Volume 15, No. 2 (2006), pp. Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

MAXISET FOR DENSITY ESTIMATION ON $\mathbb{R}$

F. Autin

CNRS-UMR 6632, Univ. Aix-Marseille 1, Centre de Mathématiques et Informatique
39 rue F. Joliot Curie, 13453 Marseille Cedex 13, France
E-mail: autin@cmi.univ-mrs.fr

The problem of density estimation on $\mathbb{R}$ is considered. Adopting the maxiset point of view, we focus on performance of adaptive procedures. Any rule which consists in neglecting the wavelet empirical coefficients smaller than a sequence of thresholds v_n will be called an elitist rule*. We prove that for such a procedure the maximal space for the rate $v_n^{\alpha p}$, with $0 < \alpha < 1$, is always contained in the intersection of a Besov space and a weak Besov space. With no assumption on compactness of the support of the density goal f, we show that the hard thresholding rule is the best procedure among elitist rules when taking the classical choice of thresholds $v_n = \mu\sqrt{n^{-1}\log(n)}$, with $\mu > 0$. Then, we point out the significance of data-driven thresholds in density estimation by comparing the maxiset of the hard thresholding rule with the one of Juditsky and Lambert-Lacroix's procedure.*

Key words: adaptive procedures, Besov spaces, data-driven thresholds, maxiset theory, minimax theory, nonparametric estimation, thresholding rules, wavelet decomposition.

2000 Mathematics Subject Classification: 62G05, 62G07, 62G20.

1. Introduction

For recent years, nonparametric estimation methods have given a renewal of interest in minimax theory, particularly, in the problem of density estimation, which consists in providing methods to estimate a density f from independent realizations $X_1, \ldots, X_n$. This basic problem has been extensively studied in the literature on nonparametric estimation and various methods and approaches proposed by Devroye [6] and Silverman [23].

Donoho, Johnstone, Kerkyacharian and Picard [9] and Cohen, De Vore, Kerkyacharian and Picard [4] studied the performance of the hard thresholding estimator. Assuming that the density f can be decomposed in a wavelet basis, this procedure reconstructs the target density f by only keeping the empirical coefficients which are greater than a specific value (threshold). This procedure has been proved to be very significant when dealing with the estimation of compactly supported densities. In particular, Donoho *et al.* [9] have shown that this adaptive procedure attains the minimax rate of convergence over Besov spaces (up to a logarithmic term). What is relevant in the paper of Cohen *et al.* [4] is that a new way of measuring the performance of statistical procedures has been proposed. It consists in investigating the maximal space (or *maxiset*), where the procedure attains a given rate of convergence. This new approach appears more optimistic than the minimax one, since it provides a functional set which is directly connected with the procedure. By assuming that the density f is compactly supported, Cohen *et al.* [4] proved that the maxiset associated with the hard thresholding procedure was the intersection of a Besov space and a weak Besov space. In this paper, we prove that the hypothesis of compactness of the support of f can be discarded.

Recently, Juditsky and Lambert-Lacroix [13] proposed a new adaptive procedure for density estimation on $\mathbb{R}$ when dealing with Hölder spaces. In their procedure, they propose to use a data-driven threshold to estimate the density function. A natural question arises here: In the maxiset context, is it relevant to replace the usual threshold by a data-driven one? In this paper we give an answer to this question, underlining the limits of usual thresholding rules in the maxiset sense.

Consequently our goal is threefold. We refer to any procedure, where the small empirical coefficients are neglected, as an *elitist rule* and prove that the maxiset of such a procedure is always contained in a Besov space. In other words, we exhibit conditions on procedures ensuring that their maxiset is contained in the intersection of a Besov space and a weak Besov space. Secondly, with no assumption about compactness of the density to be estimated, we prove that the hard thresholding procedure is the *best procedure* among elitist ones, since its maxiset is the largest one among those of elitist rules (*ideal maxiset*). Thirdly, under the maxiset approach, we compare this last procedure with the data-driven thresholding procedure provided by Juditsky and Lambert-Lacroix. Thanks to this, we succeed in pointing out the significance of the choice of data-driven thresholds in density estimation by proving that the maxiset of Juditsky and Lambert-Lacroix's procedure is larger than any elitist rule's one.

The paper is organized as follows. Section 2 recalls the problem of density estimation on $\mathbb{R}$ and defines the basic tools we shall need in the study. In Section 3, we define the maxiset point of view and the functional spaces often arising when dealing with this approach. In Section 4, we prove that the hard thresholding rule is the best procedure among elitist rules. Section 5 deals with the data-driven thresholds. We prove that the maxiset associated with the procedure of Juditsky and Lambert-Lacroix [13] is larger than any elitist rule's one. Finally, Section 6 contains the proofs of technical lemmas.

2. Density Estimation Model

We consider the problem of estimating an unknown density function f. Let $X_1, \ldots, X_n$ be n independent copies of a random variable X with density f with respect to the Lebesgue measure on $\mathbb{R}$. Instead of the local estimation setting as in Farrell [10, 11], Parzen [20], and Wahba [24], we choose to consider here the global estimation setting as Bretagnolle and Huber [3]. This paper aims at measuring the performance of estimators in a theoretical way.

To begin, let $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be compactly supported functions of $L_2(\mathbb{R})$ and denote for all $k \in \mathbb{Z}$ and all $x \in \mathbb{R}$, $\psi_{-1k}(x) = \phi(x-k)$, (resp. $\tilde{\psi}_{-1k}(x) = \tilde{\phi}(x-k)$) and for all $j \in \mathbb{N}$, $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ (resp. $\tilde{\psi}_{jk}(x) = 2^{j/2}\tilde{\psi}(2^j x - k)$).

Suppose that:

- $\{\psi_{jk};\ j \geq -1; k \in \mathbb{Z}\}$ and $\{\tilde{\psi}_{jk};\ j \geq -1; k \in \mathbb{Z}\}$ constitute a biorthogonal pair of wavelet bases of $L_2(\mathbb{R})$.
- The *reconstruction wavelet* $\tilde{\psi}$ is $\mathbf{C}^{N+1}$ for some $N \in \mathbb{N}$.
- The *wavelet* ψ is orthogonal to any polynomial of degree less than N.
- $\phi(x) = \mathbf{1}\{-\frac{1}{2} \leq x < \frac{1}{2}\}$ and support$(\psi) \subset [-\frac{m}{2}, \frac{m}{2}[$ for some $m \in \mathbb{N}^*$.

The feature of this particular basis which is intensively used throughout the paper, is that there exists $\nu > 0$ such that $|\psi(x)| \geq \nu$ on the support of ψ. Some most popular examples of such bases are given in Daubechies [5] and Donoho and Johnstone [7].

Suppose now that f can be represented as

$$f(t) = \sum_{j \geq -1} \sum_{k \in \mathbb{Z}} \beta_{jk} \tilde{\psi}_{jk}(t),$$

where $\forall j \geq -1$ for all $k \in \mathbb{Z}$:

- $\beta_{jk} = \displaystyle\int_{I_{jk}} f(t)\psi_{jk}(t)\, dt,$
- $I_{jk} = \left\{x \in \mathbb{R}; -\frac{m}{2} \leq (2^j \vee 1)x - k < \frac{m}{2}\right\}.$

Remark 2.1. As for each (j,k), the support of ψ_{jk} is contained in I_{jk}, we can easily prove that for every $j \geq -1$ and every $x \in \mathbb{R}$,

$$\#\{k;\ x \in I_{jk}\} \leq m. \tag{2.1}$$

Moreover, $\psi_{j,mi}$ and $\psi_{j,mi'}$ have disjoint supports. Thus

$$\sum_k p_{jk} = \sum_{l=1}^{m} \sum_i p_{j,mi+l} \leq \sum_{l=1}^{m} \int f(x)\, dx = m. \tag{2.2}$$

In the sequel, we shall denote:

- $p_{jk} = \displaystyle\int_{I_{jk}} f(t)\, dt, \qquad \forall j \geq -1,\ \forall k \in \mathbb{Z},$
- $\sigma_{jk}^2 = \displaystyle\int_{I_{jk}} f(t)\psi_{jk}^2(t)\, dt - \beta_{jk}^2, \qquad \forall j \geq -1,\ \forall k \in \mathbb{Z},$
- $f_j = \displaystyle\sum_{l=-1}^{j-1} \sum_{k \in \mathbb{Z}} \beta_{lk}\psi_{lk}, \qquad \forall j > -1.$

3. Maxiset Point of View

In this paper, we consider the maxiset approach to measure the performance of estimators. This new approach was introduced by Cohen *et al.* [4] and consists in finding the maximal space (maxiset), where a fixed procedure $\hat{f}$ achieves a given rate of convergence r_n. In the following subsection, we introduce the maxiset notation.

