The Annals of Statistics
1979, Vol. 7, No. 4, 890–902

EMPIRICAL BAYES ESTIMATION IN LEBESGUE-EXPONENTIAL FAMILIES WITH RATES NEAR THE BEST POSSIBLE RATE[1]

BY R. S. SINGH

University of Guelph

Asymptotically optimal (a.o.) empirical Bayes (EB) estimators are proposed. Speeds and the best possible speed at which these estimators are a.o. are investigated. The underlying component problem is the squared error loss estimation of θ based on an observation X whose conditional (on θ) pdf is of the form $u(x)C(\theta)\exp(\theta x)$. The function u could have *infinitely many discontinuities*; θ is distributed according to an *unknown* and *unspecified* G with support in Θ, and Θ could be *unbounded*.

Using n independent past experiences of the component problem, EB estimators ϕ_n for the present problem are exhibited for each integer $r > 1$. The risks $R(\phi_n, G)$ due to ϕ_n are shown to converge to the *minimum Bayes risk* $R(G)$. In particular, for each δ in $[r^{-1}, 1]$, sufficient conditions are given under which $c_1 n^{-2(r-1)/(1+2r)} \leqslant R(\phi_n, G) - R(G) \leqslant c_2 n^{-2(\delta r-1)/(1+2r)}$, where c_1 and c_2 are positive constants. The right hand-side inequality holds *uniformly* in G satisfying certain conditions, while the other holds at all degenerate G and for all large n. (Thus with δ close to one, ϕ_n achieves almost the *exact rate*.) Examples of exponential families such as normal, gamma and one with pdf's having infinitely many discontinuities are given where the conditions for the above inequalities are satisfied *uniformly* in G with $\int |\theta|^{2r\delta} dG(\theta) < \infty$.

1. Introduction. Yu [21], Lin [7] and Singh [16] considered the empirical Bayes approach, (introduced by Robbins (1955), and later developed by Johns (1957), Robbins (1963, 1964), Samuel (1963), and Johns and Van Ryzin (1971, 1972), among others), to the squared error loss estimation (SELE) in the one parameter exponential family. The *component problem* in [21], [7] and [16] is the SELE of θ based on an observation X (which could be a sufficient statistic for θ) having conditional (on θ) pdf of the form

$$p_\theta(x) = u(x)C(\theta)\exp(\theta x) \tag{1.0}$$

where for an $a \geqslant -\infty$,

$$u(x) > 0 \qquad \text{if and only if} \quad x > a, \tag{1.1}$$

and $C(\theta) = (\int \exp(\theta x)u(x)dx)^{-1}$. The parameter θ is distributed according to an *unknown* and *unspecified* G with support in Θ, a subset of the *natural parameter space* $\{\theta | C(\theta) > 0\}$. The risk of an estimator ϕ is $R(\phi, G) = E(\phi - \theta)^2$, and the

Received April 1977; revised February 1978.

[1]Part of the paper was prepared while the author was at the Indian Statistical Institute, New Delhi. The research was supported in part by National Research Council of Canada Grant A4631.

AMS 1970 *subject classifications.* Primary 62F15, 62F10; secondary 62C25.

Key words and phrases. Empirical Bayes estimation, squared error loss, rates of convergence, the best possible rate, asymptotically optimal.

estimator which achieves the *Bayes envelope* $R(G) = \inf_\phi R(\phi, G)$ is given by

$$\phi_G(X) = E(\theta|X) = \frac{\int \theta C(\theta)\exp(\theta x)dG(\theta)}{\int C(\theta)\exp(\theta x)dG(\theta)}. \tag{1.2}$$

Using n independent past experiences (of the problems identical to the above component problem) through observations $X_1, \cdots, X_n$, and the present observation X, X's being i.i.d. with pdf $p(x) = \int p_\theta(x)dG(\theta)$, asymptotically optimal (a.o.) estimators (for the *definition,* see any of the references cited above) have been exhibited in [21], [7] and [16]. These works may be thought of, to some extent, as the counterpart of the *notable* work by Johns and Van Ryzin (1972) where the empirical Bayes (EB) approach to linear loss two-action problem in the above family is considered.

Whereas Lin's (1975) presentation and proofs are nice, his restrictions, among others, (i) the existence and the continuity of the rth derivative $u^{(r)}$ of u for r sufficiently large, (ii) the boundedness in x of $\sup_{0<\zeta\leqslant\varepsilon}|p^{(r)}(x+\zeta)|$ for *every* $\varepsilon > 0$ and for r sufficiently large, and (iii) $\Theta \subset [0, \infty)$, in order to get the speed of asymptotic optimality *near* $0(n^{-\frac{1}{3}})$ are seemingly *unnecessary*. Though Singh (1976) improves Lin's rate to *near* $0(n^{-\frac{2}{5}})$, his assumptions, the boundedness of Θ and the boundedness of $(p(x)/u(x))\sup_{x\leqslant t<x+\varepsilon}u(t)$ in x for *some* $\varepsilon > 0$, also appear to be *unnecessary*. The reason for this speculation is that we, in this note, are able to exhibit estimators (by using $X_1, \cdots, X_n$ and X) which are a.o. with very *high speed* (namely, arbitrarily close to $0(n^{-1})$) *without* requiring any assumption on the smoothness of u, or, for that matter, any of the above sort of restrictions at all. More precisely, for every integer $r > 1$, estimators ϕ_n based on $X_1, \cdots, X_n$ and X are exhibited such that, for some constants c, c',

$$cn^{-2(r-1)/(1+2r)} \leqslant R(\phi_n, G) - R(G) \leqslant c'n^{-2(r-1)+/(1+2r)} \tag{1.3}$$

where the right-hand side inequality holds *uniformly* in G satisfying certain conditions and for each $n \geqslant 1$; and the left-hand side inequality holds at each degenerate G and for all n large enough. In fact, in many exponential families, including normal, gamma and one with pdf having *infinitely many discontinuities*, (1.3) holds uniformly in G satisfying $\int|\theta|^{2r-}dG(\theta) < \infty$.

