Bernoulli **10**(2), 2004, 187–220

On minimax density estimation on $\mathbb{R}$

ANATOLI JUDITSKY and SOPHIE LAMBERT-LACROIX

LMC-IMAG, BP 53, F-38041 Grenoble Cedex 9, France. E-mail: juditsky@inrialpes.fr Sophie.Lambert@imag.fr

The problem of density estimation on $\mathbb{R}$ on the basis of an independent sample $X_1 \ldots, X_N$ with common density f is discussed. The behaviour of the minimax L_p risk, $1 \leqslant p \leqslant \infty$, is studied when f belongs to a Hölder class of regularity s on the real line. The lower bound for the minimax risk is given. We show that the linear estimator is not efficient in this setting and construct a wavelet adaptive estimator which attains (up to a logarithmic factor in N) the lower bounds involved. We show that the minimax risk depends on the parameter p when $p < 2 + 1/s$.

Keywords: adaptive estimation; minimax estimation; nonparametric density estimation

1. Introduction

Let $X_1 \ldots, X_N$ be a vector of independent realizations of a random variable X with cumulative distribution function F which possesses a density $f(\cdot)$ with respect to Lesbesgue measure on the real line. Our objective is to recover the unknown density function $f : \mathbb{R} \to \mathbb{R}^+$ given the sample $X_1 \ldots, X_N$.

This is a basic problem which has been extensively studied in the literature on nonparametric estimation; for an overview of various methods and approaches, see Devroye (1987) and Silverman (1986). When constructing an estimation algorithm, it is generally supposed that the estimated density f has certain regularity properties. In other words, f belongs to some functional class $\mathcal{F}$. This a priori knowledge allows us to form an estimator f_N (a measurable function of observations) of f. However, its statistical properties can only be studied asymptotically (when the sample size N tends to ∞). Then in order to derive the properties of the proposed estimator for finite N we have to establish the properties of the maximal risk over $\mathcal{F}$. This explains the common use of the so-called minimax approach.

In this set-up the risk

$$\rho(f_N, f) = \mathrm{E}_f \|f_N - f\|$$

is associated with an estimator f_N, where $\|\cdot\|$ is a functional norm or a seminorm. Then the minimax estimator f_N^* is the minimizer (over the set of all estimators) of the maximum risk $R(f_N, \mathcal{F})$ over the class $\mathcal{F}$:

$$R(f_N, \mathcal{F}) = \sup_{f \in \mathcal{F}} \rho(f_N, f).$$

Thus in the minimax framework f_N^* is the optimal estimator with accuracy $R_N(\mathcal{F}) = R(f_N^*, \mathcal{F})$. $R_N(\mathcal{F})$ is also referred to as the *minimax risk*. The principal question in the

1350–7265

minimax framework is how to form a minimax estimator and what is the value of the minimax risk $R_N(\mathcal{F})$.

We consider the following estimation problem. Suppose that the density $f(x)$, $f:\mathbb{R}\to\mathbb{R}^+$, belongs to a Hölder class $\mathcal{F}=\mathcal{F}(s,L)$, that is, the derivative $f^{(k)}$ of f, $k=\max\{i\in\mathbb{N}|i<s\}$, exists and

$$[f]_s+\|f\|_\infty\leqslant L,\qquad [f]_s=\sup_{x\neq y}\frac{|f^{(k)}(x)-f^{(k)}(y)|}{|x-y|^{s-k}}.$$

Our objective is to estimate f given independent observations $X_1\ldots,X_N$ with common density f.

The results obtained can be summarized as follows. Consider the minimax risk

$$R_N^{(p)}(\mathcal{F}(s,L))=\inf_{f_N}\sup_{f\in\mathcal{F}(s,L)}[\mathrm{E}_f\|\hat{f}_N-f\|_p^2]^{1/2},\qquad 1\leqslant p\leqslant\infty;$$

here the infimum is taken over all estimates f_N of f. We show that there exist 'universal' constants c and C depending only on the regularity parameters s and p such that, for $1\leqslant p<\infty$, the minimax risk

$$c\varphi(N)\leqslant R_N^{(p)}(\mathcal{F}(s,L))\leqslant C(\ln N)^\theta\varphi(N),\tag{1}$$

where

$$\varphi(t)=\begin{cases}L^{(p-1)/(p(s+1))}t^{-s/(2s+1)} & \text{if } 2+\dfrac{1}{s}<p<\infty,\\ L^{(p-1)/(p(s+1))}t^{-s(p-1)/(p(s+1))} & \text{if } 1\leqslant p\leqslant 2+\dfrac{1}{s},\end{cases}$$

with $\theta=\theta(s,p)>0$. Further, when $p=\infty$,

$$c\varphi\left(\frac{N}{\ln N}\right)\leqslant R_N^{(\infty)}(\mathcal{F}(s,L))\leqslant C\varphi\left(\frac{N}{\ln N}\right).$$

This result can be compared with the minimax rates for linear estimators. In the latter case we consider the minimax risk

$$R_N^{(l,p)}(\mathcal{F}(s,L))=\min_{\hat{f}_N^{(l)}}\max_{f\in\mathcal{F}}(\mathrm{E}_f\|\hat{f}_N^{(l)}-f\|_p^2)^{1/2},$$

where the minimum is taken over the class of linear estimators $\hat{f}_N^{(l)}$. Then we obtain

$$c\rho(N)\leqslant R_N^{(l,p)}(\mathcal{F}(s,L))\leqslant C\rho(N),$$

where

$$\rho(t)=\begin{cases}L^{(p-1)/(p(s+1))}t^{-s(1-1/p)/(2s(1-1/p)+1)} & \text{if } 2<p<\infty,\\ \varphi(t) & \text{if } 1\leqslant p\leqslant 2.\end{cases}$$

Note that the linear estimator is minimax (up to a constant) for $1\leqslant p\leqslant 2$. However, for $2<p<\infty$, the rate of the linear estimator is worse than that of a general nonlinear estimator.

These results deserve some comment.

Remark 1. From the large literature on minimax density estimation it is known that, as far as the estimation of a density on [0, 1] is concerned (cf. Ibragimov and Khas'minskij 1981), the minimax risk $R_n^{(p)}(\mathcal{F}(s, L))$ satisfies

$$R_N^{(p)}(\mathcal{F}(s, L)) \asymp L^{1/(2s+1)} N^{-s/(2s+1)} \qquad \text{for } 1 \leqslant p \leqslant \infty,$$

and

$$R_N^{(\infty)}(\mathcal{F}(s, L)) \asymp L^{1/(2s+1)} \left(\frac{\ln N}{N}\right)^{s/(2s+1)}.$$

Except for $p = \infty$, this rate of convergence does not depend on p. Furthermore, let the regularity class $\mathcal{F}$ be that of densities of 'spatially inhomogeneous smoothness', for instance, let $\mathcal{F}$ be a Besov class $\mathcal{F}(s, p', q, L)$ with small p'; see Donoho *et al*. (1996) for details. In this case the rate of convergence starts to deteriorate when $p > (2s+1)p'$ and depends heavily on p.

When the estimated density is supported on $\mathbb{R}$, the known results are as follows. Bretagnolle and Huber (1979) studied the behaviour of the maximal risk

$$R^{(p)}(\hat{f}_N, \mathcal{F}) = \sup_{f \in \mathcal{F}} [\mathrm{E}_f \|\hat{f}_N - f\|_p^2]^{1/2}$$

for a family of density classes $\mathcal{F} = \mathcal{G}(s, p, L)$, $2 \leqslant p < \infty$, of finite *jauge*. That is, a density $f \in \mathcal{G}(s, p, L)$ if its *jauge* $\rho_{s,p}(f) = \|f^{(s)}\|_p^{p/(2s+1)} \|f\|_{p/2}$ is bounded by L. It is shown that the maximal risk of the kernel estimator f_N is of order $N^{-s/(2s+1)}$ on $f \in \mathcal{G}(s, p, L)$. Note that here the parameter p is the same in the definition of the risk and of the functional class. Ibragimov and Khas'minskij (1980) established minimax rates of convergence for Sobolev classes on $\mathbb{R}$, $\mathcal{F} = \mathcal{F}(s, p, L)$, $2 \leqslant p \leqslant \infty$, and the risk $R_N^{(p)}$. They showed that in this set-up the maximal risk

$$R_N^{(p)}(\mathcal{F}) \asymp N^{-s/(2s+1)}.$$

When $p = \infty$ an extra logarithmic factor appears in the minimax risk:

$$R_N^{(\infty)}(\mathcal{F}) \asymp \left(\frac{\ln N}{N}\right)^{s/(2s+1)}.$$

Golubev (1992) gave the exact asymptotics of the minimax risk in this set-up when $p = 2$.

On the other hand, Donoho *et al*. (1996) studied the behaviour of the risk $R_N^{(p)}(\mathcal{F})$, $1 \leqslant p \leqslant \infty$, for a family of Besov functional classes $\mathcal{F} = \mathcal{F}(s, p', q, L)$ (here p' and p can be different). The result from their paper which is relevant to our study can be stated as follows: when $2 \leqslant p' \leqslant p$, the minimax rates of convergence $R_N^{(p)}(\mathcal{F}(s, p', q, L))$ for the density estimation on $\mathbb{R}$ are the same (up to a constant) as the minimax rates for the Besov class $\mathcal{F}(s, p', q, L)$ on [0, 1]. However, the problem of minimax density estimation on $\mathbb{R}$ when $p < p'$ remains open: minimax rates of convergence and minimax estimators are unknown in this case.

Remark 2. We observe in (1) that the minimax risk $R_N^{(p)}(\mathcal{F})$ on a Hölder class $\mathcal{F} = \mathcal{F}(s, L)$ on $\mathbb{R}$ is cardinally different when compared to that for $\mathcal{F}(s, L)$ on a compact. When $p > 2 + 1/s$, the minimax risk $R_N^{(p)}(\mathcal{F})$ is of the same order as in the equivalent estimation problem on [0, 1]. However, the behaviour of the minimax risk changes dramatically when p becomes smaller than the critical value $2 + 1/s$. In this zone the minimax risk depends heavily on p. To the best of our knowledge, the phenomenon observed is new.

Remark 3. The lower and the upper bound differ by a logarithmic factor. We suppose that the extra logarithm of N in the upper bound is due to the specific estimator we use. Note that in the case $p = \infty$, the extra logarithm also appears in the lower bound, and in this case the upper and the lower bound are equivalent up to a constant.

In fact, an interesting question can be asked: why do convergent nonparametric estimators of a density on a real line exist at all? Note that we wish to estimate a function on an infinite domain given only a finite number of observations. Then why would the expected L_p error, $1 \leqslant p \leqslant \infty$, be small in this situation? The general (and sloppy) answer to this question is quite simple: the function we estimate is a probability density. Therefore, the function f not only belongs to a 'regularity class' $\mathcal{F}$, but also satisfies the conditions $f(x) \geqslant 0$ and $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$, that is, $f \in \mathcal{F} \cap W_1^+$, where W_1^+ is the intersection of an L_1 ball of radius 1 with the positive octant. This condition provides an additional constraint when maximizing the risk $R^{(p)}$, $p > 1$ over $\mathcal{F}$. Note that this also provides an intuitive 'explanation' of the negative result in Devroye and Györfi (1985) on the rates of convergence for the $R^{(1)}$ risk: this extra constraint is of no use when the error is measured in the L_1 norm.

Note that the class $\mathcal{F} \cap W_1^+$ is that of functions of 'non-homogeneous' smoothness. It is well known (cf. Donoho *et al*. 1995) that linear estimators are suboptimal on such a class.

The rest of this paper is organized as follows. The lower bound for the minimax risk in (1) is given in Section 2. In Section 3 we study the properties of linear estimators, and in Section 4 we construct a wavelet adaptive estimator $\hat{f}_n$ of f which provides the upper bound in (1). The proofs of the results are gathered together in Section 5.

2. Lower bound for density estimation

Our objective here is to establish the lower bound for the minimax risk on the Hölder class $\mathcal{F}(s, L)$.

Theorem 1. *There is a positive constant $c_0 = c_0(s, p)$ such that, for any estimate $\hat{f}_N$ of f, the maximal risk $R^{(p)}(\hat{f}_N, \mathcal{F}(s, L))$ satisfies*

$$R^{(p)}(\hat{f}_N, \mathcal{F}(s, L)) \geqslant \begin{cases} c_0 L^{(p-1)/(p(s+1))} N^{-s(p-1)/(p(s+1))}, & \text{for } 1 \leqslant p \leqslant 2 + \dfrac{1}{s}, \\ c_0 L^{(p-1)/(p(s+1))} N^{-s/(2s+1)}, & \text{for } 2 + \dfrac{1}{s} < p < \infty. \end{cases} \tag{2}$$

Moreover, when $p = \infty$*, we have*

$$R^{(\infty)}(\hat{f}_N, \mathcal{F}(s, L)) \geqslant c_0 L^{1/(s+1)} \left(\frac{\ln N}{N}\right)^{s/(2s+1)}. \tag{3}$$

Our next objective is to provide an upper bound for the risk $R_N(\mathcal{F})$. We start with a linear density estimator.

3. Linear estimation

We recall some basic properties of a biorthogonal wavelet basis.