3.1. Maxiset notation. For a loss function ρ, we denote by

$$MS(\hat{f}, \rho, r_n) = \{f;\ \sup_{n\in\mathbb{N}^*} r_n^{-1}\rho(\hat{f}, f) < \infty\}$$

the maxiset associated with the procedure $\hat{f}$ and the rate r_n.

Let us remark that the maxiset point of view is more optimistic than the minimax one, since it provides the strong connection between an estimation procedure and a functional space.

3.2. Functional spaces. In this subsection, we introduce the sequence spaces often arising when dealing with the maxiset approach (see Cohen *et al.* [4] and Kerkyacharian and Picard [17, 18]).

Definition 3.1. Let $0 < s < N+1$ and $1 \leq p, q \leq \infty$. We say that a density f of $L_p(\mathbb{R})$ belongs to the Besov space $\mathcal{B}^s_{p,q}$ if and only if

$$\big(2^{j(s-\frac{1}{p}+\frac{1}{2})}\|\beta_{j\cdot}\|_{l_p};\ j \geq -1\big) \in l_q.$$

Remark 3.1. It is clear, using the definition above, that a density f belongs to $\mathcal{B}^s_{p,\infty}$ if and only if

$$\sup_{J\in\mathbb{N}} 2^{Jsp} \sum_{j\geq J} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p < \infty. \tag{3.1}$$

The Besov spaces are of statistical interest, since they model important forms of spatial inhomogeneity. These spaces have been proved to play a prominent role when dealing with the maxiset approach. Indeed, Kerkyacharian and Picard [16] have proved that the maximal space, where any linear procedure attains the rate of convergence $n^{-sp/(1+2s)}$ for the $\mathbb{L}_p$-risk, $p \geq 2$, is contained in the Besov space $\mathcal{B}^s_{p,\infty}$. Let us recall that the scale of Besov spaces includes the Hölder spaces ($C^s = \mathcal{B}^s_{\infty,\infty}$) and the Hilbert–Sobolev spaces ($H^s_2 = \mathcal{B}^s_{2,2}$).

Definition 3.2. Let $0 < r < p < \infty$. We say that f belongs to the weak Besov space $W(r,p)$ if and only if

$$\sup_{\lambda>0} \lambda^r \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\{|\beta_{jk}| > \lambda\} < \infty,$$

which is equivalent to (see Cohen *al.* [4])

$$\sup_{\lambda>0} \lambda^{r-p} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{|\beta_{jk}| \leq \lambda\} < \infty.$$

These spaces naturally appear when studying the maximal spaces of thresholding rules (see Cohen *al.* [4] and Kerkyacharian and Picard [17, 18]). Weak Besov spaces constitute a large class of functions since, using Markov's inequality, it is easy to prove that for $r < p$, the Besov space $\mathcal{B}^s_{rr} \subset W(r,p)$ when $s \geq \frac{p}{2r} - \frac{1}{2}$. Under the maxiset approach, we prove in Section 4 that weak Besov spaces are directly connected with a large family of procedures called *elitist rules*.

Definition 3.3. Let $0 < r < p < \infty$. We say that f belongs to the space $W^*(r,p)$ if and only if

$$\sup_{\lambda>0} \lambda^r \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k \sigma^p_{jk} \mathbf{1}\left\{ \frac{|\beta_{jk}|}{\sigma_{jk}} > \lambda \right\} < \infty,$$

which is equivalent to (see Kerkyacharian and Picard [17])

$$\sup_{\lambda>0} \lambda^{r-p} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\left\{ \frac{|\beta_{jk}|}{\sigma_{jk}} \leq \lambda \right\} < \infty.$$

Note that $W(r,p)$ and $W^*(r,p)$ are natural spaces to measure the sparsity of a sequence by controlling the proportion of nonnegligible β_{jk}'s. In Section 5, we illustrate the strong link between the spaces $W^*(r,p)$ and procedures based on data-driven thresholds.

Definition 3.4. We say that a function f belongs to the space $\chi(r,p)$ if and only if

$$\sup_{\lambda>0} \lambda^{r-p} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{p_{jk} \leq \lambda^2\} < \infty.$$

These functional spaces constitute a large family of functions. To be more precise, consider the following proposition dealing with embeddings of functional spaces.

Proposition 3.1. *For any* $0 < \alpha < 1$ *and any* $1 \leq p < \infty$*, we have the following inclusions of spaces:*

$$\begin{aligned} \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p,p) &\subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap \chi((1-\alpha)p,p), \\ \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p,p) &\subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap \chi((1-\alpha)p,p). \end{aligned} \tag{3.2}$$

Moreover, if $\alpha p > 2$*, then*

$$\mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p,p) \subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p,p). \tag{3.3}$$

Proof. Here and later, C represents any constant we need and can be different from one line to another.

Denote $K_\psi = \|\phi\|_\infty \vee \|\psi\|_\infty$. Let $\lambda > 0$ and let u be the integer such that $2^u \leq \lambda^{-2} < 2^{1+u}$.

Clearly, if $\lambda^2 \geq \frac{\nu^2}{2K^2_\psi}$, then for any f that belongs to $\mathcal{B}^{\alpha/2}_{p,\infty}$

$$\sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{p_{jk} \leq \lambda^2\} \leq \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \leq C\lambda^{\alpha p}.$$

Suppose now that $\lambda^2 < \frac{\nu^2}{2K_\psi^2}$. Since for any (j,k), $|\beta_{jk}| \le K_\psi 2^{j/2} p_{jk}$, we have for any $j < u$:

$$p_{jk} \le \lambda^2 \Longrightarrow |\beta_{jk}| \le K_\psi \lambda$$

and

$$\sigma_{jk}^2 \ge 2^j \nu^2 p_{jk} - 2^j K_\psi^2 p_{jk}^2 = 2^j p_{jk}(\nu^2 - K_\psi^2 p_{jk}) \ge 2^{j-1}\nu^2 p_{jk}.$$

So, if f belongs to $W((1-\alpha)p, p)$ (resp. $W^*((1-\alpha)p, p)$), then

$$\begin{aligned}
&\sum_{j\ge -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{p_{jk} \le \lambda^2\} \\
&\quad \le \sum_{j<u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{p_{jk} \le \lambda^2\} + \sum_{j\ge u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\quad \le C \sum_{j=-1}^{u-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{|\beta_{jk}| \le K_\psi \lambda\} + C 2^{-\frac{\alpha}{2}up} \le C\lambda^{\alpha p},
\end{aligned}$$

resp.,

$$\begin{aligned}
&\sum_{j\ge -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{p_{jk} \le \lambda^2\} \\
&\quad \le \sum_{j<u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{p_{jk} \le \lambda^2\} + \sum_{j\ge u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\quad \le C \sum_{j=-1}^{u-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{|\beta_{jk}| \le K_\psi 2^{j/2} p_{jk}\} + C 2^{-\frac{\alpha}{2}up} \\
&\quad \le C \sum_{j=-1}^{u-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{|\beta_{jk}| \le \frac{\sqrt{2}K_\psi}{\nu} \lambda \sigma_{jk}\} + C 2^{-\frac{\alpha}{2}up} \le C\lambda^{\alpha p}.
\end{aligned}$$

We conclude that $f \in \chi((1-\alpha)p, p)$. So (3.2) is satisfied. Now, (3.3) is clearly satisfied since for every $1 \le p < \infty$ and every α such that $\alpha p > 2$, it is easy to prove that $f \in \mathcal{B}_{p,\infty}^{\alpha/2} \Longrightarrow \sup_{j,k} \sigma_{jk} < \infty$. $\square$

4. Ideal Maxiset for Elitist Rules

Here we focus on the maxiset performance of adaptive procedures (i.e., those which do not depend on the parameter α).