Hannan and Macky (1971) suggested an EB estimator in the above exponential families which are a.o. if $\int\theta^2 dG(\theta) < \infty$. No specific rate of convergence, however, is emphasized. Further comparisons of this work with those of others are given in Section 7.

Notice that the rates in (1.3) can be made arbitrarily close to $0(n^{-1})$ by taking r sufficiently large. In view of the existing literature and our efforts in obtaining the results here, we conjecture that (i) a rate $0(n^{-1})$ or better is not possible for *any* EB estimator in *any* Lebesgue-exponential family even though Θ is *bounded*, and (ii) no specific rate of convergence for *any* EB estimator in *any* exponential family is possible *without* any moment condition on G.

Our plan is as follows: in Section 2 we prove a basic lemma. In Section 3 we exhibit our EB estimator, first by exhibiting mean square consistent estimators of

$f = p/u$, and its first derivative $f^{(1)}$. In Sections 4 and 5 we obtain our main results giving, respectively, right-hand side and left-hand side of (1.3). In Section 6 we give examples where (1.3) holds *uniformly* in G satisfying $\int|\theta|^{2r-}dG < \infty$. We conclude the paper with a few remarks in Section 7.

2. A basic lemma. For any estimator ϕ^*, the excess in risk due to ϕ^* over the Bayes envelope is $R(\phi^*, G) - R(G)$. The following lemma, which reduces the problem of search of an a.o. estimator to the one of mean square consistent estimation of the Bayes estimate ϕ_G, has been found very useful in the context of EB SELE. The conclusion of the lemma is well known, but the lemma under the present assumption was suggested to me by Professor James Hannan; it is valid, not only for p_θ of the form (1.0), but for any *arbitrary* p_θ.

LEMMA 2.1. *If $R(G)$ is finite, then for any estimator ϕ^*,*

$$R(\phi^*, G) - R(G) = E(\phi^* - \phi_G)^2. \tag{2.0}$$

REMARK 2.1. In the published literature, (e.g. Johns (1957) and most recently Lin (1975)), (2.0) is proved under the *stronger* assumption that $E|\theta|^2 < \infty$. This implies the hypothesis of the lemma. The converse, however, is not true. For example, let $p_\theta(x) = (2\pi)^{-\frac{1}{2}} \exp(-(x-\theta)^2/2)$. Then, since $E|X - \theta|^2 = 1$, $R(G) \leqslant 1$ regardless of the nature of G.

PROOF OF THE LEMMA. Let E_* denote the conditional expectation operator given X and all other random variables involved in the definition of ϕ^*. Then

$$E_*(\phi^* - \theta)^2 = (\phi^* - \phi_G)^2 + E_*\{(\phi_G - \theta)^2 + 2(\phi^* - \phi_G)(\phi_G - \theta)\}. \tag{2.1}$$

But, since $R(G) = E\{E_*(\phi_G - \theta)^2\} < \infty$, $E_*(\phi_G - \theta)^2$ (and hence $E_*|\phi_G - \theta|$) is finite w.p. 1. Therefore the second term on the right-hand side of (2.1) can be written as $E_*(\phi_G - \theta)^2 + 2(\phi^* - \phi_G)E_*(\phi_G - \theta)$ which is simply $E_*(\phi_G - \theta)^2$ since $E_*\theta = \phi_G$. Thus, $E_*(\phi^* - \theta)^2 = (\phi^* - \phi_G)^2 + E_*(\phi_G - \theta)^2$. This identity, combined with the fact that $R(\phi^*, G) = E(E_*(\phi^* - \theta)^2)$, gives (2.0). ☐

3. Proposed class of empirical Bayes estimators.

3.1. *Introduction.* Let $f(x) = \int C(\theta)\exp(\theta x)dG(\theta)$. By Theorem 2.9 of Lehmann (1959), $f^{(1)}(x) = \int \theta C(\theta)\exp(\theta x)dG(\theta)$. Therefore, by (1.2), ϕ_G can be written as

$$\phi_G = f^{(1)}/f. \tag{3.0}$$

Thus estimation of ϕ_G amounts to the estimation of $f^{(i)}$ for $i = 0, 1$, *where* $f^{(0)} = f$. Notice that f is *not* a pdf, and therefore estimators of a pdf or its derivatives available in the existing literature cannot be directly used here. As a result, we will now exhibit a class of estimators of $f^{(i)}$ for $i = 0$ and 1, and prove mean square consistency of these estimators. Then based on these estimators we will define estimators of ϕ_G.

3.2. *Estimation of f and its derivative $f^{(1)}$.* Let $r > 1$ be a fixed integer. For $i = 0, 1$, let $\mathcal{K}_i^r$ be the class of all Borel-measurable real valued bounded functions

K_i vanishing off (0, 1) such that

$$\int y^t K_i(y)dy = 1 \quad \text{if} \quad t = i \qquad (3.1)$$
$$= 0 \quad \text{if} \quad t \neq i, \qquad t = 0, 1, \cdots, r-1.$$

(Polynomials vanishing off (0, 1) and satisfying (3.1) can be constructed, e.g., see Singh (1978, 1979). The set $\mathcal{K}_i^r$ also contains the $(i+1)$st element of the dual basis for the subspace of $L_1(0, 1)$ with basis $\{1, y/1, \cdots, y^r/r!\}$.) Kernel functions of the type K_i above are used by Johns and Van Ryzin (1972) and by Singh (1977a, 1978, 1979), among others, in construction of estimators of a pdf and its derivative.

