lemma 5.3, we have $|w_n(x) - w(x)| \leq 1/\sqrt{n}$ for all $x \in [c_{1n}, c_{2n}]$. If $w(x) > d_n$ and n is large,

$$\frac{w_n(x)}{w(x)} \geq \frac{w(x) - d_n + d_n - |w_n(x) - w(x)|}{w(x) - d_n + d_n} \geq \frac{d_n - |w_n(x) - w(x)|}{d_n} \geq \frac{1}{2}.$$

Similarly, we can prove that $w(x) < -d_n \implies w_n(x) \leq w(x)/2$.

Acknowledgement

This research was supported in part by US Army Research Office, Grant DAAD19-00-1-0502 at Purdue University. The authors thank the Editor, the associate editor, and a referee for helpful comments which improved the presentation of the paper.

References

- Brown, L. D., Cohen, A. and Strawderman W. E. (1976). A complete class theorem for strict monotone likelihood ratio with applications. *Ann. Statist.* **4**, 712-722.
- Gasser, T., Muller, H-G and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47**, 238-252.
- Gupta, S. S. and Li, J. (1999). An empirical Bayes procedure for selecting good populations in positive exponential family. Technical Report #99-25, Department of Statistics, Purdue University.
- Gupta, S. S. and Li, J. (2001a). Optimal rate of convergence of monotone empirical Bayes tests for a normal mean. Technical Report #01-05, Department of Statistics, Purdue University.
- Gupta, S. S. and Li, J. (2001b). Empirical Bayes tests based on kernel sequence estimation. Technical Report #01-03, Department of Statistics, Purdue University.
- 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.
- Karunamuni, R. J. (1996). Optimal rates of convergence of monotone empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **43**, 934-947.
- Karunamuni, R. J. and Yang, H. (1995). On convergence rates of monotone empirical Bayes tests for the continuous one-parameter exponential family. *Statist. Decisions* **13**, 181-192.
- Liang, T. (2000). On an empirical Bayes test for a normal mean. *Ann. Statist.* **28**, 648-655.
- Petrov, V. V. (1995). *Limit Theorems of Probability Theory*. Clarendon Press, Oxford.
- Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, 157-163, University of California Press, Berkeley.
- Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1-20.
- Stijnen, T. (1985). On the asymptotic behaviour of empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **13**, 403-412.
- Van Houwelingen, J. C. (1976). Monotone empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **4**, 981-989.

Department of Statistics, Purdue University, 1399 Mathematical Sciences Building, West Lafayette, IN 47907-1399, U.S.A.

E-mail: li8@stat.purdue.edu

Department of Statistics, Purdue University, 1399 Mathematical Sciences Building, West Lafayette, IN 47907-1399, U.S.A.

E-mail: sgupta@stat.purdue.edu

(Received February 2001; accepted March 2002)