WAVELET ESTIMATORS OF A DENSITY CONSTRUCTED FROM OBSERVATIONS OF A MIXTURE

UDC 519.21

D. POKHYL'KO

ABSTRACT. We construct projective estimators of a density by using a wavelet basis for the data being a sample from a mixture of several components whose concentrations vary with observations. We construct linear and adaptive estimators and prove that they converge in the mean square norm. We also prove that the linear estimator converges in the uniform norm.

1. INTRODUCTION

In this paper, we construct projective estimators of a density by using a wavelet basis for the data being a sample from a mixture of several components whose concentrations vary with observations. An estimator (the so-called empirical measure) of the distributions of components of a mixture with varying concentrations was constructed and studied in [1]. General problems related to projective estimators of a density were considered in [3]. Applications of wavelets for the problem of the estimation of a density for independent identically distributed random variables are given in [2] and [4]. Basic results of the wavelet analysis as well as an extensive literature is given in [2] and [5]. Modern references on wavelet analysis are included in [6] and [7].

In Section 4, we construct a linear estimator, prove its convergence in the mean square norm for the case where the densities of the components belong to some Besov spaces, and obtain the convergence of the estimator in the uniform norm for the case where the densities of the components belong to Sobolev spaces.

In Section 5, we construct an adaptive estimator and prove its convergence in the mean square norm for the case where the densities of the components belong to Besov spaces.

2. SETTING OF THE PROBLEM

Let t_j be the time when an observation j is made, $j = 1, \dots, N$. The random variables ξ_{t_j} , $j = 1, \dots, N$, have the distribution densities

$$(1) \quad p_{t_j}(x) = \sum_{k=1}^M w_k(t_j) p_k(x), \quad j = 1, \dots, N,$$

2000 *Mathematics Subject Classification.* Primary 62G07; Secondary 42C40.

Key words and phrases. Wavelets, projective estimators of a density, estimation of components in a mixture.

respectively, where the concentration $w_k(t)$ of the component k at the time t is known and satisfies the following conditions:

$$0 \leq w_k(t) \quad \text{and} \quad \sum_{k=1}^M w_k(t) = 1$$

for all t . For the sake of brevity we write ξ_j instead of ξ_{t_j} . We also assume that the random variables ξ_j are jointly independent for a fixed N .

Problem. Let X_j be observations of the random variables ξ_j , $j = 1, \dots, N$. Assume that the functions $w_k(t)$ are known. It is necessary to construct an estimator $\hat{p}_l(x)$ of the density $p_l(x)$ of a component l of the mixture by using X_j , $j = 1, \dots, N$.

3. AUXILIARY RESULTS

The definitions of the father wavelet, mother wavelet, scale-function, weak differentiability and weak derivative, Sobolev function spaces $W_p^n(\mathbf{R})$ and $\tilde{W}_p^n(\mathbf{R})$ equipped with the norm $\|\cdot\|_{W_p^n}$, and Besov spaces $B_p^{sq}(\mathbf{R})$ equipped with the norm $\|\cdot\|_{spq}$ can be found in [2].

In what follows $\|\cdot\|_p$ denotes either $\mathcal{L}_p(\mathbf{R})$ - or ℓ_p -norm depending on the topic we discuss.

Here are some necessary conditions posed on ϕ .

Condition θ . Let $\theta_\phi(x) := \sum_{k \in \mathbf{Z}} |\phi(x - k)|$. Then

$$\text{ess sup}_x \theta_\phi(x) < \infty$$

where $\text{ess sup}_x f(x) = \inf\{y: m(\{x: f(x) > y\}) = 0\}$ and m is Lebesgue measure in the space $\mathbf{R}$.

Condition S . There exists a bounded nonincreasing function $\Phi: \mathbf{R}_+ \rightarrow \mathbf{R}$ such that

$$\int_{\mathbf{R}} \Phi(|u|) du < \infty$$

and $|\phi(u)| \leq \Phi(|u|)$ almost everywhere.

Condition $S(Q)$. Condition S holds and $\int_{\mathbf{R}} \Phi(|u|)|u|^Q du < \infty$.

Put $\varphi_{sk}(x) = 2^{s/2}\varphi(2^s x - k)$ for $s, k \in \mathbf{Z}$ and let $V_s := \text{c.l.s.}\{\varphi_{sk}, k \in \mathbf{Z}\}$.

Definition 1. Let ϕ be a scale function that satisfies condition θ . A kernel

$$K(x, y) = \sum_{k \in \mathbf{Z}} \phi(x - k) \overline{\phi(y - k)}$$

is called orthogonal projective kernel associated with ϕ .

Let Pr_V be the projective operator mapping the space $\mathcal{L}_2(\mathbf{R})$ to the subspace V .

Proposition A (see [2, p. 83]). *Let ϕ be a scale function. If condition θ holds for ϕ , then*

$$(\text{Pr}_{V_0} f)(x) = (Kf)(x) \equiv \int_{\mathbf{R}} K(x, y) f(y) dy$$

for all $f \in \mathcal{L}_2(\mathbf{R})$.