Let $I_1(x) = I(x \geqslant 1)$ and $I_2(x) = I(a < x < 1)$ where a is given by (1.1), and $I(S)$ is the indicator function of the set S. For $i = 0, 1$ and $j = 1, 2$ *denote* $f^{(i)}(x)I_j(x)$ by $f_j^{(i)}(x)$. For $x > a$, let $0 < h(x) = h_n(x)$ be such that for each x, $h \to 0$ as $n \to \infty$. Let K_i be a fixed element of $\mathcal{K}_i^r$. At $x > a$, define

$$T_i(x, h(x)) = n^{-1}(h(x))^{-1-i}\Sigma_{t=1}^n\left\{K_i\left(\frac{X_t - x}{h(x)}\right)\Big/ u(X_t)\right\}. \qquad (3.2)$$

For $i = 0, 1$ and $j = 1, 2$, our *proposed estimators* of $f_j^{(i)}$ are $\hat{f}_j^{(i)}$ defined as

$$\hat{f}_j^{(i)}(x) = T_i(x, (-1)^j h(x))I_j(x). \qquad (3.3)$$

In Theorem 3.1 below we will prove mean square consistency of $\hat{f}_j^{(i)}$ as an estimator of $f_j^{(i)}$. For the sake of brevity in writing we introduce for $j = 1, 2$

$$\alpha_j(\theta, x) = \exp\{(j-1)h(x)\theta\}I(\theta \geqslant 0) + \exp\{(j-2)h(x)\theta\}I(\theta < 0) \qquad (3.4)$$

and

$$\beta_j(\theta, x) = \int_0^1 \frac{\exp\{(-1)^j h(x)\theta y\}}{u(x + (-1)^j h(x)y)}\, dy. \qquad (3.5)$$

Note that α_j and β_j depend on n through h. Let M be the common bound of K_0 and K_1.

THEOREM 3.1. *For $i = 0, 1$ and $j = 1, 2$, and for every $0 < t \leqslant 2$,*

$$E|\hat{f}_j^{(i)} - f_j^{(i)}|^t \leqslant (Mf)^t I_j(x)\left[\{h^{r-i}E_{X=x}(|\theta|^r\alpha_j(\theta))\}^t + \{nh^{1+2i}f/E_{X=x}\beta_j(\theta)\}^{-t/2}\right] \qquad (3.6)$$

where the argument x in $\hat{f}_j^{(i)}, f_j^{(i)}, f_j, \alpha_j, \beta_j$ and h is indicated by omission.

PROOF. We will indicate the proof of the inequality (3.6) for $i = 0$ and $j = 1$ only. The proofs for others follow similarly.

Abbreviate $f^{(0)}I_1$ to f_1 and $\hat{f}_1^{(0)}$ to $\hat{f}_1$. Since $f^{(r)}(z)$ is $\int \theta^r C(\theta)e^{\theta z}dG(\theta)$ by Theorem 2.9 of Lehmann (1959) and since

$$\sup_{0 \leqslant t \leqslant 1}|f^{(r)}(x - h(x)t)| \leqslant f(x)E_{X=x}(|\theta|^r\alpha_1(\theta, x)) \qquad (3.7)$$

where α_1 is given in (3.4), by arguments similar to those used for (3.6) of Singh (1977a), we have

(3.8) $$|E\hat{f}_1 - f_1| \leqslant Mh^r f E_{X=x}(|\theta|^r \alpha_1(\theta)).$$

And the usual method of bounding the variance of the average of i.i.d. random variables followed by (3.5) gives

(3.9) $$\operatorname{Var}(\hat{f}_1) \leqslant M^2(nh)^{-1} f E_{X=x}\beta_1(\theta).$$

Now the desired result follows by Hölder inequality since $(\hat{f}_1 - f_1)^2$ is (left-hand side of (3.8))^2 + left-hand side of (3.9). □

3.3. *The proposed class of empirical Bayes estimators.* In view of the mean square consistent estimators $\hat{f}^{(i)} = \hat{f}_1^{(i)} + \hat{f}_2^{(i)}$ of $f^{(i)} = f_1^{(i)} + f_2^{(i)}$ for $i = 0, 1$, Lemma 2.1 and (3.0) our proposed (EB) estimators for the present problem are

(3.10) $$\phi_n(X) = (\hat{f}^{(1)}(X)/\hat{f}(X))_{h^{-1}(X)}$$

where for $c > 0$, $(b)_c$ *is* $-c$, b *or* c *according as* $b < -c$, $|b| \leqslant c$ *or* $b > c$. Notice that no information about G, ϕ_G or Θ is required to define ϕ_n above; and they depend only on the past observations $X_1, \cdots, X_n$ and the present observation X.

4. An upper bound for $R(\phi_n, G) - R(G)$ and rates of asymptotic optimality of the EB estimators ϕ_n. We recall from the preceding section that our EB estimators ϕ_n introduced in (3.10) depend on the integer $r > 1$ involved in the estimators of $f^{(i)}$ through the kernels K_0 and K_1. In this section we will obtain a bound for the excess risk $R(\phi_n, G) - R(G)$ for each $n \geqslant 1$, and show that for any $\varepsilon > 0$ EB estimators can be exhibited which are a.o. at the rate $0(n^{-1+\varepsilon})$. In the next section we will show that our estimators *cannot* be a.o. with rates better than $0(n^{-1+\varepsilon'})$ for some $\varepsilon' > 0$.

Theorem 4.1 below gives a bound for $R(\phi_n, G) - R(G)$. The conditions of the theorem, though looking a little stringent, *reduce to a single moment condition* on G in several exponential families, including normal, gamma and a family with pdf's having infinitely many discontinuity points. Let $c_0, c_1, \cdots$ below denote absolute constants.

THEOREM 4.1. *Recall the definitions of* α_j *and* β_j, $j = 1, 2$, *from* (3.4) *and* (3.5) *respectively. Let there exist a positive function* $c(x)$ *on* (a, ∞) *which, with*

(4.0) $$h(x) = h_n(x) = c(x)\varepsilon_n \quad \text{where} \quad \varepsilon_n = c_0 n^{-1/(2r+1)} < 1,$$

gives a δ *in* $[r^{-1}, 1]$ *such that for some* $t > 1$,

(4.1) $$(E|\theta|^{2t})^{1/(t-1)} E\{c(X)E_X|\theta|\}^{2t(r\delta-1)/(t-1)} < \infty$$

and, for each $j = 1, 2$,

(4.2) $$E\big[I_j(X)(c(X))^{2(r\delta-1)}E_X^{2\delta}|\theta^r\alpha_j(\theta, X)|\big] < \infty$$

and

(4.3) $$E\big[I_j(X)(c(X))^{-2-\delta}E_X^{\delta}\beta_j(\theta, X)/f^{\delta}(X)\big] < \infty.$$

Let ϕ_n *be given by* (3.10) *with* h *as in* (4.0). *Then*

$$0 \leqslant R(\phi_n, G) - R(G) \leqslant c_1 n^{-2(\delta r - 1)/(2r+1)}. \tag{4.4}$$

A special case of the result of Lemma 4.2 below simplifies and shortens the proof of the theorem. The proof of Lemma 4.2 in turn, is simplified by the following *general* lemma due to Singh (1974).

