The Annals of Statistics
1976, Vol. 4, No. 5, 981-989

MONOTONE EMPIRICAL BAYES TESTS FOR THE CONTINUOUS ONE-PARAMETER EXPONENTIAL FAMILY

BY J. C. VAN HOUWELINGEN

University of Utrecht, The Netherlands

Let θ be the natural parameter of a continuous one-parameter exponential family. An empirical Bayes test is constructed for testing $\theta \leq 0$ against $\theta > 0$ with a piecewise linear loss function. Since the problem is monotone, Bayes tests for a given prior distribution can be characterized by a single parameter, e.g., the size of the test under $\theta=0$. Therefore the construction of an empirical Bayes test can be reduced to the construction of an estimator of this parameter. Such an estimator is constructed and the convergence rate of its mean squared error is investigated. The empirical Bayes test constructed in this way has not only nice asymptotic properties, but it can also be applied to small samples because of its (weak) admissibility.

0. Introduction. Johns and van Ryzin (1972) constructed empirical Bayes tests for the continuous one-parameter exponential family and investigated the convergence rates of the corresponding Bayes risks. It will be shown that the monotonicity of the problem can be used to improve the empirical Bayes tests and to simplify the conditions on the prior distribution under which a given convergence rate can be achieved.

1. Description of the problem. The Bayes decision problem is considered that can be described as follows. Consider the pair (X, Θ) of real-valued random variables. The random variable X corresponds to the observable variable, the random variable Θ corresponds to the unknown parameter.

The conditional distribution $F(\cdot \mid \theta)$ of X given $\Theta = \theta$ has density function

$$f(x \mid \theta) = m(x)e^{x\theta}h(\theta) . \tag{1}$$

The functions m and h are known. They determine which exponential family is dealt with.

The random variable Θ takes values in the natural parameter space

$$\Omega = \{\theta \mid h(\theta)^{-1} = \int e^{x\theta}m(x)\, dx < \infty\} \subset \mathbb{R} .$$

The distribution G of Θ is called the prior distribution of the parameter.

The marginal distribution $F(\cdot \mid G)$ of X has density function

$$f(x \mid G) = m(x) \int e^{x\theta}h(\theta)\, dG(\theta) . \tag{2}$$

Without loss of generality it may be assumed that $0 \in \Omega$ and that $\int m(x)\, dx = 1$, or equivalently $h(0) = 1$. Moreover it is assumed that m has the following

Received March 1974; revised March 1976.

AMS 1970 *subject classifications*. Primary 62C10; Secondary 62F05, 62F15.

Key words and phrases. Empirical Bayes, monotone test, convergence rate.

properties:

(i) there exists an interval (a, b) $(-\infty \leqq a < b \leqq \infty)$ such that

$$m(x) > 0 \quad \text{if} \quad x \in (a, b)$$
$$(3) \qquad = 0 \quad \text{if} \quad x \notin (a, b)\,,$$

(ii) m is r-times differentiable on (a, b)

$|m^{(r)}(x)|$ is bounded on each interval (a', b') $(a < a' < b' < b)$.

It is wished to test $H_0\colon \theta \leqq 0$ against $H_1\colon \theta > 0$. The action space $A = \{a_0, a_1\}$. The action a_i corresponds to "accept H_i." The loss function L is defined by

$$L(\theta, a_0) = \max(\theta, 0)\,, \qquad L(\theta, a_1) = \max(-\theta, 0)\,.$$

This loss function is very common in Bayes and empirical Bayes testing. From a practical point of view it is reasonable and it has great mathematical advantages.

Since Bayes tests are essentially nonrandomized, attention can be restricted to nonrandomized tests, i.e. functions on (a, b) to A. The *Bayes risk* of a test ϕ is defined by

$$r(G, \phi) = EL(\Theta, \phi(X))\,.$$

A test is said to be a *Bayes test* w.r.t. G if

$$(4) \qquad r(G, \phi) = r(G) \equiv \inf_{\phi'} r(G, \phi')\,.$$

A Bayes test w.r.t. G can be constructed as follows. Define

$$(5) \qquad \Delta_G(x) = \int_\Omega \{L(\theta, a_0) - L(\theta, a_1)\} f(x\,|\,\theta)\, dG(\theta) = f(x\,|\,G) E\{\Theta\,|\,X = x\}\,.$$

Define

$$(6) \qquad \phi_G(x) = a_1 \quad \text{if} \quad \Delta_G(x) > 0$$
$$= a_0 \quad \text{if} \quad \Delta_G(x) \leqq 0\,.$$

It is easy to show that ϕ_G is a Bayes test w.r.t. G.

