The Annals of Statistics
1996, Vol. 24, No. 1, 212–231

OPTIMAL RATES OF CONVERGENCE OF EMPIRICAL BAYES TESTS FOR THE CONTINUOUS ONE-PARAMETER EXPONENTIAL FAMILY[1]

BY ROHANA J. KARUNAMUNI

University of Alberta

The empirical Bayes linear loss two-action problem in the continuous one-parameter exponential family is studied. Previous results on this problem construct empirical Bayes tests via kernel density estimates. They also obtain upper bounds for the unconditional regret at some prior distribution. In this paper, we discuss the general question of how difficult the above empirical Bayes problem is, and why empirical Bayes rules based on kernel density estimates are useful. Asymptotic minimax-type lower bounds are obtained for the unconditional regret, and empirical Bayes rules based on kernel density estimates are shown to possess a certain optimal asymptotic minimax property.

1. Introduction. We investigate the following component decision problem: let $\theta \sim G$ and consider testing H_0: $\theta \le \theta_0$ against H_1: $\theta > \theta_0$ based on an observation X, with X, given θ, being distributed according to the exponential family

$$f(x|\theta) = m(x)h(\theta)e^{x\theta}, \qquad -\infty \le a < x < b \le \infty, \tag{1.1}$$

where m is positive on (a, b). 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. The parameter θ is distributed according to a completely unknown prior distribution G on the natural parameter space $\Omega = \{\theta\colon\ (h(\theta))^{-1} = \int m(x)e^{x\theta}\,dx < \infty\}$.

We study empirical Bayes (EB) tests for the above problem when a sequence of past observations is available. Let $X_1, \dots, X_n$ denote the observations from n independent past experiences: $X_1, \dots, X_n$ are i.i.d. random variables with (marginal) density $f_G(x) = \int m(x)e^{x\theta}h(\theta)\,dG(\theta)$. Let X denote the observation in the present experience. Then the *conditional Bayes risk* of an EB test ϕ_n is defined as $\widehat{R}(G, \phi_n) = E[L(\theta, \phi_n(X_1, \dots, X_n; X))|X_1, \dots, X_n]$, and the *unconditional Bayes risk* is defined as $R(G, \phi_n) = E\widehat{R}(G, \phi_n)$. The minimal attainable risk, or the Bayes risk envelope, which is achieved by a Bayes test, is denoted by $R(G)$. Then the *conditional regret* (excess risk) is defined by $\widehat{\Delta}_n = \widehat{R}(G, \phi_n) - R(G)$, and the *unconditional regret* is defined by $\Delta_n = E\widehat{\Delta}_n$. An EB rule ϕ_n is said to be asymptotically optimal [Robbins (1956, 1964)] w.r.t. G if $\lim_{n\to\infty} \Delta_n = 0$.

Received September 1992; revised September 1995.

[1]This research was supported in part by the Natural Sciences and Engineering Research Council of Canada Grant GP7987.

AMS 1991 *subject classifications*. Primary 62C12; secondary 62F03, 62C20.

Key words and phrases. Empirical Bayes, monotone tests, asymptotically optimal, rates of convergence.

Johns and Van Ryzin (1972) constructed asymptotically optimal EB tests for the above component problem and investigated upper bounds for the corresponding unconditional regret. See also Yu (1970) and Karunamuni (1989). Van Houwelingen (1976) showed that the EB tests of Johns and Van Ryzin can be readily improved if the monotonicity of the problem is used. The upper bound on the unconditional regret of his EB tests is of the form $O(n^{-2r/(2r+3)}\log^2 n)$ asymptotically, where $r \geq 1$ is an integer. Stijnen (1985) studied both monotone and nonmonotone EB tests and obtained limiting distributions of the corresponding conditional regrets. Recently, Karunamuni and Yang (1995) also studied a monotone EB rule and established the limiting distribution of the corresponding conditional regret. They were able to obtain an improved upper bound (i.e., a faster rate of convergence) of the regret assuming some auxiliary information on the prior distribution.

In all of the above work, the particular EB rules that have been investigated are based on kernel estimates of the density and its derivatives. Furthermore, upper bounds on the corresponding regrets have been obtained only for some specific prior distributions. Also, more general issues relating to the inherent difficulty of the problem in general have not been discussed. Clearly, it is of theoretical and practical interest to ask the following questions. How well can the EB testing problem described above be solved by any procedure (preferably by a monotone procedure)? In terms of rates of convergence what are the best EB testing rules? What are the optimal rates of convergence? Why use EB testing rules based on kernel density estimates? The purpose of this paper is to attempt to answer these questions systematically. In a decision-theoretic framework, we can formulate these questions as an asymptotic minimax problem. The more general question of how well any (monotone) EB rule can be performed is discussed in Section 2. Specifically, a lower bound on the unconditional regret is obtained in an asymptotic minimax sense. The best achievable minimax rates are established. Section 3 considers the question of why one should use EB rules based on kernel estimates. This section specifically focuses on a certain optimal asymptotic minimax property of such EB rules: they achieve the optimal minimax rate. The optimal rates of convergence are established for two specific distributions of the exponential family (1.1), namely, the normal $(\theta, 1)$ family and the scale exponential family.

2. Minimax lower bounds on the regret. For convenience, the notations Ω and (a, b) under the integral signs that follow are suppressed whenever they are clear from the context. Further, we assume that Ω is either $(-\infty, \infty)$ or $(0, \infty)$. It is easy to show that a Bayes test (w.r.t. G) for the component problem described in Section 1 is of the form [see Johns and Van Ryzin (1972) and van Houwelingen (1976)]

$$\phi_G(x) = \begin{cases} 1, & \text{if } l_G(x) > 0, \\ 0, & \text{if } l_G(x) \leq 0, \end{cases}$$

where

$$
\begin{aligned}
l_G(x) &= \int \{L(\theta, 0) - L(\theta, 1)\} f(x|\theta)\, dG(\theta) \\
&= \int (\theta - \theta_0) f(x|\theta)\, dG(\theta),
\end{aligned} \tag{2.1}
$$

or, equivalently, ϕ_G can be written as

$$
\phi_G(x) = \begin{cases} 1, & \text{if } E_G(\theta|X = x) > \theta_0, \\ 0, & \text{if } E_G(\theta|X = x) \le \theta_0. \end{cases} \tag{2.2}
$$

We tacitly assume that $E_G(\theta|X = x)$ is well defined. This will be the case in applications of the theory to the two-action problem in exponential families with linear losses. Of course, $E_G|\theta| < \infty$ is a sufficient condition. Since the class $\{f(x|\theta)$: $f(x|\theta)$ is given by (1.1)$\}$ has monotone likelihood ratio in x, $E_G(\theta|X = x)$ is nondecreasing on (a, b). Hence the rule (2.2) is a monotone decision rule [see Berger (1985), Section 8.4, for the definition and some results for monotone decision rules]. To avoid degeneracy, we shall assume that

$$
\lim_{x \downarrow a} E_G(\theta|X = x) < \theta_0 < \lim_{x \uparrow b} E_G(\theta|X = x). \tag{2.3}
$$

Let $\mathscr{G}$ denote the class of all prior distributions G on Ω such that (2.3) is satisfied. This notation will be employed throughout the remainder of the paper without further mention. The following consequences of (2.3) can easily be verified:

$$
\begin{aligned}
&\text{(i) } G \text{ is nondegenerate;} \\
&\text{(ii) } E_G(\theta|X = x) \text{ is strictly increasing on } (a, b); \\
&\text{(iii) there exists a unique } c_G \in (a, b) \text{ such that} \\
&\qquad E_G(\theta|X = c_G) = \theta_0.
\end{aligned} \tag{2.4}
$$

Hence the Bayes rule (2.2) can be written as

$$
\phi_G(x) = \begin{cases} 1, & \text{if } x > c_G, \\ 0, & \text{if } x \le c_G. \end{cases} \tag{2.5}
$$

Let $R(G)$ denote the Bayes envelope value of the component w.r.t. G. Then $R(G) = E[L(\theta, \phi_G(X))]$, where ϕ_G is given by (2.5).