LEMMA 4.1. (*Singh* (1974)). *Let* y, y', Y, Y' *and* $0 < L$ *be reals. Further, let* Y, Y' *be random. Then for every* $t > 0$,

$$E\left(\left|\frac{y'}{y} - \frac{Y'}{Y}\right| \wedge L\right)^t \leqslant \min\{L^t, A\}, \tag{4.5}$$

$$\text{where} \quad A = 2^{t+(t-1)^+}|y|^{-t}\{E|y' - Y'|^t + (|y'/y|^t + 2^{-(t-1)^+}L^t)E|y - Y|^t\}.$$

PROOF. The proof is given in the Appendix of Singh (1977b). □

LEMMA 4.2. *As before, abbreviate* $\hat{f}^{(0)}$ *to* $\hat{f}$. *Then, for every* $s > 0$ *and* $0 \leqslant t \leqslant s$,

$$2^{-(s-1)^+}E|\phi_n - \phi_G|^s \leqslant (1 + 2^s)|\phi_G|^s I(|\phi_G| > h^{-1}) + \min\{(2h^{-1})^s, [h^{t-s}2^{s+(t-1)^+}(f)^{-t}] \cdot [E|\hat{f}^{(1)} - f^{(1)}|^t + (1 + 2^{t-(t-1)^+})h^{-t}E|\hat{f} - f|^t]\}. \tag{4.7}$$

PROOF. Recall that $\phi_G = f^{(1)}/f$ and $\phi_n = (\hat{f}^{(1)}/\hat{f})_{h^{-1}}$. For this proof only, let $\psi_1 = (\phi_G)_{h^{-1}}$ and $\psi_2 = \phi_G - \psi_1$. Then, since $|\phi_n - \psi_1| \leqslant 2h^{-1}$, by triangle inequality the left-hand side in (4.7) is exceeded by $|\psi_2|^s + (2h^{-1})^s I[|\phi_G| > h^{-1}]$ plus a quantity equal to

$$\begin{aligned} E|\phi_n - \psi_1|^s I(|\phi_G| \leqslant h^{-1}) &\leqslant \{|(\hat{f}^{(1)}/\hat{f}) - (f^{(1)}/f)| \wedge (2h^{-1})\}^s I(|\phi_G| \leqslant h^{-1}) \\ &\leqslant (2h^{-1})^{s-t}\{|(\hat{f}^{(1)}/\hat{f}) - (f^{(1)}/f)| \wedge (2h^{-1})\}^t \\ &\quad \times I(|\phi_G| \leqslant h^{-1}). \end{aligned}$$

The proof is complete by Lemma 4.1 and by the relation

$$\max\{|\psi_2|, h^{-1}I[|\phi_G| > h^{-1}]\} \leqslant |\phi_G| I(|\phi_G| > h^{-1}). \qquad \square$$

PROOF OF THE THEOREM. Lemmas 4.1 and 4.2 have greatly simplified and shortened the proof of the theorem. To complete it, let δ be as given in the theorem. Holder inequality followed by Markov inequality gives, for a $t > 1$ and $q = 2(r\delta - 1)t(t - 1)^{-1}$,

$$\begin{aligned} E\{|\phi_G(X)|^2 I(|\phi_G(X)| > h^{-1}(X)\} &\leqslant \{(E|\phi_G(X)|^{2t})^{1/t} \\ &\quad \cdot (E|h(X)\phi_G(X)|^q)^{(t-1)/t}\} \leqslant \{(E|\theta|^{2t})^{1/t} \\ &\quad \cdot (E|c(X)\phi_G(X)|^{2(r\delta-1)t(t-1)^{-1}})^{(t-1)/t}\varepsilon_n^{2(r\delta-1)}\} \text{ by (4.0).} \end{aligned} \tag{4.8}$$

Also, since $h(x) = c(x)\epsilon_n$ for $i = 0, 1$, $|\hat{f}^{(i)} - f^{(i)}|^t = \Sigma_{j=1,2}|\hat{f}_j^{(i)} - f_j^{(i)}|^t$ for every $t \geqslant 0$, by (3.6)

$$(4.9)\quad E\Big[h^{2\delta i-2}(X)\{|\hat{f}^{(i)}(X) - f^{(i)}(X)|/f(X)\}^{2\delta}\Big]$$
$$\leqslant c_2\epsilon_n^{2(\delta r-1)}E\Big[\Sigma_{j=1,2}I_j(X)\Big\{(c(X))^{2(r\delta-1)}E_X^{2\delta}|\theta^r\alpha_j(\theta, X)|$$
$$+ (c(X))^{-2-\delta}(E_X\beta_j(\theta, X)/f(X))^{\delta}\Big\}\Big].$$

Now (2.0) followed by (4.7) with $s = 2$ and $t = 2\delta$, (4.8), (4.9) and the hypothesis of the theorem completes the proof. □

5. A lower bound for $R(\phi_n, G) - R(G)$ and the best possible rate of asymptotic optimality of ϕ_n. In the preceding section we gave sufficient conditions under which $R(\phi_n, G) - R(G) = 0(n^{-2(r-1)+/(1+2r)})$. (Examples of exponential families satisfying these conditions are given in the next section.) With $c_3, c_4, \cdots$ denoting absolute constants we will now show that for all n large enough $R(\phi_n, G) - R(G) \geqslant c_3 n^{-2(r-1)/(1+2r)}$ at every G degenerate at any point in Θ, thus proving that ϕ_n could achieve almost exact rate.

Throughout the remainder of this section, let G be degenerate at an arbitrary but fixed (unknown) point θ in Θ, and denote p_θ and p_θ/u by p and f respectively.