3.1. Biorthogonal wavelet basis

Let the tuple $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be such that $\{\phi(x-k),\ \psi(2^j x - k),\ j \geqslant 0,\ k \in \mathbb{Z}\}$ and $\{\tilde{\phi}(x-k), \tilde{\psi}(2^j x - k), j \geqslant 0, k \in \mathbb{Z}\}$ constitute a biorthogonal pair of bases for $L_2(\mathbb{R})$. Some popular examples of such bases are given in Daubechies (1992). We require the *reconstruction wavelet* $\tilde{\psi}$ and $\tilde{\phi}$ to be $\mathbf{C}^{M+1}$ for some $M \in \mathbb{N}$, $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ to have compact support and the *analysis wavelet* ψ to be orthogonal to polynomials of degree no greater than M.

This implies that any function $f \in L_2(\mathbb{R})$ can be represented as

$$f(x) = \sum_k \alpha_k \tilde{\phi}_k(x) + \sum_{j \geqslant 0} \sum_k \beta_{jk} \tilde{\psi}_{jk}(x),$$

where

$$\alpha_k = \int f(x)\phi_k(x)\mathrm{d}x, \qquad \beta_{jk} = \int f(x)\psi_{jk}(x)\mathrm{d}x.$$

For technical reasons, in the wavelet estimator below we use a specific choice of the biorthogonal basis $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$, due to Donoho (1994). This is the basis generated using the function $\phi(x) = 1_{-1/2 \leqslant x < 1/2}$, and, following Donoho (1994), we call it a *blocky biorthogonal basis*. The functions ψ, $\tilde{\phi}$ and $\tilde{\psi}$ are compactly supported. We denote by δ_{jk} (δ_k) the support set of $\psi_{jk}(x)$ $(\phi_k(x))$:

$$\delta_{jk} = \left\{x \in \mathbb{R} : -\frac{m}{2} \leqslant 2^j x - k < \frac{m}{2}\right\}, \qquad \delta_k = \left\{x \in \mathbb{R} : -\frac{1}{2} \leqslant x - k < \frac{1}{2}\right\}.$$

for some $m \in \mathbb{R}$, $m \geqslant 1$.

The feature of this particular basis which is intensively used in the proof of Theorem 4 below is that there exists $\nu > 0$ such that the analysis wavelet $\psi(x)$ satisfies

$$|\psi(x)| \geqslant \nu \qquad \text{for } -\frac{m}{2} \leqslant x < \frac{m}{2}, \tag{4}$$

that is, $|\psi(x)| \geqslant \nu$ on the support of ψ.

3.2. Estimation algorithm

Consider the following estimation algorithm (cf. the linear wavelet estimator of f, suggested in Donoho *et al*. 1996):

Algorithm 1.

1. *Let* j_0 *satisfy*

$$\begin{array}{ll} L^{1/(s+1)}N^{1/(2s(1-1/p)+1)} \leqslant 2^{j_0} \leqslant 2L^{1/(s+1)}N^{1/(2s(1-1/p)+1)}, & \text{for } 2 < p < \infty, \\ L^{1/(s+1)}N^{1/(s+1)} \leqslant 2^{j_0} \leqslant 2L^{1/(s+1)}N^{1/(s+1)}, & \text{for } 1 \leqslant p \leqslant 2. \end{array} \tag{5}$$

2. *Compute empirical wavelet coefficients*

$$y_{j_0 k} = \frac{1}{N}\sum_{i=1}^{N} \phi_{j_0 k}(X_i) = \frac{2^{j_0/2}}{N}\sum_{i=1}^{N} 1_{2^{-j_0}k < X_i \leqslant 2^{-j_0}(k+1)},$$

and form the estimator

$$\hat{f}_N(x) = \sum_k y_{j_0 k}\tilde{\phi}_{j_0 k}(x).$$

Theorem 2. *Let* $\mathcal{F}(s, L)$*,* $s < M + 1$*, be a Hölder class. The linear wavelet estimator* $\hat{f}_N$ *above satisfies, for N large enough,*

$$\sup_{f\in\mathcal{F}} [\mathrm{E}_f \|\hat{f}_N - f\|_p^2]^{1/2} \leqslant \rho_l(s, p, N, L),$$

where

$$\rho_l(s, p, N, L) = \begin{cases} C(s, p)L^{(p-1)/(p(s+1))}N^{-s(1-1/p)/(s+1)}, & \text{for } 1 \leqslant p \leqslant 2, \\ C(s, p)L^{(p-1)/(p(s+1))}N^{-s(1-1/p)/(2s(1-1/p)+1)}, & \text{for } 2 < p < \infty. \end{cases}$$

Remark 4. We observe that when $1 \leqslant p \leqslant 2$ the maximal risk of the estimator $\hat{f}_N$, computed by Algorithm 1, corresponds up a constant to the lower bound (2) of Theorem 1. Otherwise, for $2 < p < \infty$, the rate of convergence of such an estimator is much worse than that suggested by the corresponding lower bound. It is important to note that this is a property not of a particular wavelet estimator, but of the whole class of *linear estimators* $\hat{f}_N^{(l)}(x)$ such that

$$\hat{f}_N^{(l)}(x) = \int T(x, y)\mathrm{d}\hat{F}_N(y) = \frac{1}{N}\sum_{i=1}^{N} T(x, X_i) \tag{6}$$

(we call the estimator linear if it is a linear functional of the empirical cdf $\hat{F}_N$). We have the following lower bound for *any* estimator of that kind:

Theorem 3. *Let* $p \geqslant 2$. *There exists* $c = c(s, p)$ *such that, for N large enough and any linear estimator* $\hat{f}_N^{(l)}$,

$$\sup_{f\in\mathcal{F}(s,L)} [\mathrm{E}_f \|\hat{f}_N^{(l)} - f\|_p^2]^{1/2} \geqslant c\rho_l(s,\ p,\ N,\ L).$$

4. Adaptive wavelet estimator

Let $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be a blocky biorthogonal wavelet basis as defined above. We suppose that ψ is orthogonal to polynomials of degree no greater than M. Consider the following algorithm:

Algorithm 2.

1. *Choose the parameters*

$$\rho \geqslant 10\frac{\|\psi\|_\infty^2}{\nu^2}\left(\frac{128\|\psi\|_\infty^2}{\nu^2} + \frac{16}{3}\right) \quad \textit{and} \quad \lambda \geqslant 14, \tag{7}$$

where ν *is defined in* (4)*. Set* $m_N = \rho \ln N$ *and compute* $j_1 \geqslant 0$ *such that*

$$2^{j_1} \leqslant \frac{N}{m_N} < 2^{j_1+1}. \tag{8}$$

2. *For* $0 \leqslant j \leqslant j_1$*, compute empirical wavelet coefficients*

$$y_{jk} = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{i=1}^N \psi_{jk}(X_i), & \textit{if } \#\delta_{jk} \geqslant m_N, \\ 0, & \textit{if } \#\delta_{jk} < m_N, \end{cases} \qquad z_k = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{i=1}^N \phi_k(X_i), & \textit{if } \#\delta_k \geqslant m_N, \\ 0, & \textit{if } \#\delta_k < m_N; \end{cases}$$

here $\#\delta_{jk} = \sum_{I=1}^N 1_{\{x_i\in\delta_{jk}\}}$ *is the 'cardinality' of the support set* δ_{jk} *of the wavelet* ψ_{jk}*. Then, for* j *and* k *such that* $\#\delta_{jk} \geqslant m_N$*, compute empirical estimates* $\hat{\sigma}_{jk}^2$ *of the variance of* y_{jk}*:*

$$\hat{\sigma}_{jk}^2 = \frac{1}{N^2}\sum_{i=1}^N (\psi_{jk}^2(X_i) - y_{jk}^2).$$

3. *Compute shrunk wavelet coefficients*

$$\hat{\beta}_{jk} = y_{jk} 1_{|y_{jk}| \geqslant \hat{\gamma}_{jk}}, \tag{9}$$

where

$$\hat{\gamma}_{jk} = \lambda\sqrt{\ln N}\,\hat{\sigma}_{jk}.$$

4. *Compute the estimate*

$$\hat{f}_N(x) = \sum_k z_k \tilde{\phi}_{0k}(x) + \sum_{j=0}^{j_1}\sum_k \hat{\beta}_{jk}\tilde{\psi}_{jk}(x).$$

The properties of the estimator $\hat{f}_N$, computed by the above algorithm, are summarized in the following result:

Theorem 4. *Let* $\mathcal{F}^{(M)} = \{\mathcal{F}(s, L), 0 < s < M+1, 0 < K < \infty\}$ *be a family of Hölder classes. Then, for any class* $\mathcal{F}(s, L) \in \mathcal{F}^{(M)}$, *there exists a constant* $C = C(s, p)$ *such that, for N large enough,*

$$\sup_{f \in \mathcal{F}(s,L)} \mathrm{E}_f \|\hat{f}_N - f\|_p \leqslant C \begin{cases} L^{(p-1)/(p(s+1))} \left(\dfrac{\ln N}{N}\right)^{s(p-1)/(p(s+1))} & \text{for } 1 \leqslant p \leqslant 2 + \dfrac{1}{s}, \\ L^{1/(2s+1)} \ln N \left(\dfrac{\ln N}{N}\right)^{s/(2s+1)} & \text{for } p = 2 + \dfrac{1}{s}, \\ L^{(p-1)/(p(s+1))} \left(\dfrac{\ln N}{N}\right)^{s/(2s+1)} & \text{for } p > 2 + \dfrac{1}{s}. \end{cases}$$

Remark 5. The choice of the parameters ρ and λ in (7) is extremely conservative. We impose the lower bounds on ρ and λ in order to obtain very rough estimates of probabilities of moderate deviations which underlie the proof of Theorem 4 (see Lemma 8 and Proposition 1). When implementing the method an appropriate choice of these parameters would be $\lambda \in [\sqrt{2}, 2]$ and $\rho \in [1, 2]$.

Note that the wavelet shrinkage estimator described above is closely related to that of Donoho *et al*. (1996). For the problem of adaptive estimation on the Besov classes on [0, 1], the proposed estimator attains the same performance as the wavelet adaptive estimator in the latter paper.

The main difference between the adaptive estimator of Donoho *et al*. and the estimator given in Algorithm 2 is the implemenation of a data-driven thresholding procedure. The idea of data-driven thresholds for wavelet estimators is not new and has been used, for instance, in Birgé and Massart (2000), Donoho and Johnstone (1995) and Juditsky (1997), among many others. However, it is implemented differently in Algorithm 1, where the thresholds are computed individually for each wavelet coefficient. In other words, in order to take the decision to keep or to cut the empirical coefficient y_{jk} it is compared to the estimate $\hat{\sigma}_{jk}$ of its standard deviation. A closely related notion of spacing selection for *B*-splines Bayesian density estimator was implemented in Ciesielski and Kamont (1999) – cf. also the adaptive window selection procedure for kernel estimators in Juditsky and Nazin (2001) and Butucea (2000).

Note that another implementation of the same idea is provided by the celebrated $\sqrt{f}$-estimator of a density (cf. Anscombe 1948; Nussbaum 1996), when the empirical wavelet coefficients are 'normalized' to stabilize the value of σ_{jk}.

One may observe that the estimator $\hat{f}_N$ is adaptive. Indeed, the parameters of the estimation algorithm do not depend on a particular functional class $\mathcal{F}(s, L)$, but the maximal risk of $\hat{f}_N$ over $\mathcal{F}(s, L)$ coincides up to a logarithmic factor with the lower bound (2) of Theorem 1. The extra logarithm factor is the price often paid in adaptation procedures (cf. Lepskij 1992; Goldenshluger and Nemirovski 1997). However, we think that in our case (L_p risks and Hölder function classes) this extra factor is due to the particular construction of the estimator. Note that in the density estimation problem on [0, 1] one can get rid of the extra logarithm (Juditsky 1997). Nevertheless, at present we do not know of an estimator of f which attains the lower bound in (2).

5. Proofs

In the proofs below, C, C' and C'' stand for positive constants whose values may depend only on s, p and the wavelet parameters.

5.1. Proof of Theorem 1

The lower bound for the minimax risk $R^{(p)}(\hat{f}_N, \mathcal{F}(s, L))$ when $p > 2 + 1/s$ can easily be obtained using the construction of Theorem 5.1 in Ibragimov and Khas'minskij (1981). Our objective here is to show the bound in (2) in the case $1 \leqslant p \leqslant 2 + 1/s$.