4.1. Definition of ideal maxiset.

Let us introduce the definition of the ideal maxiset of a family of procedures.

Definition 4.1. Let ρ be a loss function, r_n a rate, and $\mathcal{M}_n$ a family of estimation procedures. We say that the functional space V is the *ideal maxiset* of the family $\mathcal{M}_n$ for the rate r_n if the two following properties hold:

1. $\forall \hat{f} \in \mathcal{M}_n,\ \ MS(\hat{f}, \rho, r_n) \subset V$,
2. $\exists \hat{f}^* \in \mathcal{M}_n,\ \ MS(\hat{f}^*, \rho, r_n) = V$.

In the next subsection, we provide the ideal maxiset of a large family of procedures, namely, the *elitist rules*, which consist in discarding small empirical coefficients in the reconstruction of the target density f.

4.2. Definition of elitist rules. Fix $r > 0$. Let $(v(n))_{n\in\mathbb{N}^*}$ be a decreasing sequence of strictly positive real numbers tending to zero as $n \to \infty$. Denote j_n the integer such that $2^{j_n} \le v(n)^{-r} < 2^{1+j_n}$ and let E_n be a sequence of statistical experiments such that for any f we can estimate β_{jk} by $\hat{\beta}_{jk}$ for any j and any k.

Let us consider the sub-family $\mathcal{F}_K^{'}$ of keep-or-kill procedures defined by:

$$\mathcal{F}_K^{'} = \Big\{ \hat{f}(\cdot) = \sum_{j<j_n} \sum_k \omega_{jk} \hat{\beta}_{jk} \tilde{\psi}_{jk}(\cdot);\ \omega_{jk} \in \{0,1\} \text{ measurable} \Big\}.$$

Definition 4.2. We say that $\hat{f} \in \mathcal{F}_K^{'}$ is an *elitist rule* if and only if for any j and any $k \in \mathbb{Z}$,

$$|\hat{\beta}_{jk}| \le v(n) \implies \omega_{jk} = 0.$$

This definition means that the 'small' coefficients will be neglected.

In the sequel, the choice of the loss function will be the Besov norm. A possible alternative could be to use the L_p-norm, but this choice leads to technical difficulties avoided by choosing the Besov norm.

4.3. Upper bound for maxisets of elitist rules. The goal of this subsection is to prove that the maximal space, where any elitist rule attains the rate of convergence $v(n)^{\alpha p}$, is contained in the intersection of a Besov space and a weak Besov space. We have the following theorem:

Theorem 4.1. *Let $\hat{f}$ be an elitist rule. Then, for any $1 \le p < \infty$,*

$$\sup_n v(n)^{-\alpha p}\, \mathbb{E}\, \|\hat{f} - f\|^p_{\mathcal{B}^0_{p,p}} < \infty \implies f \in \mathcal{B}^{\alpha/r}_{p,\infty} \cap W((1-\alpha)p, p),$$

that is to say, using the maxiset notation,

$$MS\big(\hat{f}, \|\cdot\|^p_{\mathcal{B}^0_{p,p}}, v(n)^{\alpha p}\big) \subset \mathcal{B}^{\alpha/r}_{p,\infty} \cap W((1-\alpha)p, p). \tag{4.1}$$

Proof. Fix $1 \le p < \infty$ and let f be such that $\sup_{n>1} v(n)^{-\alpha p}\, \mathbb{E}\, \|\hat{f} - f\|^p_{\mathcal{B}^0_{p,p}} < \infty$. On the one hand, for all $n > 1$, we have:

$$\begin{aligned}
&\sum_{j\ge j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\quad\le \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \big|\beta_{jk} - \hat{\beta}_{jk} \mathbf{1}\{\omega_{jk} = 1\}\big|^p + \sum_{j\ge j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\quad= \mathbb{E}\, \|\hat{f} - f\|^p_{\mathcal{B}^0_{p,p}} \le C v(n)^{\alpha p} \le C 2^{-j_n \frac{\alpha p}{r}}.
\end{aligned}$$

From (3.1), it follows that $f \in \mathcal{B}^{\alpha/r}_{p,\infty}$.

On the other hand, since

$$|\beta_{jk}|\mathbf{1}\Big\{|\beta_{jk}| \le \frac{v(n)}{2}\Big\} \le \big|\beta_{jk} - \hat\beta_{jk}\mathbf{1}\{\omega_{jk}=1\}\mathbf{1}\{|\hat\beta_{jk}| > v(n)\}\big|,$$

we have

$$\begin{aligned}
&\sum_{j\geqslant -1} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \le \frac{v(n)}{2}\Big\}\\
&\quad \le \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \le \frac{v(n)}{2}\Big\} + \sum_{j\geqslant j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\\
&\quad \leqslant \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \big|\beta_{jk} - \hat\beta_{jk}\mathbf{1}\{\omega_{jk}=1\}1\{|\hat\beta_{jk}| > v(n)\}\big|^p\\
&\qquad + \sum_{j\geqslant j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\\
&\quad = \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \big|\beta_{jk} - \hat\beta_{jk}\mathbf{1}\{\omega_{jk}=1\}\big|^p + \sum_{j\geqslant j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\\
&\quad = \mathbb{E}\,\|\hat f - f\|^p_{\mathcal{B}^0_{p,p}} \leqslant C v(n)^{\alpha p}.
\end{aligned}$$

As v_n tends to 0 when n goes to $+\infty$, we conclude that $f \in W((1-\alpha)p, p)$. □

In the next subsection, we aim at proving that the space $\mathcal{B}^{\alpha/r}_{p,\infty} \cap W((1-\alpha)p,p)$ is the *ideal maxiset* of the family of elitist rules when dealing with a fixed rate $v_n^{\alpha p}$. Let us note that, according to Theorem 4.1, it suffices to provide an elitist rule having this functional space as the maxiset to conclude that this particular set is the ideal one.

4.4. Ideal elitist rule. We decompose the study into two parts. In the first one, we recall the main result about maxisets of Cohen *et al.* [4] when dealing with estimation for *compactly supported densities* (see Theorem 4.2). In the second one, we extend it to *noncompactly supported densities* (see Theorem 4.4). The final outcome of this section is proving that the hard thresholding rule, which is clearly an elitist rule, is optimal in the maxiset sense among the family of elitist rules. In the sequel, we suppose that $v(n) = \mu\sqrt{\frac{\log(n)}{n}}$, for some $\mu > 0$ and that $r = 2$.

4.4.1. *Compactly supported densities.* Cohen *et al.* [4] have studied the maximal space of hard thresholding rules. They obtained the following result:

Theorem 4.2 (Cohen *et al.* [4]). *For any $a > 0$, let $I = [-a, a]$, and let j_n be the integer such that $2^{j_n} \le \frac{n}{\log(n)} < 2^{j_n+1}$.*

Denote $\hat\beta_{jk} = \frac{1}{n}\sum_{i=1}^n \psi_{jk}(X_i)$ and consider the following hard thresholding estimator:

$$\hat f_\mu = \sum_{j<j_n}\sum_k \hat\beta_{jk}\mathbf{1}\Big\{|\hat\beta_{jk}| > \mu\sqrt{\frac{\log(n)}{n}}\Big\}\tilde\psi_{jk}, \tag{4.2}$$

where μ *is a large enough constant. We have for any* $1 < p < \infty$:

$$MS\big(\hat{f}_\mu, \|\cdot\|^p_{L_p}, v_n^{\alpha p}\big) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p). \tag{4.3}$$

The proof of this theorem uses the unconditional nature of the wavelet basis $\{\tilde{\psi}_{jk};\, j \geq -1;\, k \in \mathbb{Z}\}$. In the same way, it would be easy to prove the following similar result.

Theorem 4.3. *Let* $1 \leq p < \infty$*. Under the same assumptions and definitions as in Theorem* 4.2*, we get for any* $1 \leq p < \infty$:

$$MS\big(\hat{f}_\mu, \|\cdot\|^p_{\mathcal{B}^0_{p,p}}, v_n^{\alpha p}\big) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p). \tag{4.4}$$

Thus, using Theorem 4.1, we conclude that $\mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p)$ is the *ideal maxiset* of the family of elitist rules. Moreover, we can conclude that the hard thresholding procedure is optimal in the maxiset sense within the family of elitist rules dealing with compactly supported functions.