THEOREM 5.1. *Let ϕ_n be as given in Theorem* 4.1. *Let there exist an $\eta > 0$ and a finite $l > 0$ such that Lebesgue*-inf *and Lebesgue*-sup *of the restriction to $(l, l + \eta)$ of both $c(x)$ in* (4.0) *and $u(x)$ are, respectively, positive and finite. Then for all n large enough,*

$$(5.0)\qquad R(\phi_n, G) - R(G) \geqslant c_3 n^{-2(r-1)/(1+2r)}.$$

PROOF. By our hypothesis,

$$(5.1)\qquad 0 < \textstyle\int_l^{l+\eta} u(x)dx < \infty$$

and, since θ is in Θ,

$$(5.2)\qquad 0 < \inf_{l<t<l+\eta} f(t) \leqslant \sup_{l<t<l+\eta} f(t) < \infty.$$

Since G is degenerate at θ, $\phi_G \equiv \theta$ and by Lemma 2.1 $R(\phi_n, G) - R(G) = E(\phi_n - \theta)^2 \geqslant E^2|\phi_n - \theta|$. Thus, by (3.10), for a β (could be unknown) with $\beta > \theta$, $(R(\phi_n, G) - R(G))^{\frac{1}{2}}$ is no more than

$$(5.3)\quad E|\phi_n(X) - \theta| \geqslant \textstyle\int_0^{\beta-\theta} P[\phi_n(X) - \theta > t]dt$$
$$\geqslant E\{I(l < X < l + (\eta/2))\textstyle\int_0^{\beta-\theta} P_X[\hat{f}^{(1)} - \theta\hat{f} > t|\hat{f}|]dt$$

where the argument X in $\hat{f}^{(1)}$ and $\hat{f}$ is abbreviated by omission, and P_X stands for the conditional probability given X.

Suppose $l \geqslant 1$, then at $X \geqslant 1$, $\hat{f}^{(i)}$ is $\hat{f}_1^{(i)}$ introduced in Section 3. For X in $(l, l + (\varepsilon/2))$ and t in $(0, \beta - \theta)$ and for $j = 1, \cdots, n$, let

$$u(X_j)Y_j = \{h^{-1}K_1 + \theta K_0 + t|K_0|\}\left(\frac{x - X_j}{h}\right)$$

where K_0 and K_1 are the kernels used in the definitions of $\hat{f}_1$ and $\hat{f}_1^{(1)}$ in (3.3) and (3.2). Since $X_1, \cdots, X_n$ are (marginally) i.i.d., so are $Y_1, \cdots, Y_n$. Let $\mu = EY_1$ and $\sigma^2 = \text{var}(Y_1)$. Then, by arguments given in Singh (1974), pages 79–80, for all n large enough,

$$\text{(5.4)} \qquad \mu \geqslant -c_4 h(h^r + h^{r-1} + t) \quad \text{and} \quad \sigma^2 \geqslant c_5 h^{-1}.$$

Thus, since the probability under the integral sign on the extreme right-hand side of (5.3) is $P_X[\Sigma_1^n Y_j > 0] \geqslant P_X[\Sigma_1^n (Y_j - \mu) \geqslant c_4 nh(h^r + h^{r-1} + t)]$ by (5.4); by Lemma 3 on page 47 of Lamperty (1966), for a $\xi > 0$ and for all sufficiently large n,

$$\text{(5.5)} \qquad P_X\big[\hat{f}^{(1)} - \theta\hat{f} > t|\hat{f}|\big] \geqslant \exp\left\{ -\frac{nh^2c_4^2(h^r + h^{r-1} + t)^2}{2\sigma^2}(1 + \xi)\right\}$$
$$\geqslant \exp\left\{ -c_6 nh^3(h^r + h^{r-1} + t)^2\right\} \text{ by (5.4).}$$

(If $l < 1$, then choosing η small enough, we make $l + \eta < 1$. At $x < 1$, $\hat{f}^{(i)}$ is $\hat{f}_2^{(i)}$, $i = 1, 0$, and, by similar arguments, we get (5.5).) Now making the transformation $(c_6 nh^3)^{\frac{1}{2}}\, (h^r + h^{r-1} + t) = v$ we get from (5.3) and (5.5),

$$\text{(5.6)} \quad (R(\phi_n, G) - R(G))^{\frac{1}{2}} \geqslant (c_6 n)^{-\frac{1}{2}} E\left\{h^{-\frac{3}{2}} I(l < X < l + (\eta/2)) \textstyle\int_b^{b'} e^{-t^2} dt\right\}$$

where $b = (c_6 nh^3)^{\frac{1}{2}}\, (h^r + h^{r-1})(X) \to c_6^{\frac{1}{2}}$ as $n \to \infty$ and $b' = (c_6 nh^3)^{\frac{1}{2}}(h^r + h^{r-1} + \beta - \theta)(X) \to \infty$ as $n \to \infty$ uniformly in $l < X < l + (\eta/2)$ from the definition of h in (4.0) and from the hypothesis on $c(\cdot)$ in the definition of h. Consequently, from (5.1) and (5.2), the right-hand side of (5.6) is $\geqslant c_3^{\frac{1}{2}} n^{-(r-1)/(1+2r)}$. □

6. Examples and sub-theorems. We will now give examples of some important exponential families where all conditions of Theorem 4.1 reduce to a single condition of the type $E|\theta|^{2r\delta} < \infty$.