Now consider the EB problem when G is completely unknown. Let $X_1, \ldots, X_n$ denote observations from n repetitions: $X_1, \ldots, X_n$ are assumed i.i.d. with (marginal) density

$$
f_G(x) = m(x) \int e^{x\theta} h(\theta)\, dG(\theta). \tag{2.6}
$$

Then, motivated by (2.5), EB rules can be constructed by defining

$$
\psi_n(X_1, \ldots, X_n; x) = \begin{cases} 1, & \text{if } x > c_n, \\ 0, & \text{if } x \le c_n, \end{cases} \tag{2.7}
$$

where $c_n = c_n(X_1, \dots, X_n)$ denotes any estimate of c_G. It is easy to show that the conditional regret $\widehat{R}(G, \psi_n) - R(G)$ of EB rules of the form (2.7) satisfy [van Houwelingen (1976)]

$$0 \le \widehat{R}(G, \psi_n) - R(G) = \int_{c_G}^{c_n} l_G(x)\, dx = S_G(c_n), \tag{2.8}$$

where $l_G(x)$ is given by (2.1) and $S_G(y)$ is defined by

$$S_G(y) = \int_{c_G}^{y} l_G(x)\, dx, \qquad a < y < b.$$

Note that $l_G(x) > 0$ if $x > c_G$ and $l_G(x) \le 0$ if $x \le c_G$. If the second derivative of $m(\cdot)$ [see (1.1)] exists, then $S_G(\cdot)$ has the following properties:

(2.9)

(i) $0 = S_G(c_G) \le S_G(y) \le \int_a^b |l_G(x)|\, dx \le E_G|\theta| + \theta_0$;

(ii) S_G has second derivative on (a, b) and $|S_G^{(2)}|$ is bounded on each interval (a_1, b_1) with $a < a_1 < b_1 < b$;

(iii) $S_G^{(1)}(c_G) = 0$;

(iv) for any $\varepsilon > 0$, there exists $\rho_G > 0$ s.t. $S_G^{(2)}(y) \ge \rho_G$ for all $y \in (c_G - \varepsilon, c_G + \varepsilon)$.

Then, by Taylor expansion of order 2 of $S_G(y)$ about c_G, we obtain from (2.8) and (2.9) that

$$\widehat{R}(G, \psi_n) - R(G) = S_G^{(2)}(c_n^*)(c_n - c_G)^2, \tag{2.10}$$

where c_n^* is an intermediate value between c_n and c_G. Therefore, if c_n lies in a neighborhood of c_G, the rate of convergence of $\widehat{R}(G, \psi_n) - R(G)$ is determined by that of $(c_n - c_G)^2$. If $|c_n - c_G| \ge \varepsilon$ for some $\varepsilon > 0$ and $S_G^{(2)}(c_n^*) \ne 0$ asymptotically, then clearly the corresponding EB rule is not asymptotically optimal, though the lower-bound result (2.19) below will be trivially satisfied. This is the situation with the cases $c_n = \pm\infty$. However, EB rules which are not asymptotically optimal are practically less attractive. Therefore, in this paper we study the following class of asymptotically optimal EB rules: for each $G \in \mathscr{G}$, define

$$\mathscr{F}_G = \{\psi_n\colon \psi_n \text{ is defined by (2.7) with an estimator } c_n \text{ such that } c_n \to c_G \text{ as } n \to \infty, \text{ w.p.1}\},$$

and we let $\mathscr{F} = \cup_{G \in \mathscr{G}} \mathscr{F}_G$ denote the class of EB rules of interest.

A general method of constructing desirable c_n's is as follows: let f_n and $f_n^{(1)}$ denote any estimators of f_G and $f_G^{(1)}$ (the first derivative of f_G), respectively, where f_G is given by (2.6). Let

$$V_n(x) = f_n^{(1)}(x) m(x) - f_n(x) m^{(1)}(x) - \theta_0 f_n(x) m(x). \tag{2.11}$$

Define an estimator c_n of c_G by

$$c_n = \min\{x\colon\ V_n(x) = 0,\ a < x < b\}. \tag{2.12}$$

[In applications, any finite solution of the equation $V_n(x) = 0$ could be employed.] Note that (2.12) is motivated by the fact that c_G is the unique solution of the equation $V_{f_G}(x) = 0$, where

$$V_{f_G}(x) = f_G^{(1)}(x)m(x) - f_G(x)m^{(1)}(x) - \theta_0 f_G(x)m(x); \tag{2.13}$$

see (2.4)(iii). With further conditions, convergence of c_n to c_G can be established; see Section 3 for such a situation.

For the main results of this paper that concern the asymptotic minimax lower bounds on (2.10), the following construction is essential: let G_0 be a distribution function in $\mathscr{G}$ with density function g_0 satisfying $g_0(0) > 0$ and $g_0(\theta) \neq 0$ for all $\theta \in \Omega$. In the case $\Omega = (0, \infty)$, the condition $g_0(0) > 0$ is replaced by $g_0(0^+) > 0$. Let f_{G_0} denote the corresponding marginal density of X, that is,

$$f_{G_0}(x) = \int m(x)e^{x\theta}h(\theta)\,dG_0(\theta). \tag{2.14}$$

For brevity, we denote f_{G_0} by f_0 in what follows. Let $\{\delta_n\}$ be a sequence of positive numbers such that $\delta_n \to 0$ as $n \to \infty$. For some positive number k, define the function g_n on Ω by

$$g_n(\theta) = g_0(\theta) + \delta_n^k H(\theta\delta_n), \tag{2.15}$$

where H is a bounded continuous function on Ω satisfying $\int_\Omega H(\theta)\,d\theta = 0$. Then, by suitable choice of (the tail of) H and g_0 such that $g_n(\theta) \geq 0$ for small δ_n, g_n will be a density function on Ω. Let G_n denote its distribution function. Then $G_n \in \mathscr{G}$. Let f_n^* denote the marginal density of X w.r.t. G_n, that is,

$$f_n^*(x) = f_{G_n}(x) = \int m(x)e^{x\theta}h(\theta)\,dG_n(\theta). \tag{2.16}$$

The pair (f_0, f_n^*) is known as the "least-favorable" pair (when G_0 and H are properly chosen) in the class of density functions of interest in this context [Donoho and Liu (1991a, b)].

THEOREM 2.1. *Suppose that $m^{(2)}$ exists and $m^{(i)}$, $i = 0, 1, 2$, $m^{(0)} \equiv m$, are bounded on (a, b). Further, suppose that $m(c_0) \neq 0$, where c_0 is the solution of the equation $V_{f_0}(x) = 0$ with V_{f_0} and f_0 defined by* (2.13) *and* (2.14), *respectively. Let $f_{n,0}^{(i)}$ denote any estimator of $f_0^{(i)}$ (ith derivative of f_0, $f^{(0)} \equiv f$) such that, with probability* 1, $\lim_{n\to\infty} \sup_x |f_{n,0}^{(i)}(x) - f_0^{(i)}(x)| = 0$ *for* $i = 0, 1, 2$. *Suppose that the functions H, g_0 and the sequence $\{\delta_n\}$ in* (2.15) *are chosen such that g_n is nonnegative and, for some real numbers $k > j > 0$,*

$$\int (\theta - \theta_0)e^{c_0\theta}h(\theta)H(\theta\delta_n)\,d\theta = O(\delta_n^{-j}) \quad \text{as } n \to \infty, \tag{2.17}$$

and

$$\int \left(f_n^*(x) - f_0(x)\right)^2 f_0^{-1}(x)\,dx \le \frac{d_1}{n} \tag{2.18}$$

for some constant $d_1 > 0$, *where* f_n^* *is given by* (2.16). *Then for any EB rule* $\psi_n \in \mathscr{F}$ [*for the definition of* $\mathscr{F}$, *see circa* (2.10) *and* (2.11)], *we have*

$$\sup_{G \in \mathscr{G}^*} E\left(\widehat{R}(G, \psi_n) - R(G)\right) > d_2 \delta_n^{2(k-j)} \tag{2.19}$$

for all sufficiently large n *and some positive constant* $d_2 > 0$ (d_2 *is also independent of* ψ_n), *where* $\mathscr{G}^*$ *denotes any subset of* $\mathscr{G}$ *such that* $\{G_0, G_n\} \subseteq \mathscr{G}^*$.

The explicit forms of δ_n, which is determined by the restriction (2.18), are exhibited below in two examples. The condition (2.17) can usually be verified for any sequence $\delta_n \to 0$.

REMARK 2.1. Theorem 2.1 above sheds some light on the general question of how much EB rules can be improved whenever G is completely unknown. The inequality (2.19) essentially implies that

$$\liminf_{n \to \infty} \inf_{\psi_n \in \mathscr{F}} \sup_{G \in \mathscr{G}^*} \delta_n^{-2(k-j)} E\left(\widehat{R}(G, \psi_n) - R(G)\right) > 0.$$

The preceding asymptotic minimax result explains that no sequence of EB rules in $\mathscr{F}$ has regret that converges to 0 faster than $\delta_n^{2(k-j)}$ uniformly over the class $\mathscr{G}^*$. Thus, $\delta_n^{2(k-j)}$ is a lower bound on the best achievable minimax rate for the class $\mathscr{F}$. In the next section, we shall show, in fact, that $\delta_n^{2(k-j)}$ can be attained by certain EB rules, thus proving the fact that $\delta_n^{2(k-j)}$ is the optimal minimax rate of convergence for the class $\mathscr{F}$. If, however, $\mathscr{G}^*$ is partially known, then the rates can be improved. For example, if the form of each $G \in \mathscr{G}^*$ is known and only the hyperparameters are unknown, then, of course, the optimal rate of convergence is n^{-1}. When G is completely unknown, the problem becomes an "infinite-parameter" or nonparametric problem. Slow convergence rates, compared to n^{-1}, in nonparametric problems are well known to occur.