To this end, consider a density $f_0 \in \mathcal{F}(s, L/2)$ such that $f_0(x) = c_1(s)L^{1/(s+1)}N^{-s/(s+1)}$ for $0 \leqslant x \leqslant L^{-1/(s+1)}N^{s/(s+1)}$ for some $c_1(s) > 0$. Now let $\gamma = (LN)^{-1/(s+1)}$ and $\gamma_k =](k-1)\gamma, k\gamma]$ for $k = 1, \ldots, N$; and let ψ_0 be a finite function such that

$$\psi_0(x) = 0, \quad \forall x \notin [-\tfrac{1}{2}, \tfrac{1}{2}], \qquad \|\psi_0\|_\infty = 1,$$

$$\psi_0(-x) = \psi_0(x), \quad \forall x \in [-\tfrac{1}{2}, \tfrac{1}{2}], \qquad \int \psi_0(x)\mathrm{d}x = 0.$$

Consider the set Ξ_N of 2^N binary vectors $\xi = (\xi_1, \ldots, \xi_N)$, $\xi_k \in \{-1, 1\}$, $k = 1, \ldots, N$. For each vector ξ, we define the function $f^{(\xi)}$ in the following way:

$$f^{(\xi)}(x) = f_0(x) + \sum_{k=1}^{N} \xi_k \psi_k(x), \qquad \psi_k(x) = \psi(x - (k - \tfrac{1}{2})\gamma),$$

where $\psi(x) = \alpha(s)L\gamma^s\psi_0(x/\gamma)$, in which $\alpha = \alpha(s)$ is a poitive constant small enough to ensure that ψ belongs to $\mathcal{F}(s, L/4)$ and that

$$|\psi(x)| \leqslant c_1(s)L^{1/(s+1)}N^{-s/(s+1)}.$$

Note that $\int f^{(\xi)}(x)\mathrm{d}x = 1$, so such a function is really a density. Further, due to the definition of ψ, the function $f^{(\xi)} - f_0$ belongs to $\mathcal{F}(s, L/2)$. This immediately implies $f^{(\xi)} \in \mathcal{F}(s, L)$.

Now let $\rho_H(\xi, \xi')$ be the Hamming distance between two vectors of Ξ_N, namely

$$\rho H(\xi, \xi') = \#\{k : 1 \leqslant k \leqslant N, \xi_k \neq \xi'_k\}.$$

There exist (see Korostelev and Tsybakov 1993, Lemma 2.7.4, p. 79) $M = [2^{N/8}]$ vectors $\xi^1, \ldots, \xi^M$ such that $\rho_H(\xi^j, \xi^k) \geqslant N/16$, $1 \leqslant j \leqslant k \leqslant M$. We denote by $\mathcal{F}_M$ the set of functions $f^{(\xi^1)}, \ldots, f^{(\xi^M)}$. Note that the $\|\cdot\|_p$ distance between two distinct functions f and g of $\mathcal{F}_M$ is at least $C(p)N^{1/p}\|\psi\|_p$. The problem of proving the lower bound over $\mathcal{F}(s, L)$ can be reduced to that over $\mathcal{F}_M$, that is,

$$\sup_{f\in\mathcal{F}(s,L)} [\mathrm{E}_f\|\hat{f}_N - f\|_p^2]^{1/2} \geqslant \sup_{f\in\mathcal{F}_M} [\mathrm{E}_f\|\hat{f}_N - f\|_p^2]^{1/2}.$$

We associate with any estimator $\hat{f}_N$ a method $\mathcal{M}$ for distinguishing between the M hypotheses, the kth of the them stating that the observations $X_1, \ldots, X_N$ are drawn from the kth element of the set $\mathcal{F}_M$. This method $\mathcal{M}$ is as follows: given observations, use an estimator $\hat{f}_N$ to find the closest element in $\mathcal{F}_M$ to $\hat{f}_N$ in L_p norm (any one of them in the non-uniqueness case) and claim that this is the density which underlies the observations.

Assume that the true hypothesis is associated with $f \in \mathcal{F}_M$. Note that if the method $\mathcal{M}$ fails to recognize the true density, this implies that $\hat{f}_N$ is at least at the same L_p distance from f as from other $g \in \mathcal{F}_M$. In other words,

$$\|\hat{f}_N - f\|_p \geqslant \|g - f\|_p/2 \geqslant C'(p)N^{1/p}\|\psi\|_p.$$

On the other hand, the Fano inequality states that the probability of the wrong choice among M hypotheses is not less than

$$1 - \frac{N\max_{f,g\in\mathcal{F}_M} K(f,\ g) + \ln 2}{\ln M},$$

where $K(f,\ g)$ is the Kullback distance between f and g (cf. Birgé 1983). Otherwise,

$$\sup_{f\in\mathcal{F}_M} \mathrm{E}_f\|\hat{f}_N - f\|_p^2 \geqslant C\left(1 - \frac{N\max_{f,g\in\mathcal{F}_M} K(f,\ g) + \ln 2}{\ln M}\right)N^{2/p}\|\psi\|_p^2. \tag{10}$$

We have the following lemma.

Lemma 1. *There exists $\alpha > 0$ such that*

$$\frac{N}{\ln M}\max_{f^{(\xi^j)},f^{(\xi^K)}\in\mathcal{F}_M} K(f^{(\xi^j)},\ f^{(\xi^k)}) \leqslant \frac{1}{2}.$$

Proof. Recall that the Kullback distance between f and g is defined by

$$K(f,\ g) = \int f(x)\ln\frac{f(x)}{g(x)}\,\mathrm{d}x.$$

Then for $f^{(\xi^j)}$, $f^{(\xi^k)} \in \mathcal{F}_M$, we have

$$
\begin{aligned}
K(f^{(\xi^j)}, f^{(\xi^k)}) &= \sum_{l=1}^{N} \int_{\gamma_l} [f_0(x) + \xi_l^j \psi_l(x)] \ln \frac{f_0(x) + \xi_l^j \psi_l(x)}{f_0(x) + \xi_l^k \psi_l(x)}, \\
&\leqslant \sum_{l=1}^{N} \int_{\gamma_l} [f_0(x) + \xi_l^j \psi_l(x)] \frac{(\xi_l^j - \xi_l^k)\psi_l(x)}{f_0(x) + \xi_l^k \psi_l(x)} \, \mathrm{d}x, \\
&\leqslant 2N \int_{\gamma_l} \frac{1 + \alpha/c_1(s)}{1 - \alpha/c_1(s)} \frac{\alpha}{c_1(s)} f_0(x) \mathrm{d}x, \\
&\leqslant C \frac{1 + \alpha/c_1(s)}{1 - \alpha/c_1(s)} \frac{\alpha}{c_1(s)}.
\end{aligned}
$$

Further, note that $N/\ln M = 8/\ln 2$. Thus a positive α can be found such that the quantity

$$
C \frac{N}{\ln M} \frac{1 + \alpha/c_1(s)}{1 - \alpha/c_1(s)} \frac{\alpha}{c_1(s)} \leqslant \frac{1}{2}.
$$

□

Hence, from (10) and Lemma 1 we conclude that

$$
\sup_{f \in \mathcal{F}_M} [\mathrm{E}_f \| \hat{f}_N - f \|_p^2]^{1/2} \geqslant C N^{1/p} \|\psi\|_p \geqslant c_0 L^{(p-1)/(p(s+1))} N^{-s(p-1)/(p(s+1))}.
$$

5.2. Translation into the sequence space

In what follows we will use some properties of the blocky biorthogonal multi-resolution analysis $\{\phi(x-k),\ \psi(2^j x - k),\ j \geqslant 0,\ k \in \mathbb{Z}\}$ and $\{\tilde{\phi}(x-k),\ \tilde{\psi}(2^j x - k),\ j \geqslant 0,\ k \in \mathbb{Z}\}$, described in Section 3.1.

Let $f \in L_2(\mathbb{R})$. Let $\{\alpha_k, \beta_{jk}, j \geqslant 0, k \in \mathbb{Z}\}$ be the wavelet coefficients of f. Then for $0 < p, q \leqslant \infty$, $1/p - 1 < s < M + 1$, the quantity

$$
\|f\|_{spq} = \|\alpha\|_p + \left(\sum_{j \geqslant 0} 2^{qj(s+1/2-1/p)} \|\beta_{j\cdot}\|_p^q \right)^{1/q}
$$

is equivalent to the norm $\|f\|_{B^s_{pq}}$ of the Besov space B^s_{pq} (Donoho 1994; Delyon and Juditsky 1997).

When using classical injection theorems (see, for instance, Triebel 1992), we conclude that there exist C_i, which may depend only on s, p, such that:

$$
\|f\|_1 \leqslant 1 \text{ implies that } \|\alpha\|_1 \leqslant 1 \text{ and } \sup_{j \geqslant 0} 2^{-j/2} \|\beta_{j\cdot}\|_1 \leqslant C_1; \tag{11}
$$

$$
\text{for any } f \in \mathcal{F}(s, L), \qquad \|\alpha\|_\infty + \sup_{j \geqslant 0} 2^{j(s+1/2)} \|\beta_{j\cdot}\|_\infty \leqslant C_2 L; \tag{12}
$$

$$\|f\|_p \leqslant C_3\left[\|\alpha\|_p^u + \sum_{j\geqslant 0} 2^{uj(1/2-1/p)}\|\beta_{j\cdot}\|_p^u\right]^{1/u}, \qquad u = \min(2,\ p). \tag{13}$$

The latter inequality implies, in particular, that if

$$\hat{f}_N(x) = \sum_k \hat{\alpha}_k \phi_k(x) + \sum_{j\geqslant 0}\sum_k \hat{\beta}_{jk}\psi_{jk}(x),$$

then

$$\|\hat{f}_N - f\|_p \leqslant C\left[\left(\sum_k |\hat{\alpha}_k - \alpha_k|^p\right)^{u/p} + \sum_{j\geqslant 0} 2^{uj(1/2-1/p)}\left(\sum_k |\hat{\beta}_{jk} - \beta_{jk})|^p\right)^{u/p}\right]^{1/u} \tag{14}$$

Otherwise, when $p \geqslant 2$,

$$\|f\|_p \geqslant C_4\left[\|\alpha\|_p^p + \sum_{j\geqslant 0} 2^{j(p/2-1)}\|\beta_{j\cdot}\|_p^p\right]^{1/p}. \tag{15}$$

5.3. Proof of Theorem 3

In order to prove the lower bound we implement the following idea which is due to Nemirovskii (1986) (cf. also the proof of Theorem 2 in (Donoho *et al*. 1996)). We construct a family of densities $\mathcal{G} \subset \mathcal{F}$ and a probability measure P on $\mathcal{G}$. Then we replace the original problem with the equivalent parameter one, that of estimating the vector of wavelet coefficients (β_{jk}) of $f \in \mathcal{G}$ by those of the linear estimator $\hat{f}_N^{(l)}$, that is,

$$\hat{\beta}_{jk} = \int \hat{f}_N^{(l)}(x)\psi_{jk}(x)\mathrm{d}x = \frac{1}{N}\sum_{i=1}^N \int T(x,\ X_i)\psi_{jk}(x)\mathrm{d}x. \tag{16}$$

We use the Cramér–Rao inequality to show that the Bayesian risk of any esimator of that type on the family $(\mathcal{G},\ P)$ is bounded from below.

Let j satisfy

$$L^{1/(s+1)}N^{1/(2s(1-1/p)+1)} \leqslant 2^j \leqslant 2L^{1/(s+1)}N^{1/(2s(1-1/p)+1)}. \tag{17}$$

Consider a density $v_0 \in \mathcal{F}(s,\ L/2)$ such that $v_0(x) = C_1(s)2^{-js}L$ for $0 < x \leqslant 2^{js}/L$ with some $C_1(s) > 0$. Now let u_0 be a density in $\mathcal{F}(s,\ L/2)$ such that $u_0(x) = C_2(s)L^{1/(s+1)}$, $C_2(s)L^{1/(s+1)}$, $C_2(s) > 0$, for $0 \leqslant x \leqslant L^{-1/(s+1)}$; it is a polynomial of degree $[s]$ when $x < 0$ and $x > L^{-1/(s+1)}$. Let $l^* = [N^{s/(2(s-1/p)+1)}]$. We set, for $l = 0,\ \ldots,\ l^* - 1$,

$$g_l(x) = \tfrac{1}{2}\left(v_0(x) + u_0(x - L^{-1/(s+1)}l)\right). \tag{18}$$

One can easily verify that $g_l \in \mathcal{F}(s,\ L/2)$. Let η be a random variable such that $P(\eta = l) = 1/l^*$ for $l = 0,\ \ldots,\ l^* - 1$. We set $r = [2^{j(s+1)}/L]$. Now consider a random vector $\xi = (\xi_1,\ \ldots,\ \xi_r)$ with independent components such that $P(\xi_k = 1) = P(\xi_k = -1) = \frac{1}{2}$ for $k = 1,\ \ldots,\ r$. For each realization of $(\eta,\ \xi)$ we define the function

$$f^{(\eta\xi)}(x) = g_\eta(x) + \sum_{k=1}^{r} \xi_k \beta \tilde{\psi}_{jk},$$

where the coefficient

$$\beta = C_3(s) L 2^{-j(s+1/2)} \tag{19}$$

is chosen to ensure that $f^{(\eta\xi)} \in \mathcal{F}(s, L)$ along with the condition

$$f^{(\eta\xi)}(x) \leqslant \tfrac{1}{2} g_\eta(x).$$

We consider the family $\mathcal{G}$ of functions $f^{(\eta\xi)}$ with the associated probability $\mathbf{P} = P_\eta \otimes P_\xi$ on $\mathcal{G}$. We denote by E_ξ (E_η) the expectation over the distribution of the vector ξ (η) and by $\mathbf{E}$ the expectation associated with $\mathbf{P}$.

Now let

$$\hat{f}_N^{(l)}(x) = \frac{1}{N} \sum_{i=1}^{N} T(x, X_i)$$

be a linear estimator of $f \in \mathcal{G}$, and

$$\hat{\beta}_{jk} = \frac{1}{N} \sum_{i=1}^{n} \int T(x, X_i) \psi_{jk}(x) \mathrm{d}x$$

be the corresponding 'estimate' of the wavelet coefficient $\beta_{jk} = \xi_k \beta$ for $k = 1, \ldots, r$.

