Ann. Inst. Statist. Math.
Vol. 54, No. 1, 83–99 (2002)

LOCALLY ADAPTIVE WAVELET EMPIRICAL BAYES ESTIMATION OF A LOCATION PARAMETER *

MARIANNA PENSKY

Department of Mathematics, University of Central Florida, Orlando, FL 32816-1364, U.S.A.

(Received December 10, 1999; revised October 18, 2000)

Abstract. The traditional empirical Bayes (EB) model is considered with the parameter being a location parameter, in the situation when the Bayes estimator has a finite degree of smoothness and, possibly, jump discontinuities at several points. A nonlinear wavelet EB estimator based on wavelets with bounded supports is constructed, and it is shown that a finite number of jump discontinuities in the Bayes estimator do not affect the rate of convergence of the prior risk of the EB estimator to zero. It is also demonstrated that the estimator adjusts to the degree of smoothness of the Bayes estimator, locally, so that outside the neighborhoods of the points of discontinuities, the posterior risk has a high rate of convergence to zero. Hence, the technique suggested in the paper provides estimators which are significantly superior in several respects to those constructed earlier.

Key words and phrases: Empirical Bayes estimation, adaptive estimation, wavelet, posterior and prior risks.

1. Introduction

Statistical estimation based on wavelet approximation has become more and more popular in the last decade. One of the reasons for this popularity is that wavelets provide an accurate and a rapidly converging approximation for a function of an unknown degree of smoothness. The present paper develops wavelet techniques for the solution of an empirical Bayes (EB) estimation problem.

The traditional EB model is as follows. One observes independent two-dimensional random vectors $(X_1, \Theta_1), \ldots, (X_n, \Theta_n)$, where each Θ_i is distributed according to some unknown prior density function g and, given $\Theta_i = \theta$, X_i has a known conditional density function $q(x \mid \theta)$, $\theta \in (a, b) \subseteq (-\infty, \infty)$, $x \in (c, d) \subseteq (-\infty, \infty)$. In each pair, the first component is observable, but the second is not. After the $(n+1)$-th observation, $y \equiv X_{n+1}$, is made, the objective is to estimate $t \equiv \Theta_{n+1}$.

In the case of continuous random variables X and θ, the majority of practical applications of the EB model deal with the situation where θ is a location parameter or a scale parameter. The present paper considers the case when θ is a location parameter, i.e.

$$(1.1) \qquad q(x \mid \theta) = q(x - \theta), \qquad x, \theta \in (-\infty; \infty).$$

Note that the case of a scale parameter θ can be reduced to the case of a location parameter by a simple reparametrization.

*This work was supported in part by National Science Foundation (NSF), grant DMS-0004173.

Denote by $\boldsymbol{E}_q$, $\boldsymbol{E}_g$, $\boldsymbol{E}_p$ and $\boldsymbol{E}$ the mathematical expectations with respect to the densities $q(x-\theta)$, $g(\theta)$,

$$p(x) = \int_{-\infty}^{\infty} q(x-\theta)g(\theta)d\theta \tag{1.2}$$

and $\prod_{i=1}^{n} p(x_i)$, respectively, and

$$\Phi(x) = \int_{-\infty}^{\infty} (x-\theta)q(x-\theta)g(\theta)d\theta. \tag{1.3}$$

If we knew the prior density $g(\theta)$, then the Bayes estimator of θ_{n+1} under the squared loss would have the form

$$\beta(y) = \frac{\int_{-\infty}^{\infty} \theta q(y-\theta)g(\theta)d\theta}{p(y)} = y - \frac{\Phi(y)}{p(y)}. \tag{1.4}$$

As the prior density $g(\theta)$ is unknown, we construct an empirical Bayes estimator $\hat{\beta}_n(y) \equiv \hat{\beta}_n(y; X_1, X_2, \ldots, X_n)$ as the estimator of (1.4) based on observations $X_1, X_2, \ldots, X_n$.

An EB estimator $\hat{\beta}_n(y)$ may be characterized by its posterior risk

$$R(y; \hat{\beta}_n) = (p(y))^{-1}\boldsymbol{E}\int_{-\infty}^{\infty} (\hat{\beta}_n(y) - \theta)^2 q(y \mid \theta)g(\theta)d\theta,$$

or by its prior risk

$$R(\hat{\beta}_n) = \boldsymbol{E}_p R_n(y) = \int_{-\infty}^{\infty} R(y; \hat{\beta}_n)p(y)dy.$$

Note that both $R(y; \hat{\beta}_n)$ and $R(\hat{\beta}_n)$ can be partitioned into sums of two components. The first components of these sums are, respectively, the posterior risk $R(y; \beta) = (p(y))^{-1} \int_{-\infty}^{\infty} (\beta(y) - \theta)^2 q(y-\theta)g(\theta)d\theta$ and the prior risk $R(\beta) = \int_{-\infty}^{\infty} R(y; \beta)p(y)dy$ of the Bayes estimator (1.4), and they both are independent of $\hat{\beta}_n(y)$. Thus, the quality of EB estimators is measured by the second components

$$R_n(y) = \boldsymbol{E}(\hat{\beta}_n(y) - \beta(y))^2 \tag{1.5}$$

and

$$R_n = \int_{-\infty}^{\infty} R_n(y)p(y)dy. \tag{1.6}$$

In what follows, we will make no parametric assumptions on $g(\theta)$, which means that we restrict ourselves to nonparametric EB estimation. The nonparametric EB model was previously studied extensively and EB estimators were constructed for different classes of conditional densities $q(x \mid \theta)$. These include: a one-parameter exponential family (Singh (1979)), a location-parameter family (Bickel and Klaassen (1986), Singh and Prasad (1989), Penskaya (1992), Pensky (1997*a*, 1999)), a scale-parameter family (Singh and Wei (1992), Pensky (1996)), a family of uniform density functions (Nogami (1988), Huang (1997)), a family of Pareto densities (Tiwari and Zalkikar (1990)), etc. Some attempts were also made to work out an approach to the problem which is suitable for various classes of conditional densities (see, for example, Walter (1981), Robbins (1983), Pensky (1997*b*), also, Maritz and Lwin (1989) and Carlin and Louis (1996)).

The present paper develops the results of the author (see Pensky (1997c, 1998, 1999)) on construction of EB estimators using wavelet expansions. In the paper by Pensky (1997c) Meyer-type wavelets were used for construction of adaptive EB estimators in the case when $g(\theta)$ has a bounded norm in some Sobolev space. It was shown that the posterior risk (1.5) of these estimators converges to zero at a rate which is $(\ln n)$ times greater than the optimal convergence rate. This logarithmic factor is the "adaptation price" which is almost inevitable in many situations (see Brown and Low (1996)). Moreover, it was demonstrated that the estimators adjust automatically to the case when $g(\theta)$ is supersmooth (i.e. the Fourier transform $\tilde{g}(\omega)$ of $g(\theta)$ has an exponential rate of descent as $|\omega| \to \infty$).

However, although the estimators obtained by Pensky (1997c) have these very useful properties, they have a serious shortcoming. For instance, consider the situation when $q(x)$ is infinitely differentiable everywhere except for one point where it has a jump discontinuity (the simplest example of this kind is the exponential distribution). Assume also that $g(\theta)$ is a function of the similar kind. Then the EB estimators based on Meyer-type wavelets fail to provide high rate of convergence, since ANY discontinuity affects the convergence rate of an expansion based on wavelets with infinite supports. In fact, in the situation when both $q(x)$ and $g(\theta)$ are exponential, the posterior risk (1.5) of an adaptive wavelet estimator tends to zero at a rate of $O(n^{-3/4} \ln n)$. However, it is easy to see that, in this case, both $p(y)$ and $\Phi(y)$ are infinitely differentiable everywhere except at the point $y = 0$, where the first derivative of $p(y)$ and the second derivative of $\Phi(y)$ are discontinuous. Therefore, everywhere except for the neighborhood of the point $y = 0$, the EB estimator would achieve a higher rate of convergence than $O(n^{-3/4} \ln n)$ if wavelets with bounded supports, for example, Daubechies wavelets or B-spline wavelets (see Daubechies (1992), Meyer (1993), Walter (1994), Hernández and Weiss (1996), Vidakovic (1999)) were used for construction of the EB estimator. Of course, the situation described above can arise only if $q(x)$ has a finite degree of smoothness, which implies that $\tilde{q}(\omega)$ has a polynomial descent as $|\omega| \to \infty$.