REMARK 2.2. The left-hand side of (2.18) is known as the "χ^2-distance" in the literature on rates of convergence. But other distances can also be implemented; see Donoho and Liu (1991a, b). The inequality (2.19) gives a lower bound on the rate of convergence of unconditional regrets of EB rules in $\mathscr{F}$. However, a similar lower bound in terms of convergence in probability can be established for conditional regrets also.

The basic idea in the proof of Theorem 2.1 is borrowed from Donoho and Liu (1991a, b) where they showed that the difficulty in establishing lower bounds on the asymptotic minimax risk in the "full infinite-dimensional problem" is no greater than that of the "hardest one-dimensional subproblem" in many cases.

Their arguments are heavily based on the use of Le Cam's (1972) theory [see also Hájek (1972)] of convergence of experiments and asymptotic efficiency.

PROOF OF THEOREM 2.1. Define a class $\mathscr{U}$ by

$$\mathscr{U} = \{f_G\colon\ f_G \text{ is defined by (2.6)},\ G \in \mathscr{G}\},$$

that is, $\mathscr{U}$ is the class of marginal densities of X generated by $\mathscr{G}$ for the exponential family (1.1). Then f_0 and f_n^* belong to $\mathscr{U}$, where f_0 and f_n^* are defined by (2.14) and (2.16), respectively. Let $x_0 \in (a, b)$ be a fixed point. For $f \in \mathscr{U}$, define

$$T(f) = V_f(x_0) = f^{(1)}(x_0)m(x_0) - f(x_0)m^{(1)}(x_0) - \theta_0 m(x_0)f(x_0)$$

[a functional of f of interest to us—compare with (2.13)]. We suppose that $m(x_0) \neq 0$. Let T_n denote any estimator of $T(f)$ based on a random sample of size n from f. If δ_n is chosen such that (2.18) holds, then it is proved by Donoho and Liu (1991a, b) that a lower bound for estimating $T(f)$ by any estimator T_n is

$$\sup_{f\in\{f_0, f_n^*\}} P_f\{|T_n - T(f)| > |T(f_0) - T(f_n^*)|/2\} > d \tag{2.20}$$

for some positive constant d. A direct consequence of (2.20) is that

$$\sup_{f\in\{f_0, f_n^*\}} E_f\big(T_n - T(f)\big)^2 > \frac{d}{4}\,|T(f_0) - T(f_n^*)|^2, \tag{2.21}$$

and hence

$$\sup_{f\in\mathscr{U}^*} E_f\big(T_n - T(f)\big)^2 > \frac{d}{4}\,|T(f_0) - T(f_n^*)|^2 \tag{2.22}$$

for any subset $\mathscr{U}^*$ of $\mathscr{U}$ such that $\{f_0, f_n^*\} \subseteq \mathscr{U}^*$. In other words, the order of $|T(f_0) - T(f_n^*)|$ provides a lower bound for estimating $T(f)$ by any estimator based on a sample of size n. It is proved below that $|T(f_0) - T(f_n^*)|$ is of order $O(\delta_n^{k-j})$ under the assumptions of the theorem. Then one only needs to find δ_n as large as possible such that (2.18) holds. From (2.15) and (2.16), we obtain

$$f_n^*(x) = f_0(x) + \delta_n^k m(x)\int e^{x\theta}h(\theta)H(\theta\delta_n)\,d\theta, \tag{2.23}$$

and hence [using Theorem 2.9 of Lehmann (1959)],

$$T(f_n^*) - T(f_0) = \delta_n^k(m(x_0))^2\int(\theta - \theta_0)e^{x_0\theta}h(\theta)H(\theta\delta_n)\,d\theta. \tag{2.24}$$

Now take $x_0 = c_0$ and $T_n = V_n(c_0)$ [where c_0 is as defined in the theorem and V_n is defined by (2.11)], and then use (2.17), (2.21) and (2.24) to obtain

$$\sup_{f\in\{f_0, f_n^*\}} E_f\big(V_n(c_0) - V_f(c_0)\big)^2 > d_2\delta_n^{2(k-j)} \tag{2.25}$$

for some finite constant $d_2 > 0$. Observe that $V_{f_0}^{(1)}(c_0) < 0$. To see this, note that $w(x) = E_{G_0}(\theta|X = x) - \theta_0 = V_{f_0}(x)/f_{G_0}(x)m(x)$ is a strictly increasing function of $x \in (a, b)$. Thus, $w^{(1)}(c_0) > 0$. But $w^{(1)}(c_0) = -V_{f_0}^{(1)}(c_0)/f_{G_0}(c_0)m(c_0)$ since $V_{f_0}(c_0) = 0$. Therefore, $V_{f_0}^{(1)}(c_0) < 0$ since $m(c_0)f_{G_0}(c_0) > 0$. Now, using the mean value theorem, one obtains

$$V_n(c_0) = V_n(c_n) + (c_n - c_0)V_n^{(1)}(c_n^*), \tag{2.26}$$

where c_n^* is an intermediate value between c_0 and c_n. The assumptions that $\sup_x |f_{n,0}^{(i)}(x) - f_0^{(i)}(x)| \to 0$ as $n \to \infty$, w.p.1, and $m^{(i)}$ are bounded for $i = 0, 1, 2$ yield $\sup_x |V_{n,0}^{(j)}(x) - V_{f_0}^{(j)}(x)| \to 0$ as $n \to \infty$, w.p.1, for $j = 0, 1$, where $V_{n,0}$ is defined by (2.11) with $f_n^{(i)}$ replaced by $f_{n,0}^{(i)}$, $i = 0, 1$, and $V^{(1)}$ denotes the first derivative of V. But $|V_{n,0}(c_n) - V_{f_0}(c_n)| \le \sup_x |V_{n,0}(x) - V_{f_0}(x)|$. Therefore, from the preceding results we obtain $V_{n,0}(c_n) = V_{f_0}(c_n) + o(1)$ as $n \to \infty$, w.p.1. Furthermore, as $n \to \infty$, $V_{f_0}(c_n) = V_{f_0}(c_0) + o(1)$ (since $c_n \to c_0$ and V_{f_0} is continuous) and $V_{n,0}^{(1)}(c_n^*) = V_{f_0}^{(1)}(c_n^*) + o(1) = V_{f_0}^{(1)}(c_0) + o(1)$, w.p.1. The latter statement follows from $|V_{n,0}^{(1)}(c_n^*) - V_{f_0}^{(1)}(c_n^*)| \le \sup_x |V_{n,0}^{(1)}(x) - V_{f_0}^{(1)}(x)| \to 0$ and $c_n^* \to c_0$ as $n \to \infty$, w.p.1.

The above results together with a result similar to (2.26) for $V_{n,0}$ imply that $V_{n,0}(c_0) = V_{f_0}(c_0) + (c_n - c_0)V_{f_0}^{(1)}(c_0) + o(1)$ as $n \to \infty$, w.p.1. Similarly, it is easy to show that $V_{n,*}(c_0) = V_{f_n^*}(c_0) + (c_n - c_0)V_{f_0}^{(1)}(c_0) + o(1)$ as $n \to \infty$, w.p.1, where $V_{n,*}$ is defined by (2.11) with $f_n^{(i)}(x)$ replaced by $f_{n,*}^{(i)}$, $i = 0, 1$, where $f_{n,*}(x) = f_{n,0}(x) + \delta_n^k m(x) \int e^{x\theta} h(\theta) H(\theta S_n)\, d\theta$ [an estimator of f_n^* defined by (2.23)]. Now, combining the above results and (2.25), we obtain

$$\sup_{f \in \{f_0, f_n^*\}} E_f(c_n - c_0)^2 > d_3 \delta_n^{2(k-j)} \tag{2.27}$$

for some finite constant $d_3 > 0$ and any c_n such that $c_n \to c_0$ as $n \to \infty$, w.p.1. Then from (2.10) and (2.27) we obtain that, for any $\psi_n \in \mathscr{F}_G$ with $G \in \{G_0, G_n\}$,

$$\sup_{G \in \{G_0, G_n\}} E\big(\widehat{R}(G, \psi_n) - R(G)\big) \ge d_4 \delta_n^{2(k-j)} \tag{2.28}$$

for some finite constant $d_4 > 0$, since $S_G^{(2)}(y) > \rho_0 > 0$ in a neighborhood of c_0 for $G \in \{G_0, G_n\}$. [For the definition of G_0 and G_n, see circa (2.14) and (2.15), respectively.] The inequality (2.19) now follows from (2.28), since $\sup_{G \in \mathscr{G}^*} E(\cdot) \ge \sup_{G \in \{G_0, G_n\}} E(\cdot)$. This completes the proof. $\square$

We now formally discuss two examples and exhibit the form of δ_n (as a function of n) in each case.