Lemma 2. *Let $\hat{\beta}_{jk}$ be an estimator of β_{jk}, $k = 1, \ldots, r$, as above and $f \in \mathcal{G}$. Then*

$$\mathrm{E}_f(\hat{\beta}_{jk} - \beta_{jk})^2 \geqslant C|\lambda_{jk}|^2 N^{-1} \min_{x \in \delta_{jk}} g_\eta(x) + |\mathrm{E}_f \hat{\beta}_{jk} - \beta_{jk}|^2,$$

where δ_{jk} is the support of ψ_{jk} and

$$\lambda_{jk} = \frac{\partial \mathrm{E}_f \hat{\beta}_{jk}}{\partial \beta_{jk}} = \int T(x, y) \psi_{jk}(x) \tilde{\psi}_{jk}(y) \mathrm{d}x\, \mathrm{d}y.$$

Proof. Let $f(x) = f^{(\eta\xi)}(x)$ be a function in $\mathcal{G}$. The Cramér–Rao inequality, applied to any estimate $\hat{\beta}_{jk}$, gives

$$\mathrm{E}_f(\hat{\beta}_{jk} - \beta_{jk})^2 \geqslant \frac{(\partial \mathrm{E}_f \hat{\beta}_{jk} / \partial \beta_{jk})^2}{N I_{jk}} + (\mathrm{E}_f \hat{\beta}_{jk} - \beta_{jk})^2,$$

where I_{jk} is the Fisher information of the density f with respect to the parameter β_{jk}:

$$I_{jk} + \int \frac{(\partial f(x) / \partial \beta_{jk})^2}{f(x)} \mathrm{d}x.$$

Let us compute a bound for I_{jk}. Note that

$$\frac{\partial f(x)}{\partial \beta_{jk}} = \frac{\partial}{\partial \beta_{jk}}\left(g_\eta(x) + \sum_{k=1}^{r} \beta_{jk}\tilde{\psi}_{jk}(x)\right) = \tilde{\psi}_{jk}(x)$$

and

$$I_{jk} = \int \frac{\tilde{\psi}_{jk}^2(x)}{f(x)}\,\mathrm{d}x \leqslant \left(\min_{x\in\delta_{jk}} f(x)\right)^{-1} \leqslant 2\left(\min_{x\in\delta_{jk}} g_\eta(x)\right)^{-1},$$

where δ_{jk} is the support of $\tilde{\psi}_{jk}$. Further,

$$\frac{\partial \mathrm{E}_f\hat{\beta}_{jk}}{\partial \beta_{jk}} = \int \psi_{jk}(x)T(x,\, y)\frac{\partial f(x)}{\partial \beta_{jk}}\,\mathrm{d}x\,\mathrm{d}y = \int \psi_{jk}(x)T(x,\, y)\tilde{\psi}_{jk}(y)\mathrm{d}x\,\mathrm{d}y = \lambda_{jk}.$$

Putting all these results together gives the statement of the lemma. □

Lemma 3. $\mathrm{E}_\xi|\mathrm{E}_f\hat{\beta}_{jk} - \beta_{jk}|^2 \geqslant \beta^2|\lambda_{jk} - 1|^2$.

Proof. First, note that if

$$f_k^+(x) = g_\eta(x) + \sum_{l\neq k}^{r} \xi_l\beta\tilde{\psi}_{jl}(x) + \beta\tilde{\psi}_{jk}(x),$$

$$f_k^-(x) = g_\eta(x) + \sum_{l\neq k}^{r} \xi_l\beta\tilde{\psi}_{jl}(x) - \beta\tilde{\psi}_{jk}(x),$$

then

$$\begin{aligned}\mathrm{E}_{f_k^+}\hat{\beta}_{jk} - \mathrm{E}_{f_k^-}\hat{\beta}_{jk} &= \int T(x,\, y)\psi_{jk}(x)[f_k^+(y) - f_k^-(y)]\mathrm{d}x\,\mathrm{d}y \\ &= 2\beta\int T(x,\, y)\psi_{jk}(x)\tilde{\psi}_{jk}(y)\mathrm{d}x\,\mathrm{d}y = 2\beta\lambda_{jk}.\end{aligned}$$

As $|x|^2 + |y|^2 \geqslant \frac{1}{2}|x - y|^2$, when averaging over the distribution of ξ_k we obtain

$$\begin{aligned}\mathrm{E}_{\xi_k}|\mathrm{E}_f\hat{\beta}_{jk} - \beta_{jk}|^2 &= \tfrac{1}{2}[|\mathrm{E}_{f_k^+}\hat{\beta}_{jk} - \beta|^2 + |\mathrm{E}_{f_k^-}\hat{\beta}_{jk} + \beta|^2] \\ &\geqslant \tfrac{1}{4}|\mathrm{E}_{f_k^+}\hat{\beta}_{jk} - \beta - \mathrm{E}_{f_k^-}\hat{\beta}_{jk} - \beta|^2 = \beta^2|\lambda_{jk} - 1|^2.\end{aligned}$$

□

Let us now bound from below the risk of the estimate $\hat{f}_N^{(l)}$ on the family $\mathcal{G}$. Recall (cf. (15)) that for $p \geqslant 2$,

$$\|f\|_p^p \geqslant C\sum_{i=0}^{\infty} 2^{i(p/2-1)}\|\beta_{i\cdot}\|_p^p \geqslant 2^{j(p/2-1)}\|\beta_{j\cdot}\|_p^p$$

for any $j \geqslant 0$. Using Lemmas 2 and 3, by the Minkowski inequality,

$$\begin{aligned}\mathbf{E}\mathrm{E}_f\|\hat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^2 &\geqslant C2^{j(1-2/p)}\mathbf{E}\mathrm{E}_f\left(\sum_{k=1}^{r}|\hat{\beta}_{jk} - \beta_{jk}|^p\right)^{2/p} \\ &\geqslant C2^{j(1-2/p)}\left[\sum_{k=1}^{r}(\mathbf{E}\mathrm{E}_f|\hat{\beta}_{jk} - \beta_{jk}|^2)^{p/2}\right]^{2/p} \\ &\geqslant C'2^{j(1-2/p)}\left[\sum_{k=1}^{r}\left(|\lambda_{jk}|^p N^{-p/2}\left(\mathrm{E}_\eta \min_{x\in\delta_{jk}} g_\eta(x)\right)^{p/2} + \beta^p|\lambda_{jk} - 1|^p\right)\right]^{2/p}.\end{aligned} \tag{20}$$

Note that by definition of the family $\mathcal{G}$ (cf. (18)),

$$\mathrm{E}_\eta \min_{x\in\delta_{jk}} g_\eta(x) \geqslant \frac{CL^{1/(s+1)}}{l^*},$$

and

$$N^{-p/2}\left(\mathrm{E}_\eta \min_{x\in\delta_{jk}} g_\eta(x)\right)^{p/2} \geqslant L^{p/(2(s+1))}N^{-p(s+1/2)/(2s(1-1/\mathrm{p})+1)} \geqslant C\beta^p$$

since, by definition of β (cf. (19)) and j,

$$\beta = CL2^{-j(s+1/2)} \leqslant C'L^{1/(2(s+1))}N^{-(s+1/2)/(2s(1-1/p)+1)}.$$

Thus, from (20) we obtain

$$\mathbf{E}\mathrm{E}_f\|\hat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^2 \geqslant C2^{j(1-2/p)}\left[\sum_{k=1}^{r}\beta^p(|\lambda_{jk}|^p + |\lambda_{jk} - 1|^p)\right] \geqslant C'2^{j(1-2/p)}r^{2/p}\beta^2.$$

Substituting the latter inequality into the bound for 2^j from (17), one finally obtains

$$\mathbf{E}\mathrm{E}_f\|\hat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^2 \geqslant CL^{2(1-1/p)/(s+1)}N^{-2s(1-1/p)/(2s(1-1/p)+1)}.$$

Since

$$\sup_{f\in\mathcal{F}(s,L)} \mathrm{E}_f\|\hat{f}_N^{(l)} - f\|_p^2 \geqslant \sup_{f\in\mathcal{G}} \mathrm{E}_f\|\hat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^2 \geqslant \mathbf{E}\mathrm{E}_f\|\hat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^2,$$

this implies the desired bound.

5.4. Proof of Theorem 2

Let $p > 2$. We have the following simple lemma:

Lemma 4. *Let* $f \in \mathcal{F}(s, L)$ *be a density on* $\mathbb{R}$. *Then there exists* $C = C(s, p) < \infty$ *such that*

$$\|f\|_{s(1-1/p),p,\infty} \leqslant CL^{1-1/p},$$

that is, f belongs to the ball of radius $CL^{1-1/p}$, $C < \infty$, of the Besov space $B^{s(1-1/p)}_{p\infty}$.

Proof. We have $\|\alpha\|_p \leqslant \|\alpha\|_1$ for any $p \geqslant 1$. Further, since $\|\beta_{j\cdot}\|_p^p \leqslant \|\beta_{j\cdot}\|_1 \|\beta_{j\cdot}\|_\infty^{p-1}$, we have from (11) and (12),

$$\|\beta_{j\cdot}\|_p \leqslant \|\beta_{j\cdot}\|_1^{1/p} \|\beta_{j\cdot}\|_\infty^{1-1/p} \leqslant C2^{j/2p} L^{1-1/p} 2^{-j(s+1/2)(1-1/p)} = CL^{1-1/p} 2^{-j(s(1-1/p)+1/2-1/p)},$$

what implies the statement of the lemma. □

Now the upper bound of Theorem 2 when $p > 2$ follows from Theorem 1 of Donoho *et al*. (1996).

Now let $1 \leqslant p \leqslant 2$. Let δ_{jk} be the support bin of ϕ_{jk} and $p_{jk} = \int_{\delta_{jk}} f(x)\mathrm{d}x$. Note that

$$\alpha_{jk} = \mathrm{E}\phi_{jk} = 2^{j/2} p_{jk} \quad \text{and} \quad \mathrm{E}\phi_{jk}^2(X_1) = 2^j \int_{\delta_{jk}} f(x)\mathrm{d}x = 2^j p_{jk}.$$

As δ_{jk} and $\delta_{jk'}$ are disjointed for $k \neq k'$, we conclude that $\sum_k p_{jk} = 1$.

We have the following bound for the error of the estimation:

$$\|\hat{f}_N - f\|_p \leqslant C2^{j_0(1/2-1/p)} \|y_{j_0\cdot} - \alpha_{j_0\cdot}\|_p + C\sum_{j=j_0}^{\infty} 2^{j(1/2-1/p)} \|\beta_{j\cdot}\|_p = \delta_N^{(1)} + \delta_N^{(2)}.$$

The bound for the second term is an immediate consequence of (11) and (12):

$$\begin{aligned} \delta_n^{(2)} &\leqslant C\sum_{j=j_0}^{\infty} 2^{j(1/2-1/p)} \left(\|\beta_{j\cdot}\|_\infty^{p-1} \sum_{\mathrm{k}} |\beta_{jk}| \right)^{1/p} \\ &\leqslant C' \sum_{j=j_0}^{\infty} 2^{j(1/2-1/p)} L^{1-1/p} 2^{-j(s+1/2)(1-1/p)} 2^{j/2p} \leqslant C'' L^{1-1/p} 2^{-j_0 s(1-1/p)}. \end{aligned} \tag{21}$$

Let us now estimate $\delta_N^{(1)}$. To this end, we decompose the general term of the sum in $\delta_N^{(1)}$:

$$|y_{j_0 k} - \alpha_{j_0 k}| = |y_{j_0 k} - \alpha_{j_0 k}|(1_{p_{j_0 k} \geqslant \kappa/N} + 1_{P_{J_0 K} < \kappa/N}),$$

where $\kappa = \mathrm{e}^{-p/(p-1)}$. First, notice that

$$[\mathrm{E}|y_{j_0 k} - \alpha_{j_0 k}|^2 1_{p_{j_0 k} \geqslant \kappa/N}]^{p/2} \leqslant \left(\frac{2^{j_0} p_{j_0 k}}{N} \right)^{p/2} 1_{p_{j_0 k} \geqslant \kappa/N} \leqslant 2^{j_0 p/2} \frac{p_{j_0 k}}{\kappa N^{p-1}}. \tag{22}$$

Further, let $\#\delta_{jk}$ stand for the number of data points which hit δ_{jk}. We have

$$\begin{aligned} |y_{j_0 k} - \alpha_{j_0 k}| 1_{p_{j_0 k} < k/N} &\leqslant |y_{j_0 k}| 1_{p_{j_0} k < \kappa/N} + |\alpha_{j_0 k}| 1_{p_{j_0 k} < \kappa/N} \\ &\leqslant |y_{j_0 k}| 1_{p_{j_0 k} < \kappa/N,\, \#\delta_{j_0 k} = 1} + |y_{j_0 k}| 1_{p_{j_0 k} < \kappa/N,\, \#\delta_{j_0 k} \geqslant 2} + 2^{j/2} p_{j_0 k}. \end{aligned} \tag{23}$$