Moreover, the following equality holds

$$(7) \qquad r(G, \phi) - r(G) = \int_{\{x\,|\,\phi(x) \neq \phi_G(x)\}} |\Delta_G(x)|\, dx \qquad \text{for each test} \quad \phi\,.$$

The empirical Bayes method (see, for instance, Robbins (1955) and Robbins (1964)) considers the situation where the prior distribution G is unknown but information about G can be obtained from previous comparable experiments. This can be formalized as follows. Let $X_1, \cdots, X_n$ be a sequence of i.i.d. rv's, independent of (X, Θ), with the same marginal distribution as X. Instead of a test based on X alone, a test is used based on $X_1, \cdots, X_n, X$. Such a test procedure is called an *empirical Bayes test* (e.B. test). Formally, an e.B. test is a sequence $(\phi_n)_{n \geqq 1}$ of functions on $(a, b)^{n+1}$ to A. Its Bayes risk is defined by

$$r_n(G, \phi_n) = EL(\Theta, \phi_n(X_1, \cdots, X_n, X))\,.$$

An e.B. test $\phi = (\phi_n)_{n \geqq 1}$ is said to be *asymptotically optimal* (a.o.) for G if

$$\lim_{n\to\infty} \{r_n(G, \phi_n) - r(G)\} = 0 .$$

Notice that $r_n(G,\phi_n) - r(G) \geqq 0$ by definition (4) of $r(G)$.

A practically useful e.B. test should be a.o. for all G in a wide class $\mathscr{G}$ of prior distributions. Moreover, $r_n(G, \phi_n) - r(G)$ should converge to zero rapidly.

Johns and van Ryzin (1972) constructed an estimator Δ_n of Δ_G and defined $\phi = (\phi_n)_{n \geqq 1}$ by

$$\begin{aligned} \phi_n(x_1, \cdots, x_n, x) &= a_1 \qquad \text{if} \quad \Delta_n(x_1, \cdots, x_n, x) > 0 \\ &= a_0 \qquad \text{elsewhere.} \end{aligned}$$

Using the results of Robbins (1964) they could easily prove that ϕ is a.o. for all G satisfying

$$E_G|\Theta| < \infty . \tag{8}$$

They also gave results concerning the convergence rate of $r_n(G, \phi_n) - r(G)$.

Van Houwelingen (1973) showed that the e.B. tests of Johns and van Ryzin can be readily improved if the monotonicity of the problem is used and the results of Karlin and Rubin (1956) are applied.

This method yields an e.B. test that can be written as

$$\begin{aligned} \phi_n(x_1, \cdots, x_n, x) &= a_1 \qquad \text{if} \quad x > c_n(x_1, \cdots, x_n) \\ &= a_0 \qquad \text{elsewhere.} \end{aligned} \tag{9}$$

Such an e.B. test is weakly admissible in the sense that $\phi_n(x_1, \cdots, x_n, \cdot)$ is an admisible test for the non-empirical problem for all $x_1, \cdots, x_n$ and all n. The e.B. test of Johns and van Ryzin lacks this weak admissibility property.

Notice that this kind of weak admissibility does not necessarily coincide with the strong admissibility defined by Meeden (1972). In his definition $\phi = (\phi_n)_{n \geqq 1}$ is admissible if there does not exist an e.B. test $\psi = (\psi_n)_{n \geqq 1}$ such that

$$r_n(G, \psi_n) \leqq r_n(G, \phi_n) \qquad \text{for all} \quad G \quad \text{and all} \quad n ,$$

with strict inequality for at least one G and one n.

In the next sections it will be shown how an e.B. test of type (9) can be constructed. Moreover, its convergence rate will be investigated.