In what follows, we consider the case when both $q(x)$ and $g(\theta)$ have finite degrees of smoothness. We construct a nonlinear wavelet EB estimator based on wavelets with bounded supports. We show that these estimators provide local adaptivity to the unknown smoothness of $\beta(y)$. Moreover, we demonstrate that a finite number of jump discontinuities of the functions $p(y)$ and $\Phi(y)$ do not influence the rate of convergence of the prior risk R_n to zero. This phenomenon was reported earlier by Hall and Patil (1995) for estimators of a probability density function. Application of the techniques suggested in the paper allows one to obtain estimators with significantly higher convergence rate than those derived by Singh and Prasad (1989).

It should be mentioned that the present paper continues the line of papers which apply wavelet methods to curve estimation (see, for example, Kerkyacharian and Picard (1992), Antoniadies *et al.* (1994), Donoho (1994), Masry (1994), Donoho and Johnstone (1995), Hall and Patil (1995), Donoho *et al.* (1996), Efromovich (1997), Abramovich and Silverman (1998), Hall *et al.* (1998)).

2. Construction of the empirical Bayes estimators in the location parameter case

Throughout the paper we use the notation $\tilde{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx$ for the Fourier transform of a function $f(x)$. Denote the sets of all integer numbers by $\boldsymbol{Z}$, and let $\|f\|_c = \sup_y |f(y)|$ and $\|f\|_{L_k} = \{\int_{-\infty}^{\infty} |f(x)|^k dx\}^{1/k}$. Assume that $g(\theta)$ is square integrable, and that $q(x)$ satisfies the following conditions

C1. $\|(|y|+1)q(y)\|_c < \infty$;

C2. $|\tilde{q}'(\omega)| \le C_q |\tilde{q}(\omega)|$, $Q = \max(C_q, 1)$.

Since the Bayes estimator (1.4) is presented via the ratio of two unknown functions, we break the estimation process into two steps. First, we estimate $\Phi(y)$ and $p(y)$, and after that we construct an estimator of the ratio based on the estimators of the numerator $\Phi(y)$ and the denominator $p(y)$.

For the sake of construction of wavelet approximations of $\Phi(y)$ and $p(y)$, consider a real-valued scaling function $\varphi(x)$ and a corresponding real-valued wavelet $\psi(x)$ in $L^2(-\infty, \infty)$. Assume that both $\varphi(x)$ and $\psi(x)$ are bounded and compactly supported, namely,

$$\operatorname{supp}\varphi \subset [-A, A], \qquad \operatorname{supp}\psi \subset [-B, B]. \tag{2.1}$$

Assume also that the wavelet is ϱ-regular, i.e. both $\varphi(x)$ and $\psi(x)$ are ϱ times continuously differentiable. The former implies the following zero moment conditions

$$\int x^k \psi(x) dx = 0, \qquad 0 \le k \le \varrho, \quad \varrho > 2. \tag{2.2}$$

Note that the last condition implies that the Fourier transforms of $\Phi(y)$ and $p(y)$ are absolutely integrable.

To simplify the derivation of the estimators of $p(y)$ and $\Phi(y)$ and subsequent calculation of the risks (1.5) and (1.6) of the EB estimator, denote

$$f_i(y) = \int_{-\infty}^{\infty} (y-\theta)^i q(y-\theta) g(\theta) d\theta, \qquad i = 0, 1, \tag{2.3}$$

so that $f_0(y) = p(y)$, $f_1(y) = \Phi(y)$. Observe that $f_i(y)$, $i = 0, 1$, can be written as

$$f_i(y) = \sum_{k \in \boldsymbol{Z}} a_{m,k}^{(i)} \varphi_{m,k}(y) + \sum_{j=m}^{\infty} \sum_{k \in \boldsymbol{Z}} b_{j,k}^{(i)} \psi_{j,k}(y), \qquad i = 0, 1. \tag{2.4}$$

Here

$$\varphi_{m,k}(x) = 2^{m/2} \varphi(2^m x - k), \qquad \psi_{j,k}(x) = 2^{j/2} \psi\left(2^j x - k\right), \tag{2.5}$$

with $k \in \boldsymbol{Z}$ and $j \ge m$, and the coefficients in (2.4) have the forms

$$a_{m,k}^{(i)} = \int_{-\infty}^{\infty} f_i(x) \varphi_{m,k}(x) dx, \qquad b_{j,k}^{(i)} = \int_{-\infty}^{\infty} f_i(x) \psi_{j,k}(x) dx. \tag{2.6}$$

Let us construct the estimators of $a_{m,k}^{(i)}$, $i = 0, 1$. Notice that if we found a function $u_{m,k}^{(i)}(x)$ satisfying the equation

$$\int_{-\infty}^{\infty} q(x-\theta) u_{m,k}^{(i)}(x) dx = \int_{-\infty}^{\infty} (x-\theta)^i q(x-\theta) \varphi_{m,k}(x) dx, \qquad i = 0, 1, \tag{2.7}$$

for any θ, then

$$a_{m,k}^{(i)} = \boldsymbol{E}_p u_{m,k}^{(i)}(X), \qquad i = 0, 1. \tag{2.8}$$

Applying the Fourier transform to both sides of the equation (2.7), we obtain

$$u_{m,k}^{(i)}(x) = 2^{m/2} U_m^{(i)}(2^m x - k), \tag{2.9}$$

where $U_m^{(0)}(x) = \varphi(x)$ and $U_m^{(1)}(x)$ is the inverse Fourier transform of the function

$$\tilde{U}_m^{(1)}(\omega) = i\tilde{q}'(-2^m\omega)[\tilde{q}(-2^m\omega)]^{-1}\tilde{\varphi}(\omega), \tag{2.10}$$

where i is the imaginary unit and $\tilde{q}'(-2^m\omega)$ is the value of $\tilde{q}'(\cdot)$ at the point $(-2^m\omega)$. The inverse Fourier transform of the function (2.10) exists since $\tilde{\varphi}(\omega)$ has a bounded support and $|\tilde{q}'(\omega)/\tilde{q}(\omega)|$ is bounded. Combining the formulae (2.8)–(2.10), we estimate the coefficients $a_{m,k}^{(i)}$ by

$$\hat{a}_{m,k}^{(i)} = n^{-1}2^{m/2}\sum_{l=1}^{n} U_m^{(i)}(2^m X_l - k). \tag{2.11}$$

Similarly, introducing a function $V_j^{(i)}(x)$, where $V_j^{(0)}(x) = \psi(x)$ and $V_j^{(1)}(x)$ is the inverse Fourier transform of the function

$$\tilde{V}_j^{(1)}(\omega) = i\tilde{q}'(-2^j\omega)[\tilde{q}(-2^j\omega)]^{-1}\tilde{\psi}(\omega), \qquad i = \sqrt{-1}, \tag{2.12}$$

we can estimate $b_{j,k}^{(i)}$ by

$$\hat{b}_{j,k}^{(i)} = n^{-1}\sum_{l=1}^{n} 2^{j/2} V_j^{(i)}(2^j X_l - k), \qquad i = 0, 1. \tag{2.13}$$

Combining (2.4)–(2.13) together, we arrive at the nonlinear wavelet estimator of $f_i(y)$

$$\hat{f}_i(y) = \sum_{k\in \boldsymbol{Z}} \hat{a}_{m,k}^{(i)}\varphi_{m,k}(y) + \sum_{j=m}^{m+r}\sum_{k\in \boldsymbol{Z}} \hat{b}_{j,k}^{(i)} I(|\hat{b}_{j,k}^{(i)}| > \delta_n)\psi_{j,k}(y), \qquad i = 0, 1. \tag{2.14}$$

Here $\hat{f}_0(y)$ is the estimator of $p(y) = f_0(y)$ and $\hat{f}_1(y)$ is the estimator of $\Phi(y) = f_1(y)$.