A natural question arises here: Is the hard thresholding procedure still optimal within this class of rules, without assuming compact support of the target density f? The answer is YES. We shall prove it in the next subsection.

4.4.2. *Noncompactly supported densities.* Let us introduce the following quantities:

- $m_n = \dfrac{\mu}{K_\psi}\left(1 \wedge \dfrac{\mu\nu^2}{2K_\psi}\right)\log(n)$,
- $n_{jk} = \displaystyle\sum_{i=1}^n \mathbf{1}\{X_i \in I_{jk}\}$,
- $\hat{\beta}_{jk} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n \psi_{jk}(X_i)$.

The following theorem can be viewed as a generalization of Theorem 4.3 when dealing with density estimation on $\mathbb{R}$.

Theorem 4.4. *Let* $0 < \alpha < 1$ *and* $1 \leq p < \infty$ *be such that* $\alpha p > 2$*. If* μ *is large enough, then*

$$MS\big(\hat{f}_\mu, \|\cdot\|^p_{\mathcal{B}^0_{p,p}}, v_n^{\alpha p}\big) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p). \tag{4.5}$$

From Theorem 4.1 and Theorem 4.4, we deduce an immediate corollary:

Corollary 4.1. *Let* $0 < \alpha < 1$ *and* $1 \leq p < \infty$ *be such that* $\alpha p > 2$*. The ideal maxiset of elitist rules for the rate* $v_n^{\alpha p}$ *is* $\mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p)$*. Moreover, the hard thresholding rule is the best elitist procedure in the maxiset sense.*

Proof of Theorem 4.4. "$\subset$" It suffices to apply Theorem 4.1.

"$\supset$" Let $f \in \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p,p)$. The Besov-risk of $\hat{f}_\mu$ can be decomposed as follows:

$$\mathbb{E}\,\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} = \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \big|\beta_{jk} - \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > v_n\}\big|^p + \|f - f_{j_n}\|^p_{\mathcal{B}^0_{p,p}} = A_0 + A_1.$$

Since $f \in \mathcal{B}^{\alpha/2}_{p,\infty}$, from (3.1)

$$A_1 = \|f - f_{j_n}\|^p_{\mathcal{B}^0_{p,p}} \le \mathbb{E}\,\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} \le C2^{-j_n\alpha p/2} \le C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.$$

A_0 can be decomposed into two parts:

$$\begin{aligned}A_0 &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \big|\beta_{jk} - \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > v_n\}\big|^p\\ &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\{|\hat{\beta}_{jk}| \le v_n\}\\ &\quad + \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk} - \hat{\beta}_{jk}|^p\mathbf{1}\{|\hat{\beta}_{jk}| > v_n\} = A_0^{'} + A_0^{''}.\end{aligned}$$

Now,

$$\begin{aligned}A_0^{'} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\{|\hat{\beta}_{jk}| \le v_n\}\\ &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\{|\hat{\beta}_{jk}| \le v_n\}\big[\mathbf{1}\{|\beta_{jk}| \le 2v_n\} + \mathbf{1}\{|\beta_{jk}| > 2v_n\}\big]\\ &= A_{01}^{'} + A_{02}^{'}.\end{aligned}$$

Using the definition of $W((1-\alpha)p,p)$,

$$\begin{aligned}A_{01}^{'} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\{|\hat{\beta}_{jk}| \le v_n\}\mathbf{1}\{|\beta_{jk}| \le 2v_n\}\\ &\le \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\{|\beta_{jk}| \le 2v_n\} \le C(2v_n)^{\alpha p} \le C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.\end{aligned}$$

We need the following lemma:

Lemma 4.1. *Let $1 \le p < \infty$. For any $\gamma > 0$, there exists $\mu(\gamma) < \infty$ and $C < \infty$ such that for any $-1 \le j < j_n$ and any $k \in \mathbb{Z}$,*

$$\mathbb{P}\left(|\hat{\beta}_{jk} - \beta_{jk}| > \mu\sqrt{\frac{\log(n)}{n}}\right) \le \frac{C}{n^\gamma}.$$

The *proof* is clear by using the Bernstein inequality. $\square$

Choosing $\gamma \geq \frac{p}{2}$, one gets for μ large enough:

$$A_{02}^{'} = \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}| \leq v_n\}\mathbf{1}\{|\beta_{jk}| > 2v_n\}$$

$$\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{P}_f\big(|\hat{\beta}_{jk} - \beta_{jk}| > v_n\big) \leq Cn^{-\gamma} \leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.$$

Let us now consider the following lemma:

Lemma 4.2. *For any $j < j_n$ and any k, $|\hat{\beta}_{jk}| > v_n \implies n_{jk} > m_n$.*

The *proof* is given in the Appendix.

We can decompose $A_0^{''}$ into three parts:

$$\begin{aligned}
A_0^{''} &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}| > v_n\} \\
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}| > v_n\}\mathbf{1}\{n_{jk} \geq m_n\} \\
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}| > v_n\}\mathbf{1}\{n_{jk} \geq m_n\}\mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&\quad + \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}| > v_n\}\mathbf{1}\{n_{jk} \geq m_n\}\mathbf{1}\Big\{p_{jk} \geq \frac{m_n}{2n}\Big\} \\
&\qquad \times \Big[\mathbf{1}\Big\{|\beta_{jk}| \leq \frac{v_n}{2}\Big\} + \mathbf{1}\Big\{|\beta_{jk}| > \frac{v_n}{2}\Big\}\Big] \\
&= A_{01}^{''} + A_{02}^{''} + A_{03}^{''}.
\end{aligned}$$

To bound $A_{01}^{''}$, $A_{02}^{''}$ and $A_{03}^{''}$, we introduce two lemmas.

Lemma 4.3. *For any $\gamma > 0$ there exists $\mu = \mu(\gamma) < \infty$ such that for any j, k and any n large enough*:

$$\mathbb{P}_f(n_{jk} < m_n) \leq \frac{p_{jk}}{n^\gamma} \qquad \text{if} \quad p_{jk} \geq \frac{2m_n}{n},$$
$$\mathbb{P}_f(n_{jk} \geq m_n) \leq \frac{p_{jk}}{n^\gamma} \qquad \text{if} \quad p_{jk} < \frac{m_n}{2n}.$$

This lemma is a generalization of Lemma 7 of Juditsky and Lambert-Lacroix [13]. Its proof is given in the Appendix.

Lemma 4.4. *Let $1 \leq p < \infty$. Then*:

1. $\mathbb{E}\,|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leq C\left(\frac{2^j p_{jk}}{n}\right)^p \quad$ *if* $\quad p_{jk} \geq \frac{1}{n}$,
2. $\mathbb{E}\,|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leq C\left(\frac{2^j}{n^2}\right)^p np_{jk} \quad$ *if* $\quad p_{jk} < \frac{1}{n}$,
3. $\mathbb{E}\,|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leq C\left(\frac{2^j}{n}\right)^p p_{jk}$.

The *proof* is given in the Appendix.

Using Lemma 4.3, Lemma 4.4 (3), and the Cauchy–Schwarz inequality, we have for μ large enough:

$$\begin{aligned}
A_{01}^{''} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}|>v_n\}\mathbf{1}\{n_{jk}\geq m_n\}\mathbf{1}\Big\{p_{jk}<\frac{m_n}{2n}\Big\}\\
&\leq \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbf{1}\{n_{jk}\geq m_n\}\mathbf{1}\Big\{p_{jk}<\frac{m_n}{2n}\Big\}\\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}^{1/2}|\hat{\beta}_{jk}-\beta_{jk}|^{2p}\mathbb{P}_f^{1/2}(n_{jk}\geq m_n)\mathbf{1}\Big\{p_{jk}<\frac{m_n}{2n}\Big\}\\
&\leq \frac{C}{n^{\gamma/2}}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k p_{jk}\left[\frac{2^j}{n}\right]^{p/2} \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.
\end{aligned}$$

The last inequality is due to (2.2) and requires to choose $\gamma \geq 2(p-1)$.