2. Reduction to a simple estimation problem. Consider the Bayes test, defined by (6), in greater detail. By definition (5) of Δ_G, ϕ_G can be written as

$$\begin{aligned} \phi_G(x) &= a_1 \qquad \text{if} \quad E(\Theta \mid X = x) > 0 \\ &= a_0 \qquad \text{if} \quad E(\Theta \mid X = x) \leqq 0 . \end{aligned}$$

Notice that $E(\Theta \mid X = x) = \int \theta e^{x\theta} h(\theta)\, dG(\theta) / \int e^{x\theta} h(\theta) dG(\theta)$ is a nondecreasing function of x because

$$\frac{d}{dx} E(\Theta \mid X = x) = \operatorname{Var}(\Theta \mid X = x) \geqq 0 . \tag{10}$$

To avoid degeneracy the assumption is made that

$$\lim_{x \downarrow a} E(\Theta \mid X = x) < 0 < \lim_{x \uparrow b} E(\Theta \mid X = x) . \tag{11}$$

The following consequences of (11) can easily be verified.

$$\begin{aligned} &\text{(i)} \quad G \quad \text{is nondegenerate,} \\ &\text{(ii)} \quad E(\Theta \mid X = x) \quad \text{is strictly increasing on} \quad (a, b) , \\ &\text{(iii)} \quad \text{there exists a unique} \quad c_G \in (a, b) \quad \text{such that} \quad E(\Theta \mid X = c_G) = 0 , \\ &\text{(iv)} \quad \phi_G(x) = a_1 \quad \text{if} \quad x > c_G , \quad \phi_G(x) = a_0 \quad \text{if} \quad x \leqq c_G . \end{aligned} \tag{12}$$

This shows that the Bayes test ϕ_G is completely determined by the single constant c_G. Let

$$\begin{aligned} &\text{(i)} \quad M(x) = F(x \mid 0) = \int_a^x m(y)\, dy \\ &\text{(ii)} \quad \alpha_G = P_{\theta=0}\{\phi_G(X) = a_0\} = M(c_G) . \end{aligned} \tag{13}$$

Assumption (3i) implies that M has a unique inverse function M^{-1} on [0, 1] to $[a, b]$. Hence, the constant α_G also determines ϕ_G completely. Moreover, it follows from (12iii) that

$$0 < \alpha_G < 1 . \tag{14}$$

An a.o. e.B. test can be constructed by constructing a consistent estimator α_n of α_G, that takes values in [0, 1] and by defining

$$\begin{aligned} \phi_n(x_1, \cdots, x_n, x) &= a_1 \quad \text{if} \quad x > M^{-1}(\alpha_n(x_1, \cdots, x_n)) , \\ &= a_0 \quad \text{elsewhere.} \end{aligned} \tag{15}$$

Notice that ϕ_n is of type (9).

It seems more natural to estimate c_G directly, but the estimation of α_G is easier to handle.

The next lemma relates the convergence rate of $r_n(G, \phi_n) - r(G)$ to the convergence rate of the mean squared error of α_n.

LEMMA 1. *Let ϕ_n be defined by* (15). *Suppose that* (3) *and* (11) *hold and that* $E_G|\Theta| < \infty$, *then*

$$r_n(G, \phi_n) - r(G) = O(E(\alpha_n(X_1, \cdots, X_n) - \alpha_G)^2) \qquad (n \to \infty) . \tag{16}$$

PROOF. Let ϕ_α be defined by

$$\phi_\alpha(x) = a_1 \quad \text{if} \quad x > M^{-1}(\alpha) ; \qquad \phi_\alpha(x) = a_0 \quad \text{if} \quad x \leqq M^{-1}(\alpha) .$$

From (7), (12iv) and (13) it follows that

$$S(\alpha) \equiv r(G, \phi_\alpha) - r(G) = \int_{M^{-1}(\alpha_G)}^{M^{-1}(\alpha)} \Delta_G(x)\, dx .$$

(Notice that $\Delta_G(x) > 0$ for $x > M^{-1}(\alpha_G)$ and $\Delta_G(x) \leqq 0$ for $x \leqq M^{-1}(\alpha_G)$.)