EXAMPLE 6.1. *(Normal $N(\theta, 1)$-family). Let* $u(x) = e^{-x^2/2} I(-\infty < x < \infty)$. *Then* $a = -\infty$, $C(\theta) = (2\pi)^{-\frac{1}{2}} e^{-\theta^2/2}$, *and*

$$\text{(6.0)} \qquad p_\theta(x) = (2\pi)^{-\frac{1}{2}} e^{-(x-\theta)^2/2},$$
$$-\infty < \theta < \infty, \; -\infty < x < \infty.$$

SUBTHEOREM 6.1. *For the family* (6.0), *let* $\Theta \subseteq (-\infty, \infty)$. *Let* ϕ_n *be given by* (3.10) *with* $h = c_0 n^{-1/(1+2r)}$. *If, for a* $.5 < \delta < 1 - (2r)^{-1}$,

$$\text{(6.1)} \qquad E|\theta|^{2r\delta} < \infty,$$

then

$$R(\phi_n, G) - R(G) = 0(n^{-2(r\delta-1)/(1+2r)}).$$

PROOF. The function $c(x)$ in Theorem 4.1 is here identically equal to one. By Holder inequality (6.1) is equivalent to (4.1) with $t = r\delta$. Also, $E_X^{2\delta}(|\theta|^r \alpha_j(\theta, X)) \leqslant E_X(|\theta|^r \alpha_j(\theta, X))^{2\delta}$ for $2\delta \geqslant 1$, and by (3.4) $\alpha_1(\theta, X) \leqslant I(\theta \geqslant 0) + e^{-\theta} I(\theta < 0)$ and

$\alpha_2(\theta, X) \leqslant e^{\theta}I(\theta \geqslant 0) + I(\theta < 0)$. Thus, since $\exp\{2\delta|\theta| + \theta x - (\theta^2/2)\}$ is uniformly bounded in θ and x on both $\{\theta \geqslant 0, x < 1\}$ and $\{\theta < 0, x \geqslant 1\}$, we see that (4.2) is implied by (6.1).

Now we will show that for $j = 1$ (4.3) is implied by (6.1). (The proof for $j = 2$ follows on the same lines.) Since $u(x) = e^{-x^2/2}$ by (3.5) $\beta_1(\theta, x) < \exp\{(x^2 + 1)/2\}\int_0^1 \exp\{-(x + \theta)\varepsilon_n y\}\, dy \leqslant \alpha_1(\theta, x)\exp((x^2 + 1)/2)$ for $x \geqslant 1$. Thus $E_{X=x}(\beta_1(\theta, x)I(\theta \geqslant 0, x \geqslant 1)) = \exp((x^2 + 1)/2)$, and $E[I_1(X)\{E_X\beta_1(\theta, X)I(\theta \geqslant 0)/f(X)\}^{\delta}] \leqslant 2\int_1^{\infty}(u(x)f(x))^{1-\delta}dx = 2\int_1^{\infty}(p(x))^{1-\delta}\, dx$. But, for a $\xi > 0$, Holder inequality gives

$$(6.2) \qquad \int_1^{\infty}(p(x))^{1-\delta}\, dx \leqslant \left(\int_1^{\infty}x^{-1-\xi}\right)^{\delta}\left(E|X|^{(1+\xi)\delta/(1-\delta)}\right)^{1-\delta} \leqslant \text{const.}\left\{1 + \left(E|\theta|^{(1+\xi)\delta/(1-\delta)}\right)^{1-\delta}\right\}.$$

Also, since $\alpha_1(\theta, x)I(\theta < 0) = e^{-\theta}I(\theta < 0)$, we have

$$(6.3) \qquad E\left[I_1(X)(E_X\beta_1(\theta, X)I(\theta < 0)/f(X))^{\delta}\right] \leqslant 2E\left[I_1(X)(u(X)f(X))^{-\delta}(E_Xe^{-\theta}I(\theta < 0))^{\delta}\right] \leqslant \left\{EI_1(X)(u(X)f(X))^{-\delta t}\right\}^{1/t}\left(E\left[I_1(X)(E_Xe^{-\theta}I(\theta < 0))^{\delta t'}\right]\right)^{1/t'}$$

where $1 > t^{-1} > \max\{(2r)^{-1}(1 + 2r\delta), \delta\}$ and $t' = t/(t - 1)$. Then since $\delta t < 1$ and $\delta t' > 1$, the right-hand side of (6.3) is exceeded by

$$(6.4) \qquad \left(\int_1^{\infty}(p(x))^{1-\delta}\, dx\right)^{1/t}\left(EI_1(X)E_Xe^{-\delta t'\theta}I(\theta < 0)\right)^{1/t'}.$$

As in (6.2), $\int_1^{\infty}(p(x))^{1-\delta t}\, dx \leqslant$ const. $\{1 + (E|\theta|^{(1+\xi)\delta t/(1-\delta t)})^{1-\delta t}\}$. Since $\exp(\theta x - \delta t'\theta - (\theta^2/2))I(\theta < 0, x > 1)$ is uniformly bounded in θ and x,

$$E\left(I_1(X)E_Xe^{-\delta t'\theta}I(\theta < 0)\right) \leqslant \text{const.}\int_1^{\infty}e^{-x^2/2}\, dx.$$

Thus (6.4) and hence, the left-hand side of (6.3) are finite if $E|\theta|^{(1+\xi)\delta t/(1-\delta t)} < \infty$. But this and the right-hand side of (6.2) both are finite by (6.1) and the restrictions on t and δ, since $\xi > 0$ is arbitrary . □

EXAMPLE 6.2. (*Gamma $G(\theta, s)$-family*). *For a $s > 0$, let $u(x) = x^{s-1}I(x > 0)$. Then $a = 0$, $C(\theta) = (\Gamma(s))^{-1}(-\theta)^s$ and*

$$(6.5) \qquad p_{\theta}(x) = \frac{x^{s-1}(-\theta)^s}{\Gamma(s)}e^{\theta x}, \qquad x > 0, \theta < 0.$$

SUBTHEOREM 6.2. *For the family* (6.5), *let $\Theta \subseteq (-\infty, 0)$, and for a $.5 < \delta < 1$, let*

$$(6.6) \qquad E|\theta|^{2r\delta} < \infty.$$

Let ϕ_n be given by (3.10) *with $h(x) = \{xI(x \geqslant 1) + I(0 < x < 1)\}\varepsilon_n$, where $\varepsilon_n = c_0n^{-1/(1+2r)} < \min\{.5, \delta^{-1} - 1\}$. Then*

$$R(\phi_n, G) - R(G) = 0(n^{-2(r\delta-1)/(1+2r)}).$$

PROOF. We will show that for the δ and h in the subtheorem all the conditions of Theorem 4.1 are satisfied. By Holder inequality $E(I_1(X)XE_X|\theta|)^{2r\delta} \leqslant E|X\theta|^{r\delta} = \Gamma(r\delta + s)/\Gamma(s)$. Therefore (4.1) for $t = r\delta$ is implied by (6.6).