We now use the following result:

Lemma 5. *Let $\kappa \leqslant \mathrm{e}^{-p/(p-1)}$. Then there exists $C < \infty$ such that*

$$\mathrm{E}y_{jk}^2 1_{p_{jk}<\kappa/N,\#\delta_{jk}\geqslant 2} \leqslant \frac{2^j (Np_{jk})^{2/p}}{N^2\sqrt{\pi}}.$$

Proof. From the Stirling formula we have

$$P(\#\delta_{jk}=r)=\frac{N\ldots(N-r+1)}{r!}p_{jk}^r(1-p_{jk})^{N-r}\leqslant\frac{(Np_{jk})^r\mathrm{e}^r}{\sqrt{2\pi}r^{r+1/2}},$$

so that

$$\mathrm{E}\#\delta_{jk}^2 1_{p_{jk}<\kappa/N,\#\delta_{jk}\geqslant 2}\leqslant\sum_{r=2}^{\infty}\frac{(Np_{jk})^r\mathrm{e}^r}{\sqrt{2\pi}r^{r-3/2}}=(Np_{jk})^{2/p}\sum_{r=2}^{\infty}\frac{(Np_{jk})^{r-2/p}\mathrm{e}^r}{\sqrt{2\pi}r^{r-3/2}}$$

$$\leqslant\frac{(Np_{jk})^{2/p}}{\sqrt{\pi}}\sum_{r=1}^{\infty}2^{-r}=\frac{(Np_{jk})^{2/p}}{\sqrt{\pi}},$$

since $(Np_{jk})^{2-2/p}\mathrm{e}^2\leqslant 1$. To obtain the result of the lemma it suffices to note that

$$\mathrm{E}y_{jk}^2 1_{p_{jk}<\kappa/N,\#\delta_{jk}\geqslant 2}=\frac{2^j}{N^2}\mathrm{E}\#\delta_{jk}^2 1_{p_{jk}<\kappa/N,\#\delta_{jk}\geqslant 2}.$$

□

Let us now gather together the terms of the decomposition of $\delta_N^{(1)}$ (cf. (23)). By the Minkowski inequality, we have

$$\mathrm{E}(\delta_N^{(1)})^2=2^{j_0(1-2/p)}\mathrm{E}\|y_{j_0\cdot}-\alpha_{j_0\cdot}\|_p^2\leqslant 2^{j_0(1-2/p)}C\left(\sum_k[\mathrm{E}|y_{j_0k}-\alpha_{j_0k}|^2 1_{p_{j_0k}\geqslant\kappa/N}]^{p/2}\right)^{2/p}$$

$$+2^{j_0(1-2/p)}C\mathrm{E}\left(\sum_k|y_{j_0k}|^p 1_{\#\delta_{j_0k}=1}\right)^{2/p}$$

$$+2^{j_0(1-2/p)}C\left(\sum_k[\mathrm{E}|y_{j_0k}|^2 1_{p_{j_0k}<\kappa/N,\#\delta_{j_0k}\geqslant 2}]^{p/2}\right)^{2/p}$$

$$+2^{j_0(1-2/p)}C\left(\sum_k 2^{jp/2}p_{j_0k}^p 1_{p_{j_0k}<\kappa/N}\right)^{2/p}$$

$$\leqslant C'2^{j_0(1-2/p)}\left(\frac{2^{j_0p/2}}{\kappa N^{p-1}}\sum_k p_{j_0k}\right)^{2/p}\qquad\text{(by (22))}$$

$$
\begin{aligned}
&+ C'2^{j_0(1-2/p)}2^{j_0}N^{-2+2/p} \qquad \left(\text{since } y_{j_0k}1_{\#_{j_0k}=1} = 2^{j_0/2/N}\right)\\
&+ C'2^{j_0(1-2/p)}\left(\frac{2^{j_0p/2}}{N^p}\sum_k p_{j_0k}N\right)^{2/p} \qquad (\text{by Lemma 5})\\
&+ C'2^{j_0(1-2/p)}\left(\frac{2^{j_0p/2}}{N^{p-1}}\sum_k p_{j_0k}\right)^{2/p}\\
&\leqslant C''2^{2j_0(1-1/p)}N^{-2(1-1/p)}.
\end{aligned} \tag{24}
$$

Substituting into (24) and (21) the value of 2^{j_0} from (5), we obtain the bound of the theorem.

5.5. Proof of Theorem 4

We start with some technical results. Let δ_{jk} be the support bin of ψ_{jk}, and δ_k that of ϕ_k. We put

$$p_{jk} = \int_{\delta_{jk}} f(x)\mathrm{d}x,$$

$$\sigma_{jk}^2 = \mathrm{E}\xi_{jk}^2 = \frac{1}{N}(\mathrm{E}(\psi_{jk}^2(X_1) - \beta_{jk}^2),$$

where $\xi_{jk} = y_{jk} - \mathrm{E}y_{jk} = y_{jk} - \beta_{jk}$. We also denote $\upsilon_{jk} = \mathrm{E}\xi_{jk}^4$ and

$$\gamma_{jk} = \lambda\sqrt{\ln N}\sigma_{jk}.$$

Note that σ_{jk}^2 and γ_{jk} are the deterministic counterparts of the empirical values $\hat{\sigma}_{jk}^2$ and $\hat{\gamma}_{jk}$, that is,

$$\mathrm{E}\hat{\sigma}_{jk}^2 = \frac{N-1}{N}\sigma_{jk}^2, \qquad \mathrm{E}\hat{\gamma}_{jk}^2 = \frac{N-1}{N}\gamma_{jk}^2.$$

Since $2^{-j}m$ is the diameter of the support bin δ_{jk}, the wavelets $\psi_{j,mi}$ and $\psi_{j,mi'}$ have disjoint supports. Thus

$$\sum_k p_{jk} = \sum_{l=1}^{m}\sum_i p_{j,mi+l} \leqslant \sum_{l=1}^{m}\int f(x)\mathrm{d}x = m. \tag{25}$$

This relation will be often used in what follows.

Lemma 6.

$$\sigma_{jk} \leqslant \frac{1}{N}\min(L^{1/(s+1)}, 2^j\|\psi\|_\infty^2 p_{jk}), \tag{26}$$

$$\upsilon_{jk} \leqslant 2^{2j}\|\psi\|_\infty^4(p_{jk}N^{-3} + p_{jk}^2N^{-2}). \tag{27}$$

Proof. Note that

$$\mathrm{E}\psi_{jk}^2(X_1) = \int_{\delta_{jk}} \psi_{jk}^2(x) f(x)\mathrm{d}x \leqslant \|f\|_\infty \int_{\delta_{jk}} \psi_{jk}^2(x)\mathrm{d}x \leqslant \|f\|_\infty$$

and

$$\mathrm{E}\psi_{jk}^2(X_1) = \int_{\delta_{jk}} \psi_{jk}^2(x) f(x)\mathrm{d}x \leqslant \|\psi_{jk}\|_\infty^2 \int_{\delta_{jk}} f(x)\mathrm{d}x = 2^j \|\psi\|_\infty^2 p_{jk}.$$

By the Kolmogorov inequality, as $f \in \mathcal{F}(s, L)$, $\|f\|_\infty \leqslant L^{1/(s+1)}$, which implies (26).

Further,

$$\begin{aligned}\mathrm{E}\xi_{jk}^4 &= \mathrm{E}\left(\frac{1}{N}\sum_{i=1}^N \psi_{jk}(X_i) - \mathrm{E}\psi_{jk}(X_i)\right)^4 \leqslant \frac{\mathrm{E}\psi_{jk}^4(X_1)}{N^3} + \frac{(\mathrm{E}\psi_{jk}^2(X_1))^2}{N^2} \\ &\leqslant 2^{2j}\|\psi\|_\infty^4 \left(\frac{p_{jk}}{N^3} + \frac{p_{jk}^2}{N^2}\right).\end{aligned}$$

□

Lemma 7. *Let* $m_N \geqslant 30\ln N$. *Then for any j, k and any N large enough,*

$$P\left(\#\delta_{jk} < m_N,\ p_{jk} \geqslant \frac{2m_N}{N}\right) < p_{jk}N^{-4}, \tag{28}$$

$$P\left(\#\delta_{jk} \geqslant m_N,\ p_{jk} < \frac{2m_N}{2N}\right) < \frac{p_{jk}^3}{N^3}. \tag{29}$$

Proof. We start with the proof of (28). By the Bernstein inequality,

$$\begin{aligned}P(\#\delta_{jk} < m_N) &\leqslant P(\#\delta_{jk} - Np_{jk} \leqslant -Np_{jk}/2) \leqslant \exp\left(-\frac{(Np_{jk})^2}{8(Np_{jk} = \frac{1}{3}Np_{jk})}\right) \\ &\leqslant \exp\left(-\frac{Np_{jk}}{11}\right) \leqslant \exp\left(-\frac{m_N}{6}\right).\end{aligned}$$

When $m_N \geqslant 30\ln N$ the latter probability is less than N^{-5}, which implies (28).

To show (29) we consider two cases: $m_N/\mathrm{e}^2 N \leqslant p_{jk} \leqslant m_N/2N$ and $p_{jk} < m_N/\mathrm{e}^2 N$. In the first case we use the Bernstein inequality and the derivation is completely analogous to that used to show (28). In the second case, from the Stirling formula,

$$P(\#\delta_{jk} = r) = \frac{N \ldots (N - r + 1)}{r!} p_{jk}^r (1 - p_{jk})^{N-r} \leqslant \frac{N^r p_{jk}^r \mathrm{e}^r}{r^{r+1/2}}.$$

Then

$$P(\#\delta_{jk} \geqslant m_N) \leqslant \sum_{r=m_N}^{\infty} \frac{N^r p_{jk}^r \mathrm{e}^r}{r^{r+1/2}} \leqslant \frac{p_{jk}^3}{N^3} \frac{N^6 \mathrm{e}^3}{m_N^{7/2}} \sum_{r=m_N}^{\infty} \left(\frac{Np_{jk}\mathrm{e}}{r}\right)^{r-3}$$

$$\leqslant \frac{p_{jk}^3}{N^3} N^6 \sum_{r=m_N}^{\infty} \mathrm{e}^{-r} \leqslant \frac{p_{jk}^3}{N^3}.$$

□

Lemma 8. *Let the parameters* ρ *and* λ *satisfy* (7). *We have, for any* $0 \leqslant j \leqslant j_1$ *(where* j_1 *is defined in* (8)*)and* $\mathrm{k} \in \mathbb{Z}$*:*

(i) *If* $\mu \geqslant \max(1, 2L^{1/(s+1)} + \|\psi\|_\infty/54)$*, then*

$$P\left(|\xi_{jk}| > \mu\sqrt{\frac{\ln N}{N}}\right) \leqslant 2N^{-1}. \tag{30}$$

(ii) *If* $p_{jk} \geqslant \rho \ln N/2N$*, then*

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N} \psi_{jk}^2(X_i) - \mathrm{E}\psi_{jk}^2(X_1)\right| > \frac{2^j \nu^2 p_{jk}}{8}\right) \leqslant p_{jk} N^{-4}. \tag{31}$$

(iii) *Moreover, if* $\rho \ln N/2N \leqslant p_{jk} \leqslant \nu^2/2\|\psi\|_\infty^2$*, then*

$$\sigma_{jk}^2 \geqslant \frac{2^j \nu^2 p_{jk}}{2N}, \tag{32}$$

and for $\gamma_{jk} = \lambda\sqrt{\ln N}\sigma_{jk}$ *we have*

$$P\left(|\xi_{jk}| > \frac{\gamma_{jk}}{4}\right) \leqslant p_{jk} N^{-4}. \tag{33}$$

Proof. We use the Bernstein inequality. Since $\|\psi_{jk}\|_\infty \leqslant 2^{j_1/2}\|\psi\|_\infty \leqslant \|\psi\|_\infty \sqrt{N/(\rho \ln N)}$,

$$P\left(|\xi_{jk}| > \mu\sqrt{\frac{\ln N}{N}}\right) < 2\exp\left(-\frac{\mu^2 \ln N/N}{2\sigma_{jk}^2 + \frac{2}{3}\|\psi_{jk}\|_\infty N^{-1} \mu\sqrt{\ln N/N}}\right)$$

$$\leqslant 2\exp\left(-\frac{\mu^2 \ln N}{2L^{1/(s+1)} + \frac{2}{3}\|\psi\|_\infty \mu\rho^{-1/2}}\right) \qquad \text{(by (26))}$$

$$\leqslant 2\exp\left(-\frac{\mu \ln N}{2L^{1/(s+1)} + \|\psi\|_\infty/54}\right) \qquad \text{(when } \mu \geqslant 1\text{)}$$

$$= 2N^{-\mu/(2L^{1/(s+1)} + \|\psi\|_\infty/54)}$$

(note that as $\|\psi\|_\infty/\nu \geqslant 1$, $\sqrt{\rho} \geqslant 36$). this implies (i).