After the estimators $\hat{f}_i(y)$ of $f_i(y)$, $i = 0, 1$, are constructed, we need to estimate the ratio $\Phi(y)/p(y)$. Following Singh (1979), denote

$$H(x, L) = xI(|x| \le L) + L\,\mathrm{sign}\,(x)I(|x| > L), \qquad L > 0, \tag{2.15}$$

and construct the nonlinear wavelet EB estimator of θ of the form

$$\hat{\beta}_n(y) = y - H(\hat{f}_1(y)/\hat{f}_0(y), L_n). \tag{2.16}$$

3. The risks of the EB estimator

As the EB estimator is constructed, our objective is to calculate its risks (1.5) and (1.6). To derive the risks of the estimator (2.16), we first establish upper bounds for the mean squared error of the estimator (2.14). Proofs of all statements in this section are given in the Appendix.

LEMMA 1. *Assume that $q(x)$ satisfies the assumptions* C1 *and* C2 *and*

$$\delta_n = \delta_0\sqrt{\ln n/n}, \qquad m = m_0 \log_2(\ln n), \quad m + r = -2\log_2(\delta_n), \tag{3.1}$$

where

$$\delta_0 > 16(Q^2 \|q\|_c \|\psi\|_{L^2}^2 + Q\|\tilde{\psi}\|_L). \tag{3.2}$$

If $f_i(x)$ is $t_i \geq 1$ times continuously differentiable in the neighborhood $(y - 2^{1-m}B, y + 2^{1-m}B)$ of the point y, then

$$\boldsymbol{E}[\hat{f}_i(y) - f_i(y)]^2 = O(n^{-2\tau_i/(2\tau_i+1)}[\ln n]^{2\tau_i/(2\tau_i+1)}), \tag{3.3}$$

$$\boldsymbol{E}[\hat{f}_i(y) - f_i(y)]^4 = O(n^{-4\tau_i/(2\tau_i+1)}[\ln n]^{4\tau_i/(2\tau_i+1)}), \tag{3.4}$$

with $\tau_i = \min(t_i, \varrho + 1)$, $i = 0, 1$.

LEMMA 2. *Assume that $q(x)$ satisfies the conditions* C1 *and* C2 *and that parameters δ_n, m and r are defined by* (3.1) *and* (3.2). *Assume also that $f_i(x)$ is $t_i \geq 1$ times differentiable in a piecewise sense, that is, there exist points $-\infty < x_1 < x_2 < \cdots < x_N < \infty$ such that the first t_i derivatives of f_i exist and are bounded and continuous on (x_{l-1}, x_l), $l = 1, \ldots, N$, $N < \infty$, with left- and right-hand limits. Denote by I the union of all the intervals $(2^{-j}(k - B), 2^{-j}(k + B))$, $j \geq m$, that contain one of the points x_l, $l = 1, \ldots, N$. Then*

$$\boldsymbol{E}\int_I [\hat{f}_i(y) - f_i(y)]^2 dy = O(n^{-2\tau_i/(2\tau_i+1)}[\ln n]^{2\tau_i/(2\tau_i+1)}), \tag{3.5}$$

where $\tau_i = \min(t_i, \varrho + 1)$.

THEOREM 1. *Let the conditions of Lemma 1 hold for $f_0(y)$ and $f_1(y)$. Denote $\tau = \min(\tau_0, \tau_1, \varrho + 1)$ and assume that L_n is such that*

$$\lim_{n\to\infty} L_n (n^{-1}\ln n)^{2\tau_0/(2\tau_0+1)} = 0. \tag{3.6}$$

Then

$$R_n(y) = O(n^{-2\tau/(2\tau+1)}[\ln n]^{2\tau/(2\tau+1)}). \tag{3.7}$$

THEOREM 2. *Let all assumptions of Lemma 2 be valid for $f_0(y)$ and $f_1(y)$ and, besides that, the τ_i-th derivatives of $f_i(y)$, $i = 0, 1$, be uniformly bounded. Assume that for some $K > 1$ and $\lambda > 1$*

$$\int_{-\infty}^{\infty} \theta^{2K} g(\theta)d\theta < \infty, \tag{3.8}$$

$$\int_{-\infty}^{\infty} |x|^\lambda p(x)dx < \infty. \tag{3.9}$$

Let $T = 2K\lambda\tau_1 \,[2K\lambda - \lambda + 1 - 2(\lambda - 1)\tau_1]^{-1}$. If L_n has the form

$$L_n = \begin{cases} (n^{-1}\ln n)^{-(\tau_0/(2\tau_0+1))\cdot((\lambda-1)/2\lambda K)}, & \text{if } \tau_0 \leq T, \\ (n^{-1}\ln n)^{-(\tau_1/(2\tau_1+1))\cdot((\lambda-1)/(2\lambda K-\lambda+1))}, & \text{if } \tau_0 > T, \end{cases} \tag{3.10}$$

then

$$R_n = \begin{cases} O([n^{-1}\ln n]^{(\tau_0/(2\tau_0+1))\cdot((K-1)(\lambda-1)/K\lambda)}), & \text{if } \tau_0 \leq T, \\ O([n^{-1}\ln n]^{(\tau_1/(2\tau_1+1))\cdot(2(K-1)(\lambda-1)/(2K\lambda-\lambda+1))}), & \text{if } \tau_0 > T. \end{cases} \tag{3.11}$$

COROLLARY 1. *If, in the conditions of Theorem 2, $\tau_0 = \tau_1 = \tau$, and*

$$L_n = (n^{-1} \ln n)^{-(\tau/(2\tau+1))\cdot((\lambda-1)/2\lambda K)} \tag{3.12}$$

then

$$R_n = O([n^{-1} \ln n]^{(\tau/(2\tau+1))\cdot((K-1)(\lambda-1)/K\lambda)}). \tag{3.13}$$

Remark. The assumption (3.9) can be replaced by the condition

$$\int_{-\infty}^{\infty} |x|^{\mu} q(x) dx < \infty, \tag{3.14}$$

which is easy to verify. Since $|x| \leq |x - \theta| + |\theta|$, the inequalities (3.8) and (3.14) imply that (3.9) is valid with $\lambda = \min(2K, \mu)$.