Definition 2. Let $B(s, p, q, L)$ be the set defined by

$$B(s, p, q, L) = \{f \in B_p^{sq}: \|f\|_{spq} \leq L, f \text{ is a density}\}.$$

Here and in what follows α_0 and β_j denote the sequences $\{\alpha_{0k}\}_{k \in \mathbf{Z}}$ and $\{\beta_{jk}\}_{k \in \mathbf{Z}}$ of coefficients in the wavelet expansion of f , namely

$$f = \sum_{k \in \mathbf{Z}} \alpha_{0k} \phi_{0k} + \sum_{j \geq 0} \sum_{k \in \mathbf{Z}} \beta_{jk} \psi_{jk}.$$

Proposition B ([2, p. 123]). *Let ϕ be a scale function such that $\sum_{k \in \mathbf{Z}} |\hat{\phi}(\xi + 2k\pi)| = 1$ almost everywhere. Assume that there exists a 2π -periodic function $m_0 \in \mathcal{L}[0, 2\pi]$ such that $\hat{\phi}(\xi) = \hat{\phi}(\xi/2)m_0(\xi/2)$ where $\hat{\phi}$ is the Fourier transform of ϕ . Let ϕ satisfy Condition $S(Q)$ and let ϕ be a $(Q+1)$ times weakly differentiable function. If $\phi^{(Q+1)}$ satisfies Condition θ , then the norm $\|\cdot\|_{spq}$ in B_p^{sq} is equivalent to the norm $\|\cdot\|'_{spq}$ in the space of wavelet coefficients for all s, p , and q such that $0 < s < Q+1$ and $1 \leq p, q \leq \infty$ where*

$$\|f\|'_{spq} = \|\alpha_0\|_p + \left(\sum_{j=0}^{\infty} \left(2^{j(s+1/2-1/p)} \|\beta_j\|_p \right)^q \right)^{1/q}.$$

We also need the following two inequalities (see Appendix C in [2]).

Bernstein inequality. *Let Y_i be independent random variables such that*

$$|Y_i| \leq M, \quad \mathbf{E} Y_i = 0, \quad \mathbf{E} Y_i^2 < \infty$$

where M is a nonrandom constant. Put $b_n = \sum_{i=1}^n \mathbf{E} Y_i^2$. Then

$$\mathbf{P} \left(\left| \sum_{i=1}^n Y_i \right| \geq \lambda \right) \leq 2 \exp \left(- \frac{\lambda^2}{2(b_n^2 + \lambda M/3)} \right)$$

for all λ .

Rosenthal inequality. *Let $p \geq 2$. Assume that $X_1, \dots, X_n$ are independent random variables such that $\mathbf{E} X_i = 0$ and $\mathbf{E} |X_i|^p < \infty$. Then there is a number $C(p) > 0$ such that*

$$\mathbf{E} \left| \sum_{i=1}^n X_i \right|^p \leq C(p) \left(\sum_{i=1}^n \mathbf{E} |X_i|^p + \left(\sum_{i=1}^n \mathbf{E} X_i^2 \right)^{p/2} \right).$$

4. CONSTRUCTION OF A LINEAR ESTIMATOR

We consider the space $\mathcal{L}_2(\mathbf{R})$ of real square-integrable functions. We define the scalar product $\langle \cdot, \cdot \rangle_N$ in the space $\mathbf{R}^N$ for $x, y \in \mathbf{R}^N$ as follows:

$$\langle x, y \rangle_N := \frac{1}{N} \sum_{k=1}^N x_k y_k.$$

If there exists a solution

$$a_l = \{a_l(t_j), j = 1, \dots, N\}$$

of the problem

$$(2) \quad \begin{cases} \langle a_l, w_k \rangle_N = \delta_{kl}, \\ \langle a_l, a_l \rangle_N \rightarrow \min, \end{cases}$$

then we construct a linear estimator for the component l where δ_{kl} is the Kronecker delta. By $\Gamma_N = (\langle w_k, w_l \rangle_N)_{k,l=1}^M$ we denote the Gram matrix of the system of vectors w_k for the scalar product $\langle \cdot, \cdot \rangle_N$ and let γ_{lk}^N be the minor (l, k) of the matrix Γ_N . If $\det \Gamma_N > 0$, then we solve the problem by using Lagrange multipliers and obtain (see [1])

$$a_l(t) = \frac{1}{\det \Gamma_N} \sum_{k=1}^M (-1)^{l+k} \gamma_{lk}^N w_k(t).$$

Consider the father wavelet ϕ and construct the linear estimator of resolution level s for the component l :

$$(3) \quad \hat{p}_l^s(x) = \frac{1}{N} \sum_{j=1}^N a_l(t_j) K_s(x, X_j)$$

where $K_s(x, y) = 2^s K(2^s x, 2^s y)$ and $K(x, y)$ is the projective orthogonal kernel associated with ϕ .

Lemma 1. $\hat{p}_l^s(x)$ is an unbiased estimator for $p_l^s(x) \equiv (\text{Pr}_{V_s} p_l)(x)$.