EXAMPLE 2.1 [Normal $(\theta, 1)$ family]. Consider the exponential family in (1.1) with $m(x) = \exp(-x^2/2)$ and $h(\theta) = (2\pi)^{-1/2} \exp(-\theta^2/2)$; that is, for each $-\infty < \theta < \infty$, $f(x|\theta) = (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)$, $-\infty < x < \infty$. Then $\Omega = (-\infty, \infty)$. Consider testing H_0: $\theta \le 0$ against H_1: $\theta > 0$, so that $\theta_0 = 0$. Let G_0 be the prior distribution on $(-\infty,\infty)$ with density function g_0 [see circa (2.14)] defined by

$$g_0(\theta) = \rho_l/(1+\theta^2)^l, \qquad -\infty < \theta < \infty, \tag{2.29}$$

for some $1.5 > l > 1$, where ρ_l is a constant such that $\rho_l(1+\theta^2)^{-l}$ is a density function. Then $G_0 \in \mathscr{G}$ and clearly $g_0(\theta_0) = g_0(0) > 0$. Furthermore, it is easy to show that $V_{f_0}(0) = 0$ using (2.29) in the expressions (2.13) and (2.14). That is, $c_0 = 0$ is this case.

COROLLARY 2.1. *For the normal family in Example* 2.1 *and for the prior* G_0 *with density* (2.29) *and for an appropriate function H defined later in the proof*, (2.17) *holds with* $j = 4$ *for any* $\delta_n \to 0$, *and* (2.18) *holds with* $\delta_n = n^{-1/(2k-4)}$. *Thus, the lower bound in* (2.19) *is of order* $n^{-2(k-4)/(2k-4)}$, $k > 4$.

To prove the above corollary, we require the following lemma.

LEMMA 2.1. *Suppose that F is a cumulative distribution function. Let l be a real number s.t.* $1.5 > l > 1$. *Then the density*

$$\tilde{g}(x) = \int_{-\infty}^{\infty} \frac{\rho_l}{\left(1+(x-y)^2\right)^l}\, dF(y)$$

satisfies $\tilde{g}(x) \ge \beta |x|^{-2l}$ *as* $|x| \to \infty$ *for some positive constant* β, *where* ρ_l *is a constant s.t.* $\rho_l(1+x^2)^{-l}$ *is a density function on* $(-\infty,\infty)$.

PROOF OF COROLLARY 2.1. Since $m(x) = \exp(-x^2/2)$, the second derivative of m clearly exists and $m(c_0) = m(0) \ne 0$. To examine (2.17) and (2.18), we first construct an appropriate function H to use in (2.15).

Define $w(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, $-\infty < x < \infty$. Let $\phi_w(t)$ denote its characteristic function. Then $\phi_w(t) = \exp(-t^2/2)$. Let $H(x) = \rho_l[w(x) - w(x+1)]$. Then $\int_{-\infty}^{\infty} H(x)\,dx = 0$ and $\int_{-\infty}^{\infty} xH(x)\,dx \ne 0$. Furthermore, its Fourier transformation is $\phi_H(t) = \rho_l(1-e^{-it})\phi_w(t)$. Also $\int_{-\infty}^{\infty} |\phi_H(t)|^2\,dt < \infty$. With the preceding choice of H and g_0 defined by (2.29), it is easy to see that $g_n \ge 0$ for all large n, since $g_0(x) \ge |H(x)|$ as $|x| \to \infty$, where g_n is given by (2.15).

We now verify condition (2.17). Since $\theta_0 = 0$ and $c_0 = 0$, by a change of variable ($\theta\delta_n = t$) and by rearranging terms, (2.17) becomes

$$
\begin{aligned}
&\rho_l(2\pi)^{-1/2}\int_{-\infty}^{\infty}\theta e^{-\theta^2/2}H(\theta\delta_n)\,d\theta \\
&\quad = \rho_l(2\pi)^{-1/2}\delta_n^{-2}\biggl\{\int_{-\infty}^{\infty} tw(t)\biggl[\exp\biggl(\frac{-(t\delta_n^{-1})^2}{2}\biggr) \\
&\qquad\qquad - \exp\biggl(\frac{-((t-1)\delta_n^{-1})^2}{2}\biggr)\biggr]\,dt \\
&\qquad\qquad + \int_{-\infty}^{\infty}\exp\biggl(\frac{-((t-1)\delta_n^{-1})^2}{2}\biggr)w(t)\,dt\biggr\}.
\end{aligned}
\tag{2.30}
$$

Using the dominated convergence theorem (DCT), it is easy to show that, as $n \to \infty$,

$$
\int_{-\infty}^{\infty}\exp\biggl(\frac{-((t-1)\delta_n^{-1})^2}{2}\biggr)w(t)\,dt = o(\delta_n^{-2}). \tag{2.31}
$$

Expanding $\exp(-(t\delta_n^{-1})^2/2)$ and $\exp(-((t-1)\delta_n^{-1})^2/2)$ separately and then taking the difference, we obtain

$$
\int_{-\infty}^{\infty} tw(t)\biggl[\exp\biggl(\frac{-(t\delta_n^{-1})^2}{2}\biggr) - \exp\biggl(\frac{-((t-1)\delta_n^{-1})^2}{2}\biggr)\biggr]\,dt = 2\delta_n^{-2} + o(\delta_n^{-2}) \tag{2.32}
$$

for all sufficiently large n, since all moments of $w(x)$ are finite. From (2.30), (2.31) and (2.32), we then have

$$
\rho_l(2\pi)^{-1/2}\int_{-\infty}^{\infty}\theta\exp\biggl(\frac{-\theta^2}{2}\biggr)H(\theta\delta_n)\,d\theta = 2\rho_l(2\pi)^{-1/2}\delta_n^{-4} + o(\delta_n^{-4})
$$

for all sufficiently large n. Hence, (2.17) is satisfied with $j = 4$.

Now consider (2.18). Under the hypotheses of Example 2.1, observe that (2.18) is equivalent to

$$
\delta_n^{2k}\int\biggl((2\pi)^{-1/2}\int\exp\biggl(\frac{-(x-\theta)^2}{2}\biggr)H(\theta\delta_n)\,d\theta\biggr)^2 f_0^{-1}(x)\,dx \le \frac{d}{n}, \tag{2.33}
$$

where H is as defined in the beginning of the present proof and f_0 is given by (2.14) with prior density (2.29). Let $\beta_1 = \max_{x\in[-1,1]} f_0^{-1}(x)$. Then $0 < \beta_1 < \infty$.

By Parseval's identity for Fourier transforms, we obtain

$$
\begin{aligned}
I &= \int_{|x|\le 1}\left((2\pi)^{-1/2}\int_{-\infty}^{\infty}\exp\left(\frac{-(x-\theta)^2}{2}\right)H(\theta\delta_n)\,d\theta\right)^2 f_0^{-1}(x)\,dx \\
&\le \beta_1\int_{-\infty}^{\infty}\left((2\pi)^{-1/2}\int_{-\infty}^{\infty}\exp\left(\frac{-(x-\theta)^2}{2}\right)H(\theta\delta_n)\,d\theta\right)^2 dx \\
&= \beta_1\int\left|\delta_n^{-1}\exp\left(\frac{-t^2}{2}\right)\phi_H\left(\frac{t}{\delta_n}\right)\right|^2 dt \\
&= \beta_1\delta_n^{-1}\int\exp\left(\frac{-(t\delta_n)^2}{2}\right)|\phi_H(t)|^2\,dt \\
&= O(\delta_n^{-1}),
\end{aligned}
\tag{2.34}
$$

uniformly in δ_n, since $\int|\phi_H(t)|^2 dt < \infty$. By the Fourier inversion formula,