By definition of $\phi = (\phi_n)_{n \geq 1}$,

$$r_n(G, \phi_n) - r(G) = ES(\alpha_n(X_1, \cdots, X_n)) .$$

The definition of S and the differentiability properties (3 ii) of m together imply that

(i) $0 = S(\alpha_G) \leqq S(\alpha) \leqq \int_a^b |\Delta_G(x)|\,dx \leqq E_G|\Theta| = C_1 < \infty$,

(ii) S has a second derivative on $(0, 1)$, $|S''(\alpha)|$ is bounded on each interval (α_1, α_2) with $0 < \alpha_1 < \alpha_2 < 1$,

(iii) $S'(\alpha_G) = 0$.

Since $0 < \alpha_G < 1$ it is possible to choose a constant $\varepsilon > 0$ such that $I_\varepsilon \equiv [\alpha_G - \varepsilon, \alpha_G + \varepsilon] \subset (0, 1)$.

Let $C_2 = \sup_{\alpha \in I_\varepsilon} |S''(\alpha)|/2$.

Let F_n be the distribution function of $\alpha_n(X_1, \cdots, X_n)$, then

$$\begin{aligned} r_n(G, \phi_n) - r(G) &= \int_0^1 S(\alpha)\,dF_n(\alpha) = \int_{\alpha \in I_\varepsilon} S(\alpha)\,dF_n(\alpha) + \int_{\alpha \notin I_\varepsilon} S(\alpha)\,dF_n(\alpha) \\ &\leqq C_2 \int_{\alpha \in I_\varepsilon} (\alpha - \alpha_G)^2\,dF_n(\alpha) + C_1 \int_{\alpha \notin I_\varepsilon} dF_n(\alpha) \\ &\leqq (C_2 + \varepsilon^{-2} C_1) E(\alpha_n(X_1, \cdots, X_n) - \alpha_G)^2 . \end{aligned}$$

This completes the proof.

Note that the condition $E_G|\Theta| < \infty$ was given in (8) as a sufficient condition for asymptotic optimality.

From Lemma 1 it is clear that $\phi = (\phi_n)_{n \geqq 1}$ is a.o. for all G satisfying $E_G|\Theta| < \infty$, provided that α_n is a consistent estimator of α_G. It appears that attention can be restricted to the construction of an estimator α_n of α_G.

3. Construction of an estimator of α_G. In order to estimate α_G it is useful to make the transformation $Z_i = M(X_i)$. If X_i has the distribution given by (2) and (3), Z_i is a random variable on $(0, 1)$ with density function

$$u(z) = \int_\Omega e^{\theta M^{-1}(z)} h(\theta)\,dG(\theta) . \tag{17}$$

The differentiability properties (3) of m together with (12) imply that u has the following properties.

$$\begin{aligned} &\text{(i)} \quad u'(z) < 0 \quad \text{for} \quad 0 < z < \alpha_G , \\ &\qquad\quad\;\; = 0 \quad \text{for} \quad z = \alpha_G , \\ &\qquad\quad\;\; > 0 \quad \text{for} \quad \alpha_G < z < 1 . \end{aligned} \tag{18}$$

(ii) There exist constants $\varepsilon, \delta > 0$ such that $u''(z) \geqq \varepsilon$ for all $z \in [\alpha_G - \delta, \alpha_G + \delta]$.

(iii) u is (r + 1)-times differentiable on $(0, 1)$ and $|u^{(r+1)}(z)|$ is bounded on every interval (z_1, z_2) with $0 < z_1 < z_2 < 1$.

To prove these properties, notice that

$$\begin{aligned} &\text{(i)} \quad m(M^{-1}(z))u'(z) = u(z)E(\Theta \mid X = M^{-1}(z)) , \\ &\text{(ii)} \quad m(M^{-1}(z))^2 u''(z) = u(z)E(\Theta^2 \mid X = M^{-1}(z)) - m'(M^{-1}(z))u'(z) . \end{aligned} \tag{19}$$

From (i) it follows that (18 i) holds, from (ii) it follows that (18 ii) holds. Property (18 iii) can be proved by repeated differentiation.