Proof. Since X_j is a random variable with the density $p_{t_j} = \sum_{k=1}^M w_k(t_j) p_k(x)$, $K_s(x, X_j)$ is an unbiased estimator for $p_{t_j}^s(x)$ (see [2, p. 126]) and the projection operator is linear, we get

$$\mathbb{E} \hat{p}_l^s(x) = \frac{1}{N} \sum_{j=1}^N a_l(t_j) \left(\sum_{k=1}^M w_k(t_j) p_k^s(x) \right) = \sum_{k=1}^M p_k^s(x) \langle a_l, w_k \rangle_N = p_l^s(x). \quad \square$$

Lemma 2 ([2, p. 129]). Let ϕ be a father wavelet whose associated kernel $K(x, y)$ is such that there exists $F \in \mathcal{L}_2(\mathbf{R})$ for which

$$|K(x, y)| \leq F(x - y).$$

Let p be the density of the random variable X such that $p \in \mathcal{L}_2(\mathbf{R})$. Then

$$\mathbb{E} \|K_s(x, X) - (\text{Pr}_{V_s} p)(x)\|_2^2 \leq 2^s \int_{\mathbf{R}} F^2(u) du.$$

Lemma 3. Let ϕ be a father wavelet and denote its associated kernel by K . Assume that there exists $F \in \mathcal{L}_2(\mathbf{R})$ such that $|K(x, y)| \leq F(x - y)$. If $p_k \in \mathcal{L}_2(\mathbf{R})$, $k = 1, \dots, M$, and $\hat{p}_l^s$ is the estimator defined by (3), then

$$\mathbb{E} \int_{\mathbf{R}} (\hat{p}_l^s(x) - p_l^s(x))^2 dx \leq \frac{1}{N} \langle a_l, a_l \rangle_N 2^s \int_{\mathbf{R}} F^2(u) du$$

for all $l = 1, \dots, M$ and all $s \geq 0$.

Proof. Since for all $x \in \mathbf{R}$

$$Y_j(x) := K_s(x, X_j) - p_{t_j}^s(x), \quad j = 1, \dots, N,$$

are jointly independent random variables and $\mathbb{E} Y_j(x) = 0$ almost everywhere with respect to x , Lemmas 1 and 2 imply that

$$\begin{aligned} \mathbb{E} \int_{\mathbf{R}} (\hat{p}_l^s(x) - p_l^s(x))^2 dx &= \mathbb{E} \int_{\mathbf{R}} \frac{1}{N} \sum_{j=1}^N \frac{1}{N} a_l^2(t_j) Y_j^2(x) dx \\ &\leq \frac{1}{N} \langle a_l, a_l \rangle_N 2^s \int_{\mathbf{R}} F^2(u) du. \quad \square \end{aligned}$$

In what follows the relation “ $\simeq$ ” for sequences $a_n \geq 0$ and $b_n \geq 0$ means that there are $C_1 > 0$ and $C_2 > 0$ such that

$$C_1 a_n \leq b_n \leq C_2 a_n$$

for all $n \geq 1$.

Theorem 1. *Let ϕ satisfy the assumptions of Proposition B with $Q \geq 0$ and the assumptions of Lemma 3. Let $p_i \in B(s, 2, q, L)$ for some $L > 0$ and for all $i = 1, \dots, M$, $s \in (0, Q + 1)$, and $1 \leq q \leq 2$. Then*

there exists $C = C(s, q, L) > 0$ such that for all $l = 1, \dots, M$

$$\sup_{\substack{p_i \in B(s, 2, q, L) \\ i=1, \dots, M}} \mathbb{E} \left\| \hat{p}_l^j - p_l \right\|_2^2 \leq C \left(\langle a_l, a_l \rangle_N \frac{2^j}{N} + 2^{-2js} \right)$$

where the estimator $\hat{p}_l^j$ is defined by (3) and N is the size of the sample.

Proof. Since $\mathbb{E} \hat{p}_l^j = \Pr_{V_j} p_l$ and $\hat{p}_l^j \in V_j$, we have

$$\mathbb{E} \left\| \hat{p}_l^j - p_l \right\|_2^2 = \mathbb{E} \left\| \hat{p}_l^j - \mathbb{E} \hat{p}_l^j \right\|_2^2 + \left\| \mathbb{E} \hat{p}_l^j - p_l \right\|_2^2.$$

Now we estimate the second term on the right-hand side of the latter equality by applying Proposition B and Lemma 1:

$$\left\| \mathbb{E} \hat{p}_l^j - p_l \right\|_2^2 = \left\| p_l^j - p_l \right\|_2^2 = \sum_{i=j}^{\infty} \sum_{k \in \mathbf{Z}} |\beta_{ik}|^2 \leq 2^{-2js} \sum_{i=j}^{\infty} 2^{2is} \|\beta_i\|_2^2 \leq 2^{-2js} (\|p_l\|_{s22}')^2,$$

since $B_q^{s2}(\mathbf{R}) \subset B_2^{s2}(\mathbf{R})$ (see Corollary 9.2(i) in [2, p. 124]). Moreover for all $L > 0$ and all q , $1 \leq q \leq 2$, there exists a constant $C_1 = C_1(s, q, L)$ such that for all $p_l \in B(s, 2, q, L)$

$$(\|p_l\|_{s22}')^2 \leq C_1.$$

Thus $\left\| \mathbb{E} \hat{p}_l^j - p_l \right\|_2^2 \leq C_1 s^{-2js}$.