4. Examples

Example 1. *The gamma distribution.* Let

$$q(x) = [\Gamma(\alpha+1)]^{-1} \sigma^{\alpha+1} x^{\alpha} \exp(-\sigma x) I(x \geq 0), \qquad \alpha \geq 0,$$

so that $\tilde{q}(\omega) = (1 + i\omega\sigma^{-1})^{-\alpha-1}$. It is easy to see that the conditions C1 and C2 hold, and the functions $f_i(x)$ take the forms $f_i(x) = \int_{-\infty}^{x} (x-\theta)^i q(x-\theta) g(\theta) d\theta, i = 0, 1$. Let the integer part of α be equal to s: $s = \text{Ent}(\alpha)$. Then, taking the derivatives of $f_i(x)$, we obtain

$$\begin{aligned} f_i^{(j+1)}(x) &= g(x) \left[\frac{d^j}{dx^j} (x-\theta)^i q(x-\theta) \right]_{\theta = x} \\ &+ \int_{-\infty}^{x} \frac{d^{j+1}}{dx^{j+1}} \left[(x-\theta) q(x-\theta) \right] g(\theta) d\theta. \end{aligned} \tag{4.1}$$

If α is an integer, i.e. $\alpha = s$, then the non-integral term in (4.1) does not vanish for $j = s + i$, otherwise, for $j = s + i + 1$. Thus, it follows from (4.1) that all discontinuities of $g(\theta)$ are inherited by $f_i^{(s+i+1)}(x)$ if $\alpha = s$, or by $f_i^{(s+i+2)}(x)$ if $\alpha > s$.

Consider the situation when $g(\theta)$ has a finite number of jump discontinuities x_l, $l = 1, \ldots, N$, and infinitely differentiable otherwise. Choose the ϱ-regular wavelet basis with $\varrho \geq s + 3$. Then, Theorem 1 implies that in the neighborhoods of the points x_l, $l = 1, \ldots, N$, the posterior risk has the order (3.7), where $\tau = \min(\varrho + 1, s + 1)$ if $\alpha = s$, and $\tau = \min(\varrho + 1, s + 2)$ if $\alpha > s$, $s = \text{Ent}(\alpha)$. Outside the neighborhoods of the points x_l, $l = 1, \ldots, N$, the posterior risk has the form (3.7) with $\tau = \varrho + 1$.

Now, let us discuss the rate of convergence of the prior risk. Note that, if $g(\theta)$ has a finite number of jump discontinuities, the functions $f_i(x)$, $i = 0, 1$, are infinitely differentiable everywhere except for the points x_l, $l = 1, \ldots, N$. Since λ can be chosen as large as one wants, the prior risk is given by (3.13) with $\tau = \varrho + 1$ and λ being a large number of one's choice. Here L_n should be chosen according to (3.12).

Example 2. *The double-exponential distribution.* Let $q(x) = 0.5 \exp(-|x - \theta|)$. Then, $\tilde{q}(\omega) = (w^2 + 1)^{-1}$, and conditions C1 and C2 hold. Writing $f_i(x)$ as

$$f_i(x) = 0.5 \left[\int_{-\infty}^{x} (x-\theta)^i e^{-(x-\theta)} g(\theta) d\theta + \int_{x}^{\infty} (x-\theta)^i e^{-(\theta-x)} g(\theta) d\theta \right]$$

and taking the derivatives, we arrive at

$$f_0''(x) = -g(x) + 0.5 \int_{-\infty}^{\infty} e^{-|x-\theta|} g(\theta) d\theta,$$

$$f_1'''(x) = -2g(x) + 0.5 \int_{-\infty}^{\infty} e^{-|x-\theta|} (3 - |x - \theta|)\, g(\theta) d\theta.$$

The above formulae mean that $f_0''(x)$ and $f_1'''(x)$ inherit all discontinuities of $g(\theta)$. Therefore, if $g(\theta)$ has a finite number of jump discontinuities, the posterior risk is $O([n^{-1} \ln n]^{4/5})$ in the neighborhoods of the points of discontinuity of $g(\theta)$, and is $O([n^{-1} \ln n]^{2\tau/(2\tau+1)})$ with $\tau = \varrho + 1$ otherwise. The prior risk has the form (3.13) with $\tau = \varrho + 1$ and λ being a large number of our choice, provided L_n is given by (3.12).

5. Discussion

In the present paper we have constructed the EB nonlinear wavelet estimators based on wavelets with bounded supports, and established the upper bounds for the prior and the posterior risks of these estimators. The estimators suggested in the paper are locally adaptive in the sense that they provide adaptation to the degree of smoothness of $f_i(y)$, $i = 0, 1$. In the case when $q(x)$ has only few continuous derivatives, or has none, and $g(\theta)$ has jump discontinuities at a finite number of points, this technique allows one to achieve higher rates of convergence than in the case when the EB estimator is based on Meyer-type wavelets (see Pensky (1997c)) or on Parzen-type kernels (see Pensky (1997a), Singh and Prasad (1989)).

For example, consider the situation when $q(x)$ is the p.d.f. of the exponential distribution (choose $\alpha = 0$ in Example 1) and $g(\theta)$ is concentrated on a bounded interval and is infinitely differentiable everywhere except at a finite number of points where it has jump discontinuities. Then, the conditions of Theorems 1 and 2 hold for any positive K and λ. Thus, although the adaptive EB estimator based on the Meyer-type wavelets has the posterior risk of the order $O(n^{-3/4} \ln n)$ for any $y > 0$, the estimator (2.16) has the posterior risk of the order $O([n^{-1} \ln n]^{3/4})$ in the immediate neighborhoods of the points of discontinuity of $g(\theta)$, and $R_n(y) = O([n^{-1} \ln n]^{2\tau/(2\tau+1)})$ with $\tau = \varrho + 1$ elsewhere. Choosing a large value of ϱ we ensure that the estimator has a high convergence rate outside the neighborhoods of the points of discontinuity of $g(\theta)$.

The prior risk in this situation can be made as close to $O(n^{-1/2}\sqrt{\ln n})$ as one wishes by choosing sufficiently large ϱ, K and λ. Note that this is a significant improvement of the result of Singh and Prasad (1989) who constructed the EB estimator under the assumption that $g(\theta)$ has a bounded support in the case when $q(x)$ is the density function of an exponential distribution. They demonstrated that for an $(r-1)$-times differentiable $g(\theta)$, $R_n = O([n^{-r/(r+1)} \ln \ln n]^{0.5-\varepsilon})$, where ε is an arbitrary positive integer. In the situation when $g(\theta)$ is discontinuous at a finite number of points, $r = 1$, and so that the method of Singh and Prasad (1989) gives $R_n = O([n^{-1/2} \ln \ln n]^{0.5-\varepsilon})$. Therefore, in spite of the fact that Singh and Prasad (1989) obtained their estimator for $q(x)$ of a particular type, the rate of convergence which they achieved is much slower than the rate of convergence of the estimator (2.16) constructed under general assumptions. It can be shown that the rate of convergence of R_n can be further improved using the fact that $q(x)$ is the exponential function.

However, we should bring the reader's attention to the fact that the results above are asymptotic, i.e. they take place if n is sufficiently large. Note that the requirement that ϱ is large implies larger support bounds A and B. This can lead to a difficulty if the number of observations is relatively small. Really, notice that in this situation m and consequently j in Lemma 2 do not turn to infinity, and the intervals surrounding the points of discontinuity may be fairly long. To overcome this difficulty one needs to choose the value of ϱ in accordance with n for finite n. Nevertheless, it does not lead to a significant change in the rate of convergence since, as ϱ grows, the value of $2\varrho/(2\varrho+1)$ changes less and less.

We also need to say few words about adaptivity. The estimator (2.16) is locally adaptive since its posterior risk has an optimal rate of convergence to zero at any point y provided ϱ is sufficiently large and L_n satisfies (3.6). The condition (3.6) can be satisfied by taking, $L_n = \sqrt{n}$, for instance, since $\tau \geq 1$.

The situation with the prior risk is somewhat different. Although we don't need to know $t_i, i = 0, 1$, to construct the estimators $\hat{f}_i(y)$ of $f_i(y)$, $i = 0, 1$, we need to know K and τ to derive the optimal value of L_n. Since, in the majority of cases, we can predict from the nature of the problem that $f_i(y)$, $i = 0, 1$, are infinitely differentiable everywhere except for finite number of points, we can choose $\tau = \varrho + 1$. However, the knowledge of K is essential and, if K is under- or over-estimated, the rate of convergence of the prior risk drops.