Using the Cauchy–Schwarz inequality, Lemma 4.1 with $\gamma \geq 2p-1$, and Lemma 4.4 (3), one gets:

$$\begin{aligned}
A_{02}^{''} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbf{1}\{|\hat{\beta}_{jk}|>v_n\}\mathbf{1}\Big\{|\beta_{jk}|\leq\frac{v_n}{2}\Big\}\\
&\qquad\times \mathbf{1}\{n_{jk}\geq m_n\}\mathbf{1}\Big\{p_{jk}\geq\frac{m_n}{2n}\Big\}\\
&\leq \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbf{1}\Big\{|\hat{\beta}_{jk}-\beta_{jk}|>\frac{v_n}{2}\Big\}\mathbf{1}\Big\{p_{jk}\geq\frac{m_n}{2n}\Big\}\\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}^{1/2}|\hat{\beta}_{jk}-\beta_{jk}|^{2p}\,\mathbb{P}_f^{1/2}\Big(|\hat{\beta}_{jk}-\beta_{jk}|>\frac{v_n}{2}\Big)\mathbf{1}\Big\{p_{jk}\geq\frac{m_n}{2n}\Big\}\\
&\leq C\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\Big(\frac{2^j}{n}\Big)^{p/2} v_n^{\gamma-1}\sum_k p_{jk} \leq C v_n^{\alpha p}.
\end{aligned}$$

Finally, from Lemma 4.4 (1), we have:

$$\begin{aligned}
A_{03}^{''} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbf{1}\{n_{jk}\geq m_n\}\mathbf{1}\Big\{p_{jk}\geq\frac{m_n}{2n}\Big\}\\
&\qquad\times \mathbf{1}\{|\hat{\beta}_{jk}|>v_n\}\mathbf{1}\Big\{|\beta_{jk}|>\frac{v_n}{2}\Big\}\\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}\,|\hat{\beta}_{jk}-\beta_{jk}|^p \mathbf{1}\Big\{p_{jk}\geq\frac{m_n}{2n}\Big\}\mathbf{1}\Big\{|\beta_{jk}|>\frac{v_n}{2}\Big\}\\
&\leq C\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \Big(\frac{2^j p_{jk}}{n}\Big)^{p/2}\mathbf{1}\Big\{p_{jk}\geq\frac{m_n}{2n}\Big\}\mathbf{1}\Big\{|\beta_{jk}|>\frac{v_n}{2}\Big\}\\
&\leq C\frac{1}{n^{p/2}}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbf{1}\{|\beta_{jk}|>\frac{v_n}{2}\} \leq C v_n^{\alpha p}.
\end{aligned}$$

The last inequalities use the fact that $\sup_{j,k} 2^j p_{jk} < \infty$ for any $f \in \mathcal{B}_{p,\infty}^{\alpha/2}$ (with $\alpha p > 2$).

Consequently, looking at the bounds for A_0 and A_1, we conclude that

$$\sup_{n>1} v_n^{-\alpha p}\, \mathbb{E}\, \|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} < \infty.$$

So $f \in MS(\hat{f}_\mu, \|\cdot\|^p_{\mathcal{B}^0_{p,p}}, v_n^{\alpha p})$. $\square$

Until now, we have focused on nonrandom thresholds. In particular, we have proved that the hard thresholding estimator is the *best procedure* among elitist ones, in terms of the maxiset approach. It is of interest to answer the following question: Do there exist adaptive procedures which outperform hard thresholding rules in the maxiset sense? Once again, the answer is YES, by considering data-driven thresholds (see Birgé and Massart [2], Donoho and Johnstone [8], Juditsky [12], and Juditsky and Lambert-Lacroix [13]). This will be proved in the next section.

5. On the Significance of Data-Driven Thresholds

The aim of this section is to prove the significance of data-driven thresholds, in the context of estimating compactly or noncompactly supported densities.

We study the maxiset associated with the data-driven thresholding procedure described by Juditsky and Lambert-Lacroix [13]. Here, the decision to *keep* or to *kill* empirical coefficients $\hat{\beta}_{jk}$ is taken by comparing them to their standard deviation. We prove that the maxiset associated with this particular data-driven thresholding procedure is larger than the ideal maxiset of elitist rules. We shall denote:

- $\hat{\gamma}_{jk} = \mu\sqrt{\dfrac{\log(n)}{n}}\hat{\sigma}_{jk} = v_n\hat{\sigma}_{jk}$, where $\hat{\sigma}^2_{jk} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(\psi^2_{jk}(X_i) - \hat{\beta}^2_{jk})$,
- $\gamma_{jk} = \mu\sqrt{\dfrac{\log(n)}{n}}\sigma_{jk} = v_n\sigma_{jk}$.

Consider the *data-driven thresholding estimator* defined by Juditsky and Lambert-Lacroix [13]:

$$\bar{f}_n(t) = \sum_{j=-1}^{j_n-1}\sum_{k\in\mathbb{Z}} \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}\tilde{\psi}_{jk}(t),$$

with $2^{j_n} \le \frac{n}{\log(n)} < 2^{j_n+1}$.

Theorem 5.1. *Let* $0 < \alpha < 1$ *and* $1 \le p < \infty$ *be such that* $\alpha p > 2$. *If* μ *is large enough, then*

$$MS(\bar{f}_n, \|\cdot\|^p_{\mathcal{B}^0_{p,p}}, v_n^{\alpha p}) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W^*((1-\alpha)p, p). \tag{5.1}$$

Combined with (3.3) of Proposition 3.1, this theorem proves that the maxiset associated with the data-driven thresholding estimator $\bar{f}_n$ is larger than the maxiset of any elitist estimator $\hat{f}$ built with a nonrandom threshold.

Proof of Theorem 5.1. "$\subset$" Fix $1 \le p < \infty$ and let f be such that

$$\sup_{n>1} \left(\frac{n}{\log(n)}\right)^{\alpha p/2} \mathbb{E}\, \|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} < \infty.$$

On the one hand, with the same arguments as in the proof of Theorem 4.1, for all $n > 1$, we have:

$$\sum_{j \ge j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \le \mathbb{E}\, \|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} \le C\left(\frac{\log(n)}{n}\right)^{\alpha p/2} \le C 2^{-j_n \frac{\alpha p}{2}}.$$

It follows that $f \in \mathcal{B}^{\alpha/2}_{p,\infty}$.

On the other hand, for any $n > 1$ we have,

$$\begin{aligned}
&\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \le \frac{\upsilon_n \sigma_{jk}}{4}\Big\} \\
&\quad = \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \le \frac{\gamma_{jk}}{4}\Big\} \\
&\quad = \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \le \frac{\gamma_{jk}}{4}\Big\} \Big[\mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&\qquad + \mathbf{1}\Big\{\frac{m_n}{2n} \le p_{jk} \le \frac{\nu^2}{2K_\psi^2}\Big\} + \mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\}\Big] = B_0 + B_1 + B_2.
\end{aligned}$$

Let us introduce the following lemma.

Lemma 5.1. *For any $j < j_n$, any k and any n large enough, $|\hat{\beta}_{jk}| > \hat{\gamma}_{jk} \Longrightarrow n_{jk} > m_n$.*

The proof of this lemma is given in the Appendix.