Now we estimate the first term. Applying Lemma 3 we get

$$\mathbb{E} \left\| \hat{p}_l^j - p_l^j \right\|_2^2 \leq \frac{1}{N} \langle a_l, a_l \rangle_N 2^j \int_{\mathbf{R}} F^2(u) du = C_2 \langle a_l, a_l \rangle_N \frac{2^j}{N}.$$

Putting $C = \max(C_1, C_2)$, we complete the proof of the theorem. $\square$

Remark. If j is such that $2^j \simeq N^{1/(2s+1)}$, then the right-hand side of the inequality of Theorem 1 is minimal and equals $C'(\langle a_l, a_l \rangle_N + 1)N^{-2s/(2s+1)}$.

Lemma 4. *Let $f \in C^1(\mathbf{R})$ be such that $\|f\|_{\infty} < \infty$ and $\|f'\|_{\infty} < \infty$. Then*

$$\|f\|_{\infty} \leq \sqrt[3]{4 \|f'\|_{\infty} \|f\|_2^2}.$$

Proof. Indeed, for all $\varepsilon \in (0, 1)$ there exists x_0 such that $|f(x_0)| \geq \varepsilon \|f\|_{\infty}$. Then there exists $\delta > 0$ such that

$$|f(x)| \geq \varepsilon \|f\|_{\infty} / 2$$

for all $x \in (x_0 - \delta, x_0 + \delta)$ and $\delta \geq \varepsilon \|f\|_{\infty} / (2 \|f'\|_{\infty})$. Therefore

$$\|f\|_2^2 \geq \int_{x_0 - \delta}^{x_0 + \delta} |f(x)|^2 dx \geq \int_{x_0 - \delta}^{x_0 + \delta} (\varepsilon \|f\|_{\infty} / 2)^2 dx \geq \left(\frac{\varepsilon \|f\|_{\infty}}{2} \right)^2 \frac{\varepsilon \|f\|_{\infty}}{\|f'\|_{\infty}} = \frac{(\varepsilon \|f\|_{\infty})^3}{4 \|f'\|_{\infty}},$$

whence

$$\varepsilon \|f\|_{\infty} \leq \sqrt[3]{4 \|f\|_2^2 \|f'\|_{\infty}}$$

for all $\varepsilon \in (0, 1)$. Letting $\varepsilon \rightarrow 1$ we obtain the desired result. $\square$

Theorem 2. Let $B > 0$, ϕ be a father wavelet, $K(x, y)$ the orthogonal kernel associated with ϕ , and let N be the size of the sample. If there exists $Q \in \mathbb{N}$ such that $\phi \in C^Q(\mathbf{R})$ and $\text{supp } \phi$ is a compact set, then there exists $C = C(Q, B)$ such that

$$\sup_{\substack{p_i \in W_{\infty}^{Q+1}(\mathbf{R}), i=1, \dots, M, \\ \|p_i\|_{W_{\infty}^{Q+1}} \leq B,}} \mathbb{E} \left\| \hat{p}_l^j - p_l \right\|_{\infty} \leq C \left(2^{-j(Q+1)} + \frac{2^j}{N^{1/3}} \sqrt{\langle a_l, a_l \rangle_N} \right).$$

Proof.

$$\mathbb{E} \left\| \hat{p}_l^j - p_l \right\|_{\infty} \leq \mathbb{E} \left\| \hat{p}_l^j - \mathbb{E} \hat{p}_l^j \right\|_{\infty} + \mathbb{E} \left\| \mathbb{E} \hat{p}_l^j - p_l \right\|_{\infty} = \mathbb{E} \left\| \hat{p}_l^j - p_l^j \right\|_{\infty} + \left\| p_l^j - p_l \right\|_{\infty},$$

since $\hat{p}_l^j$ is an unbiased estimator of p_l^j .

Let us estimate the second term. Since $\phi \in C^Q(\mathbf{R})$, $Q \geq 1$, and $\text{supp } \phi$ is a compact set, condition $S(n)$ holds for all $n \in \mathbf{N}$ and moreover $\phi \in \tilde{W}_{\infty}^Q(\mathbf{R})$. Thus there exists a constant $C_1 = C_1(B)$ such that

$$\left\| p_l^j - p_l \right\|_{\infty} \leq C_1(B) 2^{-j(Q+1)}$$

for all $p_l \in W_{\infty}^{Q+1}(\mathbf{R})$ with $\|p_l\|_{W_{\infty}^{Q+1}} \leq B$ (see Theorem 8.1(ii) in [2, p. 76], Lemma 8.6 in [2, p. 85], and Theorem 8.2 in [2, p. 91]).