$$
\begin{aligned}
&(2\pi)^{-1/2}\int\exp\left(\frac{-(x-\theta)^2}{2}\right)H(\theta\delta_n)\,d\theta \\
&\qquad = -(2\pi)^{-1/2}\int e^{-itx}\delta_n^{-1}\exp\left(\frac{-t^2}{2}\right)\phi_H\left(\frac{t}{\delta_n}\right)dt \\
&\qquad = \delta_n^{-1}(2\pi x^2)^{-1}\int e^{-itx}\frac{d^2}{dt^2}\left\{\exp\left(\frac{-t^2}{2}\right)\phi_H\left(\frac{t}{\delta_n}\right)\right\}dt,
\end{aligned}
\tag{2.35}
$$

where the last equality follows from the result that $\mathscr{T}(\psi^{(2)}) = -x^2\mathscr{T}(\psi)$ for any integrable function ψ, with $\mathscr{T}$ denoting the Fourier transformation operator. Using the integrability of ϕ_H, $\phi_H^{(1)}$ and $\phi_H^{(2)}$, it is easy to show that

$$
\int_{-\infty}^{\infty}\left|\frac{d^2}{dt^2}\left\{\exp\left(\frac{-t^2}{2}\right)\phi_H\left(\frac{t}{\delta_n}\right)\right\}\right|dt \le K\delta_n^{-1}
\tag{2.36}
$$

for some finite constant $K > 0$. Hence, from (2.35), (2.36) and Lemma 2.1, we have

$$
\begin{aligned}
II &= \int_{|x|>1}\left((2\pi)^{-1/2}\int\exp\left(\frac{-(x-\theta)^2}{2}\right)H(\theta\delta_n)\,d\theta\right)^2 f_0^{-1}(x)\,dx \\
&\le K^2\delta_n^{-4}\int_{|x|>1}(2\pi x^2)^{-2}f_0^{-1}(x)\,dx \\
&= O(\delta_n^{-4}).
\end{aligned}
\tag{2.37}
$$

In view of (2.34) and (2.37), we see that $\delta_n^{2k}\int((2\pi)^{-1/2}\int\exp(-(x-\theta)^2/2)\cdot H(\theta\delta_n)\,d\theta)^2 f_0^{-1}(x)\,dx$ is of order $O(n^{-1})$ by taking $\delta_n = n^{-1/(2k-4)}$. This completes the proof of Corollary 2.1. □

EXAMPLE 2.2 (Scale exponential family). Consider the exponential family in (1.1) with $m(x) = I_{\{x>0\}}(x)$ and $h(\theta) = \theta$; that is, for each $\theta > 0$, $f(x|\theta) = \theta e^{-x\theta}$ if $x > 0$, and 0 elsewhere. Consider testing H_0: $\theta \le \theta_0$ against H_1: $\theta >$

θ_0 for some $\theta_0 > 0$. The parameter space $\Omega = (0, \infty)$. Let G_0 be the prior distribution on Ω with density

$$g_0(\theta) = \widetilde{\rho}_l/(1+\theta^2)^l, \qquad \theta > 0, \tag{2.38}$$

for some $l > 1$, where $\widetilde{\rho}_l$ is a constant s.t. $\widetilde{\rho}_l(1+\theta^2)^{-l}$ is a density function on $(0, \infty)$. Then again $G_0 \in \mathscr{G}$ and clearly $g_0(\theta_0) > 0$. Let c_0 be the solution of $V_{f_0}(x) = 0$ in $(0, \infty)$, where V_{f_0} and f_0 are given by (2.13) and (2.14), respectively, with G_0 defined according to (2.38).

COROLLARY 2.2. *For the scale exponential family in Example* 2.2 *and for the prior G_0 with density* (2.38) *and for an appropriate function H defined later in the proof*, (2.17) *holds with $j = 4$ for any $\delta_n \to 0$, and* (2.18) *holds with $\delta_n = n^{-1/(2k-3)}$. Hence, the order of the lower bound in* (2.19) *is $n^{-2(k-4)/(2k-3)}$, $k > 4$.*

LEMMA 2.2. *Let $f_0(x) = \int_0^\infty \theta e^{-x\theta} g_0(\theta)\, d\theta$, $x > 0$, where $g_0(\theta)$ is given by* (2.38). *Then $f_0(x)$ satisfies $f_0(x) \geq \beta_3 x^{-2}$ as $x \to \infty$ for some finite constant $\beta_3 > 0$.*

PROOF OF COROLLARY 2.2. Since $m(x) = I_{\{x>0\}}(x)$ and $c_0 > 0$, clearly $m(c_0) > 0$. Again, to study (2.17) and (2.18), we first define an appropriate function H to employ in (2.15). For $p > 1$, define

$$\widetilde{w}(x) = \begin{cases} (\Gamma(p))^{-1} x^{p-1} e^{-x}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $H(x) = \widetilde{\rho}_l[\widetilde{w}(x) - \widetilde{w}(x-1)]$ for $x > 0$ and 0 elsewhere, where $\widetilde{\rho}_l$ is as defined in (2.38). Then $\int_0^\infty H(x)\, dx = 0$, $\int_0^\infty xH(x)\, dx \neq 0$ and $g_0(x) \geq |H(x)|$ as $x \to \infty$, where g_0 is given by (2.38).

Consider now condition (2.17) with the above specifications. Note that, under the hypotheses of Example 2.2, (2.17) reads as

$$\int_0^\infty \theta^2 \exp(-c_0\theta) H(\theta\delta_n)\, d\theta - \theta_0 \int_0^\infty \theta \exp(-c_0\theta) H(\theta\delta_n)\, d\theta = O(\delta_n^{-j}). \tag{2.39}$$

By a change of variables and rearrangement of terms,

$$\begin{aligned} &\text{the first term on the l.h.s. of (2.39)} \\ &\quad = \widetilde{\rho}_l \delta_n^{-3} \biggl\{ \int_0^\infty t^2 \widetilde{w}(t)[\exp(-c_0 t\delta_n^{-1}) - \exp(-c_0(t+1)\delta_n^{-1})]\, dt \\ &\qquad\qquad - 2\int_0^\infty t \exp(-c_0(t+1)\delta_n^{-1}) \widetilde{w}(t)\, dt \\ &\qquad\qquad - \int_0^\infty \exp(-c_0(t+1)\delta_n^{-1}) \widetilde{w}(t)\, dt \biggr\}. \end{aligned} \tag{2.40}$$

Using the DCT, it is easy to show that, as $n \to \infty$,

$$\begin{aligned} &\text{(i)} \int_0^\infty t \exp(-c_0(t+1)\delta_n^{-1}) \widetilde{w}(t)\, dt = o(1), \\ &\text{(ii)} \int_0^\infty \exp(-c_0(t+1)\delta_n^{-1}) \widetilde{w}(t)\, dt = o(1). \end{aligned} \tag{2.41}$$

Expanding $\exp(-c_0 t\delta_n^{-1})$ and $\exp(-c_0(t+1)\delta_n^{-1})$ separately and then taking the difference, one obtains

$$\begin{aligned}(2.42)\qquad &\int_0^\infty t^2\widetilde{w}(t)[\exp(-c_0 t\delta_n^{-1}) - \exp(-c_0(t+1)\delta_n^{-1})]\,dt \\ &\quad = \delta_n^{-1}c_0\int_0^\infty t^2\widetilde{w}(t)\,dt + o(\delta_n^{-1})\end{aligned}$$

for all sufficiently large n, since $\int_0^\infty t^i\widetilde{w}(t)\,dt < \infty$ for all $i \geq 0$. In view of (2.39) to (2.42), we have

$$\int_0^\infty \theta^2\exp(-c_0\theta)H(\theta\delta_n)\,d\theta = \delta_n^{-4}c_0\int_0^\infty t^2\widetilde{w}(t)\,dt + o(\delta_n^{-4})$$

for all sufficiently large n. Also, the second term on the left-hand side of (2.39)= $\delta_n^{-2}\int_0^\infty t\exp(-c_0 t\delta_n^{-1})H(t)\,dt = o(\delta_n^{-2})$ for all sufficiently large n. In view of the above results together with (2.39), we see that (2.17) is satisfied for $j = 4$.