We should mention also that in the theory developed above no specific information about the structure of the function $q(x)$ has been employed. Consequently, if some specific information is used for construction of the EB estimator or evaluation of its prior risk, an even better rate of convergence of the prior risk to zero may possibly be achieved.

6. Appendix

The proofs of Lemma 1 and 2 is based on the following lemmas.

LEMMA 3. *If $\hat{a}^{(i)}_{m,k}$ and $\hat{b}^{(i)}_{j,k}$ are defined by* (2.11) *and* (2.13), *respectively, and $2^{m+r} = o(n)$ as $n \to \infty$, then*

$$\boldsymbol{E}(\hat{a}^{(i)}_{m,k} - a^{(i)}_{m,k})^{2l} \leq C_{a,l} n^{-l}, \qquad l = 1, 2; \tag{6.1}$$

$$\boldsymbol{E}(\hat{b}^{(i)}_{j,k} - b^{(i)}_{j,k})^{2l} \leq C_{b,l} n^{-l}, \qquad l = 1, 2, 4. \tag{6.2}$$

PROOF OF LEMMA 3. Without loss of generality, let us prove inequalities (6.2). Observe that by Parseval's identity

$$\begin{aligned}\boldsymbol{E}[2^{j/2} V_j^{(i)}(2^j X - k)]^{2s} &\leq 2^{j(s-1)} \|q\|_c \int_{-\infty}^{\infty} [V_j^{(i)}(z)] v^{2s} dz \\ &\leq 2^{j(s-1)} \|q\|_c \sup_x |V_j^{(i)}(x)|^{2s-2} \\ &\quad \cdot \int_{-\infty}^{\infty} |\tilde{q}'(-2^j \omega)/\tilde{q}(-2^j\omega)|^2 |\tilde{\psi}(\omega)|^2 d\omega.\end{aligned}$$

Since $\sup_x |V_j^{(i)}(x)| \leq (2\pi)^{-1} Q \int |\tilde{\psi}(\omega)| d\omega$, we derive that

$$\boldsymbol{E}[2^{j/2} V_j^{(i)}(2^j X - k)]^{2s} \leq C_{\psi,Q} 2^{j(s-1)}, \tag{6.3}$$

where $C_{\psi,Q}$ is an absolute constant depending on ψ and Q only.

Now, in the case of $l=1$,

$$\boldsymbol{E}(\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)})^2 \le n^{-1}\boldsymbol{E}[2^{j/2}V_j^{(i)}(2^jX-k)]^2,$$

which, together with (6.3) implies (6.2) with $l=1$. To establish (6.2) with $l=2$ and $l=4$, recall that, if ξ_s, $s=1,\dots,n$, are i.i.d. random variables with $\boldsymbol{E}\xi_s=0$, then for some absolute constants C_1 and C_2

$$\boldsymbol{E}\left[n^{-1}\sum_{l=1}^{n}\xi_s\right]^4 \le C_1[n^{-3}\boldsymbol{E}\xi_1^4 + n^{-2}(\boldsymbol{E}\xi_1^2)^2], \tag{6.4}$$

$$\boldsymbol{E}\left[n^{-1}\sum_{l=1}^{n}\xi_s\right]^8 \le C_2[n^{-7}\boldsymbol{E}\xi_1^8 + n^{-6}(\boldsymbol{E}\xi_1^6\boldsymbol{E}\xi_1^2) + n^{-6}(\boldsymbol{E}\xi_1^4)^2 + n^{-5}(\boldsymbol{E}\xi_1^2)^2(\boldsymbol{E}\xi_1^4) + n^{-4}(\boldsymbol{E}\xi_1^2)^4]. \tag{6.5}$$

Choosing $\xi_l = 2^{j/2}V_j^{(i)}(2^jX_l-k) - b_{j,k}^{(i)}$ and combining (6.3) with (6.4) and (6.5) in a view of $2^j = o(n)$, we obtain (6.2) with $l=2$ and $l=4$.

LEMMA 4. *If $2^jy - B \le k \le 2^jy + B$, then*

$$|b_{j,k}^{(i)}| \le 2^{-(\tau_i+0.5)j}L(f_i^{(\tau_i)}, y, 2^{-j+1}B)C(\psi,\tau_i). \tag{6.6}$$

Here,

$$C(\psi,s) = \int_{-B}^{B}|z|^s|\psi(z)|dz, \qquad L(f,y,\delta) = \sup_{|x-y|<\delta}|f(x)|. \tag{6.7}$$

PROOF OF LEMMA 4. Observe that by Taylor's expansion,

$$\begin{aligned} b_{j,k}^{(i)} &= 2^{j/2}\int_{-\infty}^{\infty}\psi(2^jx-k)p(x)dx \\ &= 2^{-j/2}\int_{-B}^{B}\psi(z)f_i(2^{-j}(z+k))dz \\ &= \sum_{l=0}^{\tau_i-1}2^{-0.5j(2l+1)}f_i^{(l)}(2^{-j}k)\int_{-B}^{B}z^l\psi(z)dz \\ &\quad +2^{-(\tau_i+0.5)j}\int_{-B}^{B}\int_0^1 f_i^{(\tau_i)}(2^{-j}(k+sz))z^l(1-s)^{\tau_i-1}\psi(z)dsdz. \end{aligned}$$

According to (2.2), all integrals $\int_{-\infty}^{\infty}z^l\psi(z)dz = 0$, $l=0,\dots,\tau_i-1$. Also, $2^jy-B \le k \le 2^jy+B$ and $0\le s\le 1$ imply $2^{-j}(k+sz) \in [y-2^{-j+1}B, y+2^{-j+1}B]$, so that (6.6) is valid.

LEMMA 5. *For δ_n defined in (3.1) with δ_0 given by (3.2)*

$$P(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n) \le 2n^{-2}. \tag{6.8}$$

PROOF OF LEMMA 5. The proof is based on Bernstein's inequality

$$P\left(\left|n^{-1}\sum_{l=1}^{n} Y_l\right| > z\right) \le 2\exp\left(-\frac{nz^2}{2(\sigma^2 + \|Y\|_\infty z/3)}\right),$$

where Y_l, $l = 1, \dots, n$, are i.i.d. with $\boldsymbol{E}Y_l = 0$, $\boldsymbol{E}Y_l = \sigma^2$ and $\|Y_l\| \le \|Y\|_\infty < \infty$. Applying Bernstein's inequality with $Y_l = 2^{j/2}V_j^{(i)}(2^j X_l - k) - b_{j,k}^{(i)}$, $z = \delta_n/2$ and observing that $\|Y\|_\infty = 2^{1+j/2}Q\|\psi\|_c$ and $\sigma^2 \le Q^2\|q\|_c \mid \|\psi\|_{L_2}^2$, we arrive at $P(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n) \le 2\exp(-\delta_0^2 \ln n[8(Q^2\|q\|_c\|\psi\|_{L_2} + 2^{j/2}\delta_n Q\|\psi\|_c/3)]^{-1})$. The last inequality together with (3.2) implies (6.8) since $2^{j/2}\delta_n \to 0$ as $n \to \infty$.