By (3.4) $\alpha_2(\theta, x) \equiv 1$, and for $x \geqslant 1$, $\alpha_1(\theta, x) = \exp(-\theta x\varepsilon_n)$. Thus, by Holder inequality, $E[I_2(X)E_X^{2\delta}(|\theta|^r\alpha_2(\theta, X))] \leqslant E|\theta|^{2r\delta} < \infty$ by (6.6), and

$$E\left[I_1(X)X^{2(r\delta-1)}E_X^{2\delta}(|\theta|^r\alpha_1(\theta, X))\right]$$
$$\leqslant \int_1^\infty\int_{-\infty}^0|x\theta|^{s+2r\delta}x^{-1}\exp\{(1-2\delta\varepsilon_n)\theta x\}\,dG(\theta)\,dx < \infty$$

since $2\delta\varepsilon_n < 1$. Thus (4.2) is also implied by (6.6).

Now we will verify (4.3). Since $\varepsilon_n < .5$, from (3.5) for $x \geqslant 1$, $\beta_1(\theta, x) \leqslant \max\{2^{s-1}, 1\} \cdot x^{1-s}\exp(-\theta x\varepsilon_n)$, and for $0 < x < 1$, $\beta_2(\theta, x) \leqslant \max\{x^{1-s}, 1.5\}$. Therefore,

$$E\left[I_2(X)E_X^\delta\beta_2(\theta, X)/f^\delta(X)\right] \leqslant 2\int_0^1(p(x))^{1-\delta}\,dx \leqslant 2\left(\int_0^1 p(x)\,dx\right)^{1-\delta} < 2,$$

where the second inequality follows by Holder inequality. Also, for $(1-\varepsilon_n\delta) > t^{-1} > \delta(\geqslant 1-\delta)$ and $t' = t/(t-1)$ the Holder inequality gives

$$E\left[I_1(X)X^{-2-\delta}E_X^\delta\beta_1(\theta, X)/f^\delta(X)\right] \leqslant \max\{2^{\delta(s-1)}, 1\}l_1 \cdot l_2 \text{ where} \tag{6.7}$$
$$l_1 = E^{1/t}\left[I_1(X)X^{-2-\delta}p^{-\delta}(X)\right]^t, \text{ and}$$
$$l_2 = E^{1/t'}\left[E_Xe^{-\theta X\varepsilon_n}\right]^{\delta t'}$$

Note that $p(x)$ is bounded on $x \geqslant 1$. Therefore, since $\int_1^\infty x^{-(2+\delta)t}\,dx < \infty$ and $0 < \delta t < 1$, $l_1 < \infty$. Also, since $1 < \delta t' < \varepsilon_n^{-1}$, by Holder inequality

$$(l_2)^{t'} \leqslant E(\exp(-\theta X\varepsilon_n\delta t')) < \infty.$$

Thus the left-hand side of (6.7) is also finite. Thus we conclude (4.3). □

EXAMPLE 6.3. (*A family of distributions with densities having infinitely many discontinuities*). *Let* $u(x) = \Sigma_0^\infty((i+1)I(i < x \leqslant i+1))$. *Then* $C(\theta) = \theta(e^\theta - 1)$, $a = 0$ *and*

$$p_\theta(x) = \theta(e^\theta - 1)e^{\theta x}\Sigma_0^\infty((i+1)I(i < x \leqslant i+1)). \tag{6.8}$$

The proof of the following subtheorem follows by arguments identical to those used in the proof of Subtheorem 6.2 with $s = 1$.

SUBTHEOREM 6.3. *Consider the family with density* (6.8). *Let* $\Theta \subseteq (\infty, 0)$. *If, for a* δ *in* $[.5, 1)E|\theta|^{2r\delta} < \infty$, *then* ϕ_n *given by* (3.10), *with* $h(x) = (xI(x \geqslant 1) + I(0 < x < 1))c_0 n^{-1/(1+2r)}$, $0 < c_0 < \min\{.5, \delta^{-1} - 1\}$, *is* a.o. *with a rate* $O(n^{-2(\delta r-1)/(1+2r)})$.

The following corollary, which is an *immediate* consequence of Theorem 5.1, shows that the rates obtained in Subtheorems 6.1–6.3 are *arbitrarily close to the best possible rate* $O(n^{-2(r-1)/(1+2r)})$.

COROLLARY 6.1. *For the families considered in Subtheorems* 6.1–6.3, *and for the* EB *estimators* ϕ_n *there,*

$$R(\phi_n, G) - R(G) \geqslant c_7 n^{-2(r-1)/(1+2r)}$$

for some positive const. c_7 *and for all degenerate* G *and all sufficiently large* n.

7. Remarks and discussions. For each integer $r > 1$ we have exhibited a class of EB estimators for the general exponential family (1.0). Theorem 4.1 gives sufficient conditions under which these estimators are a.o. with rates near $0(n^{-2(r-1)/(1+2r)})$ which is the best possible rate with our estimators according to Theorem 5.1. No assumption on the smoothness of u is made.

The conditions of Theorem 4.1 reduce to a single moment condition on G in the families (6.0), (6.5) and (6.8), among others. In contrast to this, Lin (1975) has verified his conditions only in the family where $u(x) \equiv 1$ or 0 according as $x > 0$ or $x \leqslant 0$, i.e. $p_\theta^{(x)} = -\theta e^{\theta x} I(x > 0)$; and $G(\theta)$ having a density $(\Gamma(t))^{-1}\theta^{t-1}e^{-\theta}I(\theta > 0)$, $0 < t < \infty$, (and thus having all moments finite). His estimators are shown there to be a.o. with the rate $0(n^{-t(3t+2)^{-1}+\varepsilon})$ for some $\varepsilon > 0$, which is near $0(n^{-(1/3)+\varepsilon})$ only for t sufficiently large.

O'Bryan and Susarla (1976) considered EB estimation in the family of *normal* distributions with the support of G in [0, 1]. A novel feature of their work is that they allowed the sample size to vary from one problem to another. Their rate result, however, is $0(n^{-(1/3)+\varepsilon})$ for some $\varepsilon > 0$, in spite of the boundedness of the parameter space.