Let us now examine condition (2.18). For the setup in Example 2.2, (2.18) is equivalent to

$$(2.43)\qquad \delta_n^{2k}\int_0^\infty\left(\int_0^\infty \theta e^{-x\theta}H(\theta\delta_n)\,d\theta\right)^2 f_0^{-1}(x)\,dx \leq \frac{d}{n},$$

where H is now as defined in the beginning of the present proof and f_0 is given by (2.14) with prior density (2.38). By a change of variable ($\theta\delta_n = t$) followed by another change of variable ($x\delta_n^{-1} = y$), (2.43) is equivalent to

$$\delta_n^{2k-3}\int_0^\infty\left(\int_0^\infty te^{-yt}H(t)\,dt\right)^2 f_0^{-1}(y\delta_n)\,dy \le \frac{d}{n}. \tag{2.44}$$

Since $H(x) = \widetilde{\rho}_l[\widetilde{w}(x) - \widetilde{w}(x-1)]$ for $x > 0$, an application of the Cauchy–Schwarz inequality followed by the moment inequality and some simplification yields

$$\begin{aligned}
&\left(\int_0^\infty te^{-yt}H(t)\,dt\right)^2\\
&\quad\le 2\widetilde{\rho}_l^2\int_0^\infty t^2e^{-2yt}\widetilde{w}(t)\,dt + 2\widetilde{\rho}_l^2\int_0^\infty (t+1)^2e^{-2y(t+1)}\widetilde{w}(t)\,dt\\
&\quad\le M_1\{(2y+1)^{-(p+2)}\\
&\qquad\quad + e^{-2y}[(2y+1)^{-(p+2)} + (2y+1)^{-(p+1)} + (2y+1)^{-p}]\}
\end{aligned}\tag{2.45}$$

for some finite constant $M_1 > 0$. Let $\beta_4 = \max_{0<x<1} f_0^{-1}(x)$. Then $0 < \beta_4 < \infty$. Therefore, using (2.45), we obtain

$$\begin{aligned}
J_1 &= \int_0^1\left(\int_0^\infty te^{-yt}H(t)\,dt\right)^2 f_0^{-1}(y\delta_n)\,dy\\
&\le \beta_4\int_0^1\left(\int_0^\infty te^{-yt}H(t)\,dt\right)^2 dy\\
&\le M_2\beta_4
\end{aligned}\tag{2.46}$$

for some finite constant $M_2 > 0$. Also, by Lemma 2.2 and (2.45),

$$\begin{aligned}
J_2 &= \int_1^\infty\left(\int_0^\infty te^{-yt}H(t)\,dt\right)^2 f_0^{-1}(y\delta_n)\,dy\\
&\le M_1\delta_n^2\int_1^\infty y^2\{(2y+1)^{-(p+2)} + e^{-2y}\}\,dy\\
&= O(\delta_n^2) = O(1).
\end{aligned}\tag{2.47}$$

Now, in view of (2.46) and (2.47), we conclude that the left-hand side of (2.44) is of order $O(n^{-1})$ by taking $\delta_n = n^{-1/(2k-3)}$, and the order of the lower bound in (2.19) is $n^{-2(k-4)/(2k-3)}$, $k > 4$. This completes the proof of Corollary 2.2. □

3. EB rules based on kernel density estimates. In this section, we study EB rules based on kernel estimates of density f and its derivatives $f^{(i)}$, $i = 1, 2$. We show below that EB rules of the form (2.7), with c_n obtained via kernel estimates of V_f [see (2.13)], achieve the rates established in Corollaries 2.1 and 2.2. In the following, we apply kernel function K and its νth derivative $K^{(\nu)}$ satisfying, for some positive integer $r(\geq \nu)$,

$$
\begin{aligned}
&\text{(i) support of } K^{(\nu)} = [-\tau, \tau], \qquad \tau > 0;\\
&\text{(ii) } K^{(j)}(\tau) = K^{(j)}(-\tau) = 0, \qquad j = 0, 1, \dots, \nu - 1;\\
&\text{(iii) } \int_{-\tau}^{\tau} K^{(\nu)}(x)x^j\,dx = \begin{cases} 0, & \text{if } j = 0, \dots, \nu-1, \nu+1, \dots, r-1,\\ (-1)^\nu \nu!, & \text{if } j = \nu;\end{cases}\\
&\text{(iv) } \int_{-\tau}^{\tau} K^{(\nu)}(x)x^r\,dx \neq 0.
\end{aligned}
\tag{3.1}
$$

Then the kernel estimate of $f^{(\nu)}(x)$ (νth derivative of f, $f^{(0)} \equiv f$) is defined by

$$
\widehat{f}^{(\nu)}(x) = \frac{1}{nh_n^{\nu+1}} \sum_{j=1}^{n} K^{(\nu)}\left(\frac{X_j - x}{h_n}\right),
\tag{3.2}
$$

where $X_1, \dots, X_n$ denote a sample from the density f and h_n ($h_n \to 0$ as $n \to \infty$) is the bandwidth.

In the following, it is assumed that unknown densities belong to the class

$$
\mathscr{C}_{B,r} = \left\{f\colon \sup_x |f^{(r)}(x)| \leq B\right\}
\tag{3.3}
$$

for some known finite constant $B > 0$. That is, rth derivatives of f are assumed bounded uniformly in $\mathscr{C}_{B,r}$ [this can be weakened to a Lipschitz condition on the $(r-1)$th derivative]. Some useful properties of $\widehat{f}^{(\nu)}$ are given in the following lemmas. Detailed proofs of Lemmas 3.1 and 3.2 can be found, for example, in Müller and Gasser (1979) and Singh (1977), respectively.

LEMMA 3.1. *For $r > \nu \geq 0$, let $f \in \mathscr{C}_{B,r}$. Then, for $\nu \geq 0$,*

$$
E\widehat{f}^{(\nu)}(x) - f^{(\nu)}(x) = h_n^{r-\nu} f^{(r)}(x) B_{r,\nu}(1 + o(1))
\tag{3.4}
$$

and

$$
\operatorname{var} \widehat{f}^{(\nu)}(x) = (nh_n^{2\nu+1})^{-1} f(x) A_\nu (1 + o(1)),
\tag{3.5}
$$

where

$$
B_{r,\nu} = \frac{(-1)^{r-\nu}}{(r-\nu)!} \int_{-\tau}^{\tau} K(t) t^{r-\nu}\,dt
\tag{3.6}
$$

and

$$
A_\nu = \int_{-\tau}^{\tau} (K^{(\nu)}(t))^2\,dt.
\tag{3.7}
$$

LEMMA 3.2. *Let $K^{(2)}$ be Lipschitz continuous on $[-\tau, \tau]$. Let $f \in \mathscr{C}_{B,r}$ with integer $r \geq 3$. Let $h_n = n^{-1/(2r+1)}$. Then, as $n \to \infty$, $\sup_x |\widehat{f}^{(\nu)}(x) - f^{(\nu)}(x)| = o(1)$, with probability 1, for $\nu = 0, 1, 2$.*

Lemmas 3.1 and 3.2 hold for a kernel function with a support on the interval (0, 1) as well; see, for example, Singh (1977). Such kernels have been used in the literature in order to estimate densities with left endpoints, such as those connected with Example 2.2.

For any known large finite number $A > 0$, define

$$\mathscr{G}_A = \{G \in \mathscr{G}\colon\ |c_G| \le A\}, \tag{3.8}$$

where c_G is the solution of the equation $V_{f_G}(x) = 0$ with V_{f_G} given by (2.13), $G \in \mathscr{G}$; that is, c_G satisfies (2.4)(iii). (The value of A can be chosen arbitrarily large as desired.)

Let $\varepsilon > 0$ be fixed. Since $|V^{(1)}_{f_G}(c_G)| > 0$ and $V^{(1)}_{f_G}$ is continuous, there exists $\alpha_G > 0$ s.t. $|V^{(1)}_{f_G}(x)| \ge \alpha_G$ for all $x \in N_\varepsilon(\alpha_G)$, where $N_\varepsilon(\alpha_G) = \{x\colon\ |x - c_G| \le \varepsilon\}$ is an ε-neighborhood of α_G. Define $\alpha(\mathscr{H}) = \inf\{\alpha_G\colon\ G \in \mathscr{H} \subseteq \mathscr{G}\}$ for any $\mathscr{H} \subseteq \mathscr{G}$. Let $\mathscr{H}^*$ be the largest subset of $\mathscr{G}$ s.t. $\alpha(\mathscr{H}^*) > 0$. For brevity, we denote $\alpha(\mathscr{H}^*)$ by α^* below. Let

$$\widetilde{\mathscr{H}} = \{G \in \mathscr{H}^*\colon\ f_G \in \mathscr{U} \cap \mathscr{C}_{B,r}\}, \tag{3.9}$$

where

$$\mathscr{U} = \Big\{f_G\colon\ f_G(x) = m(x)\int e^{x\theta}h(\theta)\,dG(\theta),\ G \in \mathscr{G}\Big\}, \tag{3.10}$$

and $\mathscr{C}_{B,r}$ is given by (3.3).