From (18 i) it follows that α_G can be written as

$$\alpha_G = \varepsilon + \int_\varepsilon^{1-\varepsilon} I_{(-\infty,0]}(u'(z))\,dz \qquad \text{if} \quad \alpha_G \in (\varepsilon, 1-\varepsilon) \qquad 0 < \varepsilon < \tfrac{1}{2}\,.$$

Suppose that $u_n'(z)$ is an estimator of $u'(z)$ based on $Z_1, \cdots, Z_n$ and that $(\varepsilon_n)_{n\geqq 1}$ is a decreasing sequence tending to zero.

An estimator α_n of α_G can be constructed as follows:

$$\alpha_n = \varepsilon_n + \int_{\varepsilon_n}^{1-\varepsilon_n} I_{(-\infty,0]}(u_n'(z))\,dz\,. \tag{20}$$

Let

$$\sigma_n^2(z) = E\{u_n'(Z_1, \cdots, Z_n, z) - u'(z)\}^2\,. \tag{21}$$

The next lemma relates the mean squared error of α_n to $\sigma_n^2(z)$.

LEMMA 2. *Let α_n be defined by* (20) *and σ_n by* (21). *Suppose that* (3) *and* (11) *hold. Let n_0 be such that $\alpha_G \in (\varepsilon_n, 1-\varepsilon_n)$ for $n \geqq 0$. Then,*

$$(22) \qquad E(\alpha_n(Z_1, \cdots, Z_n) - \alpha_G)^2 \leqq \left\{\int_{\varepsilon_n}^{1-\varepsilon_n} \min(1, \sigma_n(z)/|u'(z)|)\,dz\right\}^2 \qquad \textit{for} \quad n \geqq n_0\,.$$

PROOF. Define $I(z) = I_{(-\infty,0]}(u'(z))$ and $I_n(z) = I_{(-\infty,0]}(u_n'(z))$. Let $n \geqq n_0$, then

$$E(\alpha_n - \alpha_G)^2 = \int_{\varepsilon_n}^{1-\varepsilon_n}\int_{\varepsilon_n}^{1-\varepsilon_n} E[\{I_n(z_1) - I(z_1)\}\{I_n(z_2) - I(z_2)\}]\,dz_1\,dz_2\,.$$

From Schwarz' inequality and the equality

$$\{I_n(z) - I(z)\}^2 = |I_n(z) - I(z)|$$

it follows that

$$E(\alpha_n - \alpha_G)^2 \leqq \left\{\int_{\varepsilon_n}^{1-\varepsilon_n} \{E|I_n(z) - I(z)|\}^{\frac{1}{2}}\,dz\right\}^2.$$

Since $E|I_n(z) - I(z)| \leqq P(|u_n'(z) - u'(z)| \geqq |u'(z)|) \leqq \min(1, \sigma_n^2(z)/|u'(z)|^2)$,

$$E(\alpha_n - \alpha_G)^2 \leqq \left\{\int_{\varepsilon_n}^{1-\varepsilon_n} \min(1, \sigma_n(z)/|u'(z)|)\,dz\right\}^2.$$

This completes the proof.

Lemma 2 is stated because in most cases it is hard to find the distribution of $u_n'(z)$. It is only possible to give upperbounds for its bias and variance.

In order to construct an estimator $u_n'(z)$ of $u'(z)$ ideas of Schuster (1969) and Johns and van Ryzin (1972) are combined. Schuster pointed out that the derivative of a density function can be estimated by the derivative of a suitable kernel estimator. Johns and van Ryzin introduced a special type of kernel. Let K be a function on $\mathbb{R}$ such that

$$(23) \qquad \begin{aligned} &\text{(i)} && K(x) = 0 \quad \text{if} \quad x \notin (0, 1)\,, \\ &\text{(ii)} && \textstyle\int_0^1 K(x)\,dx = 1\,, \\ &\text{(iii)} && K \quad \text{is absolutely continuous,} \quad K(x) = \textstyle\int_{-\infty}^x K'(y)\,dy\,, \\ &\text{(iv)} && |K'(x)| \quad \text{is bounded on} \quad \mathbb{R}\,, \\ &\text{(v)} && \textstyle\int_0^1 x^i K(x)\,dx = 0 \quad \text{for} \quad i = 1, \cdots, r-1\,. \end{aligned}$$

For example, take $K(x) = \sum_{i=1}^{r+1} a_i x^i$ for $x \in [0, 1]$ with $a_1, \cdots, a_{r+1}$ such that $K(1) = 0$, $\int_0^1 K(x)\,dx = 1$ and $\int_0^1 x^i K(x)\,dx = 0$ for $i = 1, \cdots, r-1$.