Now we estimate the first term. Since ϕ is a smooth function with a compact support, $K(x, y)$ is a uniformly bounded and continuously differentiable function with respect to both arguments and such that there exists a constant $C_2 \in \mathbf{R}_+$ for which $|K'_x(x, y)| \leq C_2$, whence we obtain that $|(K_j)'_x(x, y)| \leq 4^j C_2$. Thus

$$|(\mathbb{E} K_j(x, X_i))'_x| = \left| \left(\int_{\mathbf{R}} K_j(x, y) p_{t_i}(y) dy \right)'_x \right| = \left| \int_{\mathbf{R}} (K_j)'_x(x, y) p_{t_i}(y) dy \right| \leq 4^j C_2.$$

Put $Y(x) := \hat{p}_l^j(x) - \mathbb{E} \hat{p}_l^j(x)$. Then for all $x \in \mathbf{R}$ the derivative $Y'(x)$ exists and

$$|Y'(x)| = \left| \frac{1}{N} \sum_{i=1}^N a_l(t_i) \left(K_j(x, X_i) - p_{t_i}^j(x) \right)'_x \right| \leq \frac{2}{N} C_2 4^j \sum_{i=1}^N |a_l(t_i)| \leq 2 C_2 4^j \sqrt{\langle a_l, a_l \rangle_N}.$$

Applying Lemma 4 we obtain for all $X_1, \dots, X_N$ that

$$\|Y\|_{\infty} \leq 2 \cdot 2^j C_3 \sum_{i=1}^N |a_l(t_i)| / N < \infty$$

where $C_3 > 0$ is such that

$$|K(x, y)| < C_3$$

for all $x, y \in \mathbf{R}$. Thus

$$\|Y\|_{\infty} \leq \left(8 C_2 4^j \sqrt{\langle a_l, a_l \rangle_N} \|Y\|_2^2 \right)^{1/3}$$

for all $X_1, \dots, X_N$. Now we check the assumptions of Lemma 3. Indeed,

$$|K(x, y)| < C_3$$

for all $x, y \in \mathbf{R}$. Moreover $K(x, y) = 0$ if $|x - y| > L := 2 \text{diam}(\text{supp } \phi) < \infty$. Thus one can take $C_3 \mathbb{1}_{[0, L]}(|z|) \in \mathcal{L}_2(\mathbf{R})$ as $F(z)$. Then

$$\begin{aligned} \mathbb{E} \left\| \hat{p}_l^j - \mathbb{E} \hat{p}_l^j \right\|_{\infty} &\leq \left(8 C_2 4^j \sqrt{\langle a_l, a_l \rangle_N} \frac{1}{N} 2^j \langle a_l, a_l \rangle_N \int_{\mathbf{R}} F^2(x) dx \right)^{1/3} \\ &= C_5 \frac{2^j}{N^{1/3}} \sqrt{\langle a_l, a_l \rangle_N}. \end{aligned}$$

Putting $C = \max(C_1, C_5)$, we complete the proof of the theorem. $\square$

Remark. Putting $2^j \simeq \sqrt[3Q+6]{N}$ we get the following rate of convergence of the estimator:

$$CN^{-(Q+1)/(3Q+6)} (\langle a_l, a_l \rangle_N + 1).$$

Consider the Haar basis. It is generated by the functions $\phi = \mathbb{1}_{[0,1)}$ (father wavelet) and

$$\psi = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}$$

(mother wavelet). The associated orthogonal kernel is $K(x, y) = \mathbb{1}_{[x]=[y]}$ where $[x]$ is the integer part of a number x . Consider the following class of functions.

Definition 3.

$\text{DLip}(M, L) := \{f : f \text{ is a density, } \text{diam supp } f \leq L, \text{ and } |f(x) - f(y)| \leq M|x - y| \text{ for all } x, y \in \mathbf{R}\}.$

Theorem 3. *Let $\phi = \mathbb{1}_{[0,1)}$ be a father wavelet. Then there exists $C = C(B, L)$ such that*

$$\sup_{\substack{p_l \in \text{DLip}(B, L) \\ l=1, \dots, M}} \mathbb{E} \left\| \hat{p}_l^j - p_l \right\|_2^2 \leq C \left(B^2 2^{-j} + \frac{\langle a_l, a_l \rangle_N}{N} 2^j \right).$$

Putting $2^j \simeq N^{1/3}$, we decrease the right-hand side up to $C' N^{-2/3} (1 + \langle a_l, a_l \rangle_N)$.

Proof. Like the proofs above we apply Lemma 1 and obtain that

$$\mathbb{E} \left\| \hat{p}_l^j - p_l \right\|_2^2 = \mathbb{E} \left\| \hat{p}_l^j - p_l^j \right\|_2^2 + \left\| p_l^j - p_l \right\|_2^2,$$

since $\hat{p}_l^j \in V_j$ and $\mathbb{E} \hat{p}_l^j = p_l^j$. Let us estimate the second term. One can show that for any function of the class $\text{DLip}(B, L)$ there exists a constant $C_1(L) > 0$ such that

$$\left\| p_l^j - p_l \right\|_2^2 \leq C_1(L) B^2 2^{-2j}.$$

To estimate the first term we apply Lemma 3 with $F(v) = \mathbb{1}_{|v| \leq 1}$:

$$\mathbb{E} \left\| \hat{p}_l^j - p_l^j \right\|_2^2 \leq C_2 \frac{\langle a_l, a_l \rangle_N}{N} 2^j.$$

Summing up the two terms and considering $C = \max(C_1(L), C_2)$ we complete the proof of the theorem. $\square$