To bound B_0, we use Lemma 4.3 with $\gamma \ge \frac{p}{2}$ and Lemma 5.1:

$$\begin{aligned}
B_0 &= \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \le \frac{\gamma_{jk}}{4}\Big\} \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&\le \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} [\mathbf{1}\{n_{jk} < m_n\} + \mathbf{1}\{n_{jk} \ge m_n\}] \\
&\le \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{n_{jk} < m_n\} \\
&\quad + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{P}_f(n_{jk} \ge m_n) \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&\le \mathbb{E}\, \|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \frac{p_{jk}}{n^\gamma}
\end{aligned}$$

$$\leq \mathbb{E}\,\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} + C\;n^{-\gamma} \leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.$$

To bound B_1, let us consider the following lemma:

Lemma 5.2. *Fix $\gamma > 0$. There exists $\mu = \mu(\gamma) < \infty$ such that*

1. *if* $p_{jk} \geq \dfrac{\mu}{2K_\psi} \cdot \dfrac{\log(n)}{n}$, *then* $\mathbb{P}_f\left(\hat{\gamma}_{jk} > \mu\sqrt{\dfrac{\log(n)}{n}}\right) \leq \dfrac{p_{jk}}{n^\gamma}$;
2. *moreover, if* $\dfrac{\mu}{2K_\psi} \cdot \dfrac{\log(n)}{n} \leq p_{jk} \leq \dfrac{\nu^2}{2K_\psi^2}$ *for n large enough, then*

$$\text{(a)} \qquad \mathbb{P}_f\left(|\hat{\gamma}_{jk} - \gamma_{jk}| > \frac{\gamma_{jk}}{2}\right) \leq \frac{2p_{jk}}{n^\gamma},$$

$$\text{(b)} \qquad \mathbb{P}_f\left(|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\right) \leq \frac{2p_{jk}}{n^\gamma}.$$

Proof. This lemma is a simple generalization of Proposition 1 in Juditsky and Lambert-Lacroix [13]. The proof is omitted, since it uses similar arguments to those used by therein. □

Since $|\beta_{jk}|\mathbf{1}\{|\beta_{jk}| \leq \frac{\hat{\gamma}_{jk}}{2}\} \leq |\beta_{jk} - \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}|$, by using Lemma 5.2 (2(a)) with $\gamma \geq \frac{p}{2}$, one gets:

$$\begin{aligned}
B_1 &= \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\Big[\mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\hat{\gamma}_{jk}}{2}\Big\} \\
&\qquad + \mathbf{1}\Big\{|\beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\Big\}\Big]\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \Big(\big|\beta_{jk} - \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}\big|^p \\
&\qquad + |\beta_{jk}|^p \mathbf{1}\Big\{\hat{\gamma}_{jk} < \frac{\gamma_{jk}}{2}\Big\}\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\}\Big) \\
&\leq \mathbb{E}\,\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\qquad \times \mathbb{P}_f\Big(|\hat{\gamma}_{jk} - \gamma_{jk}| > \frac{\gamma_{jk}}{2}\Big)\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C\left[\left(\frac{\log(n)}{n}\right)^{\alpha p/2} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \frac{p_{jk}}{n^\gamma}\right] \\
&\leq C\left[\left(\frac{\log(n)}{n}\right)^{\alpha p/2} + n^{-\gamma}\right] \leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

Now, using the fact that $\sup_{j,k} 2^{j} p_{jk} < \infty$ and $\sigma_{jk}^2 \leq 2^{j} K_\psi^2\, p_{jk}$,

$$\begin{aligned}
B_2 &= \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\gamma_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&= C v_n^p \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\sigma_{jk}|^p \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C v_n^p \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k (2^{j} p_{jk})^{p/2} \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C v_n^p \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k (2^{j} p_{jk})^{p/2} \mathbf{1}\Big\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\Big\}\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C \Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.
\end{aligned}$$

Consequently, the bounds for B_i, $0 \leq i \leq 2$, show that $f \in W^*((1-\alpha)p, p)$.

"$\supset$" Let $f \in \mathcal{B}_{p,\infty}^{\alpha/2} \cap W^*((1-\alpha)p,p)$. The Besov-risk of $\bar{f}_n$ can be decomposed as follows:

$$\begin{aligned}
\mathbb{E}\,\|\bar{f}_n - f\|_{\mathcal{B}_{p,p}^0}^p &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \big|\beta_{jk} - \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}\big|^p + \|f - f_{j_n}\|_{\mathcal{B}_{p,p}^0}^p \\
&= C_0 + C_1.
\end{aligned}$$

Using similar arguments as in the proof of Theorem 4.4, since $f \in \mathcal{B}_{p,\infty}^{\alpha/2}$,

$$C_1 = \|f - f_{j_n}\|_{\mathcal{B}_{p,p}^0}^p \leq C \Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.$$

For n large enough, using Lemma 5.1, we can decompose C_0 as follows:

$$\begin{aligned}
C_0 &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \big|\beta_{jk} - \hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}\big|^p \\
&\leq \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{n_{jk} < m_n\} \\
&\quad + \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \big|\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}\big|^p \mathbf{1}\{n_{jk} \geq m_n\} \\
&= C_0^{'} + C_0^{''}.
\end{aligned}$$

Since $f \in \mathcal{B}_{p,\infty}^{\alpha/2} \cap W((1-\alpha)p,p)$, $f \in \chi((1-\alpha)p,p)$. So, by using Lemma 4.3 with $\gamma \geq \frac{p}{2}$, one gets

$$C_0^{'} = \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{n_{jk} < m_n\}$$

$$
\begin{aligned}
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{n_{jk} < m_n\} \left[\mathbf{1}\Big\{p_{jk} < \frac{2m_n}{n}\Big\} + \mathbf{1}\Big\{p_{jk} \geq \frac{2m_n}{n}\Big\} \right] \\
&\leq \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{p_{jk} < \frac{2m_n}{n}\Big\} \\
&\quad + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{P}(n_{jk} < m_n) \mathbf{1}\Big\{p_{jk} \geq \frac{2m_n}{n}\Big\} \\
&\leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \frac{p_{jk}}{n^\gamma} \\
&\leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2} + Cn^{-\gamma} \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.
\end{aligned}
$$

We have the following decomposition for $C_0^{''}$:

$$
\begin{aligned}
C_0^{''} &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} \mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p \mathbf{1}\{n_{jk} \geq m_n\} \\
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \left|\hat{\beta}_{jk} \mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}\right|^p \mathbf{1}\{n_{jk} \geq m_n\} \\
&\quad \times \left[\mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} + \mathbf{1}\Big\{p_{jk} \geq \frac{m_n}{2n}\Big\} \right] = C_{01}^{''} + C_{02}^{''}.
\end{aligned}
$$

Now, since $|\hat{\beta}_{jk} \mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}| \leq |\hat{\beta}_{jk} - \beta_{jk}| + |\beta_{jk}|$, $C_{01}^{''}$ can be decomposed into $C_{011}^{''} + C_{012}^{''}$, with

$$
\begin{aligned}
C_{011}^{''} &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\{n_{jk} \geq m_n\} \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
C_{012}^{''} &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{n_{jk} \geq m_n\} \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\}.
\end{aligned}
$$

Using again Lemma 4.3 with $\gamma \geq 2p-1$, Lemma 4.4 (3) and 4.4 (2), we get

$$
\begin{aligned}
C_{011}^{''} &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\{n_{jk} \geq m_n\} \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{E}^{1/2} |\hat{\beta}_{jk} - \beta_{jk}|^{2p} \, \mathbb{P}_f^{1/2}(n_{jk} \geq m_n) \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\} \\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \Big(\frac{2^j}{n}\Big)^{\frac{p}{2}} \frac{p_{jk}}{\sqrt{n^\gamma}} \leq C \frac{2^{j_n(\frac{p}{2}-1)}}{n^{\gamma/2}} \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}
\end{aligned}
$$

and

$$
C_{012}^{''} = \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{n_{jk} \geq m_n\} \mathbf{1}\Big\{p_{jk} < \frac{m_n}{2n}\Big\}
$$

$$\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{p_{jk} \leq \frac{m_n}{2n}\Big\} \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.$$

The last inequality uses the fact that $f \in \chi((1-\alpha)p, p)$.
We decompose C''_{02} into two parts:

$$\begin{aligned}
C''_{02} &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p \mathbf{1}\{n_{jk} \geq m_n\}\mathbf{1}\{p_{jk} \geq \frac{m_n}{2n}\} \\
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \left|\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}\right|^p \mathbf{1}\{n_{jk} \geq m_n\} \\
&\quad \times \left[\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} + \mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\}\right] = C''_{021} + C''_{022}.
\end{aligned}$$

Let us now consider this new lemma:

Lemma 5.3. *There exists a constant $C < \infty$ such that, for any $\lambda > 0$,*

$$|\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}| \leq C\Big(|\hat{\beta}_{jk} - \beta_{jk}| + \mu\sqrt{\frac{\log(n)}{n}}\Big) + |\beta_{jk}|\mathbf{1}\Big\{\hat{\gamma}_{jk} > \mu\sqrt{\frac{\log(n)}{n}}\Big\}$$

and

$$\begin{aligned}
|\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p &\leq C\Big(|\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\Big\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\Big\} \\
&\quad + \min(|\beta_{jk}|, \gamma_{jk})^p\Big) + |\beta_{jk}|^p \mathbf{1}\Big\{\hat{\gamma}_{jk} > \frac{3\gamma_{jk}}{2}\Big\}.
\end{aligned}$$

Proof. This lemma is proved in Juditsky and Lambert-Lacroix [13]. □

Using Lemma 5.2 with $\gamma \geq \frac{p}{2}$ and Lemma 5.3, one gets for μ large enough:

$$\begin{aligned}
C''_{021} &= \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \left|\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}\right|^p \mathbf{1}\{n_{jk} \geq m_n\}\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{E}\Bigg[|\hat{\beta}_{jk} - \beta_{jk}|^p + \mu^p\sqrt{\frac{\log^p(n)}{n^p}} \\
&\quad + |\beta_{jk}|^p \mathbf{1}\Big\{\hat{\gamma}_{jk} > \mu\sqrt{\frac{\log(n)}{n}}\Big\}\Bigg]\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \Bigg[\Big(\frac{2^j p_{jk}}{n}\Big)^{\frac{p}{2}} + \sqrt{\frac{\log^p(n)}{n^p}} + \frac{2^{jp/2}}{n^\gamma}\Bigg]\mathbf{1}\Big\{p_{jk} > \frac{\nu^2}{2K_\psi^2}\Big\} \\
&\leq C\Bigg[\Big(\frac{1}{n}\Big)^{\frac{p}{2}} + \sqrt{\frac{\log^p(n)}{n^p}} + n^{-\gamma}\Bigg] \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.
\end{aligned}$$

Using again Lemma 5.3,

$$C''_{022} = \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \big|\hat{\beta}_{jk}\mathbf{1}\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}\big|^p \mathbf{1}\{n_{jk} \geq m_n\}$$

$$\times \mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\} = C\ (C''_{0221} + C''_{0222} + C''_{0223}).$$

Using the Cauchy–Schwarz inequality, Lemma 4.4 (1), and Lemma 5.2 with $\gamma \geq 3p+2$, we get

$$C''_{0221} = \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbf{1}\Big\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\Big\}\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\}$$

$$\leq K \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \Big(\frac{2^j p_{jk}}{n}\Big)^{\frac{p}{2}} \mathbb{P}_f^{\frac{1}{2}}\Big(|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\Big)$$

$$\times \mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\} \leq K \frac{2^{j_n(\frac{p}{2}-1)}}{\sqrt{n^\gamma}} \leq K\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2};$$

$$C''_{0222} = \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \min(|\beta_{jk}|, \gamma_{jk})^p \mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk}\Big\}$$

$$\leq \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{|\beta_{jk}| \leq \gamma_{jk}\}$$

$$+ \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \gamma_{jk}^p \mathbf{1}\{|\beta_{jk}| > \gamma_{jk}\} \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.$$

These inequalities are obtained using the fact that $f \in W^*((1-\alpha)p, p)$. Finally, using Lemma 5.2,

$$C''_{0223} = \mathbb{E} \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{\hat{\gamma}_{jk} > \frac{3\gamma_{jk}}{2}\Big\}\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\}$$

$$\leq \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p\, \mathbb{P}_f\Big(|\hat{\gamma}_{jk} - \gamma_{jk}| > \frac{\gamma_{jk}}{2}\Big)\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\}$$

$$\leq C \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)} \sum_k \Big(2^{\frac{j}{2}} p_{jk}\Big)^p \frac{p_{jk}}{n^\gamma}\mathbf{1}\Big\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\Big\}$$

$$\leq C\frac{2^{j_n(p-1)}}{n^\gamma} \leq C\Big(\frac{\log(n)}{n}\Big)^{\alpha p/2}.$$

Consequently, looking at the bounds of C_0 and C_1, we deduce that

$$\sup_{n>1} v_n^{-\alpha p}\, \mathbb{E}\, \|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} < \infty.$$

We conclude that $f \in MS\big(\bar{f}_n, \|\cdot\|^p_{\mathcal{B}^0_{p,p}}, v_n^{\alpha p}\big)$. $\square$

Concluding remarks

Adopting the maxiset point of view for the problem of density estimation, we extended the maxiset result of Cohen *et al.* [4] to the case of noncompactly supported densities. What is more, we proved that if the usual hard thresholding procedure has the best maxiset performance within the family of elitist rules, there is a way to construct other procedures with larger maxisets, by using random thresholds. An example of such a procedure is the Juditsky and Lambert-Lacroix's [13] one.

The maxiset approach is quite interesting, since it allows us to discriminate in a theoretical way between procedures which can have the same minimax performance. According to this new approach, *the larger is the maxiset of a procedure, the better is the procedure*.

Some other examples of procedures which outperform the hard thresholding procedure in the maxiset sense are given in [1]. In particular, this author succeeds in proving that procedures which threshold empirical coefficients by blocks often have better maxiset performance than procedures which threshold empirical coefficients term-by-term.

6. Appendix

Proof of Lemma 4.2. We have

$$\begin{aligned}\mu\sqrt{\frac{\log(n)}{n}} < |\hat\beta_{jk}| = \frac{1}{n}\Big|\sum_{i=1}^{n}\psi_{jk}(X_i)\Big| &\le \frac{1}{n}\sum_{i=1}^{n}2^{j/2}K_\psi \mathbf{1}\{X_i\in I_{jk}\}\\ &\le \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{n}{\log(n)}}K_\psi\mathbf{1}\{X_i\in I_{jk}\} \qquad (\text{since } j<j_n)\\ &\le \frac{1}{n}\sqrt{\frac{n}{\log(n)}}K_\psi n_{jk}.\end{aligned}$$

Finally, one gets

$$|\hat\beta_{jk}| > \mu\sqrt{\frac{\log(n)}{n}} \Longrightarrow n_{jk} > \frac{\mu}{K_\psi}\log(n). \qquad \square$$

Proof of Lemma 4.3. *Step 1*: Suppose that $np_{jk} \ge 2\rho\log(n)$. Since $\tau_{jk}^2 = n\operatorname{Var}_f(\mathbf{1}\{X_1\in I_{jk}\}) = np_{jk}(1-p_{jk})$, then $2\tau_{jk}^2 \le \frac{n^2p_{jk}^2}{\rho\log(n)}$. Using the Bernstein inequality, we have,

$$\begin{aligned}\mathbb{P}_f\big(n_{jk}<\rho\log(n)\big) = \mathbb{P}_f\big(np_{jk}-n_{jk} > np_{jk}-\rho\log(n)\big) &\le \mathbb{P}_f\Big(np_{jk}-n_{jk} > \frac{n}{2}p_{jk}\Big)\\ &\le \exp\Bigg(-\frac{n^2p_{jk}^2}{8\big(\tau_{jk}^2+\frac{np_{jk}^2}{6}\big)}\Bigg) \le \exp\Bigg(-\frac{n^2p_{jk}^2}{8n^2p_{jk}^2\big(\frac{1}{2\rho\log(n)}+\frac{1}{6n}\big)}\Bigg)\\ &\le \exp(-K\rho\log(n)) = n^{-K\rho} \le \frac{p_{jk}}{n^\gamma}.\end{aligned}$$

The last inequality is obtained by taking ρ such that $K\rho \ge 1+\gamma$.

Step 2: Suppose now that $\frac{1}{n^{\gamma+1}} \leq np_{jk} \leq 2\rho\log(n)$. Using the Bernstein inequality, we get

$$\begin{aligned}
&\mathbb{P}_f\big(n_{jk} \geq \rho\log(n)\big)\\
&\quad= \mathbb{P}_f\big(n_{jk} - np_{jk} \geq \rho\log(n) - np_{jk}\big) \leq \mathbb{P}_f\Big(n_{jk} - np_{jk} \geq \frac{\rho\log(n)}{2}\Big)\\
&\quad\leq \exp\bigg(- \frac{\rho^2\log(n)^2}{8\big(\tau_{jk}^2 + \frac{\rho\log(n)}{6}\big)}\bigg) \leq \exp\bigg(- \frac{\rho^2\log(n)^2}{8\big(np_{jk} + \frac{\rho\log(n)}{6}\big)}\bigg)\\
&\quad\leq \exp\big(- K\rho\log(n)\big) = n^{-K\rho} \leq \frac{p_{jk}}{n^\gamma}.
\end{aligned}$$

The last inequality requires that ρ satisfies $K\rho \geq 2(1+\gamma)$.