Let $(h_n)_{n\geqq 1}$ be a decreasing sequence of constants tending to zero ($h_1 < \frac{1}{2}$).

Define

$$(24)\qquad u_n'(z_1, \cdots, z_n, z) = -n^{-1}h_n^{-2} \sum_{i=1}^n K'((z_i - z)/h_n) \qquad \text{for} \quad 0 < z \leqq \tfrac{1}{2},$$
$$= n^{-1}h_n^{-2} \sum_{i=1}^n K'((z - z_i)/h_n) \qquad \text{for} \quad \tfrac{1}{2} < z < 1.$$

This yields an estimator of $u'(z)$. Simple calculations (see Schuster (1969) and van Houwelingen (1973)) yield that

$$(25)\qquad \sigma_n^2(z) = E(u_n'(z) - u'(z))^2 \leqq K_1 n^{-1} h_n^{-3} v_n(z) + K_2 h_n^{2r} w_n(z)^2.$$

Here,

(i) K_1 and K_2 are constants that only depend on K (and r),

$$(26)\qquad \begin{aligned} &\text{(ii)}\quad v_n(z) = \sup_{0\leqq y\leqq h_n} u(z+y) \quad \text{for} \quad 0 < z \leqq \tfrac{1}{2}, \\ &\qquad\qquad = \sup_{0\leqq y\leqq h_n} u(z-y) \quad \text{for} \quad \tfrac{1}{2} < z < 1, \\ &\text{(iii)}\quad w_n(z) = \sup_{0\leqq y\leqq h_n} |u^{(r+1)}(z+y)| \quad \text{for} \quad 0 < z \leqq \tfrac{1}{2}, \\ &\qquad\qquad = \sup_{0\leqq y\leqq h_n} |u^{(r+1)}(z-y)| \quad \text{for} \quad \tfrac{1}{2} < z < 1. \end{aligned}$$

It is obvious that the convergence rate of the right-hand side of (25) is optimal for $h_n = O(n^{-1/(2r+3)})$. For the sake of simplicity, take

$$(27)\qquad h_n = Cn^{-1/(2r+3)} \qquad 0 < C < \tfrac{1}{2}.$$

Then,

$$(28)\qquad \sigma_n^2(z) \leqq n^{-2r/(2r+3)}\{C_1 v_n(z) + C_2 w_n(z)^2\}.$$

Note that the maximal value of r is determined by the differentiability properties of m (see 3 ii). Inequality (28) suggests that it would be advantageous to take r as large as possible in the construction of $u_n'(z)$, i.e., in the choice of the kernel K. Generally this is not true, because large value of r yields an estimator $u_n'(z)$ that may be a wildly oscillating function of z, especially when n is small. So, it is advisable to use a moderate value of r, even when m is analytic on (a, b).

A combination of Lemma 2 and inequality (28) yields the main theorem, which gives a precise convergence rate statement for $E(\alpha_n - \alpha_G)^2$ and hence for $r(G, \phi_n) - r(G)$, under a simple condition on G. In order to get this convergence rate the sequence $(\varepsilon_n)_{n\geqq 1}$ must be properly chosen, namely in such a way that

$$(29)\qquad \begin{aligned} &\text{(i)}\quad \lim_{n\to\infty} \varepsilon_n = 0, \\ &\text{(ii)}\quad \sup_{z\in[\varepsilon_n, 1-\varepsilon_n]} m(M^{-1}(z)) \cdot \sup_{z\in[\varepsilon_n, 1-\varepsilon_n]} \left|\frac{d^{r+1}}{dz^{r+1}} M^{-1}(z)\right|^{r+1} \\ &\qquad = O(\log n) \qquad n \to \infty. \end{aligned}$$

THEOREM. *Suppose that* (3) *and* (11) *hold. Let* α_n *be the estimator of* α_G *constructed above* (*see* (20), (23), (24), (27) *and* (29)). *If*

$$(30)\qquad E_G|\Theta|^{r+1} < \infty,$$

then

$$(31)\qquad E(\alpha_n - \alpha_G)^2 = O(n^{-2r/(2r+3)} \log^2 n) \qquad n \to \infty.$$

REMARKS.