Now define

$$\widehat{c} = \min\{x \in [-A, A]\colon\ \widehat{V}(x) = 0,\ a < x < b\}, \tag{3.11}$$

where A is as used in (3.8) and

$$\widehat{V}(x) = \widehat{f}^{(1)}(x)m(x) - \widehat{f}(x)m^{(1)}(x) - \theta_0\widehat{f}(x)m(x), \tag{3.12}$$

with $\widehat{f}^{(\nu)}$ as defined by (3.2), $\nu = 0, 1, 2$. If $f_G \in \mathscr{U} \cap \mathscr{C}_{B,r}$, m, $m^{(1)}$ and $m^{(2)}$ are bounded, then by Lemma 3.2, as $n \to \infty$,

$$\sup_x |\widehat{V}^{(i)}(x) - V^{(i)}_{f_G}(x)| = o(1), \quad \text{w.p.1}, \tag{3.13}$$

for $i = 0, 1, 2$.

THEOREM 3.1. *Let $r > 4$ and let $B = (2\pi)^{-1/2}\sum_{j=0}^{r}|a_j|$, where a_j is the jth coefficient of the rth Hermite polynomial. Let $\widehat{\psi}$ denote the EB rule of the form* (2.7) *with $c_n = \widehat{c}$, where $\widehat{c}$ is defined by* (3.11). *Let the kernel K satisfy* (3.1) *and let $K^{(2)}$ be Lipschitz continuous on $[-\tau, \tau]$. Then, under the assumptions of Example* 2.1, *by choosing the bandwidth $h_n = n^{-1/(2r+1)}$, we have*

$$\sup_{G \in \widetilde{\mathscr{H}} \cap \mathscr{G}_A} E\big(\widehat{R}(G, \widehat{\psi}) - R(G)\big) = O(n^{-2(r-1)/(2r+1)}) \tag{3.14}$$

for all sufficiently large n.

PROOF. We first show that $\widehat{c} \to c_G$ as $n \to \infty$, w.p.1, for any $f_G \in \mathscr{U} \cap \mathscr{C}_{B,r}$, where $\widehat{c}$ is defined by (3.11). Define $\alpha_G(x) = E_G(\theta|X = x)$, $x \in (-\infty, \infty)$, $G \in \mathscr{G}$. Then α_G is a strictly increasing continuous function on $(-\infty, \infty)$; see (2.4). Note that $\alpha_G(x) = V_{f_G}(x)/m(x)f_G(x)$, since $\theta_0 = 0$, where $m(x) = \exp(-x^2/2)$ and $f_G(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp(-(x-\theta)^2/2)\, dG(\theta)$. Then

$$\alpha_G(\widehat{c}\,) - \alpha_G(c_G) = \frac{V_{f_G}(\widehat{c}\,)}{m(\widehat{c}\,)f_G(\widehat{c}\,)} - \frac{V_{f_G}(c_G)}{m(c_G)f_G(c_G)} = \frac{V_{f_G}(\widehat{c}\,)}{m(\widehat{c}\,)f_G(\widehat{c}\,)} - \frac{\widehat{V}(\widehat{c}\,)}{m(\widehat{c}\,)f_G(\widehat{c}\,)},$$

since $V_{f_G}(c_G) = 0$ and $\widehat{V}(\widehat{c}\,) = 0$. Hence,

$$|\alpha_G(\widehat{c}\,) - \alpha_G(c_G)| \le \sup_x |V_{f_G}(x) - \widehat{V}(x)| / m(\widehat{c}\,)f_G(\widehat{c}\,).$$

Thus, $\alpha_G(\widehat{c}) \to \alpha_G(c_G)$ as $n \to \infty$, w.p.1, from (3.13) and since $m(\widehat{c})f_G(\widehat{c})$ is bounded away from 0 for all n. Now $\widehat{c} \to c_G$ as $n \to \infty$, w.p.1, follows from the fact that α_G^{-1} (the inverse function of α_G) is a continuous function on $(-\infty, \infty)$.

Recall from (2.10) that, for any $G \in \mathscr{G}$,

$$\widehat{R}(G, \widehat{\psi}) - R(G) = S_G^{(2)}(c^*)(\widehat{c} - c_G)^2, \tag{3.15}$$

where c^* is an intermediate value between $\widehat{c}$ and c_G. By repeated differentiation under the integral sign, $f_G^{(r)}(x) = (-1)^r \int H_r(x-\theta) f(x|\theta)\, dG(\theta)$ for any $G \in \mathscr{G}$, where $f(x|\theta) = (2\pi)^{-1/2} \exp(-(x-\theta)^2/2)$ and H_r is the rth Hermite polynomial. Thus, $\sup_x |f_G^{(r)}(x)| \le (2\pi)^{-1/2} \sum_{j=0}^{r} |a_j|$ for any $G \in \mathscr{G}$. Furthermore, since $\theta_0 = 0$, $S_G(y) = \int_{c_G}^{y} l_G(x)\, dx = \int_{c_G}^{y} \left(\int \theta f(x|\theta)\, dG(\theta) \right) dx$ has second derivative

$$S_G^{(2)}(y) = (2\pi)^{-1/2} \left\{ \int (y-\theta)^2 \exp\left(\frac{-(y-\theta)^2}{2} \right) dG(\theta) - y \int (y-\theta) \exp\left(\frac{-(y-\theta)^2}{2} \right) dG(\theta) \right\}.$$

Therefore, for $-\infty < y < \infty$,

$$|S_G^{(2)}(y)| \le (2\pi)^{-1/2}(1 + |y|) \tag{3.16}$$

for any $G \in \mathscr{G}$. Then, since $\widehat{c} \to c_G$ as $n \to \infty$, from (3.16) we have

$$|S_G^{(2)}(c_n^*)| \le (2\pi)^{-1/2}(1 + |c_G|) \le (2\pi)^{-1/2}(1 + A) \tag{3.17}$$

for all sufficiently large n and $G \in \mathscr{G}_A$. By the mean value theorem we find [observing $\widehat{V}(\widehat{c}) = 0$]

$$\widehat{c} - c_G = \widehat{V}(c_G)/\widehat{V}^{(1)}(c_1^*), \tag{3.18}$$

where c_1^* lies between $\widehat{c}$ and c_G. Now $\widehat{c} \to c_G$ and (3.13) yield that, w.p.1,

$$\sup_{x \in N_\varepsilon(c_G)} \left| \frac{1}{\widehat{V}^{(1)}(x)} \right| = \sup_{x \in N_\varepsilon(c_G)} \left| \frac{1}{V_{f_G}^{(1)}(x)} \cdot \frac{V_{f_G}^{(1)}(x)}{\widehat{V}^{(1)}(x)} \right| \le \frac{2}{\alpha_G} \tag{3.19}$$

for all sufficiently large n and $G \in \mathscr{G}_A$ [for the definitions of α_G and $N_\varepsilon(c_G)$, see circa (3.9)]. From (3.18) and (3.19) one obtains, w.p.1,

$$|\hat{c} - c_G| \le \frac{2}{\alpha_G}\,|\widehat{V}(c_G)| \tag{3.20}$$

for all sufficiently large n and $G \in \mathscr{G}_A$. Now combine (3.15), (3.17) and (3.20) to obtain

$$E\big(\widehat{R}(G,\hat{\psi}) - R(G)\big) \le (2\pi)^{-1/2}(1+A)\left(\frac{2}{\alpha_G}\right)^2 E\big(\widehat{V}(c_G)\big)^2 \tag{3.21}$$

for all sufficiently large n and $G \in \mathscr{G}_A$. Since $V_{f_G}(c_G) = 0$, from (3.12) we have

$$\begin{aligned}&\widehat{V}(c_G) - V_{f_G}(c_G)\\ &\quad= \big(\hat{f}^{(1)}(c_G) - f_G^{(1)}(c_G)\big)m(c_G) + m^{(1)}(c_G)\big(f_G(c_G) - \hat{f}(c_G)\big).\end{aligned} \tag{3.22}$$

Now (3.21) and (3.22) together with (3.4) and (3.5) yield

$$\begin{aligned}\sup_{G\in\widetilde{\mathscr{H}}\cap\mathscr{G}_A} &E\big(\widehat{R}(G,\hat{\psi}) - R(G)\big)\\ &\le (2\pi)^{-1/2}(1+A)\left(\frac{2}{\alpha^*}\right)^2\\ &\quad\times 2\{h_n^{2(r-1)}[B\cdot B_{r,1}(1+o(1))] + (nh_n^3)^{-1}[A_1(1+o(1))]\\ &\qquad + h_n^{2r}[B\cdot B_{r,0}(1+o(1))] + (nh_n^2)^{-1}[A_0(1+o(1))]\}\\ &= O(h^{2(r-1)}) + O((nh_n^3)^{-1})\end{aligned}$$

for all sufficiently large n. The result (3.14) now follows by choosing $h_n = n^{-1/(2r+1)}$. □