5. CONSTRUCTION OF AN ADAPTIVE ESTIMATOR AND ITS PROPERTIES

Consider the so-called adaptive hard threshold estimator. If a_l is the solution of the problem (2), then the estimator of the component l is

$$(4) \quad \hat{p}_l^*(x) = \sum_{k \in \mathbf{Z}} \hat{\alpha}_{j_0 k}^* \phi_{j_0 k} + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbf{Z}} \hat{\beta}_{jk}^* \psi_{jk}(x)$$

where

$$\begin{aligned} \hat{\alpha}_{jk}^* &= \hat{\alpha}_{jk} = N^{-1} \sum_{i=1}^N a_l(t_i) \phi_{jk}(X_i), & \hat{\beta}_{jk}^* &= \hat{\beta}_{jk} \cdot \mathbb{1}_{|\hat{\beta}_{jk}| > t_j}, \\ \hat{\beta}_{jk} &= N^{-1} \sum_{i=1}^N a_l(t_i) \psi_{jk}(X_i) \end{aligned}$$

and ϕ and ψ are the father wavelet and the mother wavelet, respectively. It is easy to see that

$$\begin{aligned} \mathbb{E} \hat{\alpha}_{jk} &= \alpha_{jk} \equiv \int_{\mathbf{R}} \phi_{jk}(x) p_l(x) dx, \\ \mathbb{E} \hat{\beta}_{jk} &= \beta_{jk} \equiv \int_{\mathbf{R}} \psi_{jk}(x) p_l(x) dx. \end{aligned}$$

We fix $L > 0$.

Definition 4. Put

$$\tilde{B}^{C*}(s, r, q, M^*) = \left\{ f \text{ is a density: } \|f\|_{srq} \leq M^*, \|f\|_{\infty} \leq C^*, \text{diam}(\text{supp } f) \leq L \right\}.$$

Theorem 4. Assume that

1. The father wavelet ϕ and the mother wavelet ψ are bounded, $\text{supp } \phi$ and $\text{supp } \psi$ are compact sets, and there is $Q \in \mathbf{Z}$ such that $Q \geq 0$ and $\phi^{(Q+1)}$ is bounded.
2. $\hat{\beta}_{jk}^* = \hat{\beta}_{jk} \mathbb{1}_{|\hat{\beta}_{jk}| > t_j}$ for $t_j = c\sqrt{j/N}$.
3. The following inequalities hold:

$$\begin{aligned} 1 \leq q \leq r \leq 2, \quad Q+1 > s > \frac{1}{r}, \\ 2^{j_0} \simeq N^{\frac{1}{2s+1}}, \quad 2^{j_1} \geq N^{\frac{\alpha}{s-1/r+1/2}}, \quad \alpha = \frac{s}{2s+1}. \end{aligned}$$

Then there is a constant $c_0 > 0$ such that for all $c > c_0$ there exists a number $C > 0$ for which

$$\sup_{\substack{p_i \in \tilde{B}^{C*}(s, r, q, M^*) \\ i=1, \dots, M}} \mathbb{E} \|\hat{p}_l^* - p_l\|_2^2 \leq C N^{-2s/(2s+1)} (\ln N)^{(2-r)/2} \max(\langle a_l, a_l \rangle_N, 1)$$

where $\hat{p}_l^*$ is estimator (4) with parameters j_0, j_1 , and t_j .

Proof. We follow the method used in [2]. Put

$$\hat{p}_l^{j_0}(x) = \sum_{k \in \mathbf{Z}} \hat{\alpha}_{j_0 k} \phi_{j_0 k}(x).$$

Then

$$\begin{aligned} \mathbb{E} \|\hat{p}_l^* - p_l\|_2^2 &= \underbrace{\mathbb{E} \|\hat{p}_l^{j_0} - \mathbb{E} \hat{p}_l^{j_0}\|_2^2}_{T_1} + \underbrace{\sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} \mathbb{E} \left((\hat{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}_{|\hat{\beta}_{jk}| > t_j} \right)}_{T_2} \\ &\quad + \underbrace{\sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} \beta_{jk}^2 \mathbb{P}(|\hat{\beta}_{jk}| \leq t_j)}_{T_3} + \underbrace{\sum_{j=j_1+1}^{\infty} \sum_{k \in \Omega_j} \beta_{jk}^2}_{T_4}. \end{aligned}$$

The set $\Omega_j = \{k: \beta_{jk} \neq 0\}$ is bounded, since $\text{supp } \phi$ and $\text{supp } \psi$ are bounded. In particular, $\text{Card}(\Omega_j) \leq 2^j L + \tau$ where L is the constant involved into the definition of the class $\tilde{B}^{C*}(s, r, q)$, and

$$\tau := \max(\text{diam supp } \phi, \text{diam supp } \psi).$$

Using Lemma 3 and assumption 3 of the theorem we get

$$\text{there exists } C'_1 > 0 \text{ such that for all } p_l \in \tilde{B}^{C*}(s, r, q, M^*)$$

$$T_1 \leq C'_1 \frac{2^{j_0}}{N} \langle a_l, a_l \rangle_N \leq C_1 N^{-2s/(2s+1)} \langle a_l, a_l \rangle_N.$$

The assumptions of Lemma 3 hold, since $\text{supp } \phi$ is a compact set. Thus $K(x, y) = 0$ if $|x - y| > \text{diam supp } \phi$, and moreover there is c_{11} such that $|K(x, y)| \leq c_{11}$.