(i) For the normal case condition (29 ii) is satisfied if $1/\varepsilon_n = O((\log n)^{1/(r+1)^2})$. Generally, it is recommended to take a sequence $(\varepsilon_n)_{n\geq 1}$ with a very small first term ε_1 and with a very slow convergence to zero, e.g. $\varepsilon_n = 10^{-10}/\log\log(e+n)$.

(ii) Condition (30) offers another argument for the advisability of taking a moderate value of r in the construction of α_n through $u_n'(z)$.

(iii) The result of the theorem is closely related to the results of Johns and van Ryzin (1972). They get a covergence rate $O(n^{-\delta r/(2r+3)})$ with $0 < \delta < 2$ if m is $(r+1)$-times differentiable and if G satisfies some conditions which for special cases turn out to be equivalent to moment conditions like (30).

PROOF OF THE THEOREM. Throughout the proof C, C_1 and C_2 are constants that may have different values in different positions. Define

$$L_n(z) = \min(1, n^{-\beta}\{C_1 v_n(z)^{\frac{1}{2}} + C_2 w_n(z)\}/|u'(z)|) .$$

with $\beta = r/(2r+3)$.

From Lemma 2 and inequality (28) it follows that it suffices to show that $\int_{\varepsilon_n}^{1-\varepsilon_n} L_n(z)\,dz = O(n^{-\beta}\log n)$. Since $0 < \alpha_G < 1$ it is possible to choose a constant λ such that $0 < \lambda < \alpha_G < 1-\lambda < 1$. If n is large enough, $\varepsilon_n < \lambda$. Then, the integral can be split into three parts.

$$\int_{\varepsilon_n}^{1-\varepsilon_n} L_n(z)\,dz = \int_{\varepsilon_n}^{\lambda} L_n(z)\,dz + \int_{\lambda}^{1-\lambda} L_n(z)\,dz + \int_{1-\lambda}^{1-\varepsilon_n} L_n(z)\,dz .$$

A symmetry argument shows that it suffices to prove that

(a) $$\int_{\lambda}^{1-\lambda} L_n(z)\,dz = O(n^{-\beta}\log n) ,$$

(b) $$\int_{\varepsilon_n}^{\lambda} L_n(z)\,dz = O(n^{-\beta}\log n) .$$

PROOF OF (a). From (18 iii) it follows that v_n and w_n are bounded on $[\lambda, 1-\lambda]$. Thus $\int_{\lambda}^{1-\lambda} L_n(z)\,dz \leqq \int_{\lambda}^{1-\lambda} \min(1, Cn^{-\beta}|u'(z)|^{-1})\,dz$. From (18 i) and (18 ii) it follows that $|u'(z)|^{-1} \leqq C_1|z-\alpha_G|^{-1}$ for $z \in [\lambda, 1-\lambda]$. This yields $\int_{\lambda}^{1-\lambda} L_n(z)\,dz \leqq \int_{\lambda}^{1-\lambda} \min(1, Cn^{-\beta}|z-\alpha_G|^{-1})\,dz = C_1 n^{-\beta} + C_2 n^{-\beta}\log n$.
This completes the proof of (a).