Subtheorem 6.1 improves Corollary 2.2 of Yu (1971). For a $0 < t < 1$, if $E|\theta|^{t(1-t)^{-1}+} < \infty$, then Yu gives estimators a.o. with a rate $0(n^{-w})$ where $w = 2t(2+t)^{-1}(r-1)(2r+1)^{-1}$. Thus, if $E|\theta|^m < \infty$ for an m sufficiently large, his rates are near $0(n^{-1/3})$ whereas ours are near $0(n^{-1})$.

Subtheorem 6.2 improves and generalizes Corollary 2.1 of Yu (1971). If, for an $m \geqslant 2$, $E|\theta|^m < \infty$ then Yu gives estimators for the $G(\theta, 1)$-family (introduced in Example 6.2) which are a.o. with a rate $0(n^{-t+\eta})$ for every $\eta > 0$, where $t = (m-1.5)/(3m)$. Under this moment condition, we give estimators a.o. for the *general* $G(\theta, s)$-family, $s > 0$, with a rate $0(n^{-t^*+\eta})$ for every $\eta > 0$, where $t^* = (m-2)/(m+1)$.

Subtheorems 6.1, 6.2 and 6.3 improve, respectively, the results in Examples 4.1, 4.2 and 4.3 of Singh (1976). The supports of the prior distributions there are in a bounded interval, and rates are near $0(n^{-2/5})$, whereas, in that situation, our rates here are near $0(n^{-1})$. Moreover, in Example 4.2 of Singh (1976) $G(\theta, s)$-family only for $s \geqslant 1$ is considered.

As we have seen, in each of Examples 6.1–6.3, the conditions of Theorem 4.1 reduce to a *single* moment condition, namely, $E|\theta|^{2r\delta} < \infty$. We expect the same in other exponential families too, provided $c(x)$ in (4.0) is chosen suitably. It seems, however, that each exponential family must be treated separately if precise results there are to be obtained.

Notice that no information about Θ is required to define ϕ_n in (3.10). If, however, for known finite constants A, B, we take $\Theta \subseteq [A, B]$, our proposed estimators, instead of (3.10), would be

$$\phi_n^*(X) = (\hat{f}^{(1)}(X)/\hat{f}(X))_{A,B} \tag{7.0}$$

where $(b)_{c,d}$ is c, b or d according as $b < c$, $c \leqslant b \leqslant d$ or $b > d$. Because of this restriction on Θ, the conditions of Theorem 4.1 would greatly be simplified. For a δ satisfying the conditions of Theorem 4.1, ϕ_n^* is a.o. with rates $0(n^{-2\delta(r-1)/(1+2r)})$, which is slightly better than that given in (4.4) for ϕ_n defined in (3.10).

Acknowledgment. The author wishes to express his deep gratitude to Professor James Hannan for his various helpful comments and suggestions on the original results of this paper.

REFERENCES

[1] HANNAN, JAMES and MACKY, DAVID W. (1971). Empirical Bayes squared error loss estimation of unbounded functionals in exponential families. RM-290. Depart. Statist. Probability, Michigan State Univ.

[2] JOHNS, JR., M. V. (1957). Nonparametric empirical Bayes procedures. *Ann. Math. Statist.* **28** 649–669.

[3] JOHNS, JR., M. V. and VAN RYZIN, J. R. (1971). Convergence rates for empirical Bayes two-action problems I. Discrete case. *Ann. Math. Statist.* **42** 1521–1539.

[4] JOHNS, JR., M. V. and VAN RYZIN, J. R. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. *Ann. Math. Statist.* **43** 934–947.

[5] LAMPERTY, J. (1966). *Probability*. W. A. Benjamin, New York.

[6] LEHMANN, E. L. (1959). *Testing of Statistical Hypothesis*. Wiley, New York.

[7] LIN, P. E. (1975). Rates of convergence in empirical Bayes estimation problems: Continuous case. *Ann. Statist.* **3** 155–164.

[8] LOÈVE, MICHEL (1963). *Probability Theory, 3rd ed.* Van Nostrand, Princeton.

[9] NEYMAN, J. (1962). Two breakthroughs in the theory of statistical decision making. *Rev. Inst. Internat. Statist.* **30** 11–27.

[10] O'BRYAN, THOMAS E. and SUSARLA, V. (1976). Rates in the empirical Bayes estimation problem with nonidentical components: Case of normal distributions. *Ann. Inst. Statist. Math.* **28** 389–397.

[11] ROBBINS, HERBERT. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probability* **1** 157–163. Univ. California Press.

[12] ROBBINS, HERBERT (1963). The empirical Bayes approach to the testing of statistical hypothesis. *Rev. Inst. Internat. Statist.* **31** 195–208.

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

[14] SAMUEL, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Math. Statist.* **34** 1370–1385.

[15] SINGH, R. S. (1974). Estimation of derivatives of average of μ-densities and sequence-compound estimation in exponential families. RM-318, Depart. Statist. Probability, Michigan State Univ.

[16] SINGH, R. S. (1976). Empirical Bayes estimation with convergence rates in non-continuous Lebesgue-exponential families. *Ann. Statist.* **4** 431–439.

[17] SINGH, R. S. (1977a). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *Ann. Statist.* **5** 394–400.

[18] SINGH, R. S. (1977b). Applications of estimators of a density and its derivatives to certain statistical problems. *J. Roy. Statist. Soc. Ser. B* **39** 357–363.

[19] SINGH, R. S. (1978). Nonparametric estimation of derivatives of average of μ-*densities* with convergence rates and applications. *SIAM J. Appl. Math* **35** 637–649.

[20] SINGH, R. S. (1979). Mean squared errors of estimators of a density and its derivatives. *Biometrika* **66** No. 1.

[21] YU, BENITO. (1971). Rates of convergence in empirical Bayes two-action and estimation problems and in sequence-compound estimation problems. RM-279, Depart. Statist. Probability, Michigan State Univ.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF GUELPH
GUELPH, ONTARIO, CANADA N1G 2W1