PROOF OF LEMMA 1. Since

$$I(|\hat{b}_{j,k}^{(i)}| \ge \delta_n) \le I(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n) + I(|b_{j,k}^{(i)}| \ge 0.5\delta_n),$$
$$I(|\hat{b}_{j,k}^{(i)}| \le \delta_n) \le I(|b_{j,k}^{(i)}| \le 1.5\delta_n) + I(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n),$$

we obtain $|\hat{f}_i(y) - f_i(y)| \le \sum_{s=1}^{6} |R_s(y)|$, where

$$\begin{aligned}
R_1(y) &= \sum_{k\in\boldsymbol{Z}} (\hat{a}_{m,k}^{(i)} - a_{m,k}^{(i)})\varphi_{m,k}(y),\\
R_2(y) &= \sum_{j=m+r+1}^{\infty}\sum_{k\in\boldsymbol{Z}} b_{j,k}^{(i)}\psi_{j,k}(y),\\
R_3(y) &= \sum_{j=m}^{m+r}\sum_{k\in\boldsymbol{Z}} |b_{j,k}^{(i)}||\psi_{j,k}(y)| I(|b_{j,k}^{(i)}| \le 1.5\delta_n),\\
R_4(y) &= \sum_{j=m}^{m+r}\sum_{k\in\boldsymbol{Z}} |b_{j,k}^{(i)}||\psi_{j,k}(y)| I(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n),\\
R_5(y) &= \sum_{j=m}^{m+r}\sum_{k\in\boldsymbol{Z}} |\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}||\psi_{j,k}(y)| I(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n),\\
R_6(y) &= \sum_{j=m}^{m+r}\sum_{k\in\boldsymbol{Z}} |\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}||\psi_{j,k}(y)| I(|b_{j,k}^{(i)}| \ge 0.5\delta_n),
\end{aligned} \tag{6.9}$$

Therefore,

$$\boldsymbol{E}[\hat{f}_i(y) - f_i(y)]^{2l} \le C_l \sum_{s=1}^{6} \boldsymbol{E}[R_s(y)]^{2l}, \qquad l = 1, 2, \tag{6.10}$$

where C_l depends on l only. Let us analyze the sum (6.10) term by term.

(i) Since $m = m_0 \log_2(\ln n)$, then, according to Lemma 3,

$$\begin{aligned}
\boldsymbol{E}[R_1(y)]^{2l} &\le \sup_k \boldsymbol{E}(\hat{a}_{m,k}^{(i)} - a_{m,k}^{(i)})^{2l}\left[\sup_x \sum_{k\in\boldsymbol{Z}} |\varphi(x-k)|\right]^{2l}\\
&\le (2A+1)^{2l}\|\varphi\|_c^{2l} \sup_k \boldsymbol{E}(\hat{a}_{m,k}^{(i)} - a_{m,k}^{(i)})^{2l}\\
&= O(2^{(2l-1)m}n^{-l}) = O([\ln n]^{m_0(2l-1)}n^{-l}), \qquad l = 1, 2,
\end{aligned} \tag{6.11}$$

because $\sum_{k\in\boldsymbol{Z}}|\varphi(x-k)|$ contains at most $(2A+1)$ terms, where A is defined in (2.1).

(ii) Applying Lemma 4 and taking into account that $R_2(y)$ is not random, we obtain

$$|R_2(y)| \le \sum_{j=m+r+1}^{\infty} \sum_{k\in\boldsymbol{Z}} 2^{-j\tau_i} L(f_i^{(\tau_i)}, y, 2^{-j+1}B) C(\psi, \tau_i) |\psi(2^j y - k)|.$$

Since $\sum_{k\in\boldsymbol{Z}}$ contains at most $(2B+1)$ terms for every y and $2^{m+r} = O(\sqrt{n/\ln n})$,

$$\boldsymbol{E}[R_2(y)]^{2l} = o\left(n^{-l}\right), \qquad l=1,2. \tag{6.12}$$

(iii) To construct an upper bound for $\boldsymbol{E}[R_3(y)]^{2l}$, observe that $R_3(y)$ is non-random and consider $M = M_n$ such that

$$2^M \sim \delta_n^{-2/(2\tau_i+1)}. \tag{6.13}$$

Partition $R_3(y)$ into two sums, $R_{31}(y)$ and $R_{32}(y)$, containing components with $j \le M$ and $j > M$, respectively. Then

$$R_{31}(y) \le \sum_{j=m}^{M} \sum_{k\in\boldsymbol{Z}} 1.5\delta_n 2^{j/2} |\psi(2^j y - k)| = O(2^{M/2}\delta_n), \tag{6.14}$$

since $\sum_{k\in\boldsymbol{Z}}$ contains at most $(2B+1)$ terms for every y. To find an upper bound for $R_{32}(y)$ note that

$$\begin{aligned} R_{32}(y) &\le \sum_{j=M+1}^{m+r} 2^{j/2} \sum_{k\in\boldsymbol{Z}} |b_{j,k}||\psi(2^j y - k)| \\ &\le 2^{-M\tau_i}(2B+1)\|\psi\|_c L(f_i^{(\tau_i)}, y, 2^{-M+1}B) C(\psi, \tau_i). \end{aligned} \tag{6.15}$$

Combining (6.13)–(6.15) and observing that $R_3(y)$ is not random, we derive that

$$\sup_y \boldsymbol{E}[R_3(y)]^{2l} = O(\delta_n^{-4\tau_i l/(2\tau_i+1)}), \qquad l=1,2. \tag{6.16}$$

(iv) To find an upper bound for $\boldsymbol{E}[R_4(y)]^{2l}, l=1,2$, note that

$$\boldsymbol{E}[R_4(y)]^{2l} \le \left\{ \sum_{j=m}^{m+r} \sum_{k\in\boldsymbol{Z}} 2^{j/2} |b_{j,k}||\psi(2^j y - k)| [P(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n)]^{1/(2l)} \right\}^{2l},$$

so that, by Lemmas 4 and 5,

$$\boldsymbol{E}[R_4(y)]^{2l} = O(n^{-2}), \qquad l=1,2. \tag{6.17}$$

(v) To derive an upper bound for $\boldsymbol{E}[R_5(y)]^{2l}, l=1,2$, observe that

$$\begin{aligned} \boldsymbol{E}[R_5(y)]^{2l} \le \Bigg\{ &\sum_{j=m}^{m+r} \sum_{k\in\boldsymbol{Z}} C_3 2^{j/2} [P(|\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n)]^{1/(4l)} \\ &\cdot [\boldsymbol{E}(\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)})^{4l}]^{1/(4l)} |\psi(2^j y - k)| \Bigg\}^{4l}. \end{aligned}$$

Thus, application of Lemmas 3 and 5 lead to

$$\boldsymbol{E}[R_5(y)]^{2l} = O(n^{-l}), \qquad l = 1, 2, \tag{6.18}$$

since $n^{-1}2^{m+r} = o(1)$ as $n \to \infty$.

(vi) By Cauchy's inequality,

$$\boldsymbol{E}[R_6(y)]^{2l} \le \left\{ \sum_{j=m}^{m+r} \sum_{k \in \boldsymbol{Z}} [\boldsymbol{E}(\hat{b}_{j,k} - b_{j,k})^{2l}]^{1/(2l)} 2^{j/2} |\psi(2^j y - k)| I\left(|b_{j,k}| > 0.5\delta_n\right) \right\}^{2l}.$$

Using Lemmas 4, we obtain that $|b_{j,k}| > 0.5\delta_n$ implies that $2^{j/2} \le J_n$, where

$$J_n = O([2\delta_n^{-1} L(f_i^{(\tau_i)}, y, 2^{-j+1}B)]^{1/(2\tau_i+1)}).$$

Thus, from Lemma 3 it follows that

$$\boldsymbol{E}[R_6(y)]^{2l} \le \left\{ \sum_{j=m}^{2\log_2 J_n} \sum_{k \in \boldsymbol{Z}} 2^{j/2} [\boldsymbol{E}(\hat{b}_{j,k}^{(i)} - b_{j,k}^{(i)})^{2l}]^{1/(2l)} |\psi(2^j y - k)| \right\}^{2l}, \qquad l = 1, 2.$$

Therefore,

$$\boldsymbol{E}[R_6(y)]^{2l} = O(n^{-2l\tau_i/(2\tau_i+1)} [\ln n]^{-l/(2\tau_i+1)}), \qquad l = 1, 2. \tag{6.19}$$

To complete the proof of Lemma 1, combine (6.10), (6.11), (6.12), (6.16)–(6.19).