Let us estimate the term T_4 . Using Corollary 9.2(i), (iii) in [2, p. 124] we prove that $B_r^{sq} \subset B_2^{s'2}$ for $r \leq 2$ and $s' = s - 1/r + 1/2$. Moreover, $B^{C*}(s, r, q, M^*)$ is bounded in $B_2^{s'2}(\mathbf{R})$. It follows from Proposition B and assumption 3 that

$$T_4 = \sum_{j=j_1+1}^{\infty} \sum_{k \in \Omega_j} \beta_{jk}^2 \leq 2^{-2j_1 s'} \sum_{j=j_1+1}^{\infty} 2^{2js'} \sum_{k \in \mathbf{Z}} \beta_{jk}^2 \leq C'_2 2^{-2j_1 s'} \leq C_2 N^{-2s/(2s+1)}.$$

Here $C'_2 = \|p_l\|_2^2 \leq C_2 := \sup_{p_l \in B^{C*}(s, r, q, M^*)} \|p_l\|_2'^2 < \infty$.

To estimate $T_2 + T_3$ we note that

$$\mathbb{1}_{|\hat{\beta}_{jk}| > t_j, |\beta_{jk}| \leq t_j/2} \leq \mathbb{1}_{|\hat{\beta}_{jk} - \beta_{jk}| > t_j/2}, \quad \mathbb{1}_{|\hat{\beta}_{jk}| \leq t_j, |\beta_{jk}| > 2t_j} \leq \mathbb{1}_{|\hat{\beta}_{jk} - \beta_{jk}| > t_j/2}.$$

Further, if $|\hat{\beta}_{jk}| \leq t_j$ and $|\beta_{jk}| > 2t_j$, then $\beta_{jk}^2 \leq 4(\hat{\beta}_{jk} - \beta_{jk})^2$. Thus

$$\begin{aligned} T_2 + T_3 &\leq \underbrace{\sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} \mathbb{E} \left((\hat{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}_{|\beta_{jk}| > t_j/2} + \beta_{jk}^2 \mathbb{1}_{|\beta_{jk}| \leq 2t_j} \right)}_{T'_5} \\ &\quad + \underbrace{5 \sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} \mathbb{E} \left((\hat{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}_{|\hat{\beta}_{jk} - \beta_{jk}| > t_j/2} \right)}_{T_5}. \end{aligned}$$

Now we estimate T'_5 . It is clear that

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^2 = \text{Var } \hat{\beta}_{jk} = \frac{1}{N^2} \sum_{i=1}^N a_l^2(t_i) \text{Var } \psi_{jk}(X_i) \leq \frac{C_* \langle a_l, a_l \rangle_N}{N},$$

since p_t is a convex combination of p_r , $r = 1, \dots, M$, and hence $\|p_{t_i}\|_{\infty} \leq C_*$. It follows from the Markov inequality

$$\text{Card}(\{(j, k) : j_0 \leq j \leq j_1, |\beta_{jk}| > t_j/2\}) \leq \left(\frac{2}{t_{j_0}} \right)^r \sum_{j=j_0}^{j_1} \sum_{k \in \mathbf{Z}} |\beta_{jk}|^r$$

that

$$\sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} \mathbb{E} \left((\hat{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}_{|\beta_{jk}| > t_j/2} \right) \leq C_2 \langle a_l, a_l \rangle_N N^{-2s/(2s+1)},$$

since there is a constant C_{22} such that

$$\sum_{j=0}^{\infty} \sum_{k \in \mathbf{Z}} |\beta_{jk}|^r \leq C_{22} < \infty$$

for all $p_l \in \tilde{B}^{C*}(s, r, q, M^*)$, and there are $c'_1, c'_2 > 0$ such that

$$c'_1 \sqrt{j/N} \leq \sqrt{\ln N/N} \leq c'_2 \sqrt{j/N}, \quad 0 < j_0 \leq j \leq j_1.$$

The other part of T'_5 is estimated similarly:

$$\sum_{j=j_0}^{j_1} \sum_{k \in \mathbf{Z}} \beta_{jk}^2 \mathbb{1}_{|\beta_{jk}| \leq 2t_j} \leq (2t_{j_1})^{2-r} \sum_{j=j_0}^{j_1} \sum_{k \in \mathbf{Z}} |\beta_{jk}|^r \leq C_3 N^{-2s/(2s+1)} (\ln N)^{(2-r)/2},$$

since $r \leq 2$. Thus $T'_5 \leq C_4 N^{-2s/(2s+1)} (\ln N)^{(2-r)/2}$.