PROOF OF (b). Since $u(z)$ is a nonincreasing function of z for $0 < z < \alpha_G$, it follows from the definition of v_n that there exists a constant n_1 such that

$$v_n(z) = u(z) \geqq u(\alpha_G) \qquad \text{for} \quad 0 < z \leqq \lambda \quad \text{and} \quad n \geqq n_1 .$$

Moreover it follows from (19 i) that

$$|u'(z)|^{-1} = m(M^{-1}(z))u(z)^{-1}|E(\Theta \mid X = M^{-1}(z)|^{-1} .$$

From the monotonicity of $E(\Theta \mid X = M^{-1}(z))$ it follows that

$$|u'(z)|^{-1} \leqq CA_n u(z)^{-1} \qquad \text{for} \quad z \in [\varepsilon_n, \lambda] .$$

Here, $A_n = \sup_{z\in[\varepsilon_n, 1-\varepsilon_n]} m(M^{-1}(z))$.
The results above imply that

$$\begin{aligned} L_n(z) &\leqq n^{-\beta}\{C_1 v_n(z)^{\frac{1}{2}} + C_2 w_n(z)\}/|u'(z)| \\ &\leqq A_n n^{-\beta}\{C_1 + C_2 w_n(z)\} . \end{aligned}$$

Define $B_n = \sup_{z\in[\varepsilon_n, 1-\varepsilon_n]} |(d^{r+1}/dz^{r+1})M^{-1}(z)|^{r+1}$. From (29ii) it follows that it suffices to show that

$$\text{(c)} \qquad \int_{\varepsilon_n}^{\lambda} w_n(z)\,dz = O(B_n) \qquad n \to \infty\,.$$

PROOF OF (c). From (17) it follows that

$$|u^{(r+1)}(z)| \leqq C \sup_{1\leqq i\leqq r+1} \int |\theta|^i e^{\theta M^{-1}(z)} h(\theta)\,dG(\theta) \sup_{1\leqq i\leqq r+1} \left|\frac{d^i}{dz^i} M^{-1}(z)\right|^{r+1}.$$

Let f be an arbitrary function on $(0, a]$, then

$$|f(z)| \leqq |f(a)| + a \sup_{y\in[z,a]} |f'(y)|\,.$$

Repeated use of this inequality yields

$$|u^{(r+1)}(z)| \leqq C \sup_{1\leqq i\leqq r+1} E_i(z) B_n \qquad \text{for} \quad \varepsilon_n \leqq z \leqq \lambda\,,$$

where

$$E_i(z) = \int |\theta|^i e^{\theta M^{-1}(z)} h(\theta)\,dG(\theta)\,.$$

Thus

$$\int_{\varepsilon_n}^{\lambda} w_n(z)\,dz \leqq C B_n \sup_{1\leqq i\leqq r+1} \int_0^{\lambda} \sup_{0\leqq y\leqq h_n} E_i(z+y)\,dz\,.$$

A similar argument to the one used in the proof of (18i) shows that there exists a constant $c_i \in [0, 1]$, such that

$$E_i'(z) < 0 \qquad \text{for} \quad 0 < z < c_i\,,$$
$$E_i'(z) > 0 \qquad \text{for} \quad c_i < z < 1\,.$$

This implies that in the neighborhood of zero $E_i(z)$ is either bounded or decreasing. Moreover $\int_0^1 E_i(z)\,dz = E|\Theta|^i < \infty$.

Hence, it follows that

$$\sup_{1\leqq i\leqq r+1} \int_0^{\lambda} \sup_{0\leqq y\leqq h_n} E_i(z+y)\,dz \leqq \sup_{1\leqq i\leqq r+1} \int_0^{\lambda} \sup_{0\leqq y\leqq h_1} E_i(z+y)\,dz < \infty\,.$$

This completes the proof.

REFERENCES

[1] VAN HOUWELINGEN, J. C. (1973). On empirical Bayes rules for the continuous one-parameter exponential family. Doctoral thesis at the University of Utrecht.

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

[3] KARLIN, S. and RUBIN, H. (1956). The theory of statistical decision procedures for distributions with a monotone likelihood ratio. *Ann. Math. Statist.* **27** 272–299.

[4] MEEDEN, G. (1972). Some admissible empirical Bayes procedures. *Ann. Math. Statist.* **43** 96–101.

[5] ROBBINS, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 157–164, Univ. of California Press.

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

[7] SCHUSTER, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40** 1187–1195.

DEPARTMENT OF MATHEMATICAL STATISTICS
UNIVERSITY OF UTRECHT
BUDAPESTLAAN 6
UTRECHT, THE NETHERLANDS