REMARK 3.1. To see why the EB rule $\hat{\psi}$ defined in Theorem 3.1 attains the lower bound in (2.19) with δ_n as in Corollary 2.1, it is enough to observe that G_0 and G_n belong to $\widetilde{\mathscr{H}} \cap \mathscr{G}_A$, where G_0 and G_n are the distribution functions of densities (2.29) and (2.15), respectively, with H in (2.15) as defined in the proof of Corollary 2.1. Clearly, $G_0 \in \mathscr{G}_A$, since $c_0 = 0$. Hence, $G_n \in \mathscr{G}_A$ for all sufficiently large n. Also, we notice that $f_G \in \mathscr{U} \cap \mathscr{C}_{B,r}$ when $B = (2\pi)^{-1/2}\sum_{j=0}^r|a_j|$ for all $G \in \mathscr{G}$. Thus, $f_0 \in \mathscr{U} \cap \mathscr{C}_{B,r}$ and hence $f_n^* \in \mathscr{U} \cap \mathscr{C}_{B,r}$ for all sufficiently large n, where f_n^* is given by (2.16). Furthermore, $|V_{f_0}^{(1)}| \ge \alpha_0 > 0$ for all $x \in N_\varepsilon(0)$ for any $\varepsilon > 0$ (note that $c_0 = 0$ in this case). Thus, $G_0 \in \widetilde{\mathscr{H}}$. Hence, $G_n \in \widetilde{\mathscr{H}}$ for all sufficiently large n. Then, from (2.19) and Corollary 2.1, we obtain [by taking $3k = 8 + 4r$ in (2.19)]

$$\sup_{G\in\widetilde{\mathscr{H}}\cap\mathscr{G}_A} E\big(\widehat{R}(G,\hat{\psi}) - R(G)\big) \ge dn^{-2(r-1)/(2r+1)} \tag{3.23}$$

for some finite constant $d > 0$. The upper and lower bounds (3.14) and (3.23) show that the EB rule $\hat{\psi}$ achieves the optimal minimax rate $n^{-2(r-1)/(2r+1)}$.

THEOREM 3.2. *Let $r > 4$. Let $B > 2$ be a large but fixed number. Let $\widehat{\psi}$ be defined by* (2.7) *with $c_n = \widehat{c}$, where $\widehat{c}$ is defined by* (3.11). *Suppose that support of the kernel function K is* $(0, 1)$ *and satisfies* (3.1)(iii) *and suppose that $K^{(2)}$ is Lipschitz continuous on* $(0, 1)$. *Then, under the assumptions of Example* 2.2, *by choosing the bandwidth $h_n = n^{-1/(2r+1)}$, we have*

$$\sup_{G\in\widetilde{\mathscr{H}}\cap\mathscr{G}_0\cap\mathscr{G}_A} E\big(\widehat{R}(G,\widehat{\psi}) - R(G)\big) = O(n^{-2(r-1)/(2r+1)}) \tag{3.24}$$

for all sufficiently large n, where $\mathscr{G}_0 = \{G \in \mathscr{G}\colon \int_0^\infty \theta^r\, dG(\theta) \le B\}$.

PROOF. Again, it is easy to show that $\widehat{c} \to c_G$ as $n \to \infty$, w.p.1 [the proof is exactly similar to the proof given for Theorem 3.1, since $m(\widehat{c})f_G(\widehat{c})$ is bounded away from 0 in this case as well]. By repeated differentiation under the integral sign [using Theorem 2.9 of Lehmann (1959)], the ith derivative of $f_G(x) = \int_0^\infty f(x|\theta)\, dG(\theta) = \int_0^\infty \theta e^{-\theta x}\, dG(\theta)$, $x > 0$, is $f_G^{(i)}(x) = (-1)^i \int_0^\infty \theta^i f(x|\theta)\, dG(\theta)$, $i \ge 1$. Thus, $|f^{(r)}| \le B$, and hence $G \in \mathscr{G}_0$ implies that $f_G \in \mathscr{U} \cap \mathscr{C}_{B,r}$. In this case, $S_G^{(1)}(y) = f_G^{(1)}(y) - \theta_0 f_G(y)$, $\quad y > 0$, and therefore $|S_G^{(2)}| \le M$ for some finite positive constant M depending only on θ_0 and B for all $G \in \mathscr{G}_0$. Then from (2.18), for any $G \in \mathscr{G}_0$,

$$E\big(\widehat{R}(G,\widehat{\psi}) - R(G)\big) = E\big(S_G^{(2)}(c_n^*)(\widehat{c} - c_G)^2\big) \le ME(\widehat{c} - c_G)^2. \tag{3.25}$$

As in (3.20) we have, w.p.1,

$$|\widehat{c} - c_G| \le \frac{2}{\alpha_G}\, |\widehat{V}(c_G)| \tag{3.26}$$

for all sufficiently large n and $G \in \mathscr{G}_A$. In this case, $\widehat{V}(x) = \widehat{f}^{(1)}(x) - \theta_0 \widehat{f}(x)$ and $\theta_0 > 0$. Therefore, since $V_{f_G}(c_G) = 0$,

$$\widehat{V}(c_G) - V_{f_G}(c_G) = \big(\widehat{f}^{(1)}(c_G) - f_G^{(1)}(c_G)\big) + \theta_0\big(f_G(c_G) - \widehat{f}(c_G)\big). \tag{3.27}$$

Then, from (3.25), (3.26), (3.27) together with (3.4) and (3.5), we obtain

$$\begin{aligned}
&\sup_{G\in\widetilde{\mathscr{H}}\cap\mathscr{G}_0\cap\mathscr{G}_A} E\big(\widehat{R}(G,\widehat{\psi}) - R(G)\big)\\
&\quad\le M\left(\frac{2}{\alpha^*}\right)^2 2\{h_n^{2(r-1)}[B\cdot B_{r,1}(1+o(1))]^2\\
&\qquad\qquad + (nh_n^3)^{-1}[A_1(1+o(1))] + o(h_n^{2(r-1)}) + o((nh^3)^{-1})\}\\
&\quad= O(h_n^{2(r-1)}) + O((nh_n^3)^{-1})
\end{aligned}$$

for all sufficiently large n. The result follows by taking $h_n = n^{-1/(2r+1)}$. □

REMARK 3.2. By choosing $l \geq r + 2$ in (2.38), it is easy to show that $\{G_0, G_n\} \subseteq \widetilde{\mathscr{H}} \cap \mathscr{G}_0 \cap \mathscr{G}_A$ for some sufficiently large A and for all sufficiently large n, where G_0 and G_n are the distribution functions of (2.38) and (2.15), respectively, with H in (2.15) as defined in the proof of Corollary 2.2. Then from (2.19) and Corollary 2.2 we obtain, by taking $3k = 5r + 7$ in (2.19),

$$\sup_{G \in \widetilde{\mathscr{H}} \cap \mathscr{G}_0 \cap \mathscr{G}_A} E\big(\widehat{R}(G, \widehat{\psi}) - R(G)\big) \geq d_0 n^{-2(r-1)/(2r+1)} \tag{3.28}$$

for some finite constant $d_0 > 0$. The upper and lower bounds (3.24) and (3.28) show that the EB rule $\widehat{\psi}$ achieves the optimal minimax rate $n^{-2(r-1)/(2r+1)}$ in this case.

REMARK 3.3. One can replace the constant A in (3.8) by a suitably chosen sequence of positive numbers, $\{b_n\}$, such that $b_n \uparrow b$ as $n \to \infty$. However, the result of such a generality is weaker rates of convergence at (3.14) and (3.24).

Acknowledgments. I would like to thank the associate editor and the referee for their valuable comments and suggestions which led to significant improvements in both the substance and style of this paper. Thanks are also due to TaChen Liang, B. Schmuland and J. Sheahan for useful personal communications.

REFERENCES

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667.

DONOHO, D. L. and LIU, R. C. (1991b). Geometrizing rates of convergence. III. *Ann. Statist.* **19** 668–701.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.

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. (1989). Some improvements on empirical Bayes testing in Lebesgue-exponential families. *Comm. Statist. Theory Methods* **18** 3157–3176.

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.

LEHMANN, E. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

MÜLLER, H. G. and GASSER, T. (1979). Optimal convergence properties of kernel estimates of derivatives of a density function. *Lecture Notes in Math.* **757** 144–154. Springer, Berlin.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163. Univ. California Press, Berkeley.

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

SINGH, R. S. (1977). Improvements on some known nonparametric uniformly consistent estimators of derivatives of a density. *Ann. Statist.* **5** 394–399.

STIJNEN, T. (1985). On the asymptotic behavior 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.

Yu, B. O. (1970). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. Ph.D. dissertation, Dept. Statistics and Probability, Michigan State Univ.

Department of Mathematical Sciences
University of Alberta
Edmonton, Alberta
T6G 2G1 Canada