Recall that

$$\mathrm{E}\psi_{jk}^4(X_1) = \int \psi_{jk}^4(x) f(x)\mathrm{d}x \leqslant 2^{2j}\|\psi\|_\infty^4 p_{jk}.$$

Then, by the Bernstein inequality,

$$P\left(\left|\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - \mathrm{E}\psi_{jk}^2(X_1)\right| > c\right) < 2\exp\left(-\frac{c^2 N}{2\mathrm{E}\psi_{jk}^4(X_1) + \frac{2}{3}c\|\psi_{jk}^2\|_\infty}\right)$$

$$\leqslant 2\exp\left(-\frac{c^2 N}{2^{2j+1}\|\psi\|_\infty^4 p_{jk} + \frac{2}{3}c2^j\|\psi\|_\infty^2}\right).$$

Choosing $c = 2^j \nu^2 p_{jk}/8$, we obtain, when $p_{jk} \geqslant \rho \ln N/(2N)$,

$$P\left(\left|\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - \mathrm{E}\psi_{jk}^2(X_1)\right| > \frac{2^j\nu^2 p_{jk}}{8}\right)$$

$$< 2\exp\left(-\frac{2^{2j-6}\nu^4 p_{jk}^2 N}{2^{2j+1}\|\psi\|_\infty^4 p_{jk} + \frac{1}{3}2^{2j-2}\nu^2 p_{jk}\|\psi\|_\infty^2}\right)$$

$$\leqslant 2\exp\left(-\frac{\nu^4 p_{jk} N}{2^7\|\psi\|_\infty^4 + (2^4/3)\nu^2\|\psi\|_\infty^2}\right) \leqslant 2N^{-5},$$

which implies (31).

On the other hand, we have a simple bound for β_{jk}:

$$|\beta_{jk}| \leqslant \|\psi_{jk}\|_\infty \int_{\delta_{jk}} f(x)\mathrm{d}x = 2^{j/2}\|\psi\|_\infty p_{jk}.$$

Recall that the absolute value of the blocky analysis wavelet $\psi(x)$ is bounded from below on its support by ν. This implies that

$$\mathrm{E}\psi_{jk}^2(X_1) = \int_{\delta_{jk}} \psi_{jk}^2(x) f(x)\mathrm{d}x \geqslant 2^j \nu^2 p_{jk}.$$

Therefore, for $p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty)$,

$$\sigma_{jk}^2 \leqslant \frac{1}{N}(2^j\nu^2 p_{jk} - 2^j\|\psi\|_\infty^2 p_{jk}^2) \geqslant \frac{2^j\nu^2 p_{jk}}{2N},$$

and for those j and k we have $2^j \leqslant 2N\sigma_{jk}^2/\nu p_{jk} \leqslant 4N^2\sigma_{jk}^2/\nu^2\rho \ln N$, so that

$$\frac{\|\psi_{jk}\|_\infty^2 \lambda\sqrt{\ln N}}{N} = \frac{2^{j/2}\|\psi\|_\infty \lambda\sqrt{\ln N}}{N} \leqslant \frac{2\|\psi\|_\infty \sigma_{jk}\lambda}{\nu\sqrt{\rho}}. \tag{34}$$

Now the bound (33) follows from the Bernstein inequality. Indeed,

$$P\left(|\xi_{jk}| > \frac{\lambda\sqrt{\ln N}\sigma_{jk}}{4}\right) \leqslant 2\exp\left(-\frac{\lambda^2 \ln N\sigma_{jk}^2}{16(2\sigma_{jk}^2 + \frac{1}{6}\|\psi_{jk}\|_\infty N^{-1}\lambda\sqrt{\ln N}\sigma_{jk})}\right)$$

$$\leqslant 2\exp\left(-\frac{\lambda^2 \ln N\sigma_{jk}^2}{16\sigma_{jk}^2(2 + 2\lambda\|\psi\|_\infty/6\nu\sqrt{\rho})}\right) \qquad \text{(due to (34))}$$

$$\leqslant 2\exp\left(-\frac{\lambda^2 \ln N}{32 + \lambda/6}\right) < 2N^{-5} \qquad (\text{as } \sqrt{\rho} \geqslant 36\|\psi\|_\infty/\nu)$$

for $\lambda \geqslant 14$. □

Using lemma 8, we obtain the following propostion.

Propostion 1. *Let ρ and λ be as in the description of Algorithm 2.*

(i) *If j and k are such that $p_{jk} \geqslant \rho \ln N/(2N)$, then*

$$P\left(\hat{\gamma}_{jk} > \lambda\|\psi\|_\infty\sqrt{\frac{2^{j+1}p_{jk}\ln N}{N}}\right) \leqslant p_{jk}N^{-1}. \tag{35}$$

(ii) *Moreover, if $\rho \ln N/(2N) \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)$ then, for N large enough,*

$$P(|\hat{\gamma}_{jk} - \gamma_{jk}| \geqslant \tfrac{1}{2}\gamma_{jk}) \leqslant 2p_{jk}N^{-4}, \tag{36}$$

$$P(|\xi_{jk}| \geqslant \tfrac{1}{2}\hat{\gamma}_{jk}) \leqslant 2p_{jk}N^{-4}. \tag{37}$$

Proof. (i) Using the bounds $\hat{\sigma}_{jk}^2 \leqslant (1/N^2)\sum_{i=1}^N \psi_{jk}^2(X_i)$ and $\mathrm{E}\psi_{jk}^2(X_1) \leqslant 2^j\|\psi\|_\infty^2 p_{jk}$, we obtain

$$P\left(\hat{\gamma}_{jk} > \lambda\|\psi\|_\infty\sqrt{\frac{2^{j+1}p_{jk}\ln N}{N}}\right) = P\left(\hat{\sigma}_{jk}^2 > \frac{\|\psi\|_\infty^2 2^{j+1}p_{jk}}{N}\right)$$

$$\leqslant P\left(\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) > \|\psi\|_\infty^2 2^{j+1}p_{jk}\right)$$

$$\leqslant P\left(\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - \mathrm{E}\psi_{jk}^2(X_1) > \|\psi\|_\infty^2 2^j p_{jk}\right).$$

When $p_{jk} \geqslant \rho \ln N/N$, the latter probability is bounded by $p_{jk}N^{-1}$ due to (31).

(ii) We define the following sets:

$$A^1_{jk} = \left\{ |\xi_{jk}| \leqslant \frac{\gamma_{jk}}{4} \right\},$$

$$A^2_{jk} = \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \psi^2_{jk}(X_i) - \mathrm{E}\psi^2_{jk}(X_1) \right| \leqslant \frac{N\sigma^2_{jk}}{4} \right\},$$

$$B_{jk} = A^1_{jk} \cap A^2_{jk}.$$

Then from (33), $P(A^1_{jk}) \geqslant 1 - p_{jk}N^{-4}$, and the bound in (31), with $\sigma^2_{jk} \geqslant 2^j \nu^2 p_{jk}/(2N)$, implies that $P(A^2_{jk}) \geqslant 1 - p_{jk}N^{-4}$. Thus $P(B_{jk}) \geqslant 1 - 2p_{jk}N^{-4}$. In addition, we have

$$|y^2_{jk} - \beta^2_{jk}| \leqslant |\xi_{jk}|(2|\beta_{jk}| + |\xi_{jk}|) \leqslant \frac{\gamma_{jk}}{4}\left(2|\beta_{jk}| + \frac{\gamma_{jk}}{4}\right) \qquad \text{on } A^1_{jk}.$$

Furthermore, $|\beta_{jk}| \leqslant 2^{j/2} p_{jk} \|\psi\|_\infty$ and $\sigma_{jk} > 2^{j/2}\nu\sqrt{p_{jk}/(2N)}$ by (32). Thus $|\beta_{jk}| \leqslant \sigma_{jk}\|\psi\|_\infty \sqrt{2p_{jk}N/\nu}$, and

$$|y^2_{jk} - \beta^2_{jk}| \leqslant \frac{\gamma^2_{jk}}{16}\left(1 + \frac{8\|\psi\|_\infty\sqrt{2p_{jk}N}}{\nu\lambda\sqrt{\ln N}}\right) \qquad \text{on } A^1_{jk}.$$

Then on B_{jk},

$$|\hat\sigma_{jk} - \sigma_{jk}| \leqslant \frac{|\hat\sigma^2_{jk} - \sigma^2_{jk}|}{\sigma_{jk}} \leqslant \frac{1}{N\sigma_{jk}}\left(\left| \frac{1}{N}\sum_{i=1}^{N} \psi^2_{jk}(X_i) - \mathrm{E}\psi^2_{jk}(X_1) \right| + |y^2_{jk} - \beta^2_{jk}| \right)$$

$$\leqslant \frac{\sigma_{jk}}{4} + \frac{\sigma_{jk}\lambda^2 \ln N}{16N}\left(1 + \frac{8\|\psi\|_\infty\sqrt{2p_{jk}N}}{\nu\lambda\sqrt{\ln N}}\right) \leqslant \frac{\sigma_{jk}}{2}$$

for N large enough. This establishes inequality (36). Moreover,

$$\left\{ |\xi_{jk}| > \frac{\hat\gamma_{jk}}{2} \right\} \subseteq \left\{ |\xi_{jk}| > \frac{\gamma_{jk}}{4} \right\} \cup \left\{ \hat\gamma_{jk} < \frac{\gamma_{jk}}{2} \right\}$$

and the bound (37) is a consequence of (33). □

Lemma 9. *Let* $\hat\beta_{jk}$ *be defined as in* (9)*. Then there exists* $C < \infty$ *such that*

$$|\hat\beta_{jk} - \beta_{jk}| \leqslant C\left(|\xi_{jk}| + \mu\sqrt{\frac{\ln N}{N}} \right) + |\beta_{jk}| 1_{\hat\gamma_{jk} > \mu\sqrt{\ln N/N}} \tag{38}$$

for any $\mu > 0$*, and*

$$|\hat\beta_{jk} - \beta_{jk}|^p \leqslant C\left[|\xi_{jk}|^p 1_{|\xi_{jk}| > \hat\gamma_{jk}/2} + \min(|\beta_{jk}|, \gamma_{jk})^p \right] + |\beta_{jk}|^p 1_{\hat\gamma_{jk} > 3_{\gamma_{jk}}/2}. \tag{39}$$

Proof. We have, by virtue of Lemma 2 of Delyon and Juditsky (1996),

$$|\hat\beta_{jk} - \beta_{jk}|^p \leqslant |3\xi_{jk}| 1_{|\xi_{jk}| > \hat\gamma_{jk}/2} + \min(|\beta_{jk}|, \tfrac{3}{2}\hat\gamma_{jk})^p. \tag{40}$$

This implies, in particular, that

$$|\hat{\beta}_{jk} - \beta_{jk}| \leqslant |3\xi_{jk}| + \tfrac{3}{2}\mu\sqrt{\frac{\ln N}{N}} 1_{\hat{\gamma}_{jk} \leqslant \mu\sqrt{\ln N/N}} + |\beta_{jk}| 1_{\hat{\gamma}_{jk} > \mu\sqrt{\ln N/N}}.$$

On the other hand, we have from (40),

$$\begin{aligned}|\hat{\beta}_{jk} - \beta_{jk}|^p &\leqslant |3\xi_{jk}|^p 1_{|\xi_{jk}| > \hat{\gamma}_{jk}/2} + \min(|\beta_{jk}|, \tfrac{9}{4}\gamma_{jk})^p 1_{\hat{\gamma}_{jk} \leqslant 3 y_{jk}/2} + |\beta_{jk}|^p 1_{\hat{\gamma}_{jk} > 3 y_{jk}/2} \\ &\leqslant C\Big[|\xi_{jk}|^p 1_{|\xi_{jk}| > \hat{\gamma}_{jk}/2} + \min(|\beta_{jk}|, \gamma_{jk})^p\Big] + |\beta_{jk}|^p 1_{\hat{\gamma}_{jk} > 3 y_{jk}/2}\end{aligned}$$

□

We return now to the proof of Theorem 4. Note that from (14) we have the following bound for the estimation error:

$$\|\hat{f}_N - f\|_p \leqslant \sum_{j=0}^{\infty} 2^{j(1/2-1/p)} \|\hat{\beta}_{j\cdot} - \beta_{j\cdot}\|_p + \|z_\cdot - \alpha_\cdot\|_p \equiv r_N. \tag{41}$$

We decompose r_N as follows:

$$\begin{aligned} r_N &\leqslant \sum_{j=j_1+1}^{\infty} 2^{j(1/2-1/p)} \|\beta_{j\cdot}\|_p + \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \|\hat{\beta}_{j\cdot} - \beta_{j\cdot}\|_p + \|z_\cdot - \alpha_\cdot\|_p \\ &= \sum_{j=j_1+1}^{\infty} 2^{j(1/2-1/p)} \|\beta_{\cdot j}\|_p \\ &\quad + \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} < m_N}\right)^{1/p} \\ &\quad + \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} \geqslant m_N}\right)^{1/p} \\ &\quad + \left(\sum_k |z_k - \alpha_k|^p 1_{\#\delta_k < m_N}\right)^{1/p} + \left(\sum_k |z_k - \alpha_k|^p 1_{\#\delta_k \geqslant m_N}\right)^{1/p} \\ &\leqslant \sum_{j=j_1+1}^{\infty} 2^{j(1/2-1/p)} \|\beta_{\cdot j}\|_p \\ &\quad + \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p 1_{\#\delta_{jk} < m_N} 1_{p_{jk} > 2m_N/N}\right)^{1/p} \\ &\quad + \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p 1_{\#\delta_{jk} < m_N} 1_{p_{jk} \leqslant 2m_N/N}\right)^{1/p} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=0}^{j_1} 2^{j(1/2-1/p)}\left(\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p 1_{\#\delta_{jk}\geqslant m_N} 1_{p_{jk}<m_N/2N}\right)^{1/p}\\
&+\sum_{j=0}^{j_1} 2^{j(1/2-1/p)}\left(\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p 1_{\#\delta_{jk}\geqslant m_N} 1_{p_{jk}\geqslant m_N/2N}\right)^{1/p}\\
&+\left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k<m_N}\right)^{1/p}\\
&+\left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k\geqslant m_N} 1_{p_k\leqslant m_N/2N}\right)^{1/p}+\left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k\geqslant m_N} 1_{p_k> m_N/2N}\right)^{1/p}\\
=&\sum_{i=1}^{8} r_N^{(i)}.
\end{aligned}
$$

The principal term in the above expansion is $r_N^{(5)}$. We first need to give the bounds for the rest of the terms in the sum. Note that the bounds below are valid when N is large enough:

Lemma 10. $r_N^{(1)} \leqslant CL^{1-1/p}(\ln N/N)^{s(1-1/p)}$.