PROOF OF LEMMA 2. Denote for every j, the set of intervals where the function $f_i(y)$ is discontinuous by I_j, and

$$K_j = \{k : k \in (2^j x_s - B, 2^j x_s + B), s = 1, \ldots, N\}, \qquad N < \infty.$$

Observe that it follows from above that the set $I = \cup_j I_j$ is bounded. Similarly to the proof of Lemma 1, we derive that

$$\boldsymbol{E} \int_I [\hat{f}_i(y) - f_i(y)]^2 dy \le 36 \sum_{s=1}^{6} \boldsymbol{E} \int_I |R_s(y)|^2 dy. \tag{6.20}$$

where $R_s(y), s = 1, \ldots, 6$, are defined in (6.9). Let us examine each term of the sum (6.20) separately.

(i) Since $R_1(y)$ does not depend on the degree of smoothness of $f_i(y)$, it follows from (6.11) that

$$\boldsymbol{E} \int_I |R_1(y)|^2 dy = O([\ln n]^{m_0} n^{-1}). \tag{6.21}$$

(ii) If $y \in I$, then $k \in K_j$ for every j, and

$$R_2(y) \le \left| \sum_{j=m+r+1}^{\infty} \sum_{k \in K_j} 2^{j/2} b_{j,k}^{(i)} \psi(2^j y - k) \right|.$$

Observing that $|b_{j,k}^{(i)}| \le 2^{-j/2}\|\psi\|_{L_1} \sup|f_i(x)| = O(2^{-j/2})$ for $\tau_i \ge 0$, and taking into account that the set K_j contains only finite number of terms, we obtain

$$\int_I R_2^2(y)dy \le \int_{-\infty}^{\infty} \left[\sum_{j=m+r+1}^{\infty} \sum_{k\in K_j} 2^{j/2} b_{j,k}^{(i)} \psi(2^j y - k) \right]^2 dy$$
$$= \sum_{j=m+r+1}^{\infty} \sum_{k\in K_j} [b_{j,k}^{(i)}]^2 = O\left(\sum_{j=m+r+1}^{\infty} 2^{-j} \right),$$

which implies

$$\boldsymbol{E} \int_I |R_2(y)|^2 dy = O(n^{-1}\ln n). \tag{6.22}$$

(iii) Since

$$\boldsymbol{E} \int_I |R_3(y)|^2 dy \le \frac{9}{4}\delta_n^2 \int_I \left[\sum_{j=m}^{m+r} \sum_{k\in K_j} 2^{j/2} |\psi(2^j y - k)| \right]^2 dy$$

and $\sum_{j=m}^{m+r}$ contains $r+1 = O(\ln n)$ terms,

$$\boldsymbol{E} \int_I |R_3(y)|^2 dy = O(r^2\delta_n^2) = O\left((\ln n)^2 n^{-1}\right). \tag{6.23}$$

(iv) Note that

$$\boldsymbol{E} \int_I [R_4(y)]^2 dy \le \int_I \left\{ \sum_{j=m}^{m+r} \sum_{k\in \boldsymbol{Z}} 2^{j/2} |b_{j,k}^{(i)}| |\psi(2^j y - k)| \sqrt{P(|\hat b_{j,k}^{(i)} - b_{j,k}^{(i)}| > 0.5\delta_n)} \right\}^2 dy,$$

so that, by Lemmas 4 and 5,

$$\boldsymbol{E} \int_I |R_4(y)|^2 dy = O(n^{-2} 2^{-m} r^2) = o(n^{-1}). \tag{6.24}$$

(v) Since $R_5(y)$ does not depend on the degree of smoothness of $f_i(y)$, it follows from (6.18) that

$$\boldsymbol{E} \int_I |R_5(y)|^2 dy = o(n^{-1}). \tag{6.25}$$

(vi) It is easy to see that

$$\boldsymbol{E} \int_I |R_6(y)|^2 dy \le \int_I \left[\sum_{j=m}^{m+r} \sum_{k\in K_j} \sqrt{\boldsymbol{E}(\hat b_{j,k}^{(i)} - b_{j,k}^{(i)})^2} 2^{j/2} |\psi(2^j y - k)| \right]^2 dy.$$

Applying Lemma 3 and taking into account that $\sum_{j=m}^{m+r}$ contains $O(\ln n)$ terms, we derive

$$\boldsymbol{E} \int_I |R_6(y)|^2 dy = O((\ln n)^2 n^{-1}). \tag{6.26}$$

Now, validity of Lemma 2 follows from (6.21)–(6.26).

To prove Theorems 1 and 2, we need the following Lemma.

LEMMA 6. *For any pair of random variables ξ and η and for real numbers a, $b \neq 0$, $0 < L < \infty$ and $0 < \gamma < 2$*

$$\boldsymbol{E}\left[\left|\frac{\xi}{\eta} - \frac{a}{b}\right| \wedge L\right]^{\gamma} \leq 2|b|^{-\gamma}\left[\boldsymbol{E}(\xi - a)^{\gamma} + \left(\left|\frac{\eta}{b}\right| + L\right)^{\gamma}\boldsymbol{E}(\eta - b)^{\gamma}\right], \tag{6.27}$$

$$\boldsymbol{E}\left[\left|\frac{\xi}{\eta} - \frac{a}{b}\right| \wedge L\right]^{2} \leq \frac{8}{b^2}\left[\boldsymbol{E}(\xi - a)^{2} + \left(\frac{a}{b}\right)^{2}\boldsymbol{E}(\eta - b)^{2}\right] + \frac{16L^2}{b^4}\boldsymbol{E}(\eta - b)^4. \tag{6.28}$$

Here $x \wedge y = \min(x, y)$.

The first inequality (6.27) is the content of the lemma formulated by Singh (1979), the second inequality is its slight modification.

PROOF OF THEOREM 1. Note that $|\Phi(y)/p(y)| \leq L_n$ for any y if n is sufficiently large, so that

$$|H(\hat{f}_1(y)/\hat{f}_0(y), L_n) - f_1(y)/f_0(y)| \leq 2L_n.$$

Therefore, from (3.6), (6.28) and Lemma 1, it follows that

$$\begin{aligned} \boldsymbol{E}(\hat{\beta}_n(y) - \beta(y))^2 &\leq \boldsymbol{E}\left[\left|\frac{\hat{f}_1(y)}{\hat{f}_0(y)} - \frac{f_1(y)}{f_0(y)}\right| \wedge 2L_n\right]^2 \\ &= O([n^{-1}\ln n]^{2\tau/(2\tau+1)}) + O(L_n^2[n^{-1}\ln n]^{4\tau/(2\tau+1)}) \\ &= O([n^{-1}\ln n]^{2\tau/(2\tau+1)}). \end{aligned}$$