Step 3: Consider the case $np_{jk} \leq \frac{1}{n^{\gamma+1}}$. Using simple bounds on the tails of the binomial distribution (see inequality 1 on p. 482 in Shorack and Wellner [22]), we get

$$\begin{aligned}
\mathbb{P}_f\big(n_{jk} \geq \rho\log(n)\big) &\leq \frac{(1-p_{jk})}{1 - \frac{(n+1)p_{jk}}{2}}\, \mathcal{C}_n^2\, p_{jk}^2(1-p_{jk})^{n-2}\\
&\leq \frac{n^2p_{jk}^2}{2\big(1 - \frac{(n+1)p_{jk}}{2}\big)} \leq \frac{n^2p_{jk}}{n^{\gamma+2}} = \frac{p_{jk}}{n^\gamma}. \qquad \square
\end{aligned}$$

Proof of Lemma 4.4. Parts 1 and 2: By the Rosenthal inequality, for any j,k

$$\begin{aligned}
\mathbb{E}\,(\hat\beta_{jk} - \beta_{jk})^{2p} &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \psi_{jk}(X_i) - \beta_{jk}\right)^{2p}\\
&\leq \frac{C}{n^{2p}}\bigg[\sum_{i=1}^n \mathbb{E}\,\big(\psi_{jk}(X_i) - \beta_{jk}\big)^{2p} + \Big(\sum_{i=1}^n \mathbb{E}\,(\psi_{jk}(X_i) - \beta_{jk})^2\Big)^p\bigg]\\
&\leq \frac{C}{n^{2p}}(D_0 + D_1),
\end{aligned}$$

where

$$\begin{aligned}
D_0 &= \sum_{i=1}^n \mathbb{E}\,\big(\psi_{jk}(X_i) - \beta_{jk}\big)^{2p} \leq Cn\big(\mathbb{E}\,(\psi_{jk}^{2p}(X_1)) + (\beta_{jk})^{2p}\big)\\
&\leq Cn\big(2^{jp}p_{jk} + (2^{j/2}p_{jk})^{2p}\big) \leq C2^{jp}np_{jk},
\end{aligned}$$

$$\begin{aligned}
D_1 &= \Big(\sum_{i=1}^n \mathbb{E}\,\big(\psi_{jk}(X_i) - \beta_{jk}\big)^2\Big)^p = \Big(\sum_{i=1}^n \mathrm{Var}\,\big(\psi_{jk}(X_i)\big)\Big)^p\\
&\leq \Big(\sum_{i=1}^n \mathbb{E}\,\big(\psi_{jk}^2(X_i)\big)\Big)^p \leq Cn^p(2^jp_{jk})^p \leq C2^{jp}(np_{jk})^p.
\end{aligned}$$

Now, if $np_{jk} \geq 1$, then $np_{jk} \leq (np_{jk})^p$. So

$$\mathbb{E}\,(\hat\beta_{jk} - \beta_{jk})^{2p} \leq C\Big(\frac{2^jp_{jk}}{n}\Big)^p.$$

If $np_{jk} < 1$, then $np_{jk} > (np_{jk})^p$. So

$$\mathbb{E}\,(\hat{\beta}_{jk} - \beta_{jk})^{2p} \le Cnp_{jk}\left(\frac{2^j}{n^2}\right)^p.$$

Finally, part 3 is just a consequence of parts 1 and 2. □

Proof of Lemma 5.1. Suppose that $\hat{\gamma}_{jk} < |\hat{\beta}_{jk}|$. Then,

$$\mu^2\frac{\log(n)}{n}\cdot\frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(X_i)^2 < \mu^2\frac{\log(n)}{n}\hat{\beta}_{jk}^2 + \hat{\beta}_{jk}^2 = \left(\mu^2\frac{\log(n)}{n} + 1\right)\hat{\beta}_{jk}^2.$$

By using bounds on the left- and right-hand sides, one gets for n large enough:

$$\mu^2\frac{\log(n)}{n^2}2^j\nu^2 n_{jk} < 2\hat{\beta}_{jk}^2.$$

And since $n|\hat{\beta}_{jk}| \le 2^{j/2}K_\psi n_{jk}$,

$$\mu^2\nu^2\log(n) < 2K_\psi^2 n_{jk}.$$

Finally, one gets

$$|\hat{\beta}_{jk}| > \hat{\gamma}_{jk} \implies n_{jk} > \frac{\mu^2\nu^2}{2K_\psi^2}\log(n). \qquad \square$$

Acknowledgements. I wish to express my thanks to Professor D. Picard for helpful comments and suggestions.

References

[1] F. Autin, *Maxisets for μ-thresholding rules*, Technical report, 2005.
[2] L. Birgé and P. Massart, *An adaptive compression algorithm in Besov spaces*, Constructive Approximation, 16 (2000), 1–36.
[3] J. Bretagnolle and C. Huber, *Estimation des densités: risque minimax*, Z. Wahrsch. verw. Gebiete, 47 (1979), 119–137.
[4] A. Cohen, R. De Vore, G. Kerkyacharian, and D. Picard, *Maximal spaces with given rate of convergence for thresholding algorithms*, Appl. Comput. Harmon. Anal., 11 (2001), 167–191.
[5] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
[6] L. Devroye, *A Course in Density Estimation*, Progress in Probability and Statistics, Vol. 14, Birkhaeuser, 1987.
[7] D. L. Donoho and I. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), 425–455.
[8] D. L. Donoho and I. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., 90 (1995), 1200–1224.
[9] D. L. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, *Density estimation by wavelet thresholding*, Ann. Statist., 24 (1996), 508–539.
[10] R. H. Farrell, *On the lack of a uniformly consistent sequence of estimators of a density function in certain cases*, Ann. Math. Statist., 38 (1967), 471–474.

[11] R. H. Farrell, *On the best obtainable asymptotic rates of convergence in estimation of a density function at a point*, Ann. Math. Statist., 43 (1972), 170–180.
[12] A. Juditsky, *Wavelet estimators*: *Adapting to unknown smoothness*, Math. Methods Statist., 6 (1997), 1–25.
[13] A. Juditsky and S. Lambert-Lacroix, *On minimax density estimation on* $\mathbb{R}$, Bernoulli, 10 (2004), 187–220.
[14] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelet, Approximation and Statistical Applications*, Lecture Notes in Statistics, Vol. 129, Springer-Verlag, New York, 1998.
[15] G. Kerkyacharian and D. Picard, *Density estimation in Besov spaces*, Statist. Probab. Letters, 13 (1992), 15–24.
[16] G. Kerkyacharian and D. Picard, *Density estimation by kernel and wavelets methods – Optimality of Besov spaces*, Statist. Probab. Letters, 18 (1993), 327–336.
[17] G. Kerkyacharian and D. Picard, *Thresholding algorithms, maxisets, and well concentrated bases*, Test, 9 (2001), 283–344.
[18] G. Kerkyacharian and D. Picard, *Minimax or maxisets*? Bernoulli, 8 (2002), 219–253.
[19] Y. Meyer, *Ondelettes et Opérateurs*, Hermann, Paris, 1990.
[20] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist., 33 (1962), 1065–1076.
[21] S. C. Schwartz, *Estimation of a density probability by an orthogonal series*, Ann. Math. Statist., 38 (1967), 1261–1265.
[22] G. Shorack and J. Wellner, *Empirical Processes with Applications to Statistics*, Wiley, 1986.
[23] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statist. and Appl. Probab., Chapman and Hall, 1986.
[24] G. Wahba, *Optimal convergence properties of variable Knot, kernel and orthogonal series methods for density estimation*, Ann. Statist., 3 (1975), 15–29.

[*Received May* 2006]