Proof. Inequality (12) implies that $\|\beta_{j\cdot}\|_\infty \leqslant L2^{-j(s+1/2)}$. Moreover, from (11) we conclude that, for some $C<\infty$,

$$2^{-j/2}\|\beta_{j\cdot}\|_1 \leqslant C.$$

Now

$$
\begin{aligned}
r_N^{(1)} &= \sum_{j=j_1+1}^{\infty} 2^{j(1/2-1/p)}\|\beta_{\cdot j}\|_p < \sum_{j=j_1+1}^{\infty} 2^{j(1/2-1/p)}\|\beta_{j\cdot}\|_\infty^{(p-1)/p}\|\beta_{j\cdot}\|_1^{1/p}\\
&\leqslant CL^{1-1/p}\sum_{j=j_1+1}^{\infty} 2^{j(1/2-1/p)}2^{-j(s+1/2)(1-1/p)}2^{j/(2p)}\\
&\leqslant CL^{1-1/p}\sum_{j=j_1+1}^{\infty} 2^{-js(1-1/p)} \leqslant C'L^{1-1/p}2^{-j_1 s(1-1/p)}.
\end{aligned}
$$

□

Lemma 11. $[\mathrm{E}(r_N^{(2)})^2]^{1/2} \leqslant C/N$.

Proof. Note that, for any $p\geqslant 1$, $\|\beta_{j\cdot}\|_p \leqslant \|\beta_{j\cdot}\|_1$. Thus, by the Minkowski inequality,

$$[\mathrm{E}(r_N^{(2)})^2]^{1/2} \leqslant \left[\mathrm{E}\left(\sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k |\beta_{jk}| 1_{\#\delta_{jk}<m_N} 1_{p_{jk}>2m_N/N}\right)^2\right]^{1/2}$$

$$\leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k \left[\mathrm{E}|\beta_{jk}|^2 1_{\#\delta_{jk}<m_N} 1_{p_{jk}>2m_N/N}\right]^{1/2}$$

Due to inequality (28) in Lemma 7 we obtain the bound

$$[\mathrm{E}(r_N^{(2)})^2]^{1/2} \leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k |\beta_{jk}| P^{1/2}(\#\delta_{jk} < m_N) 1_{p_{jk}>2m_N/N}$$

$$\leqslant C \sum_{j=0}^{j_1} 2^{j(1-1/p)} \max_k P^{1/2}(\#\delta_{jk} < m_N) 1_{p_{jk}>2m_N/N} \qquad (\text{as } 2^{-j/2}\|\beta_{j\cdot}\|_1 \leqslant C)$$

$$\leqslant C\left(\frac{N}{\ln N}\right)^{1-1/p} N^{-2} \leqslant \frac{C}{N}.$$

□

Lemma 12. $r_N^{(3)} \leqslant CL^{(1-1/p)/(s+1)}(\ln N/N)^{s(1-1/p)/(s+1)}$.

Proof. We split the sum $r_N^{(3)}$ into two parts: when $0 \leqslant j \leqslant j_0$ we use the bound $|\beta_{jk}| \leqslant 2^{j/2}\|\psi\|_\infty p_{jk}$. When $j > j_0$ we bound $|\beta_{jk}|_\infty$ by $CL2^{-j(s+1/2)}$ (cf. (12)). Thus

$$r_N^{(3)} \leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^p 1_{p_{jk}\leqslant 2m_N/N}\right)^{1/p}$$

$$\leqslant C \sum_{j=0}^{j_0} 2^{j(1/2-1/p)} \left(\sum_k 2^{jp/2} p_{jk}^p 1_{p_{jk}\leqslant 2m_N/N}\right)^{1/p}$$

$$+ C \sum_{j=j_0+1}^{j_1} 2^{j(1/2-1/p)} \|\beta_{j\cdot}\|_\infty^{1-1/p} \left(\sum_k |\beta_{jk}|\right)^{1/p}$$

$$\leqslant C' \sum_{j=0}^{j_0} 2^{j(1-1/p)} \left(\sum_k 2^{jp/2} \left(\frac{2m_N}{N}\right)^{p-1} p_{jk}\right)^{1/p}$$

$$+ C \sum_{j=j_0+1}^{j_1} 2^{j(1/2-1/p)} L^{1-1/p} 2^{-j(s+1/2)(1-1/p)} 2^{j/(2p)}$$

$$\leqslant C''\left[2^{j_0(1-1/p)}\left(\frac{2m_N}{N}\right)^{1-1/p}+L^{1-1/p}2^{-j_0 s(1-1/p)}\right]$$

(by (25)). Finally, choosing j_0 such that

$$\left(\frac{N}{\ln N}\right)^{1/(s+1)}L^{1/(s+1)}\leqslant 2^{j_0}<2\left(\frac{N}{\ln N}\right)^{1/(s+1)}L^{1/(s+1)},$$

we obtain the statement of the lemma. □

Lemma 13. $[\mathrm{E}(r_N^{(4)})^2]^{1/2}\leqslant C\max\{L^{(1-1/p)/(s+1)}(\ln N/N)^{s(1-1/p)/(s+1)},\ N^{-1/2}\}$.

Proof. We first remark that $|\hat{\beta}_{jk}-\beta_{jk}|\leqslant|\xi_{jk}|+|\beta_{jk}|$ and

$$|\hat{\beta}_{jk}-\beta_{jk}|^p\leqslant 2^{p-1}(|\xi_{jk}|^p+|\beta_{jk}|^p).$$

Thus

$$r_N^{(4)}\leqslant C\sum_{j=0}^{j_1}2^{j(1/2-1/p)}\left(\sum_k|\xi_{jk}|^p1_{\#\delta_{jk}\geqslant m_N}1_{p_{jk}<m_N/2N}\right)^{1/p}$$
$$+C\sum_{j=0}^{j_1}2^{j(1/2-1/p)}\left(\sum_k|\beta_{jk}|^p1_{p_{jk}<m_N/2N}\right)^{1/p}. \qquad (43)$$

We have already obtained a bound for the second term on the right-hand side of (43) in Lemma 12:

$$\sum_{j=0}^{j_1}2^{j(1/2-1/p)}\left(\sum_k|\beta_{jk}|^p1_{p_{jk}<m_N/2N}\right)^{1/p}\leqslant CL^{(1-1/p)/(s+1)}\left(\frac{\ln N}{N}\right)^{s(1-1/p)/(s+1)}.$$

Recall that, by (27),

$$\mathrm{E}|\xi_{jk}|^4=v_{jk}\leqslant\|\psi\|_\infty^4 2^{2j}(p_{jk}N^{-3}+p_{jk}^2N^{-2}).$$

We can estimate the first term on the right-hand side of (43) as follows:

$$\left[\mathrm{E}\left(\sum_{j=0}^{j_1}2^{j(1/2-1/p)}\left(\sum_k|\xi_{jk}|^p1_{\#\delta_{jk}\geqslant m_N,\,p_{jk}<m_N/2N}\right)^{1/p}\right)^2\right]^{1/2}$$

$$
\begin{aligned}
&\leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k [\mathrm{E}|\xi_{jk}|^2 1_{\#\delta_{jk}\geqslant m_N,\, p_{jk}<m_N/2N}]^{1/2} \\
&\leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k (\mathrm{E}|\xi_{jk}|^4)^{1/4} P^{1/4}(\#\delta_{jk} \geqslant m_N) 1_{p_{jk}<m_N/2N} \\
&\leqslant C \sum_{j=0}^{j_1} 2^{j(1-1/p)} \sum_k \frac{p_{jk}^{1/4}}{N^{3/4}} P^{1/4}(\#\delta_{jk} \geqslant m_N) 1_{p_{jk}<m_N/2N} \\
&\leqslant C \sum_{j=0}^{j_1} 2^{j(1-1/p)} N^{-3/2} \sum_k p_{jk} \qquad \text{(by (29))} \\
&\leqslant C \left(\frac{N}{\ln N}\right)^{1-1/p} N^{-3/2} \leqslant CN^{-1/2}.
\end{aligned}
$$

□

Lemma 14. $r_N^{(6)} \leqslant C(\ln N/N)^{1-1/p}$ *and* $[\mathrm{E}(r_N^{(7)})^2]^{1/2} \leqslant CN^{-1}$.

Proof. The proof of this lemma is completely analogous to that of Lemmas 11 and 12. □

Lemma 15. $[\mathrm{E}(r_N^{(8)})^2]^{1/2} \leqslant C\max(N^{1/p-1}, N^{-1/2})$.

Proof. When $p \geqslant 2$,

$$
\begin{aligned}
\mathrm{E}(r_N^{(8)})^2 &= \mathrm{E}\left(\sum_k |z_k - \alpha_k|^p 1_{\#\delta_k \geqslant m_N} 1_{p_{jk}>2m_N/N}\right)^{2/p} \\
&\leqslant \mathrm{E}\left(\sum_k |z_k - \alpha_k|^2\right) \leqslant \sum_k \frac{p_k}{N} \leqslant CN^{-1}.
\end{aligned}
$$

When $1 \leqslant p < 2$, we use the bound

$$
\begin{aligned}
\mathrm{E}(r_N^{(8)})^2 &\leqslant \left[\sum_k (\mathrm{E}|z_k - \alpha_k|^2)^{p/2} 1_{p_k>2m_N/N}\right]^{2/p} \\
&\leqslant \left[\sum_k \left(\frac{p_k}{N}\right)^{p/2} 1_{p_k>2m_N/N}\right]^{2/p} \leqslant CN^{-1}\left[\sum_k p_k \left(\frac{N}{m_N}\right)^{1-p/2}\right]^{2/p} \\
&\leqslant C' N^{2/p-2} m_N^{1-2/p} \leqslant C' N^{2/p-2}.
\end{aligned}
$$

□

We finally come to the principal term of the error decompostion (42):

$$r_n^{(5)} = \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} \geqslant m_N} 1_{p_{jk} \geqslant m_N/2N} \right)^{1/p}.$$

We split this again:

$$r_n^{(5)} = \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} \geqslant m_N} 1_{p_{jk} \geqslant m_N/2N} \left[1_{p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)} + 1_{p_{jk} > \nu^2/(2\|\psi\|_\infty^2)} \right] \right)^{1/p}.$$

$$\leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{p_{jk} > \nu^2/(2\|\psi\|_\infty^2)} \right)^{1/p}$$

$$+ \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{m_N/2n \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)} \right)^{1/p}. \tag{44}$$

To bound the second term on the right-hand side of (44) we use inequality (39) from Lemma 9,

$$|\hat{\beta}_{jk} - \beta_{jk}|^p \leqslant C\left[|\xi_{jk}|^p 1_{|\xi_{jk}| > \hat{\gamma}_{jk}/2} + \min(|\beta_{jk}|, \gamma_{jk})^p \right] + |\beta_{jk}|^p 1_{\hat{\gamma}_{jk} > 3y_{jk}/2},$$

to obtain

$$r_n^{(5)} \leqslant \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p 1_{p_{jk} > \nu^2/(2\|\psi\|_\infty^2)} \right)^{1/p}$$

$$+ C \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\xi_{jk}|^p 1_{|\xi_{jk}| > \hat{\gamma}_{jk}/2} 1_{m_N/2n \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)} \right)^{1/p}$$

$$+ C \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k \min(|\beta_{jk}|, \gamma_{jk})^p 1_{p_{jk} \geqslant m_N/2N} \right)^{1/p}$$

$$+ \sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}| 1_{\hat{\gamma}_{jk} > 3y_{jk}/2} 1_{m_N/2N \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)} \right)^{1/p}$$

$$= \sum_{i=1}^{4} \delta_N^{(i)}. \tag{45}$$

We start with the bound on $\delta_N^{(1)}$.

Lemma 16. $[\mathrm{E}(\delta_N^{(1)})^2]^{1/2} \leqslant \sqrt{\ln N/N}$.