PROOF OF THEOREM 2. To establish an upper bound for the prior risk (1.6), note that the conditions (3.8) and (3.9) imply that

$$\int_{-\infty}^{\infty}\left(\frac{f_1(y)}{f_0(y)}\right)^{2K} p(y)dy < \infty, \tag{6.29}$$

$$\int_{-\infty}^{\infty}[p(y)]^{1/\lambda}dy < \infty, \tag{6.30}$$

respectively. Let us partition the posterior risk into a sum of two components

$$\boldsymbol{E}[\hat{\beta}_n(y) - \beta(y)]^2 = J_1(y) + J_2(y). \tag{6.31}$$

Here

$$J_1(y) = \boldsymbol{E}[(\hat{\beta}_n(y) - \beta(y))^2 I(|\beta(y)| \leq L_n)],\ \ J_2(y) = \boldsymbol{E}[(\hat{\beta}_n(y) - \beta(y))^2 I(|\beta(y)| > L_n)],$$

and, since in this case $|\hat{\beta}_n(y) - \beta(y)| \leq \min[|\hat{f}_1(y)/\hat{f}_0(y) - f_1(y)/f_0(y)|, 2L_n]$, applying (6.27) with $0 < \gamma < 1$, and Cauchy's inequality, we obtain

$$\begin{aligned} J_1(y) \leq 2^{3-\gamma}L_n^{2-\gamma}(p(y))^{-\gamma}\{[\boldsymbol{E}(\hat{f}_1(y) - f_1(y))^2]^{\gamma/2} \\ +3^{\gamma}L_n^{\gamma}[\boldsymbol{E}(\hat{f}_0(y) - f_0(y))^2]^{\gamma/2}\}. \end{aligned} \tag{6.32}$$

The second term, $J_2(y)$, is bounded by

$$J_2(y) \le 4\beta^2(y) I(|\beta(y)| > L_n). \tag{6.33}$$

Denote by I_C the compliment of the set I in $(-\infty, \infty)$. Then, multiplying (6.33) and (6.33) by $p(y)$ and integrating, we derive that for some absolute constant C_{J1}

$$\begin{aligned} J_1 &= \int_{-\infty}^{\infty} J_1(y) p(y) dy \\ &\le C_{J1} L_n^{2-\gamma} \left\{ \left[\sup_{y \in I_C} \boldsymbol{E}(\hat{f}_1(y) - f_1(y))^2 \right]^{\gamma/2} \right. \\ &\qquad \left. + L_n^{\gamma} \left[\sup_{y \in I_C} \boldsymbol{E}(\hat{f}_0(y) - f_0(y))^2 \right]^{\gamma/2} \right\} \int_{-\infty}^{\infty} [p(y)]^{1-\gamma} dy \\ &\quad + C_{J1} L_n^{2-\gamma} \left\{ \int_I [\boldsymbol{E}(\hat{f}_1(y) - f_1(y))^2]^{\gamma/2} [p(y)]^{1-\gamma} dy \right. \\ &\qquad \left. + L_n^{\gamma} \int_I [\boldsymbol{E}(\hat{f}_0(y) - f_0(y))^2]^{\gamma/2} [p(y)]^{1-\gamma} dy \right\} \end{aligned}$$

and

$$J_2 = \int_{-\infty}^{\infty} J_2(y) p(y) dy \le 4 L_n^{2-2K} \int_{-\infty}^{\infty} [\beta(y)]^{2K} p(y) dy.$$

Combining (6.29), (6.30) and the last two inequalities with the results of Lemmas 1 and 2, we derive that

$$R_n = O\left(L_n^{2-\gamma} \left(\frac{\ln n}{n} \right)^{\tau_1 \gamma/(2\tau_1+1)} + L_n^2 \left(\frac{\ln n}{n} \right)^{\tau_0 \gamma/(2\tau_0+1)} + L_n^{2-2K} \right) \tag{6.34}$$

provided $1 - \gamma \le \lambda^{-1}$. Choosing $\gamma = (\lambda - 1)/\lambda$ and analyzing (6.34), we arrive at (3.11).

Proof of Corollary 1. It follows directly from Theorem 2.

References

Abramovich, F. and Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems, *Biometrika*, **85**, 115–129.

Antoniadies, A., Grégoire, G. and McKeague, I.W. (1994). Wavelet method for curve estimation, *J. Amer. Statist. Assoc.*, **89**, 1340–1353.

Bickel, P. J. and Klaassen, C. A. J. (1986). Empirical Bayes estimation in functional and structural models and uniformly adaptive estimation of location, *Advances in Applied Mathematics*, **7**, 55–69.

Brown, L. D. and Low M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation, *Ann. Statist.*, **24**, 2524–2535.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall, London.

Daubechies, I. (1992). *Ten Lectures in Wavelets*, SIAM, Philadelphia.

Donoho, D. L. (1994). Smooth wavelet decompositions with blocky coefficient kernels, *Recent Advances in Wavelet Analysis* (eds. L. L. Schumaker and G. Webb), 259–308, Academic Press, London.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Statist. Assoc.*, **90**, 1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Ann. Statist.*, **24**, 508–539.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error, *J. Amer. Statist. Assoc.*, **92**, 526–535.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators, *Ann. Statist.*, **23**, 905–928.

Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block thresholding rules for curve estimation using kernel and wavelet methods, *Ann. Statist.*, **26**, 922–942.

Hernández, E. and Weiss, G. (1996). *A First Course on Wavelets*, CRC Press, Boca Raton.

Huang, Su-Yun (1997). Wavelet based empirical Bayes estimation for the uniform distribution, *Statist. Probab. Lett.*, **32**, 141–146.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces, *Statist. Probab. Lett.*, **13**, 15–24.

Maritz, J., and Lwin, T. (1989). *Empirical Bayes Methods*, Chapman & Hall, London.

Masry, E. (1994). Probability density estimation from dependent observations using wavelet orthonormal bases, *Statist. Probab. Lett.*, **21**, 181–194.

Meyer, I. (1993). *Wavelets Algorithms and Applications*, SIAM, Philadelphia.

Nogami, Y. (1988). Convergence rates for empirical Bayes estimation in the uniform $U(0, \theta)$ distribution, *Ann. Statist.*, **16**, 1335–1341.

Penskaya, M. (1992). On empirical Bayes estimation of a shift parameter, *Probab. Theory Appl.*, **37**, 732–734.

Pensky, M. (1996). Empirical Bayes estimation of a scale parameter, *Math. Methods Statist.*, **5**, 316–331.

Pensky, M. (1997*a*). Empirical Bayes estimation of a location parameter, *Statist. Decisions*, **15**, 1–16.

Pensky, M. (1997*b*). A general approach to nonparametric empirical Bayes estimation, *Statistics*, **29**, 61–80.

Pensky, M. (1997*c*). Wavelet empirical Bayes estimation of a location or a scale parameter, *J. Statist. Plann. Inference* (submitted).

Pensky, M. (1998). Empirical Bayes estimation based on wavelets, *Sankhyā Ser. A*, **60**, 214–231.

Pensky, M. (1999). Nonparametric empirical Bayes estimation via wavelets, *Bayesian Inference in Wavelet Based Models*, Lecture Notes in Statist., Vol. 141 (eds. P. Müller and B. Vidakovic), 323–340, Springer, New York.

Robbins, H. (1983). Jerzy Neyman memorial lecture. Some thoughts on empirical Bayes estimation, *Ann. Statist.*, **11**, 713–723.

Singh, R. S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate, *Ann. Statist.*, **7**, 890–902.

Singh, R. S. and Prasad, B. (1989). Uniformly strongly consistent prior distribution and empirical Bayes estimators with asymptotic optimality and rates in a non-exponential family, *Sankhyā Ser. A*, **51**, 334–342.

Singh, R. S. and Wei, L. (1992). Empirical Bayes with rates and best rates of convergence in $u(x)c(\theta)$ $\exp(-\frac{x}{\theta})$ family estimation case, *Ann. Inst. Statist. Math.*, **44**, 435–449.

Tiwari, R. C. and Zalkikar, J. N. (1990). Empirical Bayes estimation of a scale parameter in a Pareto distribution, *Comput. Statist. Data Anal.*, **10**, 261–270.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, Wiley, New York.

Walter, G. G. (1981). Orthogonal series estimators of the prior distribution, *Sankhyā Ser. A*, **43**, 228–245.

Walter, G. G. (1994). *Wavelets and Other Orthogonal Systems with Applications*, CRC Press, Boca Raton.