Now we estimate T_5 . In view of the Rosenthal inequality for $p = 4$ we have

$$\begin{aligned} \mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^4 &= \mathbb{E} \left(\frac{1}{N} \sum_{i=1}^N a_l(t_i) (\psi_{jk}(X_i) - \mathbb{E} \psi_{jk}(X_i)) \right)^4 \\ &\leq C_{51} \frac{1}{N^4} \left(\sum_{i=1}^N \mathbb{E}(\psi_{jk}(X_i) - \mathbb{E} \psi_{jk}(X_i))^4 a_l^4(t_i) \right. \\ &\quad \left. + \left(\sum_{i=1}^N \mathbb{E}(\psi_{jk}(X_i) - \mathbb{E} \psi_{jk}(X_i))^2 a_l^2(t_i) \right)^2 \right). \end{aligned}$$

It is easy to see that

$$\frac{1}{N^4} \sum_{i=1}^N \mathbb{E}(\psi_{jk}(X_i) - \mathbb{E} \psi_{jk}(X_i))^4 a_l^4(t_i) \leq \frac{2^j}{N^2} \langle a_l, a_l \rangle_N^2 C_{52},$$

while the second term does not exceed $C_{53}(\langle a_l, a_l \rangle_N / N)^2$.

Therefore

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^4 \leq C_{54} 2^j \frac{\max(\langle a_l, a_l \rangle_N, 1)^2}{N^2},$$

whence

$$T_5 \leq C_{55} \frac{\max \langle a_l, a_l \rangle_N}{N} \sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} 2^{j/2} \mathbf{P}^{1/2} \left(|\hat{\beta}_{jk} - \beta_{jk}| > c \sqrt{\frac{j}{N}} \right)$$

by the Cauchy–Schwarz inequality. To estimate the latter probability we apply the Bernstein inequality. According to the definition of $\hat{\beta}_{jk}$ and β_{jk} we obtain

$$\mathbf{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > c \sqrt{\frac{j}{N}} \right) \leq 2 \exp \left(- \frac{c^2 N j}{2 (b_N^2 + c \sqrt{N j} M' / 3)} \right)$$

where

$$M' := 2 \sqrt{N \langle a_l, a_l \rangle_N} 2^{j/2} \|\psi\|_\infty, \quad b_N^2 := N \langle a_l, a_l \rangle_N C_*.$$

Moreover

$$M' \geq |a_l(t_i)| \cdot |\psi_{jk}(X_i) - \mathbb{E} \psi_{jk}(X_i)|$$

for all $i, i = 1, \dots, N$. Further $b_N^2 \geq N \langle a_l, a_l \rangle_N \|p_{t_i}\| \geq \sum_{i=1}^N a_l^2(t_i) \text{Var} \psi_{jk}(X_i)$, thus the right-hand side is not increasing and the inequality remains true. Then

$$\mathbf{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > c \sqrt{\frac{j}{N}} \right) \leq 2 \exp \left(- \frac{c^2 j}{2 \left(\langle a_l, a_l \rangle_N C_* + \frac{2c \sqrt{j \langle a_l, a_l \rangle_N} 2^{j/2} \|\psi\|_\infty}{3} \right)} \right).$$

Since the right-hand side is decreasing to 0 as $c \rightarrow \infty$, we increase c to obtain

$$\mathbf{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > c \sqrt{j/N} \right) \leq 2^{-4j}$$

for all $j, j = j_0, \dots, j_1$.

Substituting the latter estimate into the above upper bound for T_5 , we get

$$T_5 \leq \frac{C_{56} \max(\langle a_l, a_l \rangle_N, 1)}{n} \sum_{j=j_0}^{j_1} \sum_{k \in \Omega_j} 2^{-3j/2} \leq \frac{C_5 \max(\langle a_l, a_l \rangle_N, 1)}{N},$$

since there is $\tilde{C} > 0$ such that $\text{Card}(\Omega_j) \leq \tilde{C} 2^j$ (see the beginning of the proof). Summing up all the terms we complete the proof of the theorem. $\square$

BIBLIOGRAPHY

1. R. E. Maïboroda, *Estimators of components of a mixture with varying concentrations*, Ukrain. Mat. Zh. **48** (1996), no. 4, 562–566; English transl. in Ukrainian Math. J. **48** (1997) no. 4, 618–622. MR1417019 (97j:62055)
2. W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov, *Wavelets, Approximation, and Statistical Applications*, Springer, New York, 1998. MR1618204 (99f:42065)
3. L. Devroye and L. Györfi, *Nonparametric Density Estimation. The L_1 -View*, Wiley, New York, 1985. MR0780746 (86i:62065)
4. D. Donoho, I. Johnstone, G. Kerkycharian, and D. Picard, *Density estimation by wavelet thresholding*, Ann. Statist. **24** (1996), 508–539. MR1394974 (97f:62061)
5. Ingrid Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1996. MR1162107 (93e:42045)
6. Jérôme Kalifa and Stéphane Mallat, *Thresholding estimators on linear inverse problems and deconvolutions*, Ann. Stat. **31** (2003), no. 1, 59–109. MR1962500 (2003k:62229)
7. R. Averkamp and C. Houdré, *Wavelet thresholding for non-necessarily Gaussian noise: idealism*, Ann. Stat. **31** (2003), no. 1, 110–151. MR1962501 (2004a:62085)

DEPARTMENT OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS, MECHANICS AND MATHEMATICS FACULTY, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE 6, KYIV 03127, UKRAINE

E-mail address: pokhid@ukr.net

Received 14/MAR/2003

Translated by OLEG KLESOV