Proof. First, remark that as $p_{jk} \leqslant C\|f\|_\infty 2^{-j}$ for some $C < \infty$ which depends only on the wavelet ψ, the inequality $p_{jk} > \nu^2/(2\|\psi\|_\infty^2)$ implies that

$$2^j \leqslant \frac{C\|f\|_\infty}{p_{jk}} \leqslant C' \frac{\|f\|_\infty^2 \|\psi\|_\infty}{\nu^2} \leqslant C''. \tag{46}$$

Moreover, for obvious reasons, there exists C (which depends only on the wavelet ψ) such that for any j, $\sum_k p_{jk} \leqslant C$ and the number of bins such that $p_{jk} > \nu^2/(2\|\psi\|_\infty^2)$ cannot exceed $C' = 2C\|\psi\|_\infty^2/\nu^2$ at each level j. Now let j' be the maximal j which satisifies (46). Then for $0 \leqslant j \leqslant j'$ and any k,

$$\mathrm{E}|\xi_{jk}|^2 \leqslant \frac{\mathrm{E}\psi_{jk}^2(X_1)}{N} \leqslant \frac{\|\psi_{jk}\|_\infty^2}{N} \leqslant \frac{C}{N}.$$

Due to the bound (35) in Propostion 1, we have, for some $\mu < \infty$,

$$P\left(\hat{\gamma}_{jk} > \mu\sqrt{\frac{\ln N}{N}}\right) \leqslant N^{-1}.$$

Using (38), we obtain

$$\begin{aligned}
[\mathrm{E}(\delta_N^{(1)})^2]^{1/2} &\leqslant \sum_{j=0}^{j'} 2^{j(1/2-1/p)} \sum_k \left[\mathrm{E}|\hat{\beta}_{jk} - \beta_{jk}|^2 1_{p_{jk}>\nu^2/(2\|\psi\|_\infty^2)}\right]^{1/2} \\
&\leqslant C\sum_{j=0}^{j'}\sum_k \left[\mathrm{E}|\xi_{jk}|^2 + \mu\frac{\ln N}{N} + |\beta_{jk}|^2 P\left(\hat{\gamma}_{jk} > \mu\sqrt{\frac{\ln N}{N}}\right)\right]^{1/2} 1_{p_{jk}>\nu^2/(2\|\psi\|_\infty^2)} \\
&\leqslant C'\left[N^{-1} + \mu\frac{\ln N}{N} + N^{-1}\right]^{1/2} \leqslant C''\sqrt{\frac{\ln N}{N}}.
\end{aligned}$$

□

Lemma 17. $[\mathrm{E}(\delta_N^{(2)})^2]^{1/2} \leqslant CN^{-1/2}$.

Proof. We have the bound

$$\begin{aligned}
[\mathrm{E}(\delta_N^{(2)2}]^{1/2} &\leqslant C\sum_{j=0}^{j_1} \mathrm{s}^{j(1/2-1/p)} \sum_k \left[\mathrm{E}|\xi_{jk}|^2 1_{|\xi_{jk}|>\hat{\gamma}_{jk}/2} 1_{m_N/2N \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)}\right]^{1/2} \\
&\leqslant C\sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k (\mathrm{E}\xi_{jk}^4)^{1/4} P^{1/4}\left(|\xi_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\right) 1_{m_N/2N \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)} \\
&\leqslant C'\sum_{j=0}^{j_1} 2^{j(1/2-1/p)} \sum_k \frac{2^{j/2} p_{jk}^{1/2}}{N^{1/2}} p_{jk}^{1/2} N^{-1} \qquad \text{(by (37))}
\end{aligned}$$

$$\leqslant C''\frac{2^{j_1/2}}{N^{3/2}} \leqslant C^{(3)}N^{-1/2}.$$

□

Lemma 18. $[\mathrm{E}(\delta_N^{(4)})^2]^{1/2} \leqslant CN^{-1/2}$.

Proof. Using (36), we obtain

$$[\mathrm{E}(\delta_N^{(4)})^2]^{1/2} \leqslant C\sum_{j=0}^{j_1} 2^{j(1/2-1/p)}\sum_k |\beta_{jk}|P^{1/2}(\hat{\gamma}_{jk} > \tfrac{3}{2}\gamma_{jk})1_{m_N/2N \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)}$$

$$\leqslant C\sum_{j=0}^{j_1} 2^{j(1-1/p)}\max_k P^{1/2}(\hat{\gamma}_{jk} > \tfrac{3}{2}\gamma_{jk})1_{m_N/2N \leqslant p_{jk} \leqslant \nu^2/(2\|\psi\|_\infty^2)}$$

$$\leqslant C'\sum_{j=0}^{j_1} 2^{j(1-1/p)}\max_k p_{jk}^{1/2}N^{-2}1_{p_{jk} \geqslant m_N/2N}$$

$$\leqslant C''2^{j_1}N^{-3/2} \leqslant C''N^{-1/2}.$$

□

Lemma 19.

$$\delta_N^{(3)} \leqslant C\begin{cases} L^{(p-1)/(p(s+1))}\left(\dfrac{\ln N}{N}\right)^{s/(2s+1)} & \textit{for } p > 2+\dfrac{1}{s}, \\ L^{1/(2s+1)}\ln N\left(\dfrac{\ln N}{N}\right)^{s/(2s+1)} & \textit{for } p = 2+\dfrac{1}{s}, \\ L^{(p-1)/(p(s+1))}\left(\dfrac{\ln N}{N}\right)^{s(p-1)/(p(s+1))} & \textit{for } p < 2+\dfrac{1}{s}. \end{cases}$$

Proof. We consider here the case $2 \leqslant p < \infty$ (the case $1 \leqslant p < 2$ being completely analogous). Let j' and j'' satisfy

$$L^{1/(s+1)}\left(\frac{N}{\ln N}\right)^{1/(2s+1)} \leqslant 2^{j'} < 2L^{1/(s+1)}\left(\frac{N}{\ln N}\right)^{1/(2s+1)},$$

$$L^{1/(s+1)}\left(\frac{N}{\ln N}\right)^{1/(s+1)} \leqslant 2^{j''} < 2L^{1/(s+1)}\left(\frac{N}{\ln N}\right)^{1/(s+1)}.$$

Recall that $\min(|\beta_{jk}|, \gamma_{jk}) \leqslant |\beta_{jk}|^q\gamma_{jk}^{1-q}$, for any $0 \leqslant q \leqslant 1$. Further, due to (26),

$$\gamma_{jk}^2 = \lambda^2 \ln N\sigma_{jk}^2 \leqslant C\min\left(\frac{2^j p_{jk}\ln N}{N}, \frac{L^{1/(s+1)}\ln N}{N}\right).$$

Thus, we have

$$
\begin{aligned}
\delta_N^{(3)} \leqslant & \sum_{j=0}^{j'} 2^{j(1/2-1/p)} \left(\sum_k \gamma_{jk}^p \right)^{1/p} + \sum_{j=j'+1}^{j''} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^{p-2} \gamma_{jk}^2 \right)^{1/p} \\
& + \sum_{j=j''+1}^{j_1} 2^{j(1/2-1/p)} \left(\sum_k |\beta_{jk}|^{p-1} |\beta_{jk}| \right)^{1/p} \\
& \leqslant C \sum_{j=0}^{j'} 2^{j(1/2-1/p)} \left(\sum_k \left(\frac{\ln NL^{1/(s+1)}}{N} \right)^{p/2-1} \frac{\ln N2^j p_{jk}}{N} \right)^{1/p} \\
& + C \sum_{j=j'+1}^{j''} 2^{j(1/2-1/p)} \left(\|\beta_{j\cdot}\|_\infty^{p-2} \sum_k \frac{\ln N2^j p_{jk}}{N} \right)^{1/p} \\
& + C \sum_{j=j''+1}^{\infty} 2^{j(1/2-1/p)} \left(\|\beta_{j\cdot}\|_\infty^{p-1} \|\beta_{j\cdot}\|_1 \right)^{1/p} \\
& \leqslant C' \sum_{j=0}^{j'} 2^{j(1/2-1/p)} \sqrt{\frac{\ln N}{N}} L^{(p-1)/(2p(s+1))} 2^{j/p} \\
& + C' \sum_{j=j'+1}^{j''} 2^{j(1/2-1/p)} \left(L 2^{-j(s+1/2)} \right)^{1-2/p} 2^{j/p} \left(\frac{\ln N}{N} \right)^{1/p} \\
& + C' \sum_{j=j''+1}^{\infty} 2^{j(1/2-1/p)} \left(L 2^{-j(s+1/2)} \right)^{1-1/p} 2^{j/(2p)} \\
& \leqslant C'' \left[2^{j'/2} \sqrt{\frac{\ln N}{N}} L^{(p-1)/(2(s+1))} + L^{1-2/p} \left(\frac{\ln N}{N} \right)^{1/p} \sum_{j=j'+1}^{j''} 2^{-j(s-(2s+1)/p)} + L^{1-1/p} 2^{-j''s(1-1/p)} \right].
\end{aligned}
$$

When $p > (2s+1)/s$, the second term of the above decomposition can be estimated as follows:

$$
\sum_{j=j'+1}^{j''} 2^{-j(s-(2s+1)/p)} \leqslant C 2^{-j'(s-(2s+1)/p)},
$$

and when $2 \leqslant p < (2s+1)/s$,

$$
\sum_{j=j'+1}^{j''} 2^{-j(s-(2s+1)/p)} \leqslant C 2^{-j''(s-(2s+1)/p)}.
$$

Substituting the values for $2^{j'}$ and $2^{j''}$, we obtain in these two cases

$$\delta_N^{(3)} \leqslant CL^{(p-1)/(p(s+1))}\left[\left(\frac{\ln N}{N}\right)^{s/(2s+1)} + \left(\frac{\ln N}{N}\right)^{s/(s+1)}\right].$$

If $p = 2s+1$, an extra logarithmic factor appears in the sum

$$\sum_{j=j'+1}^{j''} 2^{-j(s-(2s+1)/p)} = j'' - j' \leqslant C \ln N,$$

and

$$\delta_N^{(3)} \leqslant C \ln NL^{(1-2)/p}\left(\frac{\ln N}{N}\right)^{1/p} \leqslant C' \ln NL^{1/(2s+1)}\left(\frac{\ln N}{N}\right)^{s/(2s+1)}.$$

□

Finally, to obtain the statement of the theorem it suffices to combine the bounds of Lemmas 10–19 for the terms of the decomposition (42).

References

Anscombe, F. (1948) The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, **35**, 246–254.

Birgé, L. (1983) Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsheinlichkeitstheorie Verw. Geb.*, **65**, 181–237.

Birgé, L. and Massart, P. (2000) An adaptive compression algorithm in Besov spaces. *Constr. Approx.*, **16**(1), 1–36.

Bretagnolle, J. and Huber, C. (1979) Estimation des densités: risque minimax. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **47**, 119–137.

Butucea, C. (2000) The adaptive rate of convergence in a problem of pointwise density estimation. *Statist. Probab. Lett.*, **47**(1), 85–90.

Ciesielski, Z. and Kamont, A. (1999) On the choice of the spline density estimators. *Math. Pannon.*, **10**(1), 17–28.

Daubechies, I. (1992) *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 61. Philadelphia: Society for Industrial and Applied Mathematics.

Delyon, B and Juditsky, A. (1996) On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.*, **3**(3), 215–228.

Delyon, B. and Juditsky, A. (1997) On the computation of wavelet coefficients. *J. Approx. Theory*, **88**(1), 47–79.

Devroye, L. (1987) *A Course in Density Estimation*, Progr. Probab. Statist. 14. Boston: Birkhäuser Verlag.

Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation. The* L_1 *View.* New York: Wiley.

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

Donoho, D. and Johnstone, I. (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. (1995) Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B*, **57**, 301–369.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.*, **24**(2), 508–539.

Goldenshluger, A. and Nemirovski, A. (1997) On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.*, **6**(2), 135–170.

Golubev, G. (1992) Nonparametric estimation of smooth probability densities in L_2. *Problems Inform. Transmission*, **28**(1), 44–54.

Ibragimov, I. and Khas'minskij, R. (1980) On the estimation of the distribution density. *Zap. Nauchn. Sem. Leningr. Otdel. Mat. Inst. Steklov.*, **98**, 61–85.

Ibragimov, I. and Khas'minskij, R. (1981) *Statistical Estimation. Asymptotic Theory.* Berlin: Springer-Verlag.

Juditsky, A. (1997) Wavelet estimators: Adapting to unknown smoothness. *Math. Methods Statist*., **6**(1), 1–25.

Juditsky, A. and Nazin, A. (2001) On minimax approach to non-parametric adaptive control. *Internat. J. Adapt. Control Signal Process*., **15**(2), 153–168.

Korostelev, A. and Tsybakov, A. (1993) *Minimax Theory of Image Reconstruction*, Lecture Notes in Statist. 82. Berlin: Springer-Verlag.

Lepskij, O. (1992) Asymptotically minimax adaptive estimation. II: Schemes without optimal adaption: Adaptive estimators. *Theory Probab. Appl.*, **37**(3), 433–448.

Nemirovskii, A. (1986) Nonparametric estimation of smooth regression functions. *J. Comput. System Sci.*, **23**(6), 1–11.

Nussbaum, M. (1996) Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, **24**, 2399–2430.

Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Triebel, H. (1992) *Theory of Function Spaces II*, Monogr. Math. 84. Basel: Birkhäuser Verlag.

Received September 2001 and